MicroRNAs In Extracellular Vesicles as Biomarkers for Pancreaticobiliary Cancers

Daniel Si Kit Liu

A thesis submitted in accordance with the requirements of Imperial College London for the degree of Doctor of Philosophy (PhD) in Clinical Medicine Research

Department of Surgery & Cancer
Imperial College London
Hammersmith Hospital Campus, Du Cane Road
London, W12 0HS, UK

Supervisors:

Mr. Adam Frampton
Dr. Jonathan Krell
Professor Long R. Jiao

Abstract

Pancreaticobiliary cancers including pancreatic ductal adenocarcinoma (PDAC) and cholangiocarcinoma (CCA) have an extremely poor prognosis. MicroRNAs (miRNAs) are small non-coding RNAs that have been found to be commonly deregulated in cancer, found in plasma and bile and can be encapsulated in extracellular vesicles (EVs). EVs are nanosized particles with a lipid bilayer that are secreted by cells and have been found to play a role both locally within the tumour microenvironment and on distant organs to promote the premetastatic niche.

The MIRABILE (MIcroRNAs as BILE-based biomarkers in Pancreaticobiliary Cancers) study was a prospectively recruited study at Hammersmith Hospital (2017-2020) which collected samples from patients presenting with obstruction of the bile duct leading to jaundice. Our aim was to determine novel miRNAs for the diagnosis of PDAC and CCA. We detected miR-182-5p and miR-340-5p in our cell-free RNA from bile that generated an AUC value of 0.85 in the detection of malignancy.

For the first time, we investigated EV miRNAs using a size-exclusion chromatography protocol comparing this with other methods to evaluate optimum EV isolation in bile and plasma. RNA-Sequencing of bile EVs detected the presence of let-7a and miR-21 as upregulated in patients with PDAC. Furthermore, validated differential expression of miR-200a and miR-429 ($p<0.0001$) was found in both PDAC and CCA bile EVs compared with controls. Four candidates (let-7a-5p, miR-21-5p, miR-200a-3p, and miR-429) were combined to form a subset of EV-miRNAs which could detect PDAC with an AUC of 0.96. Plasma EVs were also shown to be present in greater concentrations and upregulated in miR-200a/miR-200b/miR-200c/miR-141/miR-429. Together these 5 members of the miR-200 family were combined to exhibit a validated AUC of 0.82 which in combination with CA 19-9 generated an AUC of 0.96. In conclusion, these markers have the potential to diagnose patients with PDAC and CCA.

Statement of Originality

This is to certify that to the best of my knowledge that the content of this thesis is my own work. This report has not been submitted for any degree or other purposes. All the experimentation presented and written in this thesis has been conducted by me, apart from the following:

Bile cfRNA quantification and sample collection was performed in our lab by Mireia Mato Prado, Imperial College London.

Bile cfRNA RT-qPCR was performed in our lab by Ellie Rees, Imperial College London.

Plasma EV quantification and western blotting analysis was performed in our lab by Flora Upton, MRes, Imperial College London.

Plasma EV RT-qPCR was performed in our lab by Catherine Blaker & Kate Bennett, MRes, Imperial College London.

EV Immunogold labelling and transmission electron microscopy was performed by Dr Padmini Sarathchandra.

RNA library preparation, sequencing and differential expression analysis was performed by Dr Morten Venø and the team at Omiics (Aarhus, Denmark).

Copyright Declaration

The copyright of this thesis rests with the author. Unless otherwise indicated, its contents are licensed under a Creative Commons Attribution-Non-Commercial 4.0 International Licence (CC BY-NC 4.0).

Under this licence, you may copy and redistribute the material in any medium or format. You may also create and distribute modified versions of the work. This is on the condition that: you credit the author and do not use it, or any derivative works, for a commercial purpose.

When reusing or sharing this work, ensure you make the licence terms clear to others by naming the licence and linking to the licence text. Where a work has been adapted, you should indicate that the work has been changed and describe those changes.

Please seek permission from the copyright holder for uses of this work that are not included in this licence or permitted under UK Copyright Law.

Permissions & Rights

Page No.	Type of work:	Name of work	Source of work	Copyright holder and contact	Permission requested date	Permission: yes /no	Permission note
30	Figure	FIG. 2. Progression from early to late PanIN.	Hackeng, Wenzel M., Ralph H. Hruban, G. Johan A. Offerhaus, and Lodewijk AA Brosens. "Surgical and molecular pathology of pancreatic neoplasms." Diagnostic pathology 11, no. 1 (2016): 1-17.	© The Author(s). 2016	14/01/2023	yes	Licensed with a Creative Commons Attribution 4.0 International License
34	Figure	FIG. 4. Phylotranscriptomic tree of pancreatic cancer	Collisson, Eric A., Peter Bailey, David K. Chang, and Andrew V. Biankin. "Molecular subtypes of pancreatic cancer." Nature reviews Gastroenterology & hepatology 16, no. 4 (2019): 207-220.	© Springer Nature	25.04.15	yes	Written permission
70	Figure	Receiver Operating Characteristic (ROC) curve with False Positive Rate and True Positive Rate	https://commons.wikimedia.org/wiki/File:Roc_curve.svg#metadata	© CMGLee	14/01/2023	yes	Licensed with a Creative Commons Attribution-Share Alike 4.0 International license. (CC-BY-SA-4.0)

Acknowledgements

I would like to thank my supervisors Dr Jonathan Krell, Mr Adam Frampton and Professor Long Jiao for the opportunity to have undertaken this research and for their guidance and mentorship throughout this lengthy process. Without the funding provided by the J.P. Moulton Charitable Foundation and the Royal College of Surgeons of England this would not have been possible. I would also like to acknowledge the altruism of the patients and their families who have agreed to participate in this research and contributed to making the future a safer place.

I am especially grateful to the members of the laboratory for their assistance with this project and who gave me the practical know-how to get the project off the ground. Thanks to Mireia Mato Prado, Phil Lawton, Elena Lopez Jimenez, Ellie Rees, Flora Upton, Kate Bennett, and Catherine Blaker.

I would also like to thank the other members of the Division of Cancer who have at some point or other helped with the practical and sometimes emotional lab work. Whilst COVID-19 had a massive impact on getting to know everyone better, I will be forever grateful to Dr Sophie McNamara, Dr Paula Cunnea, Dr Lisa Del Bel Belluz, Dr Claire Fletcher, Dr Yoyo Dixon, and the students of the 4th Floor.

A special thanks to the collaborators who have helped with many of the technical aspects of the project including Dr Morten Venø, Dr Elisa Giovannetti, and Dr Padmini Sarathchandra.

Finally, I am indebted to my family and friends, particularly my Mum, Dad and sister Deborah who have cheered from the side-lines and above all to my partner Katie Briggs for her infinite patience during what has been a difficult year.

Table of contents

Abstract **2**

Statement of Originality **3**

Copyright Declaration **4**

Permissions & Rights **5**

Acknowledgements **6**

Table of contents **7**

Table of Figures **15**

Table of Tables **19**

Schematic Diagram of Studies present in the Manuscript **21**

Abbreviations **22**

Full Manuscripts Arising **26**

Chapter One – Introduction **28**

1.1 Pancreatic Cancer **28**

1.1.1 Overview 28

1.1.2 Risk factors for PDAC 30

1.1.3 The genetic and molecular basis for PDAC. 31

1.1.4 The role of pancreatic stroma in PDAC 35

1.2 Bile duct cancer **36**

1.2.1 Overview 36

1.2.2 Risk factors for CCA. 37

1.2.3 The genetic and molecular basis for CCA. 37

1.3 PDAC and CCA: Clinical Picture **38**

1.3.1 The diagnosis and evaluation of patients with suspected PDAC or CCA. 38

1.3.2 Blood-based markers for the evaluation and diagnosis of PDAC/CCA. 41

1.3.3 The Staging and Treatment of PDAC and CCA. 42

1.3.4 Surgical Approaches to PDAC and CCA 45

1.3.5 Chemotherapy in the treatment of PDAC. 46

1.3.6 The Role of Neoadjuvant Chemotherapy in PDAC. 47
1.3.7 Chemotherapy in the treatment of CCA. 48
1.3.8 The Role of Neoadjuvant Chemotherapy in CCA. 49
1.4 The role of RNA biology in PDAC and CCA **50**
1.4.1 MicroRNAs as a source of biomarkers for the detection of PDAC and CCA. 50
1.4.2 miRNA genes and nomenclature. 50
1.4.3 miRNA Biogenesis. 51
1.4.4 miRNAs in Cancer. 53
1.4.5 Tissue miRNA expression in PDAC/CCA. 54
1.4.6 Laboratory techniques used in the analysis of miRNAs. 58
1.4.7 Extracellular vesicles (EVs) as a source of extracellular miRNAs. 60
1.4.8 EV Nomenclature. 61
1.4.9 Laboratory techniques for EV Isolation 62
1.4.10 Laboratory techniques for EV Characterisation. 64
1.4.11 Clinical rationale for the use of EVs. 67
1.4.12 Bile as a source of EVs. 68
1.4.13 MiRNAs are present in bile-derived EVs. 69
1.4.14 Blood plasma as a source of EVs. 69
1.4.15 MiRNAs are present in plasma-derived EVs. 70
1.4.16 The presence and role of tRNA derived fragments (tRFs) in the extracellular space. 71
1.4.17 A note on biomarker detection with reference to sensitivity and specificity. 73
1.5 Aims, Hypothesis and Objectives **75**
Chapter Two – Materials and Methods 77
2.1 Patients and Materials **77**
2.1.1 Samples collected from surgery. 77
2.1.2 Samples collected for the MIRABILE study. 77
2.1.3 Patient demographics and pathology data collection. 78

2.2 Methods 81
2.2.1 Preparation of samples for cell-free RNA analysis 81
2.2.2 Preparation of samples for EV isolation. 81
2.2.3 EV Isolation using ultracentrifugation (UC). 81
2.2.4 EV Isolation using size exclusion chromatography (SEC). 81
2.2.5 EV isolation using a Plasma/Serum Exosome Purification mini kit. 82
2.2.6 Nanoparticle tracking analysis (NTA) 82
2.2.7 Protein quantification. 83
2.2.8 Western blotting 83
2.2.9 Transmission electron microscopy. 84
2.2.10 EV concentration and preparation for RNA extraction 85
2.2.11 Total RNA extraction of bile samples using TRIzol LS reagent. 85
2.2.12 Total and EV-RNA Extraction of plasma samples using the miRNeasy Serum/Plasma advanced kit (QIAGEN). 86
2.2.13 Library preparation and next generation RNA Sequencing (performed externally) 87
2.2.14 RNA quantification by RNA-specific Fluorometry. 88
2.2.15 RNA quality and quantity analysis using Electrophoresis. 88
2.2.16 Quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR) using target-specific stem–loop primer assays (TaqMan) to detect miRNAs. 88
2.2.17 Quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR) using universal reverse transcription (RT) and locked nucleic acid (LNA) enhanced specific primers with SYBR green I assay detection to detect miRNAs. 89
2.2.18 Calculation of relative expression. 91
2.2.19 Generation of endogenous controls for normalisation. 91
2.2.20 Statistical Analysis. 91
2.2.21 MicroRNA Functional Analysis using a microRNA enrichment analysis and annotation tool (miEAA) 92
Chapter Three – Cell-free small RNA markers for PDAC and CCA in Bile and Plasma 93

3.1 Background 93
3.2 Results 95
3.2.1 Sample Collection 95
3.2.2 Clinical differences in patient characteristics for the MIRABILE bile study 97
3.2.3 Clinical differences in patient characteristics for the MIRABILE plasma study 101
3.2.4 Cell-free RNA analysis shows differences in RNA quantity and composition... 104
3.2.5 Discovery of miRNAs present in bile and plasma cfRNA by next generation sequencing is differentially expressed in cancer cohorts. 107
3.2.6 Graphical analysis of 457 miRNAs in bile shows differential expression by cancer cohort. 108
3.2.7 Looking for statistically significant candidates within our differential expression analysis of bile miRNAs. 110
3.2.8 Small RNA sequencing also reveal differential expression of tRNA-derived fragments. 114
3.2.9 Individual bile miRNA candidates from RNA Sequencing show AUCs above 0.80. 120
3.2.10 Regression analysis potentially identifies signature in bile with outstanding diagnostic potential. 123
3.2.11 Graphical analysis of cfRNA sequencing from blood plasma miRNAs fails to show significant differential expression. 125
3.2.12 Bile and plasma cfRNA concentrations and quantity are a limitation to TaqMan RT-qPCR validation. 127
3.2.13 RT-qPCR Validation from (n=87) bile samples in our six candidates from the LASSO regression analysis. 128
3.2.14 Correlation analysis of miRNA expression 133
3.2.15 Functional Analysis of identified miRNAs. 135
3.3 Discussion 137
3.3.1 Previous literature on bile cfRNAs shows heterogenous results. 137

3.3.2 A bile cfRNA biomarker has been shown to work in a suitable population at risk for PDAC/CCA..... 138

3.3.3 Biomarker discovery using sequencing and RT-qPCR validation generates models with good predictive ability. 139

3.3.4 Novel small RNA biomarkers include tRNA fragments (tRFs) in bile and may be a source of further investigation. 140

3.3.5 Limitations and future strategies 141

Chapter Four – Small RNA markers for the detection of PDAC and CCA in Bile Extracellular Vesicles 143

4.1 Background **143**

4.2 Results **144**

4.2.1 Surgical samples used for initial optimisation of EV isolation. 144

4.2.2 Initial use of ultracentrifugation as a method of EV isolation. 144

4.2.3 Size-exclusion as a method of EV isolation improves workflow 148

4.2.4 Surgical Bile EV Characterisation by NTA and TEM 151

4.2.5 Surgical Bile EVs obtained by SEC analysed by Western Blotting. 154

4.2.6 Surgical Bile EVs obtained by SEC show no difference in particle size or EV concentration. 157

4.2.7 Surgical Bile EVs undergo RNA extraction and can be sequenced for differential expression. 159

4.2.8 Differential expression analysis undertaken for EV-miRNA identified from surgical bile samples. 162

4.2.9 Differential expression analysis undertaken for EV-tRNAs identified from surgical bile samples. 165

4.2.10 Characterisation of EVs present in endoscopic bile samples obtained prospectively (MIRABILE). 171

4.2.11 Clinical characteristics for the patients comprising the MIRABILE EVs analysis. 174

4.2.12 Workflow for the biomarker analysis of EV-RNA from the MIRABILE samples. 177

4.2.13 RNA quantification in bile-derived EVs. 178
4.2.14 Differential expression analysis of bile EV-miRNAs 180
4.2.15 Candidates chosen for RT-qPCR Validation show outstanding AUC in the discovery cohort..... 185
4.2.16 A note on RT-qPCR Normalisation. 187
4.2.17 Validation of our 12 candidates shows diagnostic potential in EV-miRNA markers for PDAC and CCA. 188
4.2.18 Combination of candidates using logistic regression improves diagnostic sensitivity. 191
4.2.19 tRNAs in bile EVs as a potential future source of biomarkers. 194
4.3 Discussion 198
4.3.1 Bile-derived EVs are an under investigated source of small RNAs which are differentially expressed in PDAC. 198
4.3.2 Optimisation on Bile-derived EVs from surgery led to SEC as a novel isolation method with high throughput and reproducibility. 199
4.3.3 Comparison of EV characterisation between surgical and endoscopic samples 199
4.3.4 Let-7a, miR-21 and miR-200 family has been detected in Bile EVs for the first time..... 200
4.3.5 Diagnostic potential of these as a panel and in combination with CA 19-9 201
4.3.6 Novel tRFs may be a future promising source of biomarkers and are found present in bile EVs for the first time. 203
Chapter Five – Small RNA markers for the detection of PDAC in Plasma Extracellular Vesicles..... 204
5.1 Background..... 204
5.2 Results..... 205
5.2.1 Samples used for plasma EV isolation. 205
5.2.2 Study design for the analysis of plasma EVs 205
5.2.3 Characterisation of EVs present in plasma samples obtained prospectively (MIRABILE)..... 206

5.2.4 Western blotting of EVs present in plasma samples obtained prospectively (MIRABILE) highlights specific differences between bile and plasma derived EVs.. 207

5.2.5 Clinical characteristics for the patients comprising the MIRABILE plasma EVs analysis. 212

5.2.6 RNA isolation from plasma EVs shows differential expression of miR-200 family. 214

5.2.7 ROC curves for each miR-200 family candidate demonstrates high diagnostic potential. 219

5.2.8 miRNAs for normalisation 220

5.2.9 Validation of our miR-200 family 5 miRNA panel shows diagnostic potential for EV-miRNA plasma markers for PDAC. 222

5.2.10 EV parameters (Concentration and particle size) improve diagnostic sensitivity. 224

5.2.11 Verification of EV-miRNAs and miR-200 against lipoprotein fractions 226

5.2.12 tRNAs in plasma EVs as a potential future source of biomarkers 228

5.2.13 Prospective investigation in longitudinal post-operative patient samples shows potential change after surgery 229

5.3 Discussion 231

5.3.1 Plasma-derived EVs are a source of small RNAs which are differentially expressed in PDAC. 231

5.3.2 miR-200 family has been previously identified as overexpressed in previous studies investigating the plasma EVs of patients with PDAC. 231

5.3.3 SEC as an isolation method for plasma-EVs 232

5.3.4 Lipoprotein as a confounding factor in the isolation of plasma EVs. 233

5.3.5 Plasma EV-RNA compared with analysis of Plasma cfRNA markers. 233

5.3.5 EV parameters in plasma EVs 234

5.3.6 tRFs in plasma EVs as a novel diagnostic marker 234

5.3.7 Analysis of pre- and post-operative miR-200 family from EV-RNA shows potential for prognostication. 235

Chapter Six – General Discussion and Conclusions 236

6.1 Discussion **236**

6.1.1 Bile EVs show upregulated expression of let-7a and miR-21 in PDAC. 236

6.1.2 Bile and Plasma EVs show upregulated expression of miR-200 family in PDAC and CCA 238

6.1.2.1 The evidence for EV-miR-200 in relation to its role in tissue. 238

6.1.2.2 The role of miR-200 as a regulator in EMT plasticity. 242

6.1.2.3 The presence of miR-200 in plasma EVs and a comparison between EV and cfRNA. 244

6.2 Limitations **247**

6.3 Future Directions **249**

6.3.1 Outstanding questions arising from this project. 249

6.3.2 Identifying the precursor gene cluster responsible for let-7a upregulation. 250

6.3.3 Determine any functional role for miR-200, miR-21 and let-7a in bile and plasma. 250

6.3.4 EV miR-200, miR-21 and let-7a within early precursor lesions. 251

6.3.5 Relationship of EV-miRNAs to molecular subtype and survival. 251

6.3.6 Novel tRFs may be a future promising source of biomarkers. 252

Bibliography **254**

Appendix **279**

Papers Published **286**

Table of Figures

Figure 1: Pancreatic and biliary tract anatomy..... 29
Figure 2: Progression from early to late PanIN..... 31
Figure 3: Phylotranscriptomic tree of pancreatic cancer..... 35
Figure 4: Receiver Operating Characteristic (ROC) curve..... 73
Figure 5: Patient Recruitment and Sample Collection..... 95
Figure 6: Patients selected for the discovery and validation cohorts used in the bile and plasma cfRNA biomarker analysis..... 99
Figure 7: Examining the relationship between serum bilirubin and CA 19-9 in the overall population for the bile cfRNA analysis (n=66)..... 101
Figure 8: Schematic diagram showing the workflow for microRNA biomarker analysis. 105
Figure 9: RNA quantification and composition analysis of bile and plasma discovery cohorts..... 106
Figure 10: Differential expression of miRNAs present in bile samples (n=81)..... 109
Figure 11: Volcano plots showing significant miRNAs present in bile cfRNA samples (n=78)..... 111
Figure 12: Further multivariate analysis showing non-significant differential expression in bile samples when compared for anatomical location, presence of stent or resection status (n=81)..... 114
Figure 13: Differential expression of tRFs present in bile cfRNA samples (n=78)..... 116
Figure 14: Volcano plots showing significant tRFs present in bile samples (n=78). 117
Figure 15: Bile cfRNA tRF expression expressed according to tRNA isotype (n=78)..... 120
Figure 16: Diagnostic potential of selected individual miRNAs identified as over-expressed in bile cfRNA in Malignant vs Benign and PDAC vs Benign (n=78). 121
Figure 17: Expression of miR-196a by RNA sequencing in bile cfRNA (n=78)..... 123
Figure 18: LASSO Regression analysis to determine optimal biomarker model (n=78) 124
Figure 19: Semi-supervised heatmap showing discovery cohort expression of markers identified by LASSO in bile samples..... 125
Figure 20: Differential expression analysis of miRNAs present in plasma samples (n=41; 26 PDAC, 5 CCA, 10 Benign). 126
Figure 21: Quantification of RNA extraction from bile by Qubit Fluorometry (n=133). 127
Figure 22: RT-qPCR Validation of 4 marker panel identified by LASSO in bile samples (n=87)..... 130

Figure 23: Validated ROC Curves for miR-182 and miR-340 in bile cfRNA (Malignant vs Benign; n=87). 132
Figure 24: Validated ROC Curves for miR-182 and miR-340 in bile cfRNA (PDAC vs Benign; n=87). 133
Figure 25: Correlation matrix for miRNAs of interest, laboratory parameters and age. 134
Figure 26: Differential ultracentrifugation protocols used in prior bile EV studies..... 145
Figure 27: Photos showing initial optimization using differential ultracentrifugation as a method for EV isolation. 147
Figure 28: Photos and diagrams showing optimization using SEC as a method for EV isolation. 149
Figure 29: Study Design for the evaluation of EVs from surgical bile samples. 151
Figure 30: Characterisation of individual fractions show EVs from surgical bile samples (n=6). 152
Figure 31: NTA of individual fractions show SEC reproducibility in surgical bile samples (n=2). 153
Figure 32: Characterisation by western blot of EVs in surgical bile samples (n=6)..... 155
Figure 33: Analysis of particle and protein concentration in surgical bile EVs obtained by SEC (n=46). 158
Figure 34: Relative distribution of small RNA reads for each bile sample from the discovery cohort (n=18)..... 161
Figure 35: Differential expression using filtered RNA sequencing data for surgical bile EVs (n=18). 163
Figure 36: RNA sequencing identifies a two miRNA signature in surgical bile EVs (n=46). 165
Figure 37: Venn diagrams showing the total number of differentially expressed tRFs in surgical bile EVs (n=18). 166
Figure 38: Volcano plots showing significant tRFs identified by small RNA sequencing as present in surgical bile EV samples (n=18)..... 167
Figure 39: Surgical Bile EV tRF expression expressed according to tRNA isotype (n=18). 170
Figure 40: Characterisation of individual EV fractions using TEM and Western blotting of the MIRABILE bile samples (n=6). 172
Figure 41: NTA and protein concentration analysis in MIRABILE bile samples (n=9; n=90). 173

Figure 42: Study Design for the evaluation of EVs from MIRABILE bile samples..... 178
Figure 43: RNA Composition of MIRABILE Bile EV-RNA (n=30)..... 179
Figure 44: Differential expression analysis of miRNAs identified from small RNA sequencing in bile EVs (n=28)..... 180
Figure 45: Volcano plots showing significant miRNAs present in bile EV (MIRABILE) samples (n=28)..... 182
Figure 46: Receiver-Operator Curve plots for selected miRNAs identified as differentially expressed in PDAC and CCA samples in the bile EV biomarker analysis. 186
Figure 47: RT-qPCR Validation of marker panel in bile EV RNA samples (PDAC vs Benign; n=64). 188
Figure 48: RT-qPCR Validation of marker panel in bile EV RNA samples (CCA vs Benign; n=54). 189
Figure 49: RT-qPCR standard curves for each miRNA candidate validated (n=1 in triplicate). 191
Figure 50: Regression analyses to validate bile EV-miRNA biomarker model (PDAC; n=64)..... 192
Figure 51: Regression analyses to validate bile EV-miRNA biomarker model (CCA; n=54)..... 193
Figure 52: Evaluation of a 2 marker model (miR-200a and miR-429) by regression analysis from bile EVs (Malignant vs Benign; n=81). 194
Figure 53: Differential expression analysis for tRFs in MIRABILE bile EVs from RNA sequencing data (n=30; 10 PDAC, 10 CCA, 10 Benign). 195
Figure 54: Volcano plots showing significant tRFs present in bile EV (MIRABILE) samples (n=30)..... 196
Figure 55: Study Design for the evaluation of EVs from plasma samples..... 206
Figure 56: Characterisation of individual and pooled EV fractions using TEM and NTA of the MIRABILE plasma samples (n=6)..... 207
Figure 57: Characterisation of individual and pooled EV fractions using western blotting of the MIRABILE plasma samples (n=6)..... 209
Figure 58: Additional western blotting of SEC EV pooled fractions in plasma samples (n=6)..... 211
Figure 59: RNA quantification and initial differential expression analysis of miRNAs present in plasma EV samples (n=20). 216
Figure 60: Differential expression of miRNAs present in plasma EV samples (n=18).. 217

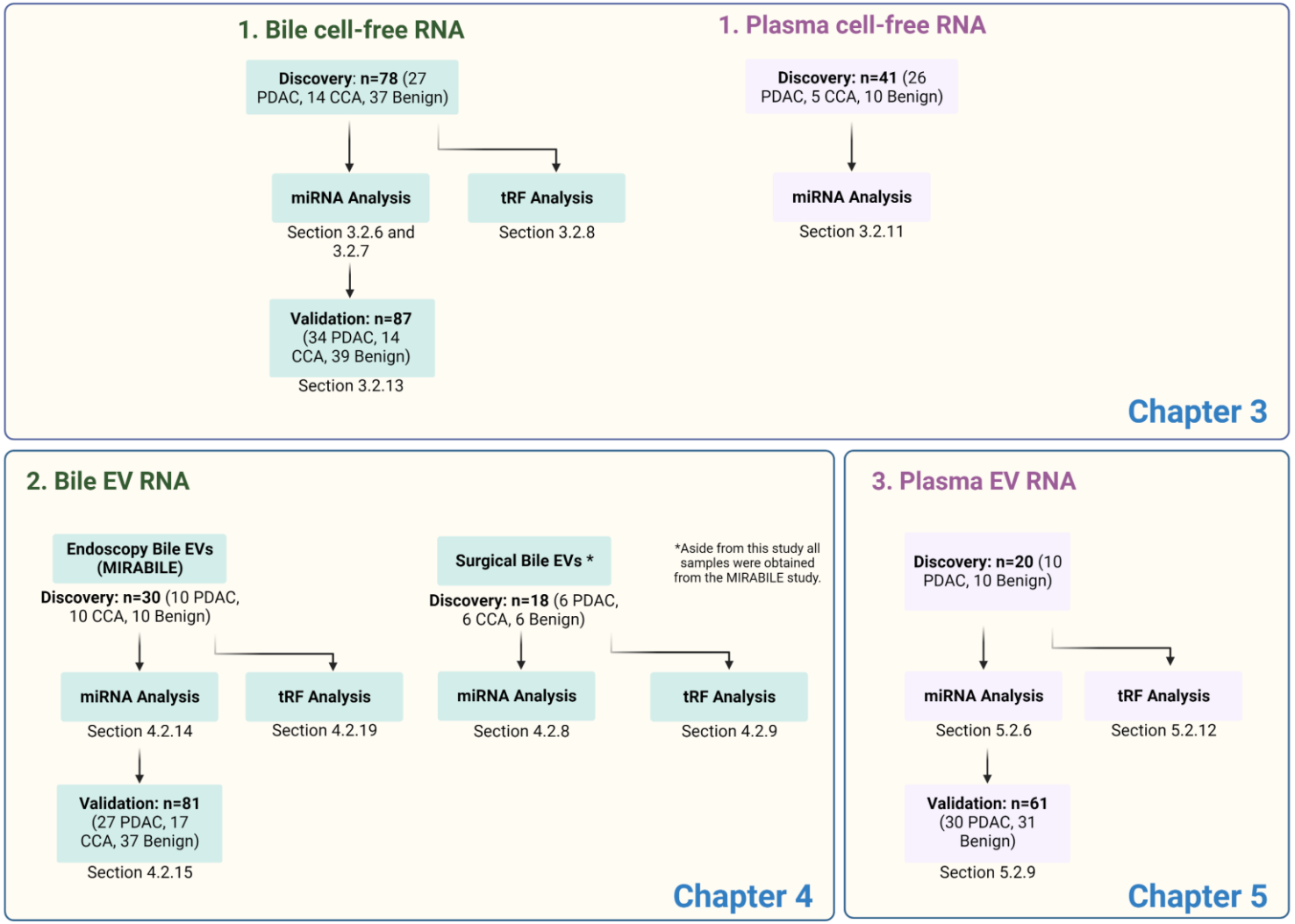
Figure 61: Comparison between circulating plasma and plasma EV differential analysis highlights improved resolution of miRNA candidates..... 219
Figure 62: Predicted ROC Curves for selected miRNAs identified as expressed in PDAC samples from the plasma EV discovery (n=18)..... 220
Figure 63: RT-qPCR Validation of marker panel identified in plasma EV RNA samples (n=61)..... 222
Figure 64: Regression analyses to validate plasma EV-miRNA biomarker model (PDAC vs Benign; n=61)..... 223
Figure 65: Regression analyses of plasma EV parameters (concentration and size) as an alternative model (PDAC vs Benign) (n=61)..... 225
Figure 66: RT-qPCR verification of marker panel in plasma EVs compared with lipoprotein and protein fractions (n=6)..... 227
Figure 67: RT-qPCR standard curves for each miRNA (n=1; triplicate). 228
Figure 68: Plasma EV tRF expression expressed according to tRNA isotype (n=18).... 229
Figure 69: Analysis of expression of miR-200 in blood plasma EVs from patients undergoing surgery for PDAC show potential to predict PDAC tumour burden (n=10). 230

Table of Tables

Table 1: PDAC Staging. 43
Table 2: Perihilar CCA Staging. 44
Table 3: Extrahepatic CCA Staging. 45
Table 4: Selected summary of previous large (>30) miRNA expression profiling studies in pancreatic cancer and pancreatic cancer cell-lines. 56
Table 5: The EV-Metric. 64
Table 6: Buffers, reagents & consumables. 79
Table 7: Components/recipes for reagents used. 80
Table 8: Antibodies. 84
Table 9: RT-qPCR Primers. 90
Table 10: Patients selected for the discovery and validation cohorts used in the bile cfRNA biomarker analysis. 98
Table 11: Patients selected for the discovery cohort used in the plasma cfRNA biomarker analysis. 103
Table 12: Top Twenty miRNAs identified as expressed in Malignant (PDAC and CCA) vs Benign bile samples. 112
Table 13: Top Ten miRNAs identified as highly expressed in PDAC vs Benign and CCA vs Benign bile samples separately (n=81). 113
Table 14: Top 10 tRFs identified as upregulated in the bile cfRNA biomarker analysis (n=78). 118
Table 15: Top performing miRNAs by AUC expressed in Malignant samples in bile cfRNA. 122
Table 16: Top performing miRNAs by AUC expressed in PDAC samples in bile cfRNA. 122
Table 17: Top performing miRNAs by AUC expressed in CCA samples in bile cfRNA. 122
Table 18: MiRNAs identified as stably expressed for normalization individually. 128
Table 19: MiRNAs identified as stably expressed for normalization as a pair. 129
Table 20: MiRNA Gene Set Enrichment Analysis. 136
Table 21: Studies identified to-date involving small RNA markers in bile. 137
Table 22: Patients selected for characterization and sequencing in the surgical bile EV microRNA biomarker analysis. 160
Table 23: Top twenty microRNAs present in bile EVs (n=18; Discovery cohort 6 PDAC, 6 CCA, 6 Benign). 162

Table 24: Significant miRNAs (p<0.05) common from each pairwise comparison alongside adjusted p-value and log2fold change. 164
Table 25: Top 10 tRFs identified as upregulated in the surgical bile EV analysis (n=18). 169
Table 26: Patients selected for the discovery and validation cohorts used in the bile EV biomarker analysis. 175
Table 27: Significant miRNAs (p<0.05) common from each pairwise comparison alongside adjusted p-value and log2fold change in MIRABILE bile EVs (n=28) 183
Table 28: Top Twenty miRNAs identified as expressed in PDAC samples in the MIRABILE bile EV biomarker analysis (n=28). 184
Table 29: Top Twenty miRNAs identified as expressed in CCA samples in the MIRABILE bile EV biomarker analysis (n=28). 184
Table 30: Top twenty miRNAs identified as expressed in Malignant samples in the MIRABILE bile EV biomarker analysis (n=28). 185
Table 31 & 32: MiRNAs identified as stably expressed for normalization in bile EVs. 187
Table 33: Top ten tRNA fragments identified as upregulated in bile EV biomarker analysis (PDAC vs Benign). 197
Table 34: Top ten tRNA fragments identified as upregulated in bile EV biomarker analysis (CCA vs Benign). 197
Table 35: Top ten tRNA fragments identified as upregulated in bile EV biomarker analysis (Malignant vs Benign) 197
Table 36: Studies identified to-date investigating bile EVs. 198
Table 37: Patients selected for the discovery and validation cohorts used in the plasma EV biomarker analysis. 213
Table 38: Significant miRNAs identified as upregulated in PDAC samples from the plasma EV biomarker analysis (n=18). 218
Table 39: MiRNAs identified as stably expressed individually for normalization in plasma EVs (n=18). 221
Table 40: MiRNA pairs identified as stably expressed for normalization in plasma EVs (n=18). 221
Table 41: Top ten tRNA fragments identified as upregulated in the plasma EV biomarker discovery analysis (n=18). 229
Table 42: Selected summary of evidence for the miR-200 family in PDAC/CCA 239
Appendix Table 1: Summary of results from overall project 279
Appendix Table 2: Summary of Plasma EV studies in PDAC 280

Schematic Diagram of Studies present in the Manuscript



Abbreviations

ADEX	-	aberrantly differentiated endocrine exocrine
ADP	-	adenosine diphosphate
adj	-	adjusted
AFM	-	atomic force microscopy
AGO	-	Argonaute
AJCC	-	American Joint Committee on Cancer
ATP	-	adenosine triphosphate
AUC	-	area under the curve
BTC	-	biliary-tract cancer
°C	-	Centigrade
CA 19-9	-	carbohydrate antigen 19-9
CAF	-	cancer-associated fibroblasts
CAP	-	capecitabine
CCA	-	cholangiocarcinoma
cDNA	-	complementary DNA
C. elegans	-	Caenorhabditis elegans
CP	-	chronic pancreatitis
CT	-	computed tomography
CSF	-	cerebrospinal fluid
dCCA	-	distal cholangiocarcinoma
DFS	-	disease free survival
DLS	-	dynamic light scattering
DNA	-	deoxyribonucleic acid
DMEM	-	Dulbecco's Modified Eagle Medium
dNTP	-	deoxyribonucleotide triphosphate
EDTA	-	ethylenediaminetetraacetic acid
EGFR	-	epidermal growth factor receptor
ELISA	-	enzyme-linked immunosorbent assay
EM	-	electron microscopy
EMT	-	epithelial-mesenchymal transition
ERCP:	-	endoscopic retrograde cholangiopancreatography
ESCRT	-	endosomal sorting complex required for transport
EUS	-	endoscopic ultrasonography

EV	-	extracellular vesicle
FBS	-	foetal bovine serum
FC	-	fold change
FFPE:	-	formalin-fixed paraffin-embedded
FNA	-	fine-needle aspiration
FOLFIRINOX	-	folinic acid, fluorouracil, irinotecan, oxaliplatin
FU	-	fluorouracil
GEM	-	gemcitabine
GEMOX	-	gemcitabine-oxaliplatin
GDP	-	guanosine diphosphate
GTP	-	guanosine triphosphate
HCC	-	hepatocellular carcinoma
HITS-CLIP	-	Argonaute High Throughput Sequencing after Cross-Linked Immunoprecipitation
HRP	-	horseradish peroxidase
iCCA	-	intrahepatic cholangiocarcinoma
IDFB	-	intraductal forceps biopsy
ILV:	-	intraluminal vesicle
IPMN:	-	intraductal papillary mucinous neoplasm
kDa	-	kilodalton
lncRNA	-	Long non-coding RNA
MISEV	-	Minimal Information for the Study of Extracellular Vesicles
MIF	-	macrophage migration inhibitory factor
miRNAs/miR	-	microRNAs
MRCP	-	magnetic resonance cholangiopancreatography
MRE	-	microRNA recognition element
MRI	-	magnetic resonance imaging
mRNA	-	messenger RNA
MVB	-	multivesicular body
MVE	-	multivesicular endosome
PDAC	-	pancreatic ductal adenocarcinoma
PanIN	-	pancreatic intraepithelial neoplasm
NaCl	-	sodium chloride
ncRNA	-	non-coding RNA
nt	-	nucleotide

NTA	-	nanoparticle tracking analysis
OS	-	overall survival
PanIN	-	pancreatic intraepithelial neoplasm
PBS	-	phosphate buffered saline
PC	-	pancreatic cancer
PCR	-	polymerase chain reaction
pCCA	-	perihilar cholangiocarcinoma
PDAC	-	pancreatic ductal adenocarcinoma
PET	-	positron emission tomography
PFA	-	paraformaldehyde
PFS	-	progression free survival
PNET	-	pancreatic neuroendocrine tumours
Pre-miRNA	-	precursor microRNA
Pri-miRNA	-	primary microRNA
PSC	-	primary sclerosing cholangitis
RT-qPCR	-	quantitative reverse transcription polymerase chain reaction
RISC	-	RNA-induced silencing complex
RNA	-	ribonucleic acid
RNase	-	ribonuclease
RNAi	-	RNA interference
RNA-Seq	-	RNA sequencing
ROC	-	receiver operating characteristic
rRNA	-	Ribosomal RNA
RT-qPCR	-	reverse transcription quantitative polymerase chain reaction
SD	-	standard deviation
SDS-PAGE	-	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SEC	-	size exclusion chromatography
RT	-	reverse transcription
siRNA	-	small interfering RNA
snRNA	-	small nuclear RNA
snoRNA	-	small nucleolar RNA
Taq	-	Thermus aquaticus
TEM	-	transmission electron microscopy
TGF-ß	-	transforming growth factor type ß
TLR	-	toll-like receptor

TNM	-	tumour, node, metastasis
tRF	-	Transfer RNA derived fragment
tRNA	-	transfer RNA
UICC	-	Union for International Cancer Control
US	-	ultrasonography
UTR	-	untranslated region

Full Manuscripts Arising

Below are complete papers that have been published during this PhD period. Only papers related to pancreatic cancer and/or extracellular vesicles have been included here; abstracts / posters / submitted papers have not been included.

The Clinical Significance of Transfer RNAs Present in Extracellular Vesicles.

Daniel S.K. Liu, Qi Zhi Clayton Yang, Mohammad Asim, Jonathan Krell, & Adam E. Frampton Int J Mol Sci. 2022 Mar 28;23(7):3692. doi: 10.3390/ijms23073692.

Size Exclusion Chromatography as a Technique for the Investigation of Novel Extracellular Vesicles in Cancer

Daniel S.K. Liu, Flora M. Upton, Eleanor Rees, Christopher Limb, Long R. Jiao, Jonathan Krell & Adam E. Frampton (2020) Cancers. 2020; 12(11):3156

The Role of Circular RNAs in Pancreatic Ductal Adenocarcinoma and Biliary-Tract Cancers

Christopher Limb, **Daniel SK Liu**, Morten T. Veno, Eleanor Rees, Jonathan Krell, Izhar N. Bagwan, Elisa Giovannetti & Adam E. Frampton 2020 Cancers 2020, 12(11), 3250 DOI: 10.3390/cancers12113250

"Open Sesame?": Biomarker Status of the Human Equilibrative Nucleoside Transporter-1 and Molecular Mechanisms Influencing its Expression and Activity in the Uptake and Cytotoxicity of Gemcitabine in Pancreatic Cancer.

Ornella Randazzo, Filippo Papini, Giulia Mantini, Alessandro Gregori, Barbara Parrino, **Daniel S. K. Liu**, Stella Cascioferro, Daniela Carbone, Godefridus J. Peters, Adam E. Frampton, Ingrid Garajova, & Elisa Giovannetti. 2020. Cancers 12, no. 11: 3206. doi: 10.3390/cancers12113206

Can circulating tumor and exosomal nucleic acids act as biomarkers for pancreatic ductal adenocarcinoma?

Daniel S.K. Liu, Mireia Mato Prado, Elisa Giovannetti, Long R. Jiao, Jonathan Krell & Adam E. Frampton Expert Review of Molecular Diagnostics, 2019 19:7, 553-558, DOI: 10.1080/14737159.2019.1622414

Pharmacogenetics of treatments for pancreatic cancer

Btissame El Hassouni, Giovanna Li Petri, **Daniel S.K. Liu**, Stella Cascioferro, Barbara Parrino, Waqar Hassan, Patrizia Diana, Asif Ali, Adam E Frampton & Elisa Giovannetti 2019 Expert Opinion on Drug Metabolism & Toxicology, 15:6, 437-447, DOI: 10.1080/17425255.2019.1620731

Chapter One – Introduction

1.1 Pancreatic Cancer

1.1.1 Overview

The pancreas is a human organ located in the retroperitoneum of the abdominal cavity, oriented in a transverse direction adjacent to the loop of the duodenum and extending to the hilum of the spleen. On average, the pancreas measures 20cm in length and weighs 85-90 grams. The vasculature adjacent to the pancreas allows anatomical separation of the organ into four parts: the head, neck, body and tail (1). The pancreas is predominately an exocrine organ which produces digestive enzymes (and precursors) such as trypsinogen, lipases, nucleases and amylases comprising 80-85% of the pancreas. This function is performed by acinar cells which secrete bicarbonate-rich fluid and enzymes into ductules and ducts that exit into a highly variable system. The main pancreatic duct (of Wirsung) most commonly drains into the duodenum at the papilla of Vater whilst an accessory pancreatic duct may occur and drains via a separate minor papilla proximal (see FIGURE 1). In most, there is a common channel termed the Ampulla of Vater where the main pancreatic duct joins the common bile duct (CBD) draining the liver.

Infiltrating pancreatic ductal adenocarcinoma (PDAC) is responsible for the majority of malignant tumours that occur in this organ (2) and originate from the exocrine part i.e. the ductal epithelium histologically. Other cancers from the exocrine system are termed Acinar cell carcinomas which can occur in up to 5% of cases identified as pancreatic cancer (3). The endocrine portion within the pancreas is composed of clusters of cells termed the Islets of Langerhans. Cancers of the endocrine cells are termed neuroendocrine tumours (NETs) and have a differing molecular profile and prognosis.

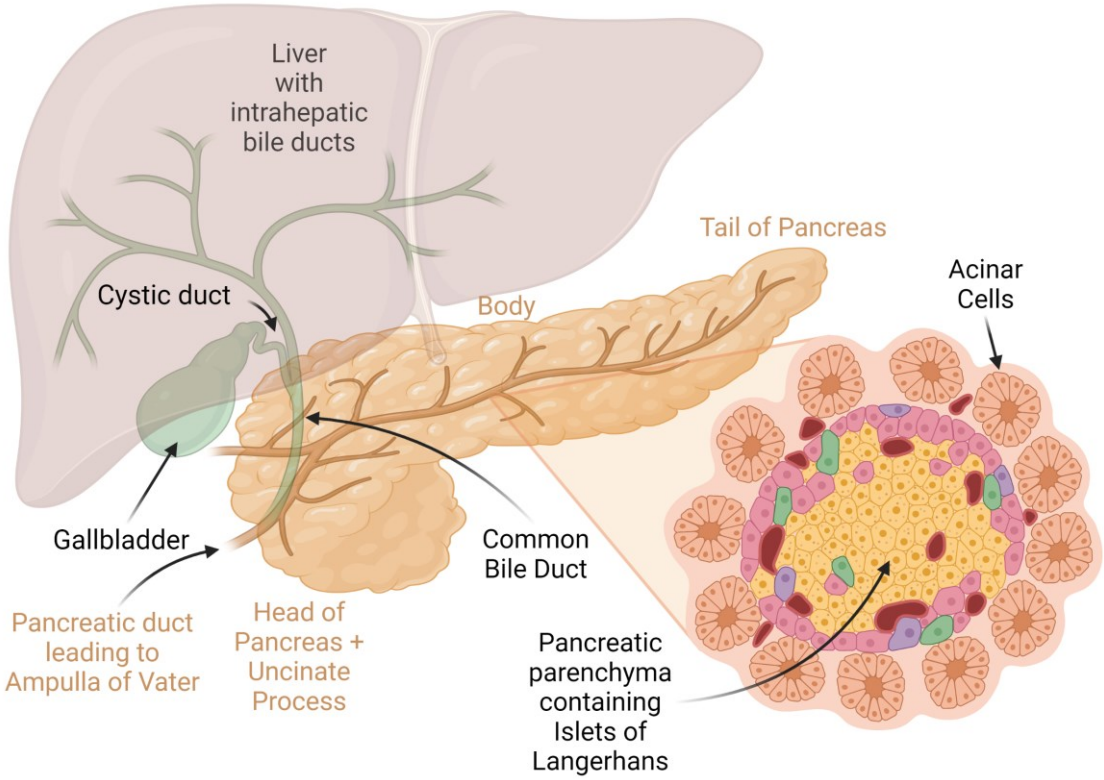


Figure 1: Pancreatic and biliary tract anatomy.
The pancreas is a retroperitoneal organ which is subdivided into three parts, the head, body, and tail. These drain into the duodenum via the main pancreatic duct (labelled) whilst an accessory pancreatic duct may occur and drains via a separate minor papilla. Here we show the drainage of bile from the liver via intrahepatic ducts (labelled above) into the common bile duct and joining the main pancreatic duct as at the ampulla of Vater. The exocrine component of the pancreas is made up of densely packed acinar glands which represent the secretory units of the pancreas made up of acinar cells surrounding a small central lumen secreting enzymes from their apical surface. The parenchyma of the pancreas is made up of mostly stroma divided by interlobular connective tissue and housing blood vessels and nerves. Within the pancreas are islets which contain the endocrine cells producing insulin, glucagon and others (somatostatin and ghrelin).

PDAC is a lethal disease with an extremely poor prognosis (5-year survival rate ~6%) (4) as 7 out of 10 patients are unsuitable for active treatment (either surgery or chemotherapy) due to late presentation with invasive and/or metastatic disease. Around 10,000 people are diagnosed every year with ~7800 deaths attributed to PDAC in the UK (4). In 2012, it was found that across the EU, pancreatic cancer was the 4th most common cause of cancer death (5). This has not changed significantly despite medical advances and a global surveillance study of cancer survival for patients across 59 countries presenting with one of the 18 most common malignancies (CONCORD-3; representing 75% of the global cancer burden) showed that survival rates for pancreatic cancer had improved by only 4-5% in the past 20 years. This included the UK, where survival rates are lower than many other countries, in Europe and the wider world.

1.1.2 Risk factors for PDAC

The reported lifetime risk for pancreatic cancer is 1 in 53 for males, and 1 in 57 for females born after 1960 in the UK (4). PDAC is primarily a disease of the elderly with 80% of cases occurring in those over 60. There is a slightly increased risk for the male gender, as well as in certain ethnic groups (Ashkenazi Jewish descent for example) (6) and those with A/B/AB blood groups (7).

Modifiable risk factors include cigarette smoking which is believed to double the risk of pancreatic cancer and is estimated to cause 25% of all cases of PDAC (8, 9). Obesity has also been associated with increased mortality from cancer of the pancreas with a risk of pancreatic cancer among obese (those with a body mass index (BMI) ≥ 30) compared with those of normal body mass index (< 25) doubled (10). A UK burden study, published in 2011, estimated that around 12.8% of pancreatic cancers in men and 11.5% in women could be attributed to this growing problem (11). Jiao *et al.* (12) additionally showed that low vegetable and fruit intake (non-Mediterranean diet), and physical inactivity also contribute to an increased risk for this disease. Diabetes mellitus can co-exist with obesity and hypertension as what is termed Metabolic Syndrome and adds additional risk for PDAC (13). Both type 1 and type 2 diabetes is reported in the literature to double the risk of pancreatic cancer (14) with new-onset diabetes in particular, a specific sign of pancreatic dysfunction and possible PDAC.

Pancreatitis is an important pre-disposing disease which increases the risk of PDAC. Chronic pancreatitis (CP) in particular, involves the irreversible destruction of exocrine parenchyma associated with fibrosis and in the late stages, endocrine function. Alcohol and gallstones are common causes of acute recurrent pancreatitis and likely to increase the risk of progression to pancreatic dysfunction, CP and subsequent PDAC (15, 16). CP is associated with up to a 11-fold elevated risk of PDAC in the first year following a diagnosis, which interestingly, decreases as the time interval goes on (6-fold after 2 years)(17). Although a clear causal role has not been identified, it is thought that recurrent inflammation leads to pancreatic stellate cell activation and carcinogenesis. This is evidenced by the increased risk in individuals genetically predisposed to recurrent pancreatitis such as those with cystic fibrosis (18). Patients with hereditary forms of CP (autosomal dominant PRSS1 gene mutation for example) also have an increased risk of PDAC (17, 19, 20).

1.1.3 The genetic and molecular basis for PDAC.

The molecular model for the development of PDAC can be understood as a histological progression from non-neoplastic normal epithelium through well-defined non-invasive lesions in ducts which eventually lead to invasive carcinoma. This classic model of progression can be identified by precursor lesions termed PanINs (pancreatic intraepithelial neoplasms; FIGURE 2) which can be found adjacent to carcinomas and show comparative molecular alterations leading to cancer (21). Other precursor lesions are also recognised such as intraductal papillary mucinous neoplasms (IPMN), and mucinous cystic neoplasms (MCNs) but these are less well defined.

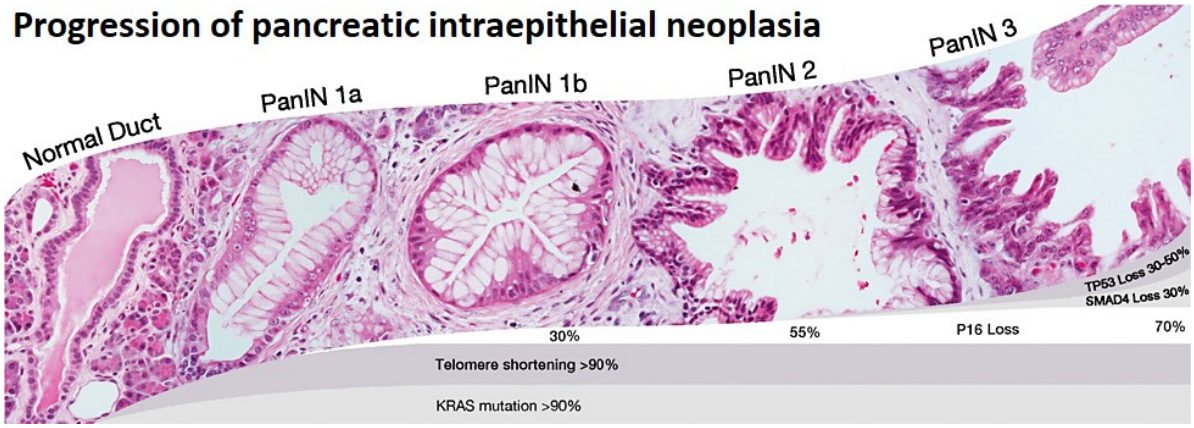


Figure 2: Progression from early to late PanIN.
Pancreatic intraepithelial neoplasms (PanINs) are a precursor lesion to PDAC and the classic model is typified by increasing degree of cytologic and architectural atypia. PanIN1 lesions can be subdivided into flat (PanIN1A) and papillary types (PanIN1B). This is followed by loss of polarity, nuclear crowding, cell enlargement, and hyperchromasia present in PanIN2s with advanced PanIN3 lesions showing luminal necrosis and manifest epithelial cell budding into the lumen of ducts. Progressive accumulation of genetic mutations from low to high grade has been identified leading to the eventual development of PDAC(22). This figure was obtained from Hackeng *et al.* (23) and was distributed under the terms of the Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/)

The alterations of multiple genes are distinct from other gastrointestinal (GI) malignancies with the average number of somatic mutations in pancreatic cancers considerably less than that found in breast or colorectal cancers (24). The four main ones can be considered KRAS, CDKN2A, SMAD4 and TP53 and are summarised below.

Around 90% of all pancreatic cancers harbour mutations in the KRAS oncogene (chromosome 12p) which has also been found in patients with CP, indicating a potential

common disease pathway (25, 26). In fact, expression of KRAS oncogenes appears to be essential for PanIN formation as well as progression to PDAC as evidenced by genetically-engineered mouse models (27). Cancer-associated RAS genes have a point mutation that encode single amino acid substitutions primarily (98%) at one of three so-called hot spots: glycine-12 (G12), glycine-13 (G13), or glutamine-61 (Q61). The single-nucleotide mutation altering a glycine GGT sequence to an aspartic acid GAT sequence (G12D) is the most common, occurring in up to 50% of KRAS mutations and indicating worse prognosis (28). This impairs the intrinsic GTP activity of the K-Ras protein resulting in a protein that is constitutively active and leads to several downstream effector signalling pathways to drive cancer growth (29).

Inactivation of the p16/CDKN2A (cyclin dependent kinase inhibitor 2A) gene on chromosome 9p also occurs in over 90% of PDAC making it another frequent genetic alteration. The CDKN2A is associated with over 10 transcript variants, of which the largest two encode for $p16^{INK4A}$ and $p14^{ARF}$ with variants affecting $p16^{INK4A}$ coding transcript more frequently observed with pancreatic cancer. 40% of pancreatic cancers occur through homozygous deletion, whilst in 40% there is an intragenic mutation coupled with loss of the second allele, and in 15% the loss is through hypermethylation of the $p16^{INK4A}$/CDKN2A gene promoter. Germline alterations in CDKN2A are most frequently associated with predisposition to melanoma and pancreatic cancer which if a family history is suspected, are an indication for clinical genetic testing (30).

DPC4 (deleted in pancreatic cancer), also known as SMAD4/MADH4, is a tumour suppressor gene on chromosome 18q which has been found to be inactivated in 55% of PDAC and encodes a protein that plays a role in signalling through the transforming growth factor type ß (TGF-ß) pathway. This results in the nuclear localisation of the Smad4 protein and growth controlling effects by regulating the expression of specific target genes. It is a relatively unique mutation in PDAC, with loss of SMAD4 only occurring biologically late in the neoplastic progression to PanIN-3 (31), and rarely in IPMN progression (3%).

P53 (chromosome 17p) is the prototype tumour suppressor gene, and is found inactivated in 50-70% of PDAC cases. This well recognised protein is a nuclear DNA-binding protein that is stimulated by DNA damage (32) and found with multiple single nucleotide polymorphisms (SNPs) some of which are correlated with an increased risk of pancreatic cancer as well as other types of cancer (33). It is the most frequently mutated gene in human cancer and acts as a cell cycle checkpoint to either trigger DNA repair

mechanisms, arrest growth at the G1/S regulation point, signal senescence or induce apoptosis. The primary effector of cell cycle arrest is through a protein p21 also known as cyclin-dependent kinase inhibitor 1. Loss of function leads to genomic instability and promotion of epithelial-to-mesenchymal transition (EMT), a key pathway in physiological development and wound healing, but exploited in cancer to promote their motility and metastasis (34). Additionally, gain-of-function mutations have also been described leading to increased proliferation in pancreatic cells and chemoresistance in PDAC (35).

EMT is an important process in embryogenesis, which involves disruption of epithelial cell homeostasis through the loss of membrane E-cadherin and the acquisition of a phenotype which can mobilise and migrate. This is controlled by canonical pathways such as WNT and transforming growth factor ß (36). Cancer progression relies on this event to proliferate and metastasise, and this is facilitated by the reversal and reacquisition of epithelial features in a process called Mesenchymal to Epithelial Transition (MET). This dynamic process is collectively referred to as epithelial to mesenchymal plasticity and the underlying quality which can be termed "stemness". It is hypothesised, although not proven, that cancers may originate from cancer stem cells (CSCs) with these properties leading to a hierarchical development of differentiation (37, 38). Three signalling pathways have been well-documented in the literature for the activation of pancreatic CSCs, namely the Notch, Wnt/ß-catenin and Hedgehog pathways. These all contribute to self-renewal and treatment resistance in PDAC (39) as well as CCA(40).

Other genes relevant to PDAC are mutated in <20% of PDACs, include oncogenes like BRAF, BRCA2, GATA-6, MYB, AKT2 and EGFR; and tumour suppressors such as ACVR1B, ACVR2A, MAP2K4/MKK4, STK11, FBXW7 and EP300 (41).

Use of whole genome sequencing has allowed us to gain insights into the complex nature of PDAC. Whilst 12 common pathways and processes were highlighted in the great majority of PDAC by Jones *et al.* (24) the heterogeneity by which these processes occur is huge. Analysis of primary tumours by Yachida *et al.* (42) identified the presence of clonal populations that give rise to distant metastases are already represented within a primary carcinoma and genetically evolved from the original parental, non-metastatic clone. They estimated by modelling the accumulation of mutations, the development of phenotypic metastatic disease takes approximately 6.8 years to occur. This window of opportunity is likely to be a period during which the patient is asymptomatic and therefore a potential area of research for non-invasive screening.

Three studies have set the scene for our molecular understanding of PDAC (see FIGURE 3). Landmark work undertaken by Collisson *et al.* used array-based mRNA expression of 27 primary epithelial archival samples micro-dissected from stroma to define three subtypes of PDAC. These were termed classical, quasi-mesenchymal and exocrine-like (43). Whilst GATA6 was recognised as a driver for the progenitor/classical subtype (44), KDM6A loss, p53 mutations and upregulation of ΔNp63/YAP1/GLI2 were associated with the squamous/basal-like subtype (45). This work was repeated by Moffitt *et al.*, who profiled over 200 bulk tumours (145 primary PDAC, 61 metastatic as well as 46 pancreas and 88 distant sites) by RNA-Seq to define two tumour subtypes (basal-like and classical) as well as stromal subtypes (normal and activated) (46). The following year, Bailey *et al.* expanded this knowledge further (47), confirming the quasi-mesenchymal/basal-like to be associated with genes highly expressed in the C2-squamous-like class of tumours of breast, bladder, lung and head and neck cancer and identified in the TCGA pan-cancer study from 2014 (48). Further analysis of the pancreatic progenitor (classical) subtype contained transcription factors PDX1, MNX1, HNF4G, HNF4A, HNF1B, HNF1A, FOXA2, FOXA3, HES1, and apomucins MUC5AC and MUC1 (which defined the IPMN-associated PDAC). Finally, the ADEX (Aberrantly Differentiated Endocrine eXocrine) subtype was defined through expression of transcriptional networks that characterised both endocrine and exocrine lineages at a late stage (rather than only one). These overlapped well with the Collisson *et al.* subtypes first mentioned, and has been confirmed by the International Cancer Genome Consortium (ICGC) study (49) with progenitor similar to classical, squamous similar to the quasi-mesenchymal and, ADEX similar to the exocrine-like subtype. Classification and prognostication of these subtypes remains an ongoing area of research but there is evidence that the subtypes do correlate with survival (classical-type leading to best prognosis)(50) (44, 51).

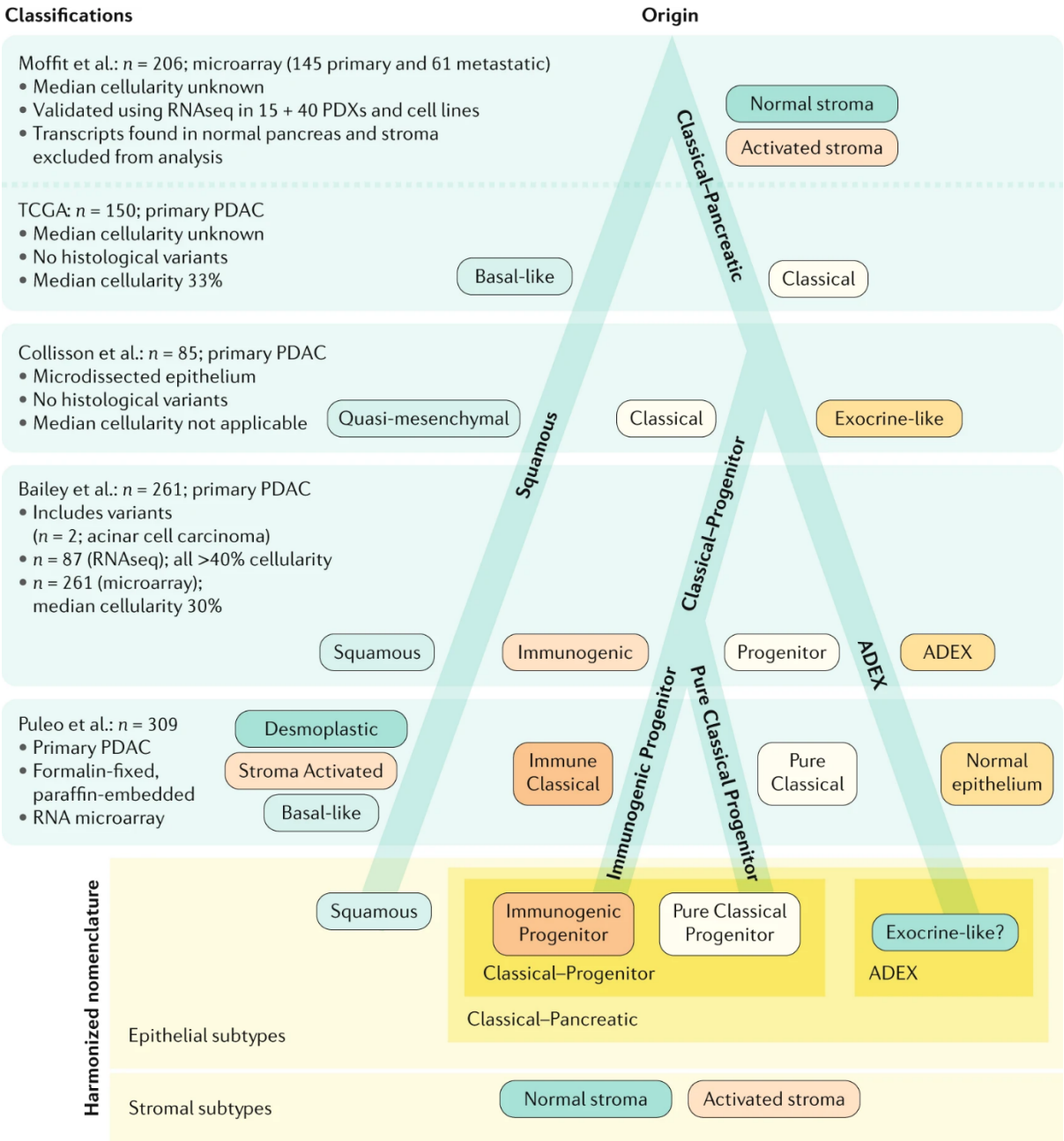


Figure 3: Phylotranscriptomic tree of pancreatic cancer.
A molecular taxonomy of PDAC based on genomic and transcriptomic data as neatly summarised by Collisson et al.(52). This summarises the recent literature in chronological order as the originally defined basal-like and classical appearances first described by Moffit *et al.* have been adapted to take into account new data from RNA sequencing and microarray studies. Reproduced with permission from Springer Nature. PDAC – pancreatic ductal adenocarcinoma, PDX - patient-derived xenograft, RNAseq - RNA sequencing, TCGA - The Cancer Genome Atlas.

1.1.4 The role of pancreatic stroma in PDAC

The stromal response to PDAC (also termed desmoplasia) is important because as much as 90% of the total tumour volume is made up of this mixture of fibroblasts, lymphocytes and extracellular matrix. PDAC is often referred to as a "cold" tumour, characterized by sparse T cell infiltrates, which may be the result of an immunosuppressive PDAC tumour

microenvironment (TME) which is dominated with myeloid cells, including TAMs (tumour-associated macrophages), granulocytes and inflammatory monocytes, actively recruited by the release of cytokines (53). Microdissection of the stromal component has also been profiled (54) but remains difficult with co-expression of epithelial genes. Activated stroma has a negative impact on prognosis (55) whilst depletion of myofibroblasts in a murine model also leads to accelerated progression (56), which means that it is far from a passive bystander in PDAC. Stroma acts as a physical barrier to efficient drug delivery and may be a niche for treatment-refractory cell populations such as the earlier mentioned CSCs (39). It is also worth mentioning that the tumour microbiome present in the TME is also of recent interest and has been shown to be associated with patient outcomes (57).

1.2 Bile duct cancer

1.2.1 Overview.

Hepatocytes (liver cells) are polyhedral in shape and form grooves along which bile, a secretion formed of water and the breakdown products of haemoglobin, drains. These form bile ductules (canals of Hering) which drain into ducts lined by cuboidal epithelium and termed cholangiocytes. Cholangiocarcinoma (CCA) is a tumour arising from these cells that line the bile duct and whilst proximally, the bile duct is histologically related to hepatocytes, distally the cell lineages are more divergent (58). Anatomically, CCA is divided into intrahepatic (iCCA) disease which is denoted as proximal to the confluence of the second order hepatic bile ducts, perihilar (pCCA) disease which involves the right and/or left hepatic ducts down to the cystic duct (known as klatskin tumour) and the rest as distal (dCCA) (59). Perihilar CCA makes up the majority (50-60%) of cases with 20-30% dCCA and 10% iCCA. Distal CCA may occur anywhere down to the Ampulla of Vater and are clinically indistinguishable from adenocarcinoma of the duodenal mucosa and PDAC.

UK-wide statistics have been complicated by changes in classification and by prior inclusion with other forms of liver cancer (in particular hepatocellular carcinoma; HCC) before 2010 but remains a rare cancer overall (affecting fewer than 6 per 100,000 people in the population) (60). Public Health England's National Cancer Intelligence Network

(NCIN) figures give the incidence in England as ~2000 cases per year of CCA whilst causing 2161 deaths in 2013 (60). Figures taken from 2012 has one-year survival for those diagnosed as approximately 20-30%, which is significantly lower than for other solid tumours.

1.2.2 Risk factors for CCA.

Lifetime risk for CCA is difficult to determine in the UK because of recent co-inclusion with liver cancer but can be estimated as <1% making it a rare cancer (61).

Worldwide, there is a higher incidence in Asian countries (due to the presence of hepatobiliary flukes *Opisthorchis viverrini* and *Clonorchis sinensis* which is endemic to the Mekong basin) with an incidence of iCCA of 113 per 100,000 in Thailand, in comparison to 0.1 per 100,000 in Australia. Other risk factors include Primary Sclerosing Cholangitis (PSC), cholelithiasis (62), exposure to chemicals such as asbestos and chronic liver disease processes such as cirrhosis or hepatitis (63).

Diseases which are associated with the development of CCA include congenital bile duct cysts (more common in females of Asian countries, especially China and Japan) (64), Caroli's disease and most importantly Primary sclerosing cholangitis (PSC) which occurs at a rate of 0.91 to 1.3 per 100 000 person-years, can be associated with IBD (inflammatory bowel disease) and more common in those of northern European descent (62).

1.2.3 The genetic and molecular basis for CCA.

Genetic mutational drivers have not been as well characterised as in PDAC as for CCA, however it appears that intrahepatic (iCCA) and extrahepatic CCA (eCCA) are both clinically and genetically distinct entities (65). Common to both are TP53, KRAS, SMAD4 mutations which also occur albeit at a lower rate than in PDAC (66). CCA tumours have been found to contain >30 genetic mutations on average, with a median 39 non-synonymous mutations per tumour in iCCA and a median 35 non-synonymous mutations per tumour for eCCA.

Mutations in the metabolic gene IDH1/2 have been found in a subset of patients with CCA with a predominance for intrahepatic disease. This is a mutation more commonly found

in glioma/glioblastomas but can be found in ~20% of iCCA tumours (67). A meta-analysis of studies looking at genetic alterations such as chromosomal imbalances within the genetic make-up of iCCA, have highlighted some similarities with HCC but further studies are warranted in this area (66).

Initially, it has been suggested that biliary duct cancers and PDAC share characteristics with intraductal papillary neoplasm of the bile duct (IPNB) and intraductal papillary mucinous neoplasm of the pancreas (IPMN) showing histological similarities (68, 69). However, molecular subtyping of CCA has revealed differences which may explain the differing responses to treatment in each disease. KRAS mutations are more prevalent in eCCA (42%) compared with iCCA (and gallbladder cancer) but is less frequent than in PDAC (70). Alterations/overexpression in ERBB2 (also known as HER2/neu) have also been noted to a greater degree in eCCA tumours which may also point towards the potential of anti-HER2 chemotherapy (71) and is relatively rare in PDAC. Extrahepatic tumours specifically show PRKACA and PRKACB fusions as well as mutations in ELF3 and ARID1B56 (72).

1.3 PDAC and CCA: Clinical Picture

1.3.1 The diagnosis and evaluation of patients with suspected PDAC or CCA.

History and examination provide the first indication that a patient may have PDAC or CCA as there is currently no available suitable screening test that can be applied safely to the wider population. In addition to investigation for any of the risk factors mentioned earlier, clinical evaluation of the abdomen may be prompted by symptoms such as diarrhoea, indigestion and abdominal distension. Unfortunately, patient reported symptoms are often non-specific, and symptoms such as weight loss, back pain, and malaise may lead the clinician away from the diagnosis of PDAC or CCA.

Additional “red flag” symptoms include jaundice, a recent diagnosis of diabetes as well as thirst and dark urine (73). It is reported that up to a fifth of patients over 60 presenting with jaundice were likely to be found to have PDAC but it is an unfortunate reality that by the time this is a detectable symptom, the disease is likely to be advanced (74). Detection rates from clinical evaluation alone is poor with a study from 2012 evaluating a clinical algorithm for the presence of risk factors and symptoms correlating with PDAC

using clinical history alone. This was validated in an independent population and at its best had a sensitivity of 62.4% (75). The SYMPTOM pancreatic study by Walter et al demonstrated the difficulties faced in early diagnosis, prospectively analysing 343 new cases of pancreatic cancer presenting to hospital having been referred from primary care to determine symptom patterns in those who were diagnosed (76). They confirmed that patients had reported multiple, often intermittent, upper gastrointestinal symptoms to their GP such as indigestion and decreased appetite prior to referral which led to a longer diagnostic waiting period.

When diagnosis is eventually suspected, computed tomography (CT) imaging is the imaging modality of choice due to speed and availability (77). In order to detect PDAC it is best performed utilising a biphasic pancreatic protocol i.e. pancreatic arterial (40–50 seconds after injection of contrast) and portal venous (65–70 seconds after contrast) phase. The reported sensitivity of CT for the detection of PDAC is 77% with a reported specificity of 63% (78). A Cochrane Review by Best *et al.* showed that CT can be sensitive in distinguishing invasive cancer lesions from benign pathology (Pooled sensitivity and specificity; 98% (95% CI 0-100%) and 76% (95% CI 2-100%)) (79) whilst Magnetic Resonance Imaging (MRI) was found to be broadly similar with sensitivity and specificity for diagnosing cancer as 80% (95% CI 58-92%) and 89% (95% CI 57-98%) respectively. MRI may also be beneficial in the detection of hepatic lesions that cannot be characterised by CT whilst MRI in the form of an MRCP (magnetic resonance cholangiopancreatography) may be preferable for cystic neoplasms of the pancreas and to evaluate biliary anatomy (77). Both modalities were found to have low sensitivity in the detection of early disease or dysplasia and other pre-malignant lesions from overt cancer and pooled figures for this show a sensitivity of 72% (95% CI 50- 87%) for CT and 75% (95% CI 30-95%) for MRI with few good quality studies identified by the review. This means that imaging alone yields poor diagnostic value, particularly in cases where there is co-existing disease and/or early small lesions.

Abnormalities of the GI tract may be imaged directly by Oesophago-gastric duodenoscopy (OGD) which when directed at the bile ducts with a side-viewing scope is termed an endoscopic retrograde cholangiopancreatography (ERCP). By doing so, it is possible to undertake a direct cholangiogram of the bile ducts by injecting radio-opaque contrast dye

into the biliary system and evaluating this radiologically. This yields a much higher degree of specificity, but with a lower sensitivity depending on the lesion size and location. Endoscopic techniques such as ERCP and endoscopic ultrasound (EUS) also allow for tissue to be obtained. Cellular samples of strictures (narrowing) of the biliary tree can be obtained through biliary brushings which provide cytological samples (no tissue architecture preserved). However these samples are often poor and a meta-analysis from 2013 by Burnett *et al.* showed that standard cytology from biliary brushings was only around 42% for the detection of cancer with a negative predictive value of 58.0% ± 3.2% (80). Specimens obtained through paediatric-sized forceps to biopsy the stricture can obtain tissue but with a greater risk of perforation. Forceps guided radiologically within the duct are best suited to intraluminal lesions such as in CCA (Intraductal Forceps Biopsy; IDFB). These may improve cellular yield, but sampling may not yield diagnostic material unless the bile duct is infiltrated by cancer (potentially locally advanced). Navaneethan *et al.* undertook a meta-analysis to compare biliary brushing with IDFB on studies up to April 2014 and found that IDFB for the diagnosis of malignant biliary strictures only yielded a slight improvement (48.1% vs 45%) in sensitivity. It recommended combining the two, increasing sensitivity to 59.4% and specificity to 100% to improve diagnosis (81).

Endoscopic ultrasound (EUS) can be used as a diagnostic adjunct to assess solid or cystic lesions directly and biopsy using a fine needle aspiration (FNA) for cytological assessment of the tumour. EUS-FNA is considered to have better sensitivity with De Moura *et al.* estimating 75% (82) whilst Sadeghi et al calculated a pooled sensitivity and specificity of 80% (95% CI 74-86%), and 97% (95% CI, 94-99%) respectively (83). In all these modalities (ERCP and EUS), specificity remains close to 100%. Additionally, EUS holds up as a diagnostic tool for solid neoplasms of the pancreas, with meta-analysis demonstrating that EUS-FNA had a pooled sensitivity for the detection of cancer of 85% (95% CI 84-86%), and a pooled specificity of 98% (95% CI, 97-99%) (84). However, neither EUS-FNA nor ERCP can reliably exclude cancer with a pooled negative likelihood ratio of 26% for EUS-FNA (95% CI, 18-38%) (83) and 54% for ERCP-guided brush and IDFB combined. Furthermore, there are some reservations about the use of EUS-FNA where transperitoneal seeding of malignancy has been reported and may ultimately risk curative resectional surgery (85). This creates diagnostic uncertainty when presented with obvious imaging abnormalities and tissue diagnosis has not been obtained.

As a result, around 10% of patients still undergo surgical resection for suspected malignancy only to have benign pathology found on subsequent histological examination (86). This is because even with a high index of clinical suspicion (i.e. symptoms of weight loss, jaundice, biochemical parameters such as a raised CA19-9 and a mass lesion on imaging), there is no current test that can be trusted to rule out a potential malignant cause (i.e. a high negative predictive power). ERCP is usually the best route to obtain tissue/cytology as well as relieving biliary obstruction so remains the mainstay in UK practice for the evaluation of potential malignant strictures (87).

1.3.2 Blood-based markers for the evaluation and diagnosis of PDAC/CCA.

Carbohydrate antigen (CA) 19-9 was first identified in 1981 in the serum of patients with colorectal and pancreatic cancer (88). Since that time however, there have been few advances in validated biomarkers for PDAC or CCA. Studies have shown that a serum CA 19-9 level of more than 100 U/mL has a sensitivity of 53% and a specificity of 92% for the diagnosis of malignancy in biliary strictures (89) whilst in PDAC it has a median sensitivity of 79% (70–90%) for the diagnosis of pancreatic cancer and median specificity of 82 (68–91%) (90). As a serum level perioperatively, it can be an informative marker of long-term prognosis with normalisation of CA19-9 within 6 months of surgery a favourable prognostic factor independent of margin status (91). However, up to 10% (92) of the Caucasian population are negative for the Lewis protein antigen (FUT3 gene) or are non-secretors (FUT2 gene) which means that CA 19-9 levels remain undetectable in this cohort of patients (83). It is also not specific for PDAC or CCA, being raised in 21% to 42% of cases of gastrointestinal (GI) cancers (93) and additionally can be elevated in the setting of cholestasis (94).

CEA is a serum glycoprotein which is also used as a tumour marker in PDAC/CCA (95). It is primarily used in decision making during treatment of colorectal cancers but is also raised in other malignancies such as PDAC. Similarly, CA-125 is a mucin glycoprotein which can be raised in serum. It is primarily useful as it detects 80% of non-mucinous ovarian carcinomas but can also be useful as a panel together with CA 19-9 in improving the sensitivity to detect cancer (96, 97).

The recent movement away from primary endoscopic biliary drainage and towards fast-track surgery in the management of T1-2 disease has further highlighted the need for

prospective blood-based markers. A retrospective case-control study by Arkadopoulos *et al.* identified a reduction in morbidity in patients with resectable PDAC undergoing surgery that did not routinely undergo a pre-operative drainage procedure (98). This was due to the serious complications that are possible during ERCP (such as haemorrhage, pancreatitis, cholangitis, and perforation of the duodenal wall) and may occur in about 1–25% of procedures. More commonly, the failure rate of ERCP or PTC to achieve successful drainage is reported as ~10% necessitating further procedures and risk of potential complications (99). A retrospective study undertaken by Pande *et al.* showed that the reduction in time to surgery could provide an additional survival benefit to those with resectable PDAC (100). The focus of this project therefore is on markers that will be useful at the time of presentation, or in primary care, improving the probability of a positive or negative diagnosis and leading to an early decision for those that would benefit from surgery.

1.3.3 The Staging and Treatment of PDAC and CCA.

Staging is the first step in determining treatment and prognosis for patients with PDAC and CCA. The primary tumour (T) size, nodal (N) involvement, and metastasis (M) staging of tumours gives a clear outline for each cancer and determining extent of disease. Both the American Joint Committee on Cancer (AJCC) and the Union for International Cancer Control (UICC) are updated regularly, and the latest editions (8th) have reflected recent evidence about predicting disease outcomes.

The 8th edition updated its T classifications in the staging of PDAC when released in 2017 to reflect staging based on size alone rather than the more difficult determination of extra-pancreatic extension in the presence of inflammation or desmoplasia as the defining characteristic of a T3 tumour. TABLE 1 shows the updated staging criteria. The nodal classification has also been changed, after it was identified in a retrospective study by Malleo *et al.* (101) that absolute number of positive lymph nodes was more informative of patient prognosis than lymph node ratios.

Table 1: PDAC Staging.

Staging criteria for pancreatic cancer according to the latest (8th) edition published in 2017 and is used clinically to determine treatment outcomes and measure progression. Tumour size determines the T stage whilst number of nodes (regionally i.e. in the area that is to be resected) determines N0,1 or 2. Metastatic disease is a binary outcome which has the poorest prognosis (Stage 4 disease). AJCC – American Joint Committee on Cancer, CA – Coeliac artery, LN – lymph node, M – metastasis, N – node, SMA – Superior Mesenteric Artery, T – tumour.

AJCC staging system for pancreatic adenocarcinoma (8th edition)

Code	Criteria
T1	Maximum tumor diameter ≤2 cm.
T2	Maximum tumor diameter >2 cm, but ≤4 cm.
T3	Maximum tumor diameter >4 cm.
T4	Tumor involves the CA or the SMA (unresectable primary tumor)
N0	No regional LN metastasis.
N1	Metastasis in 1–3 regional lymph nodes.
N2	Metastasis in ≥4 regional lymph nodes.
M0	No distant metastasis.
M1	Distant metastasis.

Code	T	N	M
Stage IA	T1	N0	M0
Stage IB	T2	N0	M0
Stage IIA	T3	N0	M0
Stage IIB	T1–T3	N1	M0
Stage 3	T4	Any N	M0
Stage 4	Any T	Any N	M1

One of the difficulties in staging and decision making in PDAC is the determination of resectability based on encasement of the adjacent vessels (coeliac artery, superior mesenteric artery (SMA), common hepatic artery, portal vein (PV), or superior mesenteric vein (SMV)). Circumferential contact with a vessel >180° or deformity in a vessel contour on CT imaging would be considered local invasion and comes under the category of borderline resectable cancer. There is considerable inter-observer variability in the assessment of resectability at CT and a centre may have additional criteria for what can be safely resected and reconstructed given local experience (102).

Table 2: Perihilar CCA Staging.

Staging criteria for perihilar cholangiocarcinoma according to the latest (8th) edition published in 2017 and is used clinically to determine treatment outcomes and measure progression. With cholangiocarcinoma difficult to measure by conventional imaging, tumour size is less relevant than location and so invasion into the liver determines the T stage. Location of metastatic nodes (regionally i.e. in the area that is to be resected) determines N0,1 or 2 whilst once again metastatic disease is a binary outcome which has the poorest prognosis (Stage 4 disease). CHA – Common hepatic artery, HA – Hepatic artery, LN – lymph node, M – metastasis, N – node, PV – Portal Vein, SMA – Superior Mesenteric Artery, T – tumour.

AJCC staging system for perihilar cholangiocarcinoma (8th edition)

Code	Criteria
T1	Confined to bile duct.
T2a	Invasion into surrounding adipose tissue.
T2b	Invasion into adjacent liver parenchyma.
T3	Invasion into branches of PV or HA.
T4	Invasion into main PV or CHA or second-order bile ducts bilaterally, or unilateral bile ducts with contralateral PV or CHA.
N0	No regional LN metastasis.
N1	Metastasis in regional lymph nodes.
N2	Metastasis to periaortic/pericaval/SMA/coeliac nodes
M0	No distant metastasis.
M1	Distant metastasis.

Code	T	N	M
Stage 0	Tis	N0	M0
Stage I	T1	N0	M0
Stage II	T2a-b	N0	M0
Stage 3A	T3	N0	M0
Stage 3B	T4	N0	M0
Stage 3C	Any T	N1	M0
Stage 4a	Any T	N2	M0
Stage 4b	Any T	Any N	M1

Table 3: Extrahepatic CCA Staging.

Staging criteria for extrahepatic cholangiocarcinoma according to the latest (8th) edition published in 2017 and is used clinically to determine treatment outcomes and measure progression. Extrahepatic cholangiocarcinoma is tumour staged according to depth of invasion through the bile duct. Location of metastatic nodes (regionally i.e. in the area that is to be resected) determines N0,1 or 2 whilst once again metastatic disease is a binary outcome which has the poorest prognosis (Stage 4 disease).CA – Coeliac artery, LN – lymph node, M – metastasis, N – node, SMA – Superior Mesenteric Artery, T – tumour.

AJCC staging system for extrahepatic cholangiocarcinoma (8th edition)

Code	Criteria
T1	Depth of invasion <5mm.
T2	Depth of invasion 5-12 mm.
T3	Depth of invasion >12 mm.
T4	Tumour involves the CA or the SMA (unresectable primary tumor).
N0	No regional LN metastasis.
N1	Metastasis in 1–3 regional lymph nodes.
N2	Metastasis in ≥4 regional lymph nodes.
M0	No distant metastasis.
M1	Distant metastasis.

Code	T	N	M
Stage 0	Tis	N0	M0
Stage I	T1	N0	M0
Stage IIA	T2	N0	M0
	T1	N1	M0
Stage IIB	T3	N0	M0
	T2-3	N1	M0
Stage 3A	T1-3	N2	M0
Stage 3B	T4	Any N	M0
Stage 4	Any T	Any N	M1

Staging for CCA depends on whether it is a pCCA or dCCA, with the American Joint Committee on Cancer (AJCC) TNM system being the most widely used. However, it also has limitations, with inter-clinician variability in determining T2/3 lesions on imaging alone as well as no additional criteria for multifocal disease (72).

1.3.4 Surgical Approaches to PDAC and CCA

For both PDAC and CCA, surgery remains the mainstay of curative treatment and this is indicated for patients with <T2/3 disease and no distant metastatic deposits. There are two surgical procedures that have the potential for a cure: the classical Kausch-Whipple operation (pancreaticoduodenectomy, PD) first described in 1939, where part of the pancreas, the gallbladder, the duodenum, the pylorus (outlet of the stomach) and the distal part of the stomach are removed, and the pylorus-preserving pancreaticoduodenectomy (PPPD), where the stomach including the pylorus are not resected with the specimen to preserve gastric function. They are applicable to PDAC and

distal CCA as they encompass a lymphadenectomy of the hepatoduodenal ligament, common hepatic artery (CHA), vena cava, inter aortocaval and right side of the superior mesenteric artery (SMA) with hepaticojejunostomy to restore bile duct continuity. No difference in survival has been identified between the two types of surgery (103). Borderline resectability may allow for surgical intervention on cancers locally invading branches of the hepatic artery or SMA, and technically there is the possibility of reconstruction on the proximal and distal margin of the veins (104). Vessel reconstruction consists of direct anastomosis, insertion of synthetic or autologous vein grafts as well as donor vessels. Avoidance of arterial resection has also been reported with the use of preoperative embolization of the CHA with the aim of collateral vessel formation in a Japanese 21 patient case series (105). However, survival was 6.6% for this group of patients at 5 years with the best prognostic factor serum CA 19-9.

After surgery, histopathology of the specimen is undertaken to evaluate tumour characteristics, resection margin and lymph node metastases. A case series by the Mayo Clinic published in 2008 showed that for those that are eligible for surgery, mean actual survival is 34 months (median 17 months) (106). Mean survival was 38 months (median 18 months) in those that were deemed to have a complete excision (R0) histologically compared with 24 months (median 15 months) for those with an incomplete excision.

1.3.5 Chemotherapy in the treatment of PDAC.

The first chemotherapy agent used successfully in the treatment of PDAC was 5-Fluorouracil (5-FU), a cytostatic antimetabolite. Introduced in 1957, 5-FU has several purported mechanisms of action, but principally functions as an inhibitor of thymidylate synthase(107). Incorporation of 5-FU leads to inhibition of RNA and protein synthesis. Since DNA, RNA and protein synthesis are essential for cell division and growth, 5-FU administration causes severe intracellular thymidine deficiency, provoking unbalanced cellular growth and triggering apoptosis. The European Study Group for Pancreatic Cancer 1 (ESPAC-1) trial in 2004 was the first study to clearly demonstrate a survival benefit in the adjuvant use of 5-FU monotherapy, this was compared to a combination of radiotherapy and 5-FU (chemoradiotherapy) which had a deleterious effect on overall survival (108).

Since then, gemcitabine (GEM), a pyrimidine analogue, and capecitabine (CAP), an oral precursor (prodrug) of 5-FU, are newer nucleoside analogues which have led to improvements in patient survival rates, both as first line therapy for those with advanced/metastatic PDAC (109) and as adjuvant therapy after surgical resection (ESPAC-4 trial) (110). Nab-paclitaxel, a liposomal formulation of paclitaxel, is a potent microtubule inhibitor which works synergistically when combined with GEM to increase the intra-tumoural concentration leading to a significant improvement in survival compared to GEM monotherapy (111, 112). The treatment with the highest clinical efficacy in PDAC, but also the highest potential toxicity, is a combination of 3 cytotoxic chemotherapy drugs, based on 5-FU combined with irinotecan and oxaliplatin named FOLFIRINOX. This is reserved for first-line treatment for patients with good performance status and advanced(unresectable) or metastatic PDAC. Modified doses have been used with similar effect for this regimen to reduce toxicity and patient side-effects. mFOLFIRINOX (for 24 weeks) is also used in adjuvant setting following R0/R1 resection and yields significantly longer survival than gemcitabine monotherapy with a median disease-free survival (DFS) of 21.4 months (113).

Recent advances in immunotherapy are now utilising our current understanding of the genomic landscape of PDAC, and account for the majority (37%) of clinical trials in PDAC (114). However, this increased understanding has not yet translated into new therapies, with many large drug trials of immunotherapy drugs proving disappointing in PDAC. Pembrolizumab, is a monoclonal antibody which targets PD-1, an immune checkpoint protein on T-cells. This is not specifically approved in Pancreatic Cancer, however does have a 'tumour agnostic' license in the US for cancer with microsatellite instability high (MSI-H) or high tumour mutational burden (TMB)(115, 116). However the response rate and outcomes are poor even for this exception, and MSI-H remains extremely rare for PDAC (< 1%) (117).

1.3.6 The Role of Neoadjuvant Chemotherapy in PDAC.

With the movement towards neoadjuvant chemotherapy for many GI cancers, recent investigations have focused on improving survival in patients that are deemed 'borderline resectable' in PDAC. The ESPAC-5 trial has published early results (December 2022)

investigating the clinical utility of short-course neoadjuvant chemotherapy in patients with borderline resectable PDAC. This multicentre, open label randomised controlled trial recruited 90 patients from 16 centres and have so far shown 1-year overall survival as over double (84% vs 39%) for patients receiving FOLFIRINOX or GEM-CAP vs immediate surgery. Locally advanced, or surgically unresectable, PDAC remains a difficult area to manage, with a recent rapid autopsy series reporting that 28% of patients with locally advanced pancreatic cancer had no evidence of metastases at the time of death.

The role of adjuvant radiotherapy appears to be still equivocal with the LAP07 (NCT0063472) and more recently the CONKO-007 (NCT01827553) trial both demonstrating no overall survival benefit (118, 119). In the UK the SCALOP-2 trial is ongoing and investigating if induction chemotherapy followed by capecitabine ± nelfinavir (a human immunodeficiency virus protease inhibitor with radio-sensitisation effects) with high- or standard-dose radiotherapy for locally advanced PDAC will change survival rates or the possibility for resection in this difficult disease group(120).

1.3.7 Chemotherapy in the treatment of CCA.

As with PDAC, surgical resection is the outcome for a minority of patients diagnosed with CCA (<30%) so for most patients who have unresectable/advanced CCA, combination chemotherapy with gemcitabine and a platinum-based agent (e.g., cisplatin) is regarded as a standard treatment for these patients based on the ABC-02 trial (121). With subsequent iterations of ABC trial showing minimal meaningful improvements, the ABC-06 clinical trial focused on 2nd line treatment following progression on GEM and cisplatin frontline treatment. The addition of 5-FU and Oxaliplatin (FOLFOX) to active symptom control improved median overall survival at 6 and 12 months compared to best supportive care (122). In Sept 2022, the FDA approved the use of durvulumab, an anti-PDL1 monoclonal antibody that functions in a similar manner to pembrolizumab (outlined previously). The results of the TOPAZ-1 trial showed that the addition of durvulumab to cisplatin and GEM chemotherapy improved overall survival by around 2 months which is clinically meaningful in the context of a median overall survival of 11.7 months (123). Despite the recent advances, evidence base has been slow for CCA with historically observation only as the standard of care for patients after surgical resection. This was altered after patients eligible for chemotherapy (performance status <2) were found to

have improved survival with CAP monotherapy in the BILCAP trial published in 2019 (121). For patients with locoregional disease, treatments to target the specific area include radiofrequency ablation, chemoembolization and radio-embolization may be considered but the data remains unclear to definitive benefit (124).

1.3.8 The Role of Neoadjuvant Chemotherapy in CCA.

With CCA, this area has been poorly investigated and to-date there have been no prospective trials investigating neoadjuvant chemotherapy and surgery remains the only curative treatment. One of the first retrospective studies, performed in 1997, was undertaken in a range of patients with resectable and unresectable disease showed that 3 out of 9 patients had a complete pathological response to neoadjuvant chemoradiotherapy with margin-negative resection (125). However, the mortality rate remained high despite the demonstration that chemotherapy could have significant anti-tumoural effects. Due to the low incidence of resectable CCA, the standard of care remains straight-to-surgery, whilst non-resectable patients who are down staged after chemotherapy (typically with a cisplatin and GEM-based regimen) are still likely to have poor outcomes. Recent retrospective evidence suggests however that this traditional notion should be challenged and that there may be a role for neoadjuvant chemotherapy in early stage CCA (126). Furthermore, with the new role for immunotherapy in the metastatic treatment of BTCs, trials investigating use of anti-PD-1 inhibitors are ongoing (NCT04308174).

1.4 The role of RNA biology in PDAC and CCA

1.4.1 MicroRNAs as a source of biomarkers for the detection of PDAC and CCA.

MicroRNAs (miRNAs) are small non-coding RNA molecules (~19-25-nt) that are distinct from longer messenger RNA molecules that have a well-established role in the genetic processes of transcription and translation. These are described as non-coding, in the sense that they do not lead to protein synthesis but play a post-transcriptional role in gene regulation. The first miRNA to be identified was lin-4 in 1993, in the model nematode *Caenorhabditis elegans* which was found to inhibit expression of the protein LIN-14 via an antisense RNA-RNA interaction (127). Despite its non-coding nature, it was found to be essential for development in the larval stage and was a conserved feature in this species. In 2000, the miRNA let-7 was subsequently identified and found also to be conserved in many other species, including the human genome (128). Since that time, a total of 1917 annotated hairpin precursors and 2656 mature miRNAs have been described (based on miRbase version 22 released in 2019) in an online publicly accessible repository for over 1493 small RNA deep sequencing datasets (129).

1.4.2 miRNA genes and nomenclature.

In miRNA nomenclature, miRNAs are generally named as miR- plus a number (e.g. miR-21). All miRNAs with a similar sequence, differing at only one or two nucleotide positions (not including the seed sequence), are given a letter suffix (e.g. miR-196a and 196b) whilst if the mature sequences are paralogs and identical, are given a number suffix. If two different mature miRNA sequences are derived from opposite arms of the same hairpin precursor, these mature sequences are named 5p for 5′ arm and 3p for the 3′ arm (e.g. miR-200a-3p). The original nomenclature system for miRNA sequences proposed by Ambros *et al.* (130) was designed to designate novel miRNAs in order of discovery but does not explain phylogeny and lineage within the genome. For example, miR-125b is the mammalian orthologue of cel-lin-4 (the first miRNA to be discovered) and occurs as two homologues in the genome coded from both chromosome 11q and 21q (127). Recognition of the genomic location from which these miRNAs are coded is relevant as 25% of miRNAs are co-localised within clusters and co-regulate multiple biological processes particularly in cancer (131).

1.4.3 miRNA Biogenesis.

The definition of what constitutes a miRNA is based on both size and the mechanism by which they are formed at a cellular level. This starts with transcription in the nucleus by RNA polymerase II to form a long primary miRNA (pri-miRNA) transcript that contains the relevant miRNA as one arm of an imperfectly aligned ~80-nt RNA hairpin (132). Thus pri-miRNA is cropped by the nuclear RNase III enzyme Drosha into what is termed a pre-miRNA which is ~65 nucleotides and contains a characteristic 2-3 nucleotide (132). Drosha does so as part of a "microprocessor complex" that contains a double-stranded RNA-binding protein named DGCR8 (DiGeorge syndrome critical region 8; also Pasha in non-human cells e.g. *C. elegans*) (133). The cleavage process is assisted by primary structural features such as presence of a basal UG motif, apical GUG motifs but also through a CNNC motif absent in nematodes and recognised by splicing factor SRp20 (SFSR3). This process is however incompletely understood as up to a fifth of pri-miRNAs lack these recognition elements (134). This Microprocessor complex is conserved throughout the animal kingdom and crucial for the overall abundance of miRNAs. Thus regulation has been investigated extensively with multiple post-translational modifications of Drosha and DGCR8 (e.g. phosphorylation by GSK3β, acetylation by p300, CBP, and GCN5 (135)) leading to localisation and stabilisation. It is furthermore self-regulated by negative feedback through a pri-miRNA-like hairpin structure found in the DGCR8 mRNA, which when cleaved reduces DGCR8 expression (136). Small microRNAs termed "mirtrons" have been described which are generated from introns after splicing and have the appropriate size and stem-loop structure to serve as a pre-miRNA. These bypass the Microprocessor as a Drosha-independent pathway and can be found in both flies and nematodes (137) as well as mammalian pathways (138).

Pre-miRNAs are identified and transported to the cytoplasm by an exportin-5/RanGTP complex (139) which is where the pathway merges with the route taken by exogenously administered RNAi. The hairpin structure of the cytoplasmic pre-miRNA is cleaved by a crucial second RNase III enzyme named Dicer (140) to form a "mature" duplex molecule (formed of two strands: sense and anti-sense). The -5p strand arises from the 5' end of the pre-miRNA hairpin whilst the -3p arises from the 3' end. Thermostability along with other factors lead to incorporation of one (or both) of the strands into what was termed the RNA-induced silencing complex (RISC). This has been found to be made up of several proteins including Dicer and Argonaute leading to post-transcriptional gene regulation by

repression, degradation or even epigenetic activity (141). The RISC is able to selectively modify RNA expression by binding and exposing bases 2-7 of the miRNA (the "seed" region) in the "guide" strand allowing target recognition (141).

Dicer however does not work alone in the preparation of dsRNA fragments and since its discovery in 2001 (142), there has been greater understanding of what is now termed the "RISC-loading complex". This involves TRBP (transactivating response RNA-binding protein), PACT (protein activator of PKR) and essentially Argonaute 2, an RNase which in some cases cleaves the passenger strand prior to the activity of Dicer (143). Some miRNAs, such as pre-miR-451 however are short (17nt) and are matured via a Dicer-independent pathway (144).

The final step in this canonical pathway is silencing either through direct mRNA cleavage (usually when the guide strand is fully complementary) or translational repression when imperfect matched to its target sequence, localised at cytoplasmic Processing-bodies (P-bodies). This target sequence is termed the miRNA response element (MRE) and is usually within the 3' untranslated region of coding mRNAs (145). The remaining passenger strand of the original duplex, dependent on stability, may be cleaved by AGO2 and degraded, leading to a bias in miRNA cellular expression. As a result, miRNA-dependent gene regulation can be seen as a multifactorial and complex picture, made up of a balance between transcription of a pri-miRNA, Drosha and Dicer activity, localisation of the miRNAs for RISC formation and availability of MREs in order to be actively cleaved or silenced. However, even this pathway is not the only mechanism of miRNA function and the presence of multiple components of the miRNA processing machinery in the nuclear compartment indicate that miRNAs may regulate both the epigenome and genome outside of the canonical cytoplasmic pathway (146).

In general, a miRNA is experimentally validated if there is evidence that confirmation its origin from a hairpin precursor (70-nt band on Northern blot for example) and has been processed by Dicer. Using these criteria, it is likely that not all 22-nt RNA fragments are miRNAs and with 1.7×10^{13} possible combinations of 22 nucleotides it is likely that many fragments remain incorrectly identified. One study concluded that less than a third of human miRNAs within miRbase are currently validated as bona fide miRNA genes and this means that caution must be taken when novel miRNA transcripts are identified (147).

Novel RNA transcripts can however sometimes lead to a re-evaluation of the canonical pathway and the recent field of epitranscriptomics has developed from a recognition of post-synthetic modifications involving the RNA chemical structure (148). New high-throughput technologies, especially deep RNA sequencing, has led to the discovery of a multitude of miRNA isoforms, modified by a single base and these have been termed IsomiRs. These can be classified by variations in nucleotides (e.g. single nucleotide polymorphisms or SNPs) or by changes in length either at the 5' or 3' end, or both. The RNase IIIs Drosha and Dicer can alter length, with alterations in secondary structures leading to adjustment of the cleavage site (148). Or they may be "nibbled" by exoribonucleases such as Nibbler or PARN (poly(A)-specific ribonuclease) .

Whilst these are called "templated" variations, as they preserve the sequence of their parent gene, the post-transcriptional addition of nucleotides to small RNA 3' ends may lead to "non-templated" changes. Mono-adenylation and mono-uridylation are the most common with at least eight nucleotidyl transferases known (149). Another form of non-templated change occurs in SNPs which occur at a genomic level but is notably rare when studied computationally (around 10% of human pre-miRNAs have documented SNPs with <1% in seed regions) (150). Additionally, base deamination may lead to a modification or editing at a post-transcriptional level and to-date, 2,885 A-to-I (adenosine to inosine) and 104 C-to-U (cytosine to uracil) unique editing events have been identified on microRNA transcripts with only 257 confirmed experimentally. This may occur to the pre-miRNA, pi0miRNA or the mature sequence through action anywhere in the biogenesis pathway (151).

1.4.4 miRNAs in Cancer.

MiRNAs have been shown to be involved in all aspects of the life cycle of a cell from growth and differentiation to survival or apoptosis. As a result, their deregulation has been linked to many human diseases including cancer. Neoplastic transformation can be regulated by miRNAs either as tumour suppressors downregulating oncogenic targets or as oncogenes (also called oncomiRs) through reducing the expression of tumour-suppressive genes. Many studies have shown that miRNAs can act both as oncogenes and tumour suppressors dependent on the context and in general act through a mechanism of translational inhibition and mRNA destabilisation (152).

One of the first examples of evidence of altered miRNA expression in human cancer was identified by Calin *et al.* in 2002 in chronic lymphocytic leukaemia (CLL). Deletions and translocations involving miR-15a and miR-16-1, two miRNAs that are clustered within chromosome 13q was found to be down-regulated in ~68% of B-cell CLL patients (153).

Germline or somatic mutations were found in 5 of 42 sequenced miRNAs in 15% of patients with CLL, but no such mutations were found in 160 control subjects without cancer. They were able to demonstrate that this miR-15/16 family negatively regulates BCL2 expression and promotes apoptosis, acting as a tumour suppressor. MiR-15a and miR-16-1 were downregulated in CLL and lead to up-regulation of BCL2 similar to in follicular lymphomas (154) . A 9 miRNA signature was furthermore shown to differentiate patients who would progress rapidly to symptomatic disease and germline mutations were found in miR-16-1, miR-27b, miR-29b-2, miR-187, and miR-206, which again were absent from patients without cancer (155). This was one of the earliest demonstrations that miRNAs could potentially be used as a marker to determine prognosis in cancer. It is reported that, through mapping the chromosomal location of 186 miRNA genes, more than 52.5% can be located at cancer-associated genomic regions or in fragile sites where deletions or amplifications commonly occur which suggest a mechanism for miRNA induced tumourigenesis (156). With other known oncogenes such as RAS discovered to be inversely correlated with expression of let-7 in lung tumour tissues (and inhibited by let-7 overexpression *in vitro*) (157) as well as C-MYC in the same fashion being regulated by the miR-17–92 cluster (also termed OncomiR-1) (158), the number of miRNAs associated with cancers has since risen exponentially (159).

1.4.5 Tissue miRNA expression in PDAC/CCA.

Pancreatic tissues consist mainly of closely packed acinar cells with intervening ducts and centroacinar cells occurring at ductule junctions. The duct system is lined by cuboidal duct cells that proximally are termed intralobular ducts with minimal collagen and that distally (forming interlobular ducts) become wider, collagenous and drain into the main pancreatic duct. The interstitium contain pancreatic stellate cells as well as the blood vessels and lymphatics that supply the pancreas (160). Tissue expression profiles in PDAC and CCA have been previously evaluated to identify the cellular source of dysregulated miRNA expression but needs to be unpicked from the underlying histological complexity of cancers arising in this organ.

One of the earliest studies by Lee *et al.* evaluated 201 miRNA precursors using high-throughput RT-qPCR methods to analyse 222 miRNAs from fresh frozen tissue samples in 28 PDAC patients (of which 15 adjacent benign samples were taken), 4 CP, 6 normal control pancreas samples and 9 PDAC cell lines (161). This was at a time when only 326 miRNAs had been discovered in humans, whereas we now believe it to be almost 7x that number. Validated miRNAs as upregulated in PDAC were let-7i, miR-221, miR-100, miR-301, miR-21, miR-181a, miR-125b and miR-212. In the same year, Bloomston *et al.* utilised a custom miRNA microarray chip to evaluate 326 human and 249 mouse miRNA genes that was published in JAMA (162). In total they found 25 differentially expressed miRNAs that could predict PDAC from Benign and 23 differentially expressed miRNAs that differentiated PDAC from CP with an impressive 93% accuracy. MiR-21, miR-221, miR-222, miR-181a, miR-181b, miR-181d, and miR-155 were all overexpressed in tumour samples, whilst they also identified 7 miRNAs (miR-99, miR-100, miR-100-1/2, miR-125a, miR-125b-1, miR-199a-1, miR-199a-2) that were co-expressed in CP and PDAC. Once again, miR-21 (and miR-155) were noted and found to be uniquely expressed in PDAC vs normal pancreas and CP. Comparisons at this time were limited by macro-dissection of tissue which would have been formalin fixed and paraffin embedded (FFPE) resulting in RNA degradation. Calatayud *et al.* reviewed the data to identify candidates which could be validated by RT-qPCR in a larger sample population of macro dissected pancreatic tissue and found that these expression profiles could be associated with prognosis. More specifically, let-7g, miR-29a-5p, miR-34a-5p, miR-205-5p, and miR-212-3p were associated with shorter OS after adjustment for sex, age and tumour stage. However there were remaining undetectable miRNAs which were assumed to be due to platform limitations due to their use of a novel high-throughput method (Fluidigm BioMark System) (163).

A meta-analysis from 2014, identified 20 studies evaluating miRNA expression in PDAC and found that miR-21 in tissue was the most consistently described, shown in Table 4 (highlighted in red) (161). Further analysis noted that raised miR-21 tumoural expression was an indicator of poor OS after resection and the pooled analysis confirmed that particular miRNAs (miR-10b, miR-34, miR-155, and miR-203) had additional prognostic significance. Increased expression of miR-21 has been subsequently associated with early precursor PanIN lesions (164), invasive IPMNs (165), and interestingly the stromal

tumour-associated fibroblasts in resected PDAC tumours (166). These fibroblasts (derived from patient resection tissue) have been found to interact through EV signalling (containing the protein Snail and miR-146a) to increase epithelial cell proliferation and potentially survival (167). More recently, next generation sequencing (NGS) has been used for PDAC tissues and has identified novel non-coding RNA molecules differentially expressed in cancer. Müller *et al.* comparing six PDAC and five control tissues identified miRNAs with no previous association with PDAC (miR-802, miR-2114, miR-561), as well as long non-coding RNA (lncRNA) transcripts. In total, 11 lncRNAs (among them LINC00152, HOTAIRM1, MIAT, GAS5) and 74 miRNAs as well as 4 small nucleolar-derived RNAs (snoRNAs) were significantly upregulated in PDAC (168). Since 2010 when microarrays were able to evaluate 200-300 miRNAs the number investigated in tissues has now risen to over 2000. Evaluating the tissue expression in PDAC has therefore had to focus on more specific compartments.

In summary, there is evidence in broad overall tissue expression that miR-21 (and to a lesser extent miR-222) is upregulated in PDAC tissue miRNA expression profiles and can be associated with prognosis. Differences in protocol i.e. preparation of samples and use of different platforms (from RT-qPCR through to microarrays and more recently RNA-Seq), normalisation strategies and use of archival FFPE samples however lead to significant study heterogeneity which limit any comparison in the search for a unifying biomarker.

Table 4: Selected summary of previous large (>30) miRNA expression profiling studies in pancreatic cancer and pancreatic cancer cell-lines.

Shown below are a summary of the studies as mentioned and miRNAs found to be upregulated in PDAC tissue. Studies shown in ascending year of publication. Highlighted in red are common miRNAs found. AC - ampullary adenocarcinoma; FFPE - formalin-fixed paraffin embedded; CP - chronic pancreatitis; NP - normal pancreas; PDAC - pancreatic ductal adenocarcinoma; RT-qPCR - reverse-transcription quantitative polymerase chain reaction; sncRNAs - small non-coding RNAs; RNA-Seq – RNA Sequencing.

Year	Author	Ref.	Number of miRNAs examined	Methodology	Sample Size	Tissue miRNAs differentially expressed in PDAC	Notes
2007	**Lee *et al.***	(161)	222 miRNAs	TaqMan RT-qPCR	28 PDAC, 15 adjacent benign tissue, 4 CP, 6 NP and 9 PDAC cell-lines	**let-7i** **miR-221** **miR-100** **miR-301** **miR-21** **miR-181a** **miR-125b** **miR-212** (upregulated in PDAC)	Relative normalisation to 18S rRNA. Noted that majority of miRNAs upregulated in PDAC.

2007	**Bloomston *et al.***	(162)	326 miRNAs	Ohio State University custom miRNA microarray chip with RT-qPCR validation	65 PDAC, 65 normal (matched), 42 CP	**miR-21** **miR-221** **miR-222** **miR-181a** **miR-181b** **miR-181d** **miR-155** (upregulated in PDAC)	Normalisation using a "per-chip" median normalization method.
2007	**Szafranska *et al.***	(169)	377 miRNAs	mirVana miRNA bioarrays and SuperTaq RT-qPCR	10 PDAC, 7 CP, 7 NP; 6 PDAC cell-lines (Compared with 33 normal human tissue)	**miR-93** **miR-196a** **miR-196b** **miR-203** **miR-205** **miR-210** **miR-221** **miR-222** **miR-224**	Used RNAlater for storage. Normalisation using the VSN method
2012	**Bauer *et al.***	(170)	863 miRNAs	Geniom Biochip miRNA microarray	94 PDAC, 19 CP, 16 NP	**122 miRNAs** (not reported)	Blood samples also investigated
2012	**Schultz *et al.***	(171)	754 miRNAs	TaqMan Array Human MicroRNA A+B Cards v2.0	170 PDAC, 107 AC, 23 CP, 28 NP	**miR-198** **miR-34c-5p** **miR-21** **miR-614** **miR-196b** **miR-210** **miR-492** **miR-31** **miR-100** **miR-222** **miR-143** **miR-205**	LASSO classifier developed with 19 microRNAs differentiating cancer cells from controls. Sensitivity - 98.5%, PPV - 97.8% and Accuracy - 96.9%.
2015	**Müller *et al.***	(168)	921 sncRNAs	Size selection using PAGE (<40nt), amplification by PCR and RNA-Seq with RT-qPCR validation	6 PDAC, 5 NP	**74 miRNAs upregulated** **Validated - miR-802** **miR-135b-5p** **miR-145-5p** **miR-103a-3p**	In silico target analysis of miR-802 revealed potential binding site in TCF4, a transcription factor involved in Wnt signalling gene
2017	**Calatayud *et al.***	(163)	46 miRNAs	TaqMan RT-qPCR using a Fluidigm BioMark System	165 PDAC, 59 AC, 6 duodenal cancer, 21 CCA, 20 gastric cancer, 39 CP, 35 NP	**miR-21-5p** **miR–23a-3p** **miR–31-5p** **miR–34c-5p** **miR–93-3p** **miR–135b-3p** **miR–155-5p** **miR–186-5p** **miR–196b-5p** **miR–203** **miR–205-5p** **miR–210** **miR–222-3p** **miR–451** **miR–492** **miR–614** **miR-622**	Normalisation not used (Raw CT values).

The latest development is the ability to perform single-cell sequencing which has the potential to achieve unparalleled insight into PDAC tumour heterogeneity and stromal crosstalk although this has yet to be translated to miRNA research (172). As an example, Peng *et al.* examined 24 PDAC tumours and 11 control pancreases, acquiring a total of 41,986 cells from PDAC samples and 15,544 cells from controls (172). They noted that the frequency of acinar cells was very low, which precluded investigation of the acinar-to-ductal-metaplasia, a hallmark of early PanIN, but that markers for ductal proliferation and loss of T-cell activation contributed to poorer prognosis. Another more recent study demonstrated the feasibility of using this technology on limited real-world samples obtained by fine needle and core biopsy (173). This revealed the presence of more than one subtype present even in relatively small samples of PDAC tissue and highlighted this as a potential limitations of bulk tumour sequencing in predicting emergent resistance (174). It is hoped that these techniques will have the potential to unlock cell-specific relationships in the PDAC miRNA transcriptome.

1.4.6 Laboratory techniques used in the analysis of miRNAs.

The first method used to detect RNA from a sample was Northern Blotting which was first developed in 1977 and involved electrophoresis through an agarose gel for the detection of specific RNA as bands which could then be imaged (175). Whilst this technique is cheap and readily available for the differential analysis of RNA expression, it had a relatively low sensitivity and a high user-dependent variability when compared with other methods.

Reverse Transcription-quantitative Polymerase Chain Reaction (RT-qPCR) is a molecular technique for the analysis of RNA based on the polymerase chain reaction (PCR) which revolutionised the study of DNA/RNA when it was invented in 1983 (176). The first step involves transcription of the miRNAs present in the sample back into DNA (i.e. reversed). This is facilitated by a primer which can be either a "stem-loop" or gene specific (TaqMan) and in some cases includes addition of a poly(A) tail. The PCR amplification process can be described in brief as based on repeated thermal cycles involving three steps. Firstly, denaturation at a high temperature (95°C), followed by annealing (50°C to 56°C) and finally extension by enzymatic activity (72°C). During denaturation the double-stranded

template DNA is separated into two single strands. At this point, the temperature is lowered to allow primers (again which can either be short molecules with specific sequence properties or universal primers with high affinity for the template DNA) to anneal. Finally, the extension phase permits the elongation of a new DNA strand by the polymerase enzyme which recognises the bound primer. These three stages are repeated 30-40 times, increasing the number of copies of DNA each time in an exponential manner. The addition of quantitative or real-time PCR enabled the monitoring of a target's amplification during the experiment and means that concentrations can be measured to a very high sensitivity. Methods used today utilise either sequence-specific RNA probes consisting of oligonucleotides that are labelled with a fluorescent reporter and can only be detected after the elongation stage, or SYBR green which binds simply to dsDNA. This can be combined with analysis of their melting temperature specific to the amplified fragment to confirm specificity. Locked nucleic acid (LNA) modified oligonucleotides are used for primers and have added higher sensitivity for small RNAs and is superior to usual TaqMan primer chemistry at low input RNA due to improved stability of the probe-target duplex (177, 178).

RT-qPCR can be used in the form of microarrays to examine differential expression of multiple genes (with the nCounter by NanoString able to detect 800 mature miRNAs) and allows for expression profiles to be determined rapidly. The main drawback of this method is the need for *a priori* knowledge of the targets and appropriate primers in order to test for expression. The use of sequencing technology has enabled the detection of novel RNA molecules such as isomiRs as well as being able to process a larger number of samples at once. Massive parallel sequencing also known as next-generation sequencing (NGS) is so-called because it provided a significant improvement in scale over Sanger sequencing which was invented in 1977 (179). Whilst the Sanger method involved electrophoresis with the random incorporation of dideoxynucleotides to allow a single stranded DNA template to be read, NGS platforms can sequence a large number of samples with short reads (<400 bases) at the same time. The process is based on two steps which can affect the quality of results. Firstly, library preparation involves generation of complementary DNA (cDNA) to the RNA samples themselves and may involve a pre-amplification or size selection leading to bias. RNA is reverse transcribed to DNA after extension by ligation or polyadenylation which introduce primer binding sites for RT according to the type of sequencing machine. Our protocol uses randomised adapters in the ligating ends and is

optimum for low-input RNA templates (180). Sequencing of the cDNA library is then undertaken which in brief involves three steps: Florescent nucleotides are bound to the DNA template strand and blocks further extension, secondly the signal is measured, and finally the terminator is cleaved allowing the steps to be repeated. Data from the cell is recorded as a sequence of nucleotides for each strand as a single read. Analysis is far more complex for sequencing than for RT-qPCR and involves mapping reads (inferred sequence of base pairs corresponding to a single DNA fragment) to the appropriate RNA molecule. For libraries generated with Illumina protocols, these small RNA-Seq reads require trimming during analysis to remove remnants of the 3′-adapter sequence.

1.4.7 Extracellular vesicles (EVs) as a source of extracellular miRNAs.

Cell-free nucleic acids (either cfDNA or cfRNA) have been identified for over 40 years (181, 182) but have become better understood with improvements in deep sequencing technology and RT-qPCR techniques. Extracellular cell-free microRNAs (cfRNAs) were first discovered in 2008 from the serum of patients with diffuse B-cell lymphoma (183). RNA has since been demonstrated as present in all other body fluids including saliva (184), urine (185), breast milk (186), seminal fluid (187), tears, amniotic fluid, colostrum, bronchial lavage, cerebrospinal fluid, peritoneal fluid, pleural fluid (188) and crucially bile fluid (189). These fluids show distinct differences in miRNA expression from one another, but share particular similarities between individuals, indicating that release is likely determined by common local production pathways and cellular export (190). Extracellular miRNAs appear to be protected from endogenous RNases and this protection from degradation is believed to be due to complexing with Argonaute proteins and/or stabilisation within small membrane EVs (191, 192). Furthermore, microRNAs have been shown to be stable through freeze-thaw cycles and extremes of pH particularly compared with large molecular weight RNAs (193, 194).

The functional role of cfRNA in circulating fluids remains unclear but it is found differentially expressed in a multitude of diseases (195). Solid tumours can be identified through cfRNA analysis of blood plasma and can correlate with tumour shedding rate and RNA expression in matched tissue (196) leading to the conclusion that these are released into the circulation during cell death or for cell signalling purposes. A recent study by Chen *et al.* identified in multiple cancers (colorectal, gastric, liver, lung and oesophageal)

both human and microbe derived RNAs present in cfRNA with cancer specific information that could be used for early diagnosis (197). Differential expression in biofluids is believed to be the result of active secretion and can be shown to have a short half-life, reflecting the current state of cells and tissues (198).

The evidence for EV-associated signalling began in 1996, with a study by Raposo *et al.* identifying vesicles from B-cells which he termed "exosomes" and could potentially present antigen via MHC class II-peptide complexes to T-cells (199). These exosomes were found to include other potential molecules such as mRNA and miRNA in a seminal EV paper from 2007 by Valadi *et al.* showing that they may have a role in intercellular signalling, transferring information between cells (200). These EV-miRNAs were resistant to RNase treatment and showed higher expression of particular miRNAs in exosomes compared to their original cells, providing indirect evidence for selective packaging. Early deep sequencing work (201) showed that the majority of total RNA isolated from EVs consisted of small RNAs <200-nt but larger than the 23-nt cut-off that is sometimes used. These comprise of all types of structural RNAs (such as vault RNA, Y-RNA, snRNA, snoRNA, SRP-RNA, and tRNA) as well as fragments deriving from long non-coding RNAs and pseudogenes. Whilst this evidence was specific to immune cell derived EVs, it could be concluded that these might be a source of cfRNA with the potential role in cancer diagnosis.

1.4.8 EV Nomenclature.

The term "extracellular vesicle" (EV) is defined by the most recent MISEV (Minimal Information for Studies of Extracellular Vesicles) position statement by the International Society for Extracellular Vesicles (ISEV) published in 2014 and updated in 2018 (202, 203). This definition is a particle that is "naturally released from the cell … delimited by a lipid bilayer and cannot replicate, i.e. does not contain a functional nucleus". EVs therefore are a diverse category that includes not just exosomes, but also other larger membrane bound structures such as apoptotic bodies and microvesicles. Use of an alternative name is based on biogenesis, differing size distribution and protein markers (204). Whilst the term exosome has been used in a broad sense to describe vesicles in the past, the canonical use of the term is generally reserved for vesicles that are 30-150nm in size and are specifically derived from late endocytic compartments, known as

multivesicular endosomes (MVEs) which fuse with the membrane of the cell (205). Without clearly demonstrating this endosomal origin, EV may be preferential as a term instead (206). Meanwhile, the field continues to discover new EVs with the field still considering the evidence for the presence of smaller 35nm particles named exomeres which have recently been obtained by asymmetric-flow field-flow fractionation (207).

1.4.9 Laboratory techniques for EV Isolation

There are multiple different methods of isolation and EV-Track, a database for the registration of EV-related studies, has identified over 1038 unique protocols from 1226 publications (208) highlighting the heterogeneity of techniques at present. A worldwide survey in 2016 collected 196 responses across 30 countries captured the heterogeneity of EV isolation methodology at an early stage of EV standardisation (209). Differential ultracentrifugation was the most popular technique at the time, and this is likely to be due to the availability of protocols, lower costs, and technical simplicity. Protocols for ultracentrifugation generally involved low speed "clearance" spins (200-2000g) to remove larger cellular debris, followed by medium spins (10,000-20,000g) to remove larger apoptotic bodies and microvesicles that have a larger size between 100-1000nm (210). A wide range of protocols also included filtration steps which utilise pore sizes of between 0.2μm to 1μm.

Surprisingly kit-based methodologies were not highly reported in this study which was circulated to ISEV members only. These kits (examples include both ExoQuick and Exo-spin) utilise polymer-based precipitation with chemicals such as polyethylene glycol used to bind or reduce the solubility of vesicles and allow them to be centrifuged or filtered at low speeds. These have a tendency to co-isolate non-EV protein from samples (211) and in plasma fail to exclude highly expressed proteins such as albumin (212). Despite this, it is often used to rapidly obtain EVs for downstream analysis and remains a popular technique in the literature for the high-throughput analysis of large numbers of samples, particularly from clinical biofluids (213, 214).

Size exclusion chromatography was first invented as a process at Queen Charlotte's Hospital in London and used potato and maize starch leading to inconsistent elution

volumes from each column (215). This was improved by the subsequent development of dextran and crosslinked agarose, and presently, gel matrices known by tradenames. These include dextran polymers (Sephadex), agarose (Sepharose), or polyacrylamide (Sephacryl or BioGel P) manufactured as beads and optimised for experimental consistency. A review by Liu *et al.* published in 2020 identified more papers year-on-year associated with the technique and reported that the benefits of SEC included short time to process, relatively inexpensive, and able to efficiently remove over 95% of blood protein which making it clinically applicable when compared with dUC or other techniques. Other papers supported this reporting that EVs isolated by chromatographic methods suffer less contamination than after UC and/or kits (216-218).

Other EV isolation methods reported in the literature include density gradient centrifugation which uses a denser than water medium such as iodixanol (Optiprep) combined with a centrifugal force to separate and purify EVs based on their buoyant density. This requires a second separation step which can be either centrifugation or filtration to isolate the EVs from solution. It is reported that exosomes have been found to float at densities ranging from 1.15 to 1.19 g/mL on continuous sucrose gradients. By comparison, vesicles purified from the endoplasmic reticulum float at 1.18 - 1.25 g/mL, and vesicles from the Golgi at 1.05 - 1.12 g/mL (219).

Many of the techniques used for EV isolation were biochemical methods that have been repurposed for the analysis of these novel structures. Asymmetrical flow field-flow fractionation (AF4) was developed in 1966 and has recently been utilised to identify structures smaller than 50nm in size from cell cultured medium (220). Whilst SEC was considered to have a resolution similar to AF4, AF4 was able to demonstrate superior resolution over a much wider size range to potentially evaluate EVs of different biogenesis. At present, there is also much interest in the development of microfluidic devices to evaluate EVs with several now having demonstrated the potential to distinguish miRNA signatures from plasma as well as cell media (221, 222).

1.4.10 Laboratory techniques for EV Characterisation.

A critical determination of EV isolation includes analysis by several methods, and this was covered in the MISEV guidelines which was updated in 2018 (203). The rapidity of progress in the field necessitated an update to take into account changes in knowledge and a growing number of separation techniques and markers chosen for characterisation. It endorsed the EV-TRACK open-source knowledgebase which used a key feature, termed an EV-METRIC to highlight nine specific areas that could be used to support critical appraisal of an EV study.

Table 5: The EV-Metric.

Shown here are the nine parameters required by the EV-Metric which is a score calculated when studies are added to the EV-TRACK database. It highlights the areas that should be used to critically appraise any EV study and can be seen as indispensable for unambiguous interpretation of any EV study (208). EV – Extracellular vesicles.

EV-Metric	
Experimental parameters related to EV separation	
Density Gradient	Type and concentrations of the gradient performed, use as validation of EVs in experiments.
EV Density	EV density should be reported if undertaken.
Ultracentrifugation Specifics	When ultracentrifugation is used, there should be a description of the g-force, duration and rotor type of these steps.
Protein Analysis	
EV-Enriched Proteins	3 or more EV-enriched proteins.
Non EV-Enriched Protein	Assessment of one or more EV-enriched protein.
Antibody Specifics	Antibodies used for immunoblotting should be reported to include clone/reference number and dilutions.
Lysate Preparation	Lysis of EVs should involve reporting of buffer composition.
Particle Analysis	
Qualitative and Quantitative Analysis	Implementation of both qualitative (e.g., electron microscopy) and quantitative methods (e.g., nanoparticle tracking analysis, high-resolution flow cytometry). Quantitative methods should include reporting of particle concentration.
Electron Microscopy Images	Inclusion of both wide-field and a close-up electron microscopy image.

Quantification of EVs by size and concentration is important to determine overall amounts, and concentrations have been associated with disease states (223). Light scattering technologies such as nanoparticle tracking analysis (NTA) utilise Brownian motion to obtain the size distribution, concentration, and zeta potential of nanoparticle samples in suspension (224). Video is captured and analysed by software utilising the Stokes-Einstein equation which determines the diffusion coefficient and from this the "hydrodynamic diameter" of the particles calibrated using standardised 100nm

polystyrene beads. This can be done directly or utilising immunofluorescence antibody tagging of specific surface markers. The machine detects all "particles" within a certain size range and therefore may overestimate if there are protein aggregates and other solute contaminants that may be within the sample. It has been reported that EVs with sizes below 50-60nm are not accurately measured compared with electron microscopy which means that NTA will underestimate small EV populations (225). Flow cytometry is a recognised alternative technique and since 2016 has been developed to a high-resolution enabling measurement with fluorescent labelling of EVs smaller than 200nm in size (226). Other methods include tuneable resistive pulse sensing (TRPS) which measures a resistance change across a nanopore (227) and atomic force microscopy (AFM) (228). In all these there are potential pitfalls such as optimal size ranges and issues with multidispersed samples which can affect measurements(203). Indirect measurements of EV quantity can also be undertaken by protein, lipid or RNA measurements.

Electron microscopy complements NTA by allowing single vesicle analysis and this can be performed by scanning (SEM) or transmission (TEM) (229). In TEM, samples are layered onto a Formvar coated copper grid which allows electrons to travel through the field and onto a sensor which captures the 2D image and visualisation can be assisted by either a negative staining material such as uranyl acetate or immunolabelling with gold particles. In SEM, a 3D picture is captured by electrons ejected or backscattered/reflected by the sample. EVs in air are circular (spherical) or characteristically "cup-shaped" in morphology due to deformation during the processing of samples (230). From this, size and the concentration can be directly estimated. Cryo-EM is a form of electron microscopy which preserves nanosized structures by rapid freezing without the need for fixation or contrast and allows for imaging of EVs in their native state (231). More recent techniques are still in development and include super-resolution imaging via photoactivated localization microscopy (PALM) or stochastic optical reconstruction microscopy (STORM). This uses photo-switchable probes and sequential imaging to obtain pictures with resolution below the diffraction-limit of light, through reconstruction from the positions of the different probes (232).

An important feature of EV characterisation is that of protein composition either of surface markers or intra-vesicular proteins that are expressed. MISEV2018 remained

adamant that "is still not possible to propose specific and universal markers of one or the other type of EVs". The updated guidelines highlighted three categories of proteins that should be analysed but did not propose markers to characterise EV subtype (203). These categories were (1) transmembrane proteins localized at the plasma membrane and/or endosomes of eukaryotic cells, (2) cytosolic proteins and (3) absence of non-EV structures/proteins e.g. albumin or proteins derived from the endoplasmic reticulum (calnexin) or mitochondria (cytochrome C). For plasma/serum EVs it was proposed that "apolipoproteins A1/2 and B (APOA1/2, APOB), and albumin (ALB) as the best negative markers to date" (233). MISEV2014 did originally specifically list transmembrane extracellular proteins as including tetraspanins (CD9, CD63, CD81), integrins (ITG) or cell adhesion molecules (CAM) and phosphatidylserine-binding MFGE8/lactadherin amongst others. It also listed cytosolic proteins such as endosome or membrane-binding proteins (TSG101, annexins, Rabs) and signal transduction or scaffolding proteins (syntenin) (202). These markers are particularly associated with exosomes as ALIX and TSG101 both relate to initial intraluminal vesicle (ILV) formation in the endosomal sorting complex required for transport (ESCRT) machinery of the cell (234).

Western blotting or flow cytometry (FACS) are the most used techniques for semi-quantitative protein analysis in order to show relative expression compared to a control (cell lysate for example). However, pure plasma exosomes are believed to contain ~1500 proteins giving a maximum yield of just ~1µg protein from 1ml of blood (235) limiting the ability to determine EV protein expression from valuable clinical samples. There is also significant heterogeneity when it comes to clinical samples, and it is not expected that all markers are constitutively expressed in EVs. A study by Ludwig *et al.* did evaluate specifically SEC-isolated tumour-derived EVs and determined that tetraspanins (CD9, CD81) were not constitutively expressed in all EVs (236). The absence of CD9 as well as other commonly used markers (TSG 101, Alix, CD63, and CD81) from EVs isolated by E10, BxPC3, and H3 cells including was also noted in a study that co-characterised by FACS (237). Protein markers for bile derived EVs were investigated as far back as 1997 in a study that undertook two-dimensional electrophoretic analysis but were not fully characterised at the time (238). Severino *et al.* confirmed the presence of TSG101, CD9, ALIX and HSP 70 compared with unextracted bile and appeared to demonstrate an increase in intensity from malignant to non-malignant samples (223). Recent work using

state-of-the-art proteomics identified syntenin-1 as a ubiquitous exosome marker (as well as ALIX and TSG101) but this has not been reflected yet in MISEV (203).

1.4.11 Clinical rationale for the use of EVs.

With the difficulties in isolation methodology and around classification, it might be wondered why the clinical focus on EVs as a source of biomarker. One of the earliest reasons given was that EV packaging was believed to provide a stable environment resistant to endonuclease digestion. This has been shown in multiple studies where EVs are treated with RNase as well as detergent (e.g., Triton X-100) to determine the relative expression of RNAs (239). Li *et al.* demonstrated that miR-21 and miR-638 were found to be stable in EVs present in bile after 24-48 hours at room temperature and after up to 3 freeze thaw cycles. Analysis by NTA of serum-derived EVs obtained by UC showed a measured decrease in concentration by 8%–20% per freeze-thaw-cycle which was only detected with the ZetaView (Particle Metrix, Germany) (225). Muller *et al.* examined plasma-derived EVs obtained by SEC which after freezing appeared to lead to an increased enrichment of free proteins (by lysis), but their function was not impaired nor was there an observed alteration in exosomal morphology (240). The overall stability of EVs has further been confirmed in other studies using SEM (241) and FACS (242).

The remaining criticism is that the proportion of RNA/DNA in EVs has been found to be a very small proportion of the extracellular cfRNA/DNA (191, 243) with the majority bound to AGO2 which also provides resistance to endonuclease activity. The detection of very small amounts of RNA (<1ng) means that the physiological role of these as paracrine signalling molecules is considered by some investigators as unlikely (189). Attempts to counter this argument have been to investigate the cellular processes that govern EV release and uptake to demonstrate their biological activity in humans. Tumour-associated EVs from cell lines have been shown in mice models to be variably taken up by organs such as the lung or liver depending on their origin (in this case breast or pancreatic cancer) by Hosino et al (244). This uptake led to upregulation of pathways such as the pro-inflammatory S100 gene (previously correlated with metastasis) in exosome-educated lung fibroblasts and has been cited as evidence for a mechanism for premetastatic niche establishment by EVs. This was also shown in PDAC with exosome education by retroorbital injection of PAN02-derived (a PDAC cell line) exosomes into wild-type mice leading to increased metastatic burden (245). EVs were predominately taken up by

Kupffer cells and led to upregulation of liver fibrosis associated factors (CTGF, EDN, IGF, PDGF and TGFβ), which was subsequently inhibited by MIF (macrophage migration inhibitory factor) knockdown. Most interestingly, these were changes that occurred in premalignant (i.e. early PanIN) murine models indicating that this process occurs prior to metastatic or even invasive disease. The role of "exosome education" in metastases has been replicated in melanoma, breast and lung cancer among others (246).

It is therefore of interest, now that a physiological role has been demonstrated for EVs that the molecules underlying it are further investigated.

1.4.12 Bile as a source of EVs.

Bile is a complex secretion originating from hepatocytes which is modified as it travels past cells of the bile duct epithelium (247). Anatomically it is either concentrated in the gallbladder or delivered directly to the intestine through the ampulla of Vater. Bile contains around 95% water within which are found bile salts, bilirubin, cholesterols and steroids as well as exogenous toxins and drug metabolites. Whilst bile is isosmotic in comparison to plasma, due to passive secretion of ions, organic anions and cations are highly concentrated in bile by active transport (ATP dependent) mechanisms at the canalicular membrane. The most prevalent organic solutes are bile salts which act on intestinal contents to emulsify dietary fats and facilitate their absorption (248). It is also the primary route for the elimination of cholesterol.

As far back as the 1980s, biliary vesicles have been isolated from human bile (249) by either ultracentrifugation or SEC (238). Bile-derived vesicles were first identified in 1985 by Sömjen et al from samples obtained from indwelling T-tubes after cholecystectomy. Supernatants after ultracentrifugation at 192,000g were analysed by what was known then as quasi-elastic light scattering and found vesicles of around 72nm in size. These could also be found in rat bile and were stained for electron microscopy (250). However, the focus of research was on the nucleation of bile cholesterol leading to the production of gallstones. It was noted that the vesicular concentration of cholesterol might be higher in the bile of patients with gallstones (251). Biliary derived vesicles were also shown to have lipid bilayers of about 6nm diameter and ranged in size 100-250nm when examined by cryo-EM (252).

1.4.13 MiRNAs are present in bile-derived EVs.

However, it was not until 2010 that the RNA composition of these vesicles was investigated. Masyuk *et al.* (253) used bile from the cannulated common bile duct (CBD) of male Sprague-Dawley rats isolated through a differential ultracentrifugation before resuspension in Phosphate-buffered saline (PBS) and further purified using 5-30% sucrose gradient obtaining exosomes with a density of 1.12–1.21 g/m. He investigated the role of cholangiocyte cilia *in vitro* and demonstrated increased expression of miR-15A, a miRNA associated with cholangiocyte proliferation as well as affecting the ERK signalling pathway.

Li et al (254) in 2014 was the first and largest investigation of human bile miRNAs in samples for patients with CCA (n=96) obtained at ERCP or PTC. They investigated EVs using an RNA microarray and selected 11 miRNAs for validation by RT-qPCR (miR-222, miR-126, miR-486-3p, miR-484, miR-19a, miR-19b, miR-16, miR-191, miR-31, miR-1274b, miR-618). They used a Multivariate Organization of Combinatorial Alterations (MOCA) algorithm to predict probabilities of CCA diagnosis and from this identified a panel of 5 miRNAs (miR-191, miR-486-3p, miR-1274b, miR-16, miR-484) with a sensitivity of 67% and specificity of 96% (model threshold tuned to maximum specificity). One of the limitations they noted was that RT-qPCR normalisation to U6 was likely to degraded in a different way to shorter miRNAs and instead they normalised to a cel-miR-39 spike-in.

1.4.14 Blood plasma as a source of EVs

Blood is a complex bodily fluid which circulates and delivers many nutrients and molecular substances around the body and removes waste products of respiration and metabolism. The peripheral blood is contained within the circulatory system of arteries and veins that flow around the human body. Blood is made up of blood cells (erythrocytes, leukocytes and platelets) within plasma which is a combination of proteins, ions, hormones and water.

For the purposes of analysis, blood can either be harvested as serum (which is activated by silica particles and left to coagulate) or plasma (which is prevented from clotting with EDTA) (255). These are then centrifuged down at speeds of around 1000 x g to remove blood cells, platelets, and clotting factors (if serum). Activation of the coagulation cascade is thought to release and increase the RNA repertoire circulating in the extracellular compartment (256, 257) so plasma is often preferred as it is the most naïve for biomarker

analysis. Extraction of the RNA from plasma to investigate extracellular miRNAs (also known as cell-free RNA or cfRNA) is a large field and there is a wealth of data on the characterisation of both these biofluids as a biomarker in CCA (258-261) and PDAC (262-264). Some of these studies also investigate the EV or exosomal compartment (259, 262) which makes comprehensive analysis of the literature very time-consuming and difficult.

Plasma has been identified as containing secreted vesicles from cellular activity since at least the 1960s (265) and the process of membrane vesicle formation appear conserved across all domains of life, being present as a mechanism in bacteria (266), fungi (267) plants as well as mammals. These were first described in plasma and had various descriptions at first such as platelet dust (268), microparticles (269), and virus-like particles (270) as efforts were made to identify their origins as either platelet-derived, or virions of some description. Eventually, the identity of a particular kind of vesicle, termed 'exosome', was identified as released from the lumen of MVEs and which is now called the exosome secretion pathway (271, 272).

1.4.15 MiRNAs are present in plasma-derived EVs.

Though plasma is relatively easy to access, it is also a complex fluid (273) and the role of EVs present in blood has been the subject of over 30 studies to-date. Exosomal miR-17-5p and miR-21 were the first initial EV-miRNAs found to be upregulated in blood serum compared to healthy controls (274). MiR-21 remains the most cited candidate having been expressed in at least 5 studies relating to blood in PDAC (275-279).

Whilst the focus of this project has been on non-coding RNAs, it is worth discussing that plasma derived EVs in PDAC were found in 2015 to contain a proteoglycan, glypican-1 (GPC-1) bound to the cell surface and determined by flow cytometry, with the relative fraction of GPC-1 positive EVs correlated with tumour burden and survival in mice and patients pre- and post-surgical tumour resection (280). This was a large study involving 190 patients with PDAC and 100 healthy donors which could detect patients with an AUC of 100% (95% CI: 98.8-100%) as well as a perfect sensitivity of 100% and specificity of 100% comparing patients with PDAC (stage I-IV) against both healthy individuals (and those with benign pancreatic disease). However, these results failed to stand up to scrutiny as a follow-up study in 2019 utilising nanoscale FACS to assess for GPC-1 positivity in 93 PDAC samples compared with 41 normal controls failed to achieve similar results (AUC of 0.5404, sensitivity of 26.7% and specificity of 87.5%) (281). Another study

did find GPC-1 detectable by ELISA in PDAC patient plasma EVs (obtained by UC) but there was no difference noted when compared to the levels in patients with benign pancreatic disease (282). Lai et al analysed 40 plasma EVs and compared GPC-1 protein levels by liquid chromatography-tandem mass spectrometry (LC-MS/MS) in a small cohort of 3 PDAC samples finding that it had a fair AUC of 0.75 whilst a 6-miRNA panel (miR-10b, miR-20a, miR-21, miR-30c, miR-106b, and miR-181a) achieved an AUC approaching 100%. Despite the results having been repeated by the original researchers with alternate antibodies (283), the failure of GPC-1 as a reproducible PDAC biomarker has been notable.

In CCA, there is little evidence of diagnostic significance with only 2 studies having looked at the transcriptome and proteome of blood derived EVs. Shen et al examined the serum-derived EVs of 48 patients (CCA n=36, Controls n=12) using the miRCURY exosome serum/plasma kit for the top twenty miRNAs identified from cell line experiments (284). They identified four miRNAs from the miR-200 family (miR-141-3p, miR-200a-3p, miR-200b-3p, and miR-200c-3p) that individually had a higher AUC than CA19-9 and more promisingly was correlated with lymph node metastases, tumour stage. Elevated serum exosomal miR-200c-3p levels was correlated with tumour free survival (p=0.005) but did not reach significance for overall survival (p=0.075). Serum EVs were evaluated for protein-based markers in a study by Arbelaiz *et al.* and published in 2017 (285). They evaluated a large cohort of 134 serum samples which included patients with CCA, HCC and PSC. They identified several biomarker candidates for the differential diagnosis of CCA versus PSC by nanoscale chromatography and in particular, FCN2, inter-alpha-trypsin inhibitor heavy chain H4, and FIBG showed higher diagnostic values with an AUC of 0.956, 0.881, and 0.881, respectively than their CA 19-9 measurements which had an AUC of 0.736 in early stage CCA (I-II) vs PSC.

1.4.16 The presence and role of tRNA derived fragments (tRFs) in the extracellular space.

Transfer RNAs (tRNA) are the most abundant RNA in the human genome (286), and was the first non-coding RNA to have been discovered (287). The canonical cytoplasmic tRNA molecule is 76-nt long and well-conserved, with a cloverleaf secondary structure and a tertiary L-shaped structure separating the anticodon triplet for recognition of template mRNA from the amino acid attachment site. Whilst tRNAs are present in all species,

tRNA genetic expression is species-specific and with over 270 different tRNA sequences present among ~450 tRNA human genes, a highly complex system in eukaryotes (288). In addition to identifying miRNAs present in our samples, small RNA sequencing allows for the discovery of short RNA molecules derived from other sources such as tRNAs.

The fragmentation products of tRNAs include tRNA-halves and tRNA-derived fragments (tRFs) and have distinct biogenesis pathways which is covered in the review by Liu et al (287). tRNA-halves are thought to be cleaved by nucleases such as angiogenin as well as Dicer and RNase Z which are upregulated in stress. The simplest fragmentation is a cleavage of the anti-codon loop forming either 5' or 3' tRNA-halves with lengths corresponding to half the full-length (~30nt) of mature tRNAs. tRFs on the other hand, includes 5' and 3'-tRFs with shorter nucleotide sequences and are named according to the cleavage site which is usually in or around the D-loop or T-loop structure of mature tRNAs. However, tRFs can also arise from 3′ nuclease cleavage of 5′ tRNA-halves and the enzymes responsible for the production of smaller tRFs remain poorly understood (289). There are also tRFs that do not correspond to the 5′ or 3′ end and are called internal tRFs (itRFs). Downstream effects of tRFs are thought to be miRNA-like regulatory and silencing activity on mRNA to facilitate post-transcriptional repression (in association with AGO proteins).

Early sequencing studies of EVs from blood plasma showed small proportions of tRNA transcripts, averaging 1.24% from 3 samples in a healthy population from samples obtained using ExoQuick (290). However, with newer databases available for the mapping of reads since the paper was published (e.g. tRFdb, MINT base) and a greater understanding of tRNA mapping methodology, it may be that these were an underrepresentation. One of the difficulties with determining the representation of tRFs in EVs is likely to be cell specific differences, as well as different EVs and RNA isolation protocols and the use of different NGS library preparation protocols and sequencing platforms (291).

1.4.17 A note on biomarker detection with reference to sensitivity and specificity.

For the purposes of clinical diagnosis, the aim is to find a model that predicts the likelihood of cancer in a test population. A receiver-operator characteristic (ROC) curve is a plot which illustrates the diagnostic ability of a test in a binary outcome (cancer or no cancer). The x-axis is the false positive rate, also known as 1- Specificity, against the y-axis which is the true positive rate also known as sensitivity. The threshold of a test is the value for which it determines its binary outcome, and the results of the test with various thresholds is plotted on this curve. A perfect test has a sensitivity of 100% (no false negatives) and 100% specificity (no false positives) (see FIGURE 4). A test which is no better than random will draw a diagonal through the plot, whilst a test that is good will have a line that is deflected towards the top left (shown in blue).

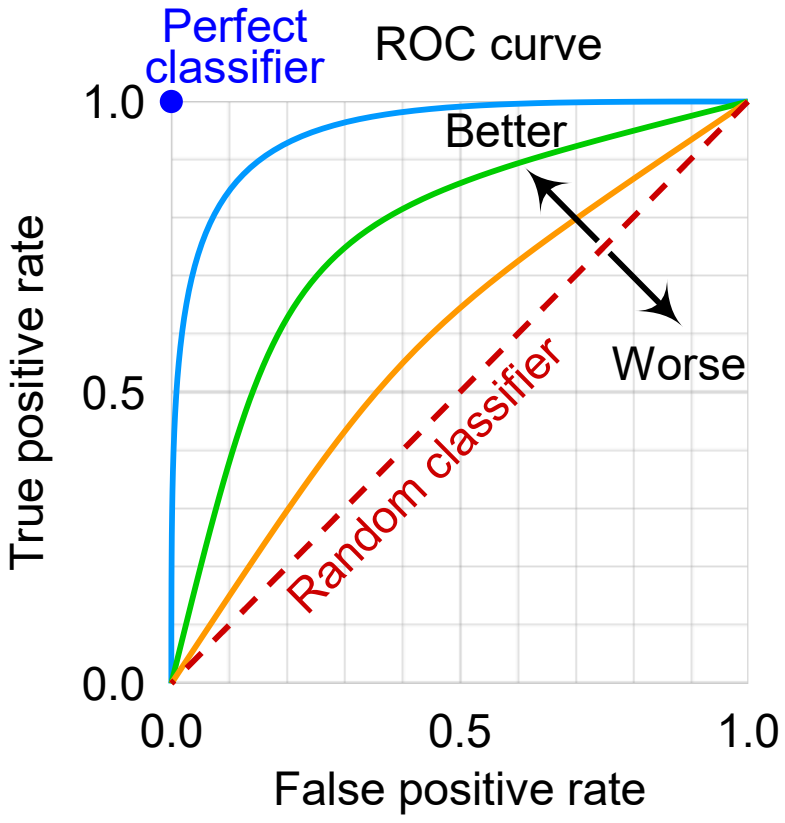


Figure 4: Receiver Operating Characteristic (ROC) curve
This diagram shows a typical ROC curve which is a graph of the true positive rate (i.e. Sensitivity) against the false positive rate (1 – Specificity). A diagnostic test with a binary outcome will draw a line through this from 0,0 to 1,1 dependent on the threshold used to determine the test result. The diagonal shows the performance of a random classifier which will be no better than chance and will therefore pass through the point 0.5,0.5 (shown in red). Three example classifiers (blue, orange, green) are shown to demonstrate possible outcomes with the area-under-the-curve (AUC) a measure of a test's overall diagnostic ability. Drawn by CMG Lee and licensed under the Creative Commons Attribution-Share Alike 4.0 International license.

Two calculations can help with determining the diagnostic value of the test. The first is the area-under-the-ROC-curve or AUC, which in a perfect test is 1.0 (or 100%) whilst a test that has no better accuracy than chance (being binary) is 0.5 (273). The second is the maximum point on the curve away from the diagonal which is determined by the Youden Index (or J statistic) and corresponds with the threshold for which the test has maximum sensitivity and specificity. This is calculated as sensitivity+specificity-1 and has a maximum of 1. All results for sensitivity and specificity unless otherwise stated are at maximum Youden Index. For the purposes of a screening test, it may be important to minimise the false negatives i.e. failing to detect a patient with cancer and it may be reasonable to maximise sensitivity at the expense of a loss of specificity. This can also be referred to as a high negative predictive value (NPV).

1.5 Aims, Hypothesis and Objectives

The original aim of the project was to investigate miRNAs as biomarkers that could aid the diagnosis of PDAC and CCA in patients with a biliary stricture. Given that they are found differentially expressed in PDAC and CCA tissues with important roles in multiple cancer pathways, my **hypothesis was that there would be differentially expressed small non-coding RNAs within bile and plasma which would be able to distinguish between patients with PDAC and CCA from those with benign disease** in this cohort of patients. Given that EVs can be found in both bile and plasma, it was therefore a secondary hypothesis that these miRNAs might be present in the EV compartment and would lead to a clinical test that could differentiate malignant and benign disease better than existing tests. Our secondary aim was then to evaluate their diagnostic potential whilst also assessing clinical utility if the study was able to do so (speed of the test, evaluation in different patient populations).

In the following chapters these are as follows:

Chapter Three – Investigate the circulating miRNAs derived from bile and blood plasma in a cohort of prospectively recruited patients presenting with benign and malignant biliary strictures.

- Identify differentially expressed miRNAs between the PDAC, CCA and Benign disease population.
- Identify other potential differentially expressed tRNA derived fragments between the PDAC, CCA and Benign disease population.
- Validate these candidates for clinical diagnosis and evaluate their diagnostic potential.

Chapter Four – Investigate the miRNAs derived from the extracellular vesicle compartment of bile from the same cohort of patients.

- Identify differentially expressed EV-miRNAs between the PDAC, CCA and Benign disease population.
- Identify other potential differentially expressed EV-tRNA derived fragments between the PDAC, CCA and Benign disease population.

- Validate these candidates for clinical diagnosis and evaluate their diagnostic potential.

Chapter Five – Investigate the miRNAs derived from the extracellular vesicle compartment of blood plasma from the same cohort of patients.

- Identify differentially expressed EV-miRNAs between the PDAC and Benign disease population.
- Identify other potential differentially expressed EV-tRNA derived fragments between the PDAC and Benign disease population.
- Validate these candidates for clinical diagnosis and evaluate their diagnostic potential.

Chapter Two – Materials and Methods

2.1 Patients and Materials

2.1.1 Samples collected from surgery.

Surgical bile samples were obtained from patients at Imperial College NHS Trust (Hammersmith Hospital, London, UK) and University Hospital of Pisa (Italy) between 4th February 2014 and 29th May 2018. These were obtained during surgery for either cholelithiasis (gallstones), PDAC or CCA. Bile was aspirated either from the gallbladder or the common hepatic duct (during pancreaticoduodenectomy) using a sterile syringe and large-gauge needle. After collection, all bile samples were transferred to the laboratory (less than one hour) and flash-frozen at -80°C until required.

All participants gave written informed consent prior to surgery and ethical approval was obtained from Imperial College Healthcare Tissue Bank for the collection and storage and analysis of patient samples: SUR_AF_17_044; "Molecular Detection and Stratification of HPB Cancers". Additional approval for samples was obtained from Camden & Islington REC: 09/H0722/77; "Early Detection of HPB Cancers using Biomarkers".

2.1.2 Samples collected for the MIRABILE study.

Bile and plasma for the "MIcroRNAs as BILE-based biomarkers in Pancreaticobiliary Cancers" (MIRABILE) study were collected from patients attending the endoscopy unit at Imperial College NHS Trust (Hammersmith Hospital, London, UK) between 2nd November 2017 and 11th March 2020. All patients were attending for evaluation of biliary obstruction by endoscopic retrograde cholangiopancreatography (ERCP). Written consent was obtained at the time of endoscopy to store samples for research purposes.

Inclusion criteria for patients recruited were age ≥ 18 years, scheduled for clinical reason to undergo an ERCP (endoscopic retrograde cholangiopancreatography), WHO performance status 0, 1 or 2, able to provide written informed consent. Exclusion criteria included patients with previous Roux-en-Y gastrojejunostomy, hepaticojejunostomy or Bilroth II, pregnancy or those with WHO performance status 3 or 4. Patients were provided with a written patient information sheet and details of the study to which they were enrolled.

All patients were starved from midnight for the procedure and blood samples were collected either through a freshly inserted 18-22G peripheral cannula or 21-23G butterfly syringe into BD K2 Ethylenediaminetetraacetic acid (EDTA) lined tubes and processed within 4 hours.

Blood was centrifuged at 2,500 x g in a Containment Level (CL) 2 room for 10 minutes at room temperature (RT) to remove cells (Centrifuge 5810R, A-4-81 rotor, Eppendorf®). The resulting platelet-free plasma was aliquoted into 1.5 mL Eppendorf safe-lock tubes within a Class 2 cell culture hood and samples were used immediately or frozen at -80°C until required.

Bile was obtained after cannulation of the common bile duct (CBD) by ERCP and before iodine contrast injection. Aspiration using a sterile 10ml syringe was performed via the sphincterotome into a sterile universal container. All bile samples were transferred to the laboratory within 4 hours and flash-frozen at -80° until required. If ERCP was not technically possible, patients underwent subsequent percutaneous transhepatic cholangiography (PTC) to decompress the biliary ducts. Bile was obtained from these patients within 12 hours of drainage and stored in a sterile universal container. These were also immediately transported to the lab and flash-frozen at -80°C until required.

2.1.3 Patient demographics and pathology data collection.

Patient electronic medical records were reviewed retrospectively, and an anonymised database was created using Microsoft Excel 365 (Redmond, USA) to document patient demographics (age at time of sampling, sex) and clinical presentation. Upon receipt of samples concurrent serum CA19-9, Bilirubin, and CRP levels were recorded (prior to therapy).

The aetiology of any strictures and therefore patient sample arms were based on diagnostic pathology from surgically resected specimens or biopsy specimens (open or laparoscopic surgical resection/biopsy, percutaneous needle, or ERCP-obtained biliary biopsy) and/or clinical criteria after ≥12 months of follow-up. Aetiology and clinical staging were determined using results of a multidisciplinary team meeting consisting of at least a consultant hepatopancreaticobiliary surgeon, consultant hepatopancreaticobiliary physician, consultant histopathologist and consultant radiologist. ERCP findings, endoscopic ultrasound findings, biliary brushing cytology, and fine needle aspiration cytology were anonymised and recorded. A clinical diagnosis of malignancy was defined

by the combination of a mass on radiographic imaging without acute cholangiopathy, and clinical or radiographic progression after ≥12 months of follow-up, or death clinically and radiographically determined to be due to cancer. For surgical resection and biopsy specimens, diagnoses and staging were rendered based on standard histo-morphological criteria. For the purposes of analysis, pathological staging was used in preference to clinical staging where possible. Any patients with a benign aetiology were either clinically determined on the basis of no further progression after ≥12 months follow-up with either documented resolution or stability of prior ductal abnormalities or no further intervention as documented in electronic hospital records at 12 months.

Selection of samples for each cohort was based on sample availability. Preference was given to the circulating bile analysis and samples distributed accordingly. Benign patients for each cohort were selected to be age and sex matched where possible to the malignant samples in the discovery cohorts to provide a control. Diagnostic and clinical follow-up information is detailed for each study arm.

Table 6: Buffers, reagents & consumables.

Reagent	Catelogue Number	Company	Storage Conditions
100% Ethanol	20821.33	VWR Chemicals	Room Temperature
10X Radioimmunoprecipitation Assay (RIPA) Buffer	#9806	Cell Signalling Technology	Room Temperature
10X Tris Buffered Saline	Lab	-	Room Temperature
Acrylamide/Bis-acrylamide, 30% Solution (Protogel)	EC8901LTR	National Diagnostics	Room Temperature
Ammonium Persulfate (APS)	A3678-25G	Merck Life Science UK Limited	4°C
Dithiothreitol (DTT) Powder	D9779-5G	Merck Life Science UK Limited	4°C
Laemmli Buffer (Pierce™ LDS 4X Sample Buffer)	84788	Thermo Fisher Scientific	Room Temperature
Minisart Syringe Filter, Polyethersulfone (PES), Pore Size 0.45 µm	16533	Sartorius	Room Temperature
miRNeasy Micro Kit	217084	QIAGEN	4°C/Room Temperature
miRNeasy Serum/Plasma Advanced Kit	217204	QIAGEN	4°C/Room Temperature
N,n,n′,n′-tetramethylethylenediamine (TEMED) (µl)	T9281-25ML	Merck Life Science UK Limited	Room Temperature
Phenylmethylsulphonyl Fluoride (PMSF)	sc-24948A	Santa Cruz	-20°C
Phosphate Buffered Saline (PBS)	806544	Merck Life Science UK Limited	Room Temperature
Pierce Bicinchoninic Acid (BCA) Protein Assay Kit	23225	Thermo Fisher Scientific	Room Temperature
Pierce Lithium Dodecyl Sulfate (LDS) Sample Buffer, Non-reducing (4X)	84788	Thermo Fisher Scientific	Room Temperature
QEVoriginal 70nm SEC Columns	SP1	Izon Science	4°C
Restore Plus Western Blot Stripping Buffer	#46430	Thermo Fisher Scientific	Room Temperature
RNaseZap	AM9780	Thermo Fisher Scientific	Room Temperature
Taqman MicroRNA Reverse Transcription (RT) Kit	4366597	Applied Biosystems	-20°C
Taqman Universal Master Mix II, No UNG	4440049	Applied Biosystems	4°C
TBS-T Buffer (Tris Buffered Saline/Tween 20)	Lab	-	Room Temperature
Tris-HCL Buffer	Lab	-	Room Temperature
Trizol LS Reagent	10296028	Invitrogen	Room Temperature
Vivaspin Turbo 4 Ultrafiltration Unit	VS04T02	Sartorius	Room Temperature

Table 7: Components/recipes for reagents used.

Reagents	Components	Final Concentrations	Storage Conditions
Sodium dodecyl sulphate polyacrylamide SDS-PAGE 11% gel	3.7 mL of Distilled Water 3.7 mL of Acrylamide protein gel (30%) 2.5 mL of Gel Buffer (1.5 M Tris-HCl pH 8.8 for resolving; 0.5 M Tris-HCl pH 6.8 for stacking) 100 µL of 10% SDS (w/v) 50 µL of 10% APS 10 µL of TEMED	11.1% (w/v) 0.375M Tris-HCl pH 8.8 or 0.125M Tris-HCl pH 6.8 1g/L (0.1%) 0.5g/L (0.05%) 0.001% (v/v)	Room Temperature
Running Buffer	1,800 mL of Milli-Q H_2O 200 mL of 10% SDS (w/v)	20g/L (2%)	Room Temperature
Transfer Buffer	1,400 mL of Milli-Q H_2O 200 mL of 1X TB Buffer 400 mL of Methanol	13 mM Tris base 4.5 mM boric acid 20% (v/v)	Room Temperature
Lysis Buffer	20 µL of 10X RIPA buffer 1 µL of PMSF (200mM) in DMSO 179 µL of Sample	0.05M Tris-HCl pH 7.4 0.15M NaCl 0.25% deoxycholic acid 1% NP-40 1mM EDTA 1mM	-20°C
TaqMan PCR	10 µL of TaqMan Universal PCR master mix 1 µL of TaqMan® Assay (20X) 5 µL of RNase-free water 4 µL of Template cDNA		-20°C
miRCURY PCR	5 µL of 2X SYBR Green Master Mix 0.5 µL of ROX Reference Dye 1 µL of miRCURY LNA PCR assay 0.5 µL of RNase free water 3 µL of Template cDNA		-20°C

2.2 Methods

2.2.1 Preparation of samples for cell-free RNA analysis.

500 µL bile samples were thawed on ice, diluted with 500 µL PBS and centrifuged at 300 x g for 10 minutes at 4°C twice, taking the supernatant and discarding the pellets. The supernatant was then centrifuged at 16,000 x g for 20 minutes at 4°C twice with the first pellet stored for later protein analysis (labelled 16k). Samples were then passed through a 0.45 µm filter (Polyethersulfone) and processed immediately. 1 mL plasma samples were thawed on ice and then to room temperature before use.

2.2.2 Preparation of samples for EV isolation.

500 µL bile samples were thawed on ice, diluted with 500 µL PBS and centrifuged at 300 x g for 10 minutes at 4°C twice, taking the supernatant and discarding the pellets. The supernatant was then centrifuged at 16,000 x g for 20 minutes at 4°C twice with the first pellet stored for later protein analysis (labelled 16k). 1 mL plasma samples were thawed on ice and then to room temperature before EV isolation.

2.2.3 EV Isolation using ultracentrifugation (UC).

Samples for ultracentrifugation were added to 10mL polypropylene oak ridge bottles (Sorvall; 03929) and diluted with 9 mL PBS. These were closed using a non-sealing cap assembly (Sorvall; 03924) and were placed in a fixed angle T-1270 titanium rotor and spun at 120,000g for 70 minutes at 4°C in a Sorvall Discovery 100S. Pellets were retained, diluted with 9 mL PBS and re-spun at 120,000g for 70 minutes at 4°C. These final pellets were retained, resuspended in 100 µL PBS and stored for later analysis.

2.2.4 EV Isolation using size exclusion chromatography (SEC).

Size exclusion chromatography (SEC) was undertaken using commercially available qEVoriginal 70nm SEC columns (Izon Science, UK) were used to isolate EVs according to the manufacturer's protocol. 1 mL of sample was loaded onto the column, followed by elution with PBS. The eluate was collected in 30 sequential fractions of 0.5 mL (labelled from 1 to 30), which were used immediately for EV characterisation and RNA extraction

or frozen at -80°C until required. For the purposes of analysis, unless otherwise stated, we deemed fractions 7-12 as EV-enriched (EV) and 22-24 as free protein-enriched (PROT) based on quantification of EVs, protein and visual inspection by TEM (see 4.2.10 Characterisation of EVs present in endoscopic bile and 5.2.3 Characterisation of EVs present in blood plasma).

2.2.5 EV isolation using a Plasma/Serum Exosome Purification mini kit.

EVs were isolated using the Plasma/Serum Exosomes Purification and RNA Isolation Mini Kit (Norgen BioTek, Ontario, Canada) according to the manufacturer's protocol. In brief, 1 mL plasma samples were thawed at 4°C and centrifuged at 400 x g for 2 minutes at RT. Following centrifugation, the plasma supernatant was mixed with 4 mL nuclease-free water and 100 µL of ExoC Buffer, followed by the addition of 200 µL of Slurry E. This was vortexed thoroughly for 10 seconds and left to rest at RT for 5 minutes. Following this, samples were centrifuged at 2000 rotations per minute (RPM) for 2 minutes at RT, the supernatant discarded and 200 µL of ExoR buffer added to the slurry pellet. Samples were left at RT to rest for 5 minutes before centrifugation once again at 500 RPM for 2 minutes at RT. The supernatant was transferred into a Mini Filter Spin column, assembled in an elution tube, and centrifuged for 1 minute at 6000 RPM at RT to collect the isolated EVs. The 200 µL EV eluate was either used immediately, or frozen at -80°C until required.

2.2.6 Nanoparticle tracking analysis (NTA).

Nanoparticle tracking analysis (NTA) was performed using a ZetaView PMX-120 (Particle Metrix, Germany) to determine particle concentration and size distribution down to a limit of 60nm. Polystyrene microspheres (100 nm) were used to configure and calibrate the instrument according to the manufacturer's protocol. Fractions were diluted 50-10,000-fold in PBS to ensure the number of particles per position were between 50-200. For each measurement, 11 positions were scanned and 30 frames per position captured under the following settings: Focus: autofocus; Camera sensitivity: 80; Shutter: 100; Scattering Intensity: detected automatically; Cell temperature: 19-24 °C; pH: 7.0. Following capture, videos were analysed by the in-built ZetaView Software 8.05.04, Camera 0.712 µm/px, with the following analysis parameters: Maximum area: 1000, Minimum area: 10; Minimum brightness; 30. Outlier measurements were automatically

detected using the software Grubb's test and excluded from the analyses. The inclusion of at least 9 camera positions was deemed a satisfactory measurement.

2.2.7 Protein quantification.

A Pierce Bicinchoninic Acid (BCA) Protein Assay Kit (Thermo Scientific, Waltham, USA) was used to quantify the protein concentration in each sample. The assay was performed according to the manufacturer's microplate protocol. In brief, triplicate volumes and PBS controls of 1-25 µL were added to 200 µL of Working Reagent in a Corning Costar 96 well plate (Fisher Scientific, Waltham, USA). The plate was mixed thoroughly for 30 seconds and incubated for 30 min at 37°C. Absorbance was measured at 562 nm using an OptiMax (Molecular Devices) tuneable microplate reader and a standard curve prepared using 2.5 to 2000 ng/µL of bovine serum albumin (BSA) according to the manufacturer's protocol. Measurements below 20 µg/mL (limit of assay) were excluded from analysis.

2.2.8 Western blotting.

Samples for analysis were concentrated using Vivaspin Turbo4 tubes with 10 kDa molecular weight cut-off (Sartorius, Göttingen, Germany) and spun at 4,000 x g for 2 hours at 4°C. Samples were then lysed on ice with the addition of 10X RIPA Buffer (Cell Signalling Technology, Danvers, USA) with protease inhibitor (See TABLE 7: Recipes). Lysed protein was quantified using a BCA Assay and equal amounts of protein (5-20 µg) were prepared with Pierce LDS 4X Sample Buffer (Thermo Scientific) and DTT as a reducing agent (0.1M final concentration) if required. Samples were heated at 95 °C for 5 minutes, separated on hand-cast 11% sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE) gels with running buffer (see TABLE 7) at 120 V for 90-120 minutes using Mini-PROTEAN Tetra Cell tanks (Bio-Rad, Hercules, USA).

These were transferred onto a nitrocellulose membrane in Transfer Buffer (see TABLE 7) using a Mini Trans-Blot cell (Bio-Rad) set to 90 V and run for 90 minutes.

Total protein bands were detected using Ponceau S solution or No-Stain Protein Labelling Reagent (Invitrogen, Waltham, USA) according to manufacturer's instruction. Membranes were blocked with 5% non-fat powdered milk in Tris-buffered Saline, 0.1% Tween-20 and incubated with primary antibodies overnight (>8 hours) at 4 °C on a roller. Membranes were then incubated with appropriate HRP-conjugated IgG secondary

antibodies (1:2500) for 1 hour at RT and washed with TBS-T. Protein expression was visualised with application of Immobilon Western Chemiluminescence HRP Substrate for 5 minutes with an iBright 1500 (Thermo Fisher Scientific). Images were analysed with Image Lab Software 6.0.1 (Bio-Rad) and membranes then stripped if necessary for repeat immunoblotting by incubating with Restore Plus Western Blot Stripping Buffer for 15 minutes at RT on a roller before re-blocking.

Table 8: Antibodies.

Antibody	Type	Order Number	Company	Storage Conditions
ALIX	Rabbit Monoclonal	#92880	Cell Signaling Technology	-20°C
ApoA1	Mouse Monoclonal	#12C8	AbCam	-20°C
CALNEXIN	Mouse Monoclonal	ab10286	AbCam	-20°C
CD63	Mouse Monoclonal	MX-49.129.5	Santa Cruz	4°C
CD81	Mouse Monoclonal	sc-166029	Santa Cruz	4°C
CD9	Mouse Monoclonal	sc-13118	Santa Cruz	4°C
GAPDH	Rabbit Monoclonal	# 5174	Cell Signaling Technology	-20°C
GPC-1	Rabbit Polyclonal	pa5-24972	Thermo Fisher/ Invitrogen	-20°C
HSP70	Mouse Monoclonal	ab2787	AbCam	-20°C
TSG101	Mouse Monoclonal	ab83	Abcam	-20°C
Anti-mouse	Goat 1nm gold conjugate	EM.GAM10/2	BBI Solutions	4°C
Anti-mouse	HRP-linked	7076	Cell Signaling Technology	-20°C
Anti-rabbit	HRP-linked	70745	Cell Signaling Technology	-20°C

2.2.9 Transmission electron microscopy.

All preparation and imaging were performed at the Electron Microscopy Centre, South Kensington Campus, Imperial College London. In brief, 8 µL samples were deposited onto 200-300 mesh glow-discharged copper grids for 2 minutes at RT, washed twice with Milli-Q water and then negatively stained with 2% uranyl acetate. The grids were observed using a Tecnai T12 TWIN (FEI Company, Hillsboro, USA) transmission electron microscope at magnification ranges 4,000-120,000x.

Immunogold labelling was performed by Dr Padmini Sarathchandra at the Magdi Yacoub Institute Heart Science Centre. In brief samples were incubated with 4% Paraformaldehyde (PFA) and 5 µL deposited on 200 mesh copper grids. Samples were washed with 0.05M PBS-glycine and blocked with 1% BSA before incubation with primary antibody. Grids were then washed twice with PBS and labelled with secondary goat anti mouse IgG antibody conjugated with 10 nm gold particles before negative staining with 2% uranyl acetate.

2.2.10 EV concentration and preparation for RNA extraction

Samples obtained by SEC for RNA extraction were concentrated using Vivaspin Turbo 4 centrifugal concentrators with 10 kDa molecular weight cut-off (Sartorius). A fresh device was used for each sample, loaded with an equal volume and spun at 4,000 g for 1 hour at 4°C. Samples were transferred immediately to a 1.5 mL Eppendorf safe lock microtube and weighed using a fine weighing scale. These values were plotted against a standard curve generated using PBS and an estimated volume calculated for subsequent reactions.

2.2.11 Total RNA extraction of bile samples using TRIzol LS reagent.

Total RNA from bile samples were extracted using the Invitrogen TRIzol LS Reagent (Thermo Fisher Scientific) according to manufacturer's protocol (with additional chloroform and ethanol wash). To avoid contamination with RNase, filter tips and sterile Eppendorf tubes were used, and benches and work gloves were prepared for RNA extraction using RNase Zap RNase decontamination solution (Thermo Fisher Scientific).

TRIzol LS Reagent was added to samples at a volumetric ratio of 3:1 and incubated for 5 minutes at RT to permit complete dissociation of the nucleoprotein complex. To this 0.2 mL of chloroform per 0.75 mL of TRIzol LS Reagent was added and incubated for 5 minutes at RT. Separation of the mixture into a lower red phenol-chloroform, a cloudy protein interphase and a colourless upper aqueous phase was assisted by centrifugation at 12 000 x g at 4°C for 15 minutes. The upper aqueous phase containing RNA was transferred to a separate Eppendorf tube and 0.2 mL of chloroform per 0.75 mL of TRIzol LS Reagent was added again with incubation for 5 minutes at RT. The two colourless phases were separated by centrifugation at 12 000 x g at 4°C for 15 minutes and the upper aqueous phase was transferred to a separate Eppendorf tube.

To precipitate the RNA, 0.5 mL of isopropanol to the aqueous phase, per 0.75 mL of TRIzol LS Reagent was added to the aqueous phase, incubated at RT for 5 minutes and stored at -80°C overnight (>8 hours). Samples were thawed at RT for 10 minutes and centrifuged at 14,000 x g for 30 minutes at 4°C. Total RNA precipitates formed a white pellet at the bottom of the tube and the supernatant carefully discarded with a micropipette. The pellet was washed with 0.25 mL of 80% ethanol per 0.75 mL of TRIzol LS Reagent, vortexed and centrifuged at 7500 x g for 5 minutes at 4°C. This wash was repeated with 0.25 mL of 80% ethanol per 0.75 mL of TRIzol LS Reagent, vortexed and centrifuged at 7500 x g for 5 minutes at 4°C. Eppendorfs containing RNA pellet was dried at RT for 20 minutes and

resuspended in 15 µL RNase-free water, transferred to a new Eppendorf tube, labelled and stored at -80°C for future analysis.

For the RNA extraction from Bile EVs, the above TRIzol method was used until separation of the aqueous phase. At this point, the protocol was optimised for RNA precipitation, using the miRNeasy micro kit (QIAGEN, Venlo, The Netherlands) according to the manufacturer's protocol. 0.3 mL of 100% ethanol per 0.75 mL of TRIzol LS Reagent was added to the aqueous phase and vortexed for 15 seconds. Up to 700 µL of the sample was added to a RNeasy MinElute spin column and centrifuged at 10,000 x g for 15 seconds at RT and the flow-through discarded. This was repeated using the remainder of the sample in 700 µL aliquots. Optional on-column DNase digestion was not performed. 700 µL of Buffer RWT was added to the RNeasy Mini spin column and centrifuged at 10,000 x g for 15 seconds at RT with the flow-through discarded. 500 µL of Buffer RPE was added to the RNeasy Mini spin column and centrifuged at 10,000 x g for 15 seconds at RT with the flow-through discarded. Finally, 500 µL of 80% ethanol was added to the RNeasy Mini spin column and centrifuged at 10,000 x g for 2 minutes at RT with the flow-through discarded. RNA remaining in the spin column was transferred to a fresh 2 mL collection tube after drying the membrane by spinning the column with the lid open at 16 000 x g for 5 minutes. Final EV-RNA was eluted with 12 µL of RNase-free water and used immediately, or frozen at -80°C until required.

2.2.12 Total and EV-RNA Extraction of plasma samples using the miRNeasy Serum/Plasma advanced kit (QIAGEN).

Total RNA from plasma samples (total and EV) were extracted using the phenol-free miRNeasy Serum/Plasma advanced kit according to manufacturer's instructions. Stations were prepared for RNA extraction using RNaseZap RNase decontamination solution (Thermo Fisher Scientific).

Lysis involved addition of proprietary Buffer RPL at a volumetric ratio of 10:3 and vortexed before incubation at RT for 3 minutes. 1 µL of Buffer RPP per 3 µL of Buffer RPL was added, vortexed for 20 seconds and incubated at RT for 3 minutes. Samples were centrifuged at 12,000 x g for 3 minutes at RT to obtain a pellet and an equal volume (Sample Volume + Buffer RPL + Buffer RPP) of isopropanol added to the supernatant. Up to 700 µL of the sample was transferred to a RNeasy UCP MinElute spin column (QIAGEN) and centrifuged at 8,000 x g for 30 seconds at RT. This was repeated using the

remainder of the sample in 700 µL aliquots. Optional on-column DNase digestion was not performed. 700 µL of Buffer RWT was added to the RNeasy UCP MinElute spin column and centrifuged at 8,000 x g for 30 seconds at RT with the flow-through discarded. 500 µL of Buffer RPE was then added to the RNeasy UCP MinElute spin column and centrifuged at 8,000 x g for 30 seconds at RT with the flow-through discarded. Finally, 500 µL of 80% ethanol was added to the RNeasy UCP MinElute spin column and centrifuged at 8,000 x g for 2 minutes at RT with the flow-through discarded. RNA remaining in the spin column was transferred to a fresh 2 mL collection tube after drying the membrane by spinning the column with the lid open at 16 000 x g for 5 minutes. RNA was eluted with 14 µL of RNase-free water and used immediately, or frozen at -80°C until required.

2.2.13 Library preparation and next generation RNA Sequencing (performed externally)

RNA samples were packaged on dry ice and sent via temperature-controlled courier (Biocair, Cambridge, UK) for small RNA sequencing by an external company. This was performed by Dr Morten Venø and the team at Omiics (Aarhus, Denmark) using Qiagen's QIAseq small RNA Library Prep kit according to manufacturer protocol. 5 µL of the RNA eluate was used in low input samples (i.e. EV-RNA). The finished libraries were quality controlled using an Agilent Bioanalyzer 2100 and quantified by use of qPCR. Libraries were pooled and sequenced on an NextSeq 500 system (Illumina, San Diego, USA) using the default single-end 75 base pair protocol to include integrated unique molecular indices (UMIs).

The raw data was processed using fastx_toolkit, and Illumina adaptor sequences were removed using cutadapt (292). Unmapped reads were mapped to miRNAs from miRBase v22 allowing zero mismatches but allowing for non-templated 3' A and T bases. Quality control was performed using FastQC (Babraham Bioinformatics) to ensure high quality data. The miRNA read counts were subjected to differential expression analysis and normalisation using DESeq2 in R (293). Volcano plots and other data visualisation for the RNA-Seq data was undertaken using R (Version 5 January 2015).

tRNA degradation/cleavage products (due to size restriction within the library preparation step) were also analysed by differential expression analysis and normalization using DESeq2 in R. All detected tRNA fragments were compared to tRNAs in MINTbase v2.0 (Mitochondrial and Nuclear tRNA fragment database, an online repository of tRNA fragments found in human tissues).

2.2.14 RNA quantification by RNA-specific Fluorometry.

Quantification of Total RNA was undertaken using Qubit 2.0 Fluorometer according to manufacturer's instructions. In brief, working reagent containing an RNA-specific dye was made by diluting the Qubit RNA broad range (BR) reagent in buffer. 500 µL polypropylene tubes (Qubit assay tubes) were prepared with 190 µL of working reagent and 10 µL of standards containing either 100 ng/µL of RNA or no RNA. 2 µL of sample was pipetted into 198 µL of working reagent and measured using the fluorometer. Concentrations were given between 0.25 to 1000 ng/µL and although the reported lower limit of the BR assay was 1 ng/µL, we have reported all expressed values.

2.2.15 RNA quality and quantity analysis using Electrophoresis.

RNA was quantified and (where appropriate) quality checked using an RNA 6000 Pico assay (sensitivity 50 – 5000 pg/µL) and the Agilent Bioanalyzer 2100 instrument according to manufacturer instructions. In brief, 9 µL of pre-prepared gel-dye aliquot (1 part RNA 6000 Pico dye concentrate to 65 parts filtered gel) were vortexed and spun down at 13 000 x g for 10 minutes at RT and loaded into the appropriate well of an RNA 6000 Pico chip. The chip was primed, and solutions added to appropriate wells (9 µL conditioning mix, 5 µL of marker, 1 µL of sample and 1 µL of RNA ladder) before vortexing for 60 seconds. Samples were denatured at 70°C for 2 minutes prior to measurement. Electrophoresis and measurement of total RNA was controlled using the 2100 Expert Software using 'Eukaryotic Total RNA Pico Series II' settings.

2.2.16 Quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR) using target-specific stem–loop primer assays (TaqMan) to detect miRNAs.

RNA was reverse transcribed using the TaqMan MicroRNA Reverse Transcription (RT) kit (Thermo Fisher Scientific formerly Applied Biosystems). Total RNA (5 ng) made up to 2.5 µL was mixed with 3.5 µL of RT master mix (according to manufacturer instructions) and 1.5 µL of microRNA-specific 5X RT TaqMan stem–loop primers (Thermo Fisher Scientific) in a 48-well PCR plate. Samples were incubated in a Tetrad 2 96 Well Thermal Cycler (MJ Research, Deltona, USA) at 16°C for 30 minutes to allow primer annealing, followed by 30 minutes at 42°C for the elongation step and 5 minutes at 85°C to inactivate

the reverse transcriptase enzyme. Specific miRNA TaqMan probes (Thermo Fisher Scientific) were used in the reverse transcription (see TABLE 7) to generate specific complementary DNA (cDNA) for each sample.

Quantitative PCR reactions were performed in duplicates using miRNA specific TaqMan probes and multiple endogenous controls following manufacturer's instructions (Thermo Fisher Scientific). In brief, 4 µL of miRNA-specific cDNA samples was added to 10 µL of TaqMan Universal PCR master mix (2X, no uracil-N-glycosylase), 1 µL of TaqMan Assay (20X) miRNA probe assay and 5 µL of RNase-free water on a MicroAmp Fast Optical 96-well reaction plate (Thermo Fisher Scientific). This was sealed using a MicroAmp Optical Adhesive Film and run through TaqMan specific PCR cycling conditions on a StepOne Plus machine (Applied Biosystems). This involved an initial step to denature the cDNA: 50°C for 2 minutes, 95°C for 10 minutes, followed by 40 cycles of (1) denaturation at 95°C for 15 seconds and (2) annealing/elongation at 60°C for 60 seconds. Measurement of fluorescence was undertaken after each cycle and the threshold cycle (C_T) determined by the built-in software. C_T values that were discordant (>0.5 apart) were repeated as a quintuplet and the average used for further analysis. MiRNAs with a C_T value > 40 were deemed not detected and set at 40 for calculation of relative expression. All plates included no-template controls with any plate with a false positive repeated to ensure accuracy.

2.2.17 Quantitative Reverse Transcription Polymerase Chain Reaction (RT-qPCR) using universal reverse transcription (RT) and locked nucleic acid (LNA) enhanced specific primers with SYBR green I assay detection to detect miRNAs.

Reverse transcription of RNA samples was undertaken using the miRCURY LNA RT kit (QIAGEN) for detection of low RNA inputs without preamplification. 1 µL Total RNA was mixed by volume with 1 µL of RT Probe Reaction Buffer (according to manufacturer's protocol, modified to use 5 µL volume), 0.25 µL of 10X miRCURY RT Enzyme Mix, 0.25 µL of UniSp6 RNA (10^8 copies/µL) spike-in and 2.5 µL of RNase-free water in a 48-well PCR plate. No-RT controls omitted the enzyme mix for RNase-free water. Samples were incubated in a Tetrad 2 96 Well Thermal Cycler at 40°C for 60 minutes to allow primer annealing/elongation, 5 minutes at 85°C to inactivate the reverse transcriptase and stored at 4°C.

Quantitative PCR reactions were performed in triplicate using the miRCURY LNA SYBR Green PCR Kit (QIAGEN) following manufacturer's instructions. In brief, cDNA samples

were diluted 1:30 with RNase-free water and 3 µL sample added to 5 µL of 2X SYBR Green Master Mix, 0.5 µL of ROX Reference Dye, 1 µL of miRCURY LNA miRNA-specific PCR assay and 0.5 µL of RNase free water in a MicroAmp Fast Optical 96-well Reaction plate (Thermo Fisher Scientific). This was sealed using a MicroAmp Optical Adhesive Film and run through miRCURY LNA specific PCR cycling conditions on a StepOne Plus machine (Applied Biosystems). This was a 'hot-start' protocol involved an initial step to denature the cDNA at 95°C for 2 minutes, followed by 40 cycles of denaturation at 95°C for 10 seconds and annealing/elongation at 56°C for 60 seconds. Measurement of fluorescence was undertaken after each cycle and the threshold cycle (C_T) determined by the built-in software. C_T values that were discordant (>0.5 apart) were repeated and the average used for further analysis. MiRNAs with a C_T value > 40 were deemed not detected and set at 40 for calculation of relative expression. All plates included no-template and no-RT controls with any plate with a false positive repeated to ensure accuracy.

Table 9: RT-qPCR Primers.

LNA – Locked Nucleic Acid

miRNA	Type	Order Number	Company	Storage Conditions
hsa-mir-26a-5p	Hydrolysis probe	000405	Applied Biosystems	-20°C
hsa-let-7b-5p	Hydrolysis probe	002619	Applied Biosystems	-20°C
hsa-mir-196a-5p	Hydrolysis probe	241070_mat	Applied Biosystems	-20°C
hsa-mir-199a-5p	Hydrolysis probe	000498	Applied Biosystems	-20°C
hsa-mir-182-5p	Hydrolysis probe	002334	Applied Biosystems	-20°C
hsa-mir-335-5p	Hydrolysis probe	000546	Applied Biosystems	-20°C
hsa-mir-340-5p	Hydrolysis probe	002258	Applied Biosystems	-20°C
hsa-mir-424-5p	Hydrolysis probe	000604	Applied Biosystems	-20°C
hsa-let-7a-5p	LNA probe	YP00205727	Qiagen	-20°C
hsa-mir-21-5p	LNA probe	YP00204230	Qiagen	-20°C
hsa-mir-23a-3p	LNA probe	YP00204772	Qiagen	-20°C
hsa-mir-26a-5p	LNA probe	YP00206023	Qiagen	-20°C
hsa-miR-93-5p	LNA probe	YP00204715	Qiagen	-20°C
hsa-miR-125a-5p	LNA probe	YP00204339	Qiagen	-20°C
hsa-mir-126-5p	LNA probe	YP00206010	Qiagen	-20°C
hsa-mir-141-3p	LNA probe	YP00204504	Qiagen	-20°C
hsa-miR-151a-3p	LNA probe	YP00204576	Qiagen	-20°C
hsa-mir-181a-2-3p	LNA probe	YP00204142	Qiagen	-20°C
hsa-mir-182-5p	LNA probe	YP00206070	Qiagen	-20°C
hsa-mir-194-5p	LNA probe	YP00204080	Qiagen	-20°C
hsa-mir-200a-3p	LNA probe	YP00204707	Qiagen	-20°C
hsa-mir-200b-3p	LNA probe	YP00206071	Qiagen	-20°C
hsa-mir-200c-3p	LNA probe	YP00204482	Qiagen	-20°C
hsa-mir-320a-3p	LNA probe	YP00206042	Qiagen	-20°C
hsa-mir-320b	LNA probe	YP02119299	Qiagen	-20°C
hsa-mir-320c	LNA probe	YP00205706	Qiagen	-20°C
hsa-mir-320d	LNA probe	YP00205667	Qiagen	-20°C
hsa-mir-429	LNA probe	YP00205901	Qiagen	-20°C
hsa-mir-451a	LNA probe	YP02119305	Qiagen	-20°C
cel-mir-39	LNA probe	YP00203952	Qiagen	-20°C
UniSp6	LNA probe	YP00203954	Qiagen	-20°C

2.2.18 Calculation of relative expression.

Fold expression levels of candidate miRNAs were calculated using the comparative ($\Delta\Delta C_T$) method and relative to the geometric mean of reference genes determined for each experiment (EndoMean). EndoMean was the geometric mean (the nth root of the product of n numbers) of values obtained for miRNAs determined as stably expressed and relative expression levels of each miRNA were calculated using the following equation:

$$2^{-\Delta C_T}, \text{ where } C_T = C_T \text{ target} - C_T \text{ EndoMean}$$

These values were further binary logarithm transformed and shown as Log fold expression.

2.2.19 Generation of endogenous controls for normalisation.

Reference genes for normalisation were used for the above calculation and determined using RNA-Seq data using the NormFinder package for R (Version 5 January 2015). The top four candidates (by stability) were compared using RT-qPCR where possible and again analysed using NormFinder package for Microsoft Excel (Version 20, May 2010). The "stability value" is briefly summarised as a number which reflects the intra- and inter-group variances of an expressed gene with the lowest number identifying the gene with the lowest overall variance and inter-group difference according to Andersen *et al.* (294).

2.2.20 Statistical Analysis.

Parametric and non-parametric data was analysed using either the Student t-test or Mann–Whitney U test respectively and used to calculate a p-value to determine statistically significant differences between groups. Categorical variables (e.g. Age > 60, Gender) for the clinical data in pairwise comparisons were analysed using Fisher's exact test. Statistical analyses including descriptive statistics were performed using GraphPad Prism v9.1.2. Volcano plots and receiver operating characteristic (ROC) curve analysis were conducted for each significantly deregulated EV miRNA to understand their ability to discriminate PDAC and CCA patients from healthy controls, leading to estimates of area under the curve (AUC) with 95% confidence intervals (CI).

Multiple logistic regression was undertaken for candidates individually and in combination using GraphPad Prism. Data was prepared as a binary outcome and all main effects are included in the model. Optimum cut-offs were determined using thresholds obtained from the ROC curve at the maximum Youden index. Youden's J statistic (also called Youden's index) is a single statistic that ranges from 0 to 1 and is determined by the formula (Specificity + Sensitivity -1). Where multiple hypotheses were tested, an appropriate Benjamini–Hochberg (False Discovery Rate) correction was applied to give an adjusted p-value. A p (adjusted) value of <0.05 was considered to be statistically significant.

Correlation analysis used Spearman Rank for non-parametric data with p-values calculated as a two-tailed analysis. Heatmaps were generated using GraphPad Prism v9.1.2.

2.2.21 MicroRNA Functional Analysis using a microRNA enrichment analysis and annotation tool (miEAA)

MiEAA is a web-based application (http://www.ccb.uni-saarland.de/mieaa_tool/) that offers a variety of commonly applied statistical tests such as over-representation analysis and miRNA set enrichment analysis, which is similar to Gene Set Enrichment Analysis (GSEA). To determine potential functional targets for the candidates identified (sorted by AUC values) we used the GSEA to evaluate pathways, Gene Ontology terms, and targets.

The input list of miRNAs was sorted in order of p-value significance according to the validation data. Validated miRNA-target interactions were obtained from miRWalk 2.0 (downloaded in 2015), HMDD2 and miRTarbase (Release 4.5; Nov. 1, 2013) converted to the current miRBase v21. P-values were computed by applying the Fisher's exact test and adjusted for multiple comparisons (Bonferroni) with a standard significance threshold of $p<0.05$.

Chapter Three – Cell-free small RNA markers for PDAC and CCA in Bile and Plasma

3.1 Background

Obstructive jaundice (defined as a laboratory value of serum bilirubin that is greater than 30 μmol/L) is a clinical sign that can indicate that there is an occlusion or narrowing of the bile ducts. The most common malignant cause for this is pancreatic ductal adenocarcinoma (PDAC) present in the head of the pancreas where the bile duct passes through to the duodenum. Other causes include cancer of the bile duct (Cholangiocarcinoma; CCA) and ampullary cancer. Patients who present with this sign undergo work up which involves blood tests, imaging by CT and/or MRI and endoscopic evaluation. Current serum biomarkers such as CA 19-9 have a sensitivity and specificity of 78.2% and 82.8% for discriminating pancreatic carcinoma from benign disease in a population of concern when a threshold of 40 U/ml is applied (295). This is comparable to the other modalities which have a sensitivity/specificity of diagnosis of malignancy of 85%/71% for MRCP and 77%/63% for CT. Endoscopic retrograde cholangiopancreatography (ERCP) adds further diagnostic information and the quoted figures for sensitivity and specificity are 85%/75%, and 79%/62% for Endoscopic Ultrasound (EUS) (296).

ERCP remains the optimal route to diagnosis enabling tissue or cells to be obtained via brush cytology, aspiration or intraductal forceps biopsy. Biomarkers from bile aspirated at this point has two benefits. Firstly, it is one of the first points of contact in the patient pathway to manage biliary obstruction due to malignancy or otherwise. And bile is intimately related to the tumour anatomically improving the potential to detect cancer related molecular markers. Our aim in this chapter was to therefore discover whether novel small RNAs for the detection of PDAC and CCA could be sourced from bile and plasma. Previous work by the lab had shown that RNA could be extracted from bile samples obtained surgically which were suitable for RNA sequencing and validated using RT-qPCR using established methods.

The MIRABILE (MIcroRNAs As BILE-based biomarkers in Pancreaticobiliary Cancers) study was a prospectively recruited study which started recruitment in 2017 at Hammersmith Hospital between 2nd November 2017 and 11th March 2020. Patients were consented for research prior to endoscopy and samples of blood plasma were taken and stored with bile obtained after cannulation of the common bile duct by ERCP. I present the work obtained from analysis of these samples which underwent RNA extraction for predictive biomarkers in pancreaticobiliary cancers.

3.2 Results

3.2.1 Sample Collection

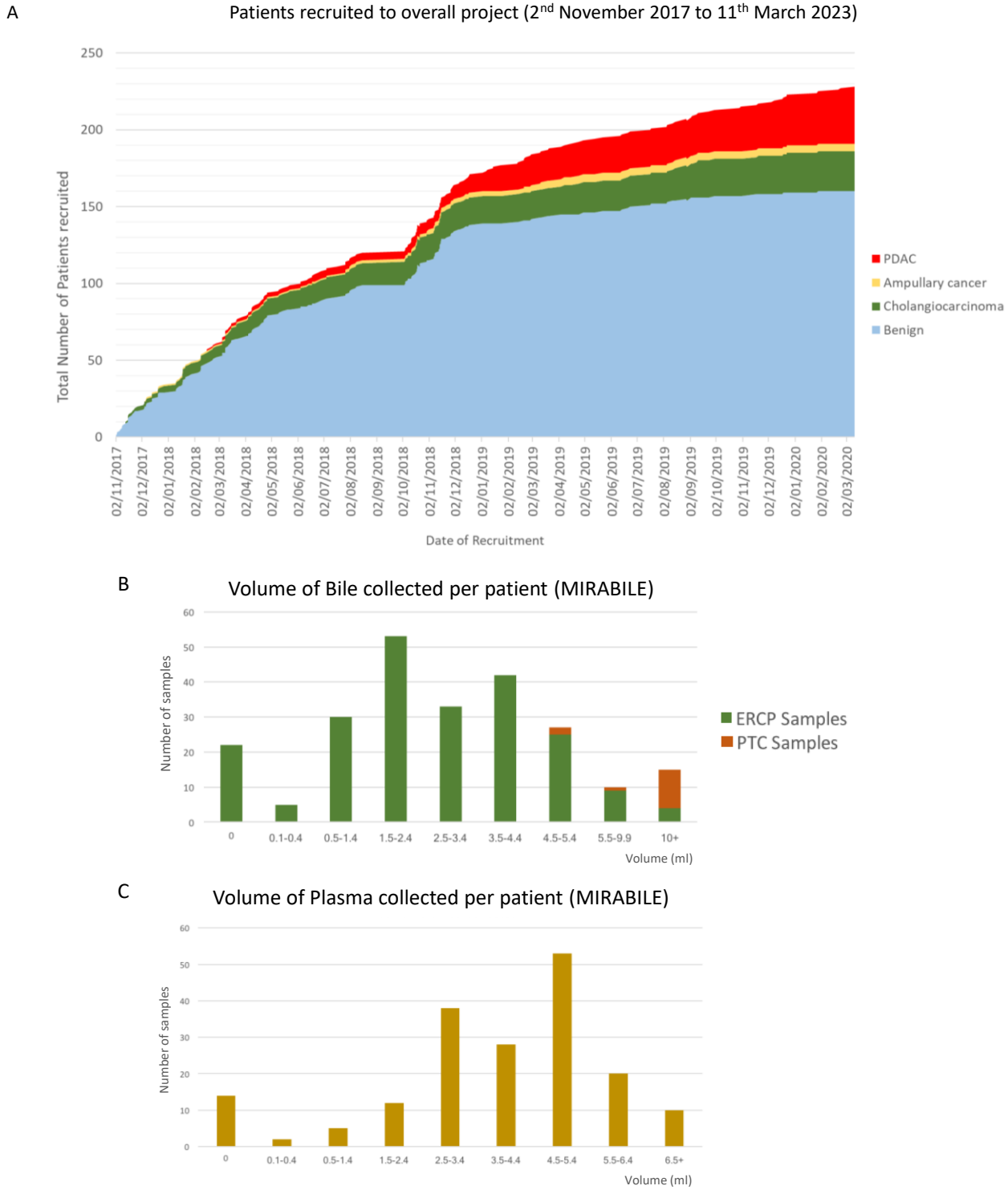


Figure 5: Patient Recruitment and Sample Collection.

A. Timeline showing prospective patient recruitment according to diagnosis for patients recruited to the MIRABILE study highlighting the large proportion of benign samples collected compared to cancer samples.
B. Bile sample collection statistics as a bar chart showing volumes in millilitres collected by endoscopy.

C. Plasma sample collection statistics as a bar chart showing volumes in millilitres collected by cannulation at the time of endoscopy.

PDAC – Pancreatic ductal adenocarcinoma, ERCP - Endoscopic retrograde cholangiopancreatography, PTC – Percutaneous transhepatic cholangiography

Samples were collected from 02/11/2017 until COVID-19 lockdown and the final sample for the project was collected 11th March 2020. Samples were collected prospectively from patients presenting to endoscopy (See METHODS AND MATERIALS) which led to a large proportion of patients recruited to the benign cohort. Figure 5A shows retrospective analysis of our prospective patient recruitment according to diagnosis for patients recruited to the MIRABILE study. The large proportion of benign patients recruited (in blue) can be compared to the smaller number of cancer patients presenting for endoscopic evaluation and as a result available for recruitment, which is typically the case (297).

A total of 241 unique patients in total were recruited, with 229 attempts to obtain bile samples and 183 attempts to obtain blood plasma samples. Figure 5B and 5C show sample collection statistics demonstrating volumes collected for bile and plasma respectively. 22 ERCPs did not result in any bile being obtained, and a further 5 samples obtained less than 0.5ml which was the minimum volume required for analysis. This meant that 12% of patients did not provide adequate sample volume for analysis. Excluding these from analysis, the average volume of bile obtained was 3.9ml. Lower than expected volumes was due to clinical limitations on the amount of time allowed within the ERCP session to aspirate bile via cannulation. An upper limit of five minutes was given to allow drainage, but this was not always achieved if patients poorly tolerated the procedure. Obtaining some samples by PTC allowed us to obtain a significantly larger volume (5-24ml) of bile depending on the adequacy of drainage and this is represented in orange (Figure 5B).

Plasma samples were obtained concurrently at the time of ERCP which meant that patients were starved for at least 6 hours prior to the sample collection. Blood was obtained in a 3.5ml collection tube containing EDTA which led to roughly 1.5ml of plasma per tube. Excluding the 16 patients that were recruited but did not have samples taken (8.7%) the average volume of plasma obtained was 4.1ml (Figure 5C).

Due to the lower-than-expected sample collection volumes, this raised difficulties for prospective RNA extraction and analysis as this would lead to unique samples with low volumes being potentially wasted early in the project. The decision was made to increase the sequencing discovery cohorts within both biofluids which therefore led to a reduction in samples available for validation.

Our *a priori* power calculation used a literature defined sensitivity of 85% for standard tests and an ideal test sensitivity of 95% for a bile-based miRNA signature (two-sided). Calculations using an α cut-off of 1% (0.01; where this indicates the likelihood of these results being due to chance and a false positive) and a ß cut-off of 10% (power=1-ß=90%; also known as the likelihood that we fail to reject the null hypothesis and generate a false negative result) estimated a minimum total sample size of 126 patients (i.e. 63 in each group).

3.2.2 Clinical differences in patient characteristics for the MIRABILE bile study

The details of the patient samples selected for each cohort or the analysis of RNA circulating markers in bile are summarised in TABLE 10. These include number of samples, age (greater than sixty), gender, and diagnosis of primary sclerosing cholangitis or chronic pancreatitis. Serum blood tests include CA 19-9, pre-endoscopic bilirubin, and C-reactive protein (CRP). Finally additional tumour parameters such as anatomical location, tumour stage, nodal status and metastatic disease are included. Determination of these values and cohort selection for analysis is detailed in the Materials & Methods.

Table 10: Patients selected for the discovery and validation cohorts used in the bile cfRNA biomarker analysis.

Overall demographic characteristics and recorded pathology results were collected retrospectively and analysed by Fisher's exact test for each cohort used for the bile cell-free microRNA biomarker analysis. PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, PSC – Primary sclerosing cholangitis

Patient Characteristics		Total (n=111)				Discovery (n=78)				Validation (n=87)			
		PDAC	CCA	Benign	P value	PDAC	CCA	Benign	P value	PDAC	CCA	Benign	P value
All cases (n, %)		**38**	**19**	**54**		**27**	**14**	**37**		**34**	**14**	**39**	
Age, *y*	<60	8 (21)	4 (21)	24 (44)		8 (30)	3 (21)	16 (43)		6 (18)	3 (21)	18 (46)	
	≥60	30 (79)	15 (79)	30 (56)	**0.0143**	19 (70)	11 (79)	21 (57)	0.1565	28 (82)	11 (79)	21 (54)	**0.0098**
	Mean (SD)	69 (11)	68 (8.5)	62 (16)	**0.0054**	68 (11)	65 (7.8)	63 (16)	0.1260	69.7 (11)	69.9 (7.6)	61.0 (16)	**0.0029**
Gender (%)	Female	13 (34)	6 (32)	21 (39)		12 (44)	4 (29)	15 (41		11 (32)	4 (29)	14 (36)	
	Male	25 (66)	13 (68)	33 (61)	0.5596	15 (56)	10 (71)	22 (59)	0.8254	23 (68)	10 (71)	25 (64)	0.6557
History of PSC (%)		0 (0)	1 (5)	4 (7)	0.2017	0 (0)	0 (0)	3 (8)	0.1105	0 (0)	1 (7)	4 (10)	0.1688
Chronic Pancreatitis (%)		4 (11)	0 (0)	11 (20)	0.0540	3 (11)	0 (0)	7 (19)	0.1858	4 (12)	0 (0)	9 (23)	0.0723
Pre-diagnostic blood parameters*	CA 19-9 (U/mL)	9341.5	3186	287	0.0770	7825	555.1	346.8	0.1023	10158	4298	287.1	0.0823
	Bilirubin (ng/mL)	248.6	148.0	42.1	**<0.0001**	243.6	109.8	45.7	**<0.0001**	253.6	148.8	48.2	**<0.0001**
	CRP (µg/mL)	50.2	56.8	47.7	0.9463	43.8	58.1	53.0	0.4931	52.2	47.4	37.9	0.7892
Anatomical location	Perihilar CCA	-	10 (53)	-		-	9 (64)	-		-	7 (50)	-	
	Distal CCA	-	7 (37)	-		-	5 (36)	-		-	5 (36)	-	
	Intrahepatic CCA	-	2 (11)	-		-	0 (0)	-		-	2 (14)	-	
	Head of Pancreas PDAC	31 (82)	-	-		21 (78)	-	-		27 (79)	-	-	
	Uncinate PDAC	3 (8)	-	-		3 (11)	-	-		3 (9)	-	-	
	Mixed PDAC (HOP/Uncinate)	4 (11)	-	-		3 (11)	-	-		4 (12)	-	-	
Tumour stage (%)	T1	0 (0)		-		0 (0)	0 (0)	-		0 (0)	0 (0)	-	
	T2	7 (18)	3 (17)	-		6 (22)	3 (21)	-		7 (21)	1 (7)	-	
	T3	10 (26)	10 (53)	-		7 (26)	8 (58)	-		8 (24)	7 (50)	-	
	T4	19 (50)	6 (33)	-		14 (52)	3 (21)	-		19 (56)	6 (43)	-	
Nodal disease (%)		23 (61)	8 (42)	-		18 (66)	6 (42)	-		23 (68)	5 (36)	-	
Metastatic disease (%)		11 (29)	1 (6)	-		6 (22)	0 (0)	-		11 (32)	1 (7)	-	

Statistical significance was determined using Fisher's exact test for categorical variables (Sex, Age >60) whilst an unpaired two-tailed Student t-test or Mann Whitney U test using Malignant (PDAC + CCA) vs Benign as the pairwise comparison. P-values <0.05 highlighted in bold. Key: HOP, Head of Pancreas; PDAC, pancreatic ductal adenocarcinoma; CA 19-9, Carbohydrate antigen 19-9; CRP, C-Reactive Protein.
*Blood results where available from patient electronic hospital records.

A total of 111 unique patient samples were used to investigate cell-free RNA (cfRNA) present in bile. The PDAC cohort included n=38 patients of which 34% were female and 66% were male. The mean age was 69 years with a standard deviation of 11 years. The CCA cohort included n=19 patients of which 32% were female and 68% were male. The mean age was 68 years with a standard deviation of 8.5 years. The Benign control cohort included n=54 patients of which 39% were female and 61% were male. The mean age was 62 years with a standard deviation of 16 years. These demographics are consistent with reported figures for PDAC (4) and CCA (298) in the UK being slightly more predominate in men. The cohort sampled appears younger than that reported nationally (peak age 85-89) which is likely due to the nature of our study (examining only those presenting to ERCP).

The discovery cohort comprised of 78 samples (27 PDAC, 14 CCA, 37 Benign) and underwent RNA sequencing. The validation cohort comprised of 87 samples (34 PDAC, 14 CCA, 39 Benign) and underwent RT-qPCR validation. 54 patients were used for both the discovery and validation analyses (23 PDAC 9 CCA 22 Benign; summarised as Venn diagrams in FIGURE 6A).

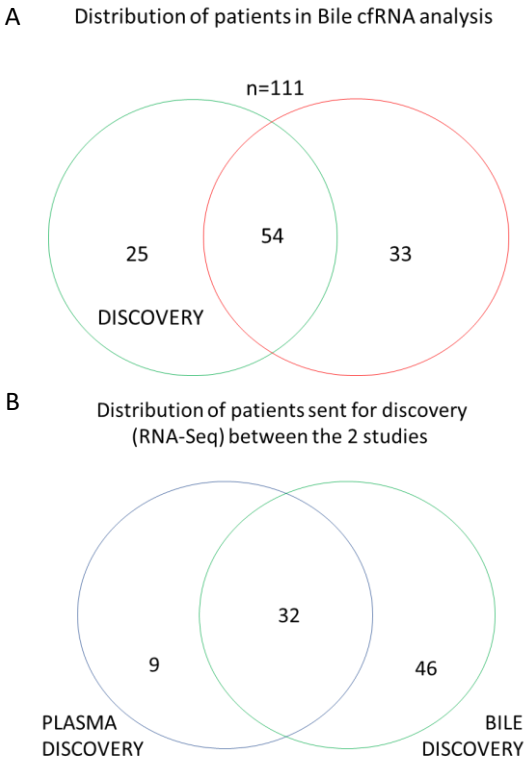


Figure 6: Patients selected for the discovery and validation cohorts used in the bile and plasma cfRNA biomarker analysis.

A. Venn diagram showing the distribution of patients between the discovery (green; sent for sequencing) and validation (red; underwent RT-qPCR) cohorts used in the analysis of bile cfRNA biomarkers.
B. Venn diagram showing the overlap of patients in each discovery cohort used in the analysis of plasma (blue circle) and bile (green circle) cfRNA biomarkers.

Malignant (PDAC+CCA) vs Benign comparison of the demographic data showed no statistical difference for age and sex in the discovery cohort as expected due to sample selection. Within the validation cohort there did appear to be a significant difference in mean age (P value 0.0029) but no difference in the sex ratio. This accounts for the overall statistically significant difference in the ages between our disease and control cohorts with an older population making up patients who present with PDAC and CCA (P 0.0054).

The mean CA 19-9 in the PDAC cohort was 9341.5U/mL, in the CCA cohort was lower at 3186U/mL and in the Benign cohort was 287U/mL. The mean serum bilirubin level in the PDAC cohort was 248.6ng/mL, in the CCA cohort was 148.0ng/mL and in the Benign cohort was 42.1ng/mL. The mean CRP in the PDAC cohort was 50.2µg/mL. in the CCA cohort was similar at 56.8µg/mL and in the Benign the mean CRP was 47.7µg/mL.

No difference was identified in the pre-therapeutic CA 19-9 overall and for the discovery cohort whilst the validation cohort did appear to note a significantly raised CA 19-9 (P value <0.0001) in the Malignant population. Bilirubin levels were significantly raised (P value <0.0001) in the Malignant samples within all cohorts whilst CRP levels were found to be no different across all patients. Shown in Figure 7 is a scatter plot of the serum bilirubin against CA 19-9 for each patient with results available. This shows visually that benign cases broadly had a lower bilirubin and clustered with a CA 19-9 <20 (7 patients). When serum bilirubin increased above 80 (dashed line) and above the ability to separate the PDAC and CCA from benign patients is not possible.

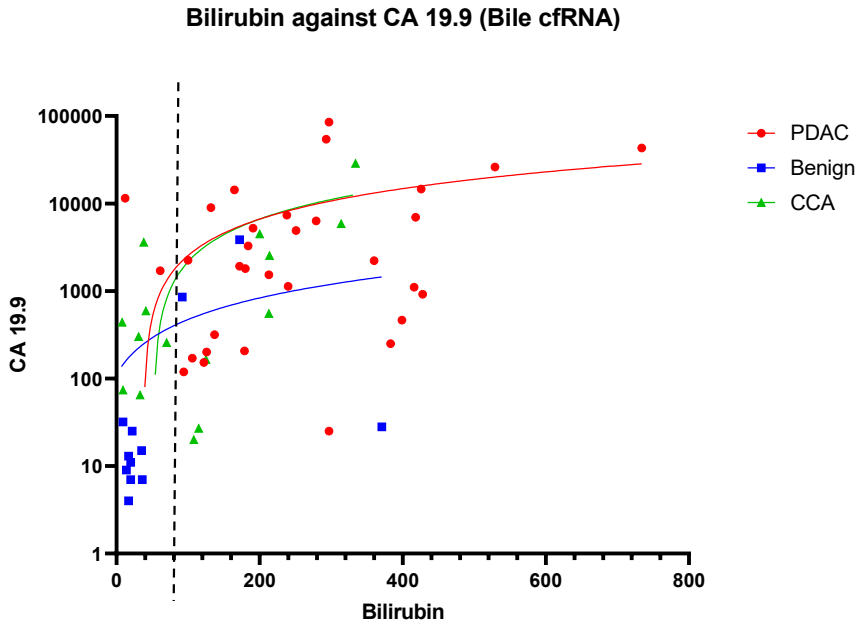


Figure 7: Examining the relationship between serum bilirubin and CA 19-9 in the overall population for the bile cfRNA analysis (n=66).

Shown is a scatter plot of the serum bilirubin against CA 19-9 (expressed logarithmically to better demonstrate spread) for each patient in the overall bile cfRNA analysis. Incomplete data was excluded (n=45) and any null values cannot be charted on a log axis (n=7). Dashed line shows serum bilirubin level of 80, beyond which the data points for each cohort cannot be separated. Trendlines were calculated using GraphPad Prism and are displayed in the colour (red for PDAC, blue for Benign and green for CCA.

Examining the cancer phenotypes of the overall population analysed showed that most cancers causing biliary obstruction anatomically were located in the head of pancreas (82%) with the remainders Uncinate (8%) or a mixture of the two (11%). CCA included mainly perihilar but also distal and intrahepatic. Tumour staging was predominantly T4 (50% in PDAC and 33% CCA) due to the nature of the patient recruitment process with a small proportion T2 (17-18%; PDAC and CCA). A high degree of nodal disease was present in all cohorts (61% in PDAC, 42% in CCA) and metastatic disease was present in 29% of PDAC patients, 6% of CCA patients.

3.2.3 Clinical differences in patient characteristics for the MIRABILE plasma study

The details of the patient samples selected for each cohort or the analysis of cfRNA in plasma are summarised in TABLE 10. A total of 41 unique patient samples were used for this and underwent RNA sequencing, corresponding to 26 patients in the PDAC cohort, 5 patients in the CCA cohort and compared to 20 age and sex matched patients in the Benign cohort.

The PDAC cohort included n=26 patients of which 35% were female and 65% were male. The mean age was 67 years with a standard deviation of 11 years. The CCA cohort included n=5 patients of which 20% were female and 80% were male. The mean age was 68 years with a standard deviation of 3 years. The Benign control cohort included n=10 patients of which 40% were female and 60% were male. The mean age was 63 years with a standard deviation of 14 years.

FIGURE 6B summarises the overlap between the patients used for the plasma analysis with those selected for the bile discovery cohort (n=76) showing that 32 patients were part of both bile and plasma studies (18 PDAC, 5 CCA, 9 Benign).

Table 11: Patients selected for the discovery cohort used in the plasma cfRNA biomarker analysis.

Overall demographic characteristics and recorded pathology results were collected retrospectively and analysed by Fisher's exact test for the discovery cohort of the plasma cfRNA study (no validation was performed).
PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, PSC – Primary sclerosing cholangitis

Patient Characteristics		MIRABILE Plasma Discovery			
		PDAC	CCA	Benign	P valu
All cases (n, %)		**26**	**5**	**10**	
Age, *y*	<60	6 (23)	0 (0)	5 (50)	
	≥60	20 (77)	5 (100)	5 (50)	0.098
	Mean (SD)	67 (11)	68 (3)	63 (14)	0.296
Gender (%)	Female	9 (35)	1 (20)	4 (40)	
	Male	17 (65)	4 (80)	6 (60)	0.711
History of PSC (%)		0 (0)	0 (0)	0 (0)	1.000
Chronic Pancreatitis (%)		2 (8)	0 (0)	0 (0)	1.000
Pre-diagnostic blood parameters*	CA 19-9 (U/mL)	8111.6	1296.0	435	0.223
	Bilirubin (ng/mL)	239.8	224.0	43.6	**<0.000**
	CRP (µg/mL)	61.5	28.5	69.2	0.588
Anatomical location	Perihilar CCA	-	3 (60)	-	
	Distal CCA	-	2 (40)	-	
	Intrahepatic CCA	-	0 (0)	-	
	Head of Pancreas PDAC	20 (77)	-	-	
	Uncinate PDAC	2 (8)	-	-	
	Mixed PDAC (HOP/Uncinate)	4 (15)	-	-	
Tumour stage (%)	T1	0 (0)	0 (0)	-	
	T2	4 (15)	2 (40)	-	
	T3	3 (12)	3 (60)	-	
	T4	19 (73)	0 (0)	-	
Nodal disease (%)		14 (54)	3 (60)	-	
Metastatic disease (%)		10 (38)	0 (0)	-	

Statistical significance was determined using Fisher's exact test for categoric variables (Sex, Age >60) whilst an unpaired two-tailed Student t-test or Mann Whitne U test using Malignant (PDAC + CCA) vs Benign as the pairwise comparison. P-valu <0.05 highlighted in bold.

Key: HOP, Head of Pancreas; PDAC, pancreatic ductal adenocarcinoma; CA 19-Carbohydrate antigen 19-9; CRP, C-Reactive Protein.

*Blood results where available from patient electronic hospital records.

The mean CA 19-9 in the PDAC cohort was 8111.6U/mL, in the CCA cohort was lower at 1296U/mL whilst in the Benign cohort was only 435U/mL. The mean serum bilirubin level in the PDAC cohort was 239.8ng/mL, in the CCA cohort was similar at 224.0ng/mL whilst the level in the Benign cohort was lowest at 43.6ng/mL. The mean CRP in the PDAC cohort was 61.5µg/mL in the CCA cohort was lowest at 28.5µg/mL whilst in the Benign cohort was slightly elevated at 69.2µg/mL.

Bilirubin levels were significantly raised (P value <0.0001) when comparing Malignant i.e. PDAC and CCA against Benign but CRP levels were not deemed significantly different between these cohorts as in the bile study. Intriguingly, CA 19-9 was elevated but not deemed significantly raised when Malignant (PDAC & CCA) were compared with Benign (P value 0.2230).

Cancer profile once again showed that anatomically the majority were located in the head of pancreas (77%) with the remainder uncinate (8%) or a mixture of the two (15%). The CCA cohort was comprised of only perihilar and distal extrahepatic cancers. Tumour staging was predominantly T4 (73% in PDAC) with a small proportion T2 (15%; PDAC). The patients selected for analysis comprising the CCA cohort was small (n=5) and consisted of a mixture of T3 (60%) and T2 (40%) disease.

3.2.4 Cell-free RNA analysis shows differences in RNA quantity and composition.

Samples once selected were taken forward for total RNA extraction using the TRIzol LS method (See Section 2.2.11) and quantity measured using both fluorometric quantification of the RNA (Qubit 2.0 Fluorometer, Thermo Fisher) and RNA electrophoresis (Agilent 2100 Bioanalyzer). These methods provide an increased degree of accuracy over UV absorbance (such as Nanodrop) due to presence of a specific RNA-binding dye which allows detection of RNA without DNA contamination in the range of 0.2 to 200 ng/µL (compared to 2 ng/µL to 15 µg/µL for the NanoDrop).

All RNA samples for the discovery cohort were analysed for RNA quantity within our workflow (FIGURE 8). A minimum of 150ng was used for RNA sequencing which correlated to a concentration of at least 7.5ng/µL. Samples analysed using Bioanalyzer

showed an average of 24.49 ng/µL, (range 0-533 ng/µl) with the limit of detection using the Pico assay of 50 pg/µl. We were not able to associate this with tumour stage or grading.

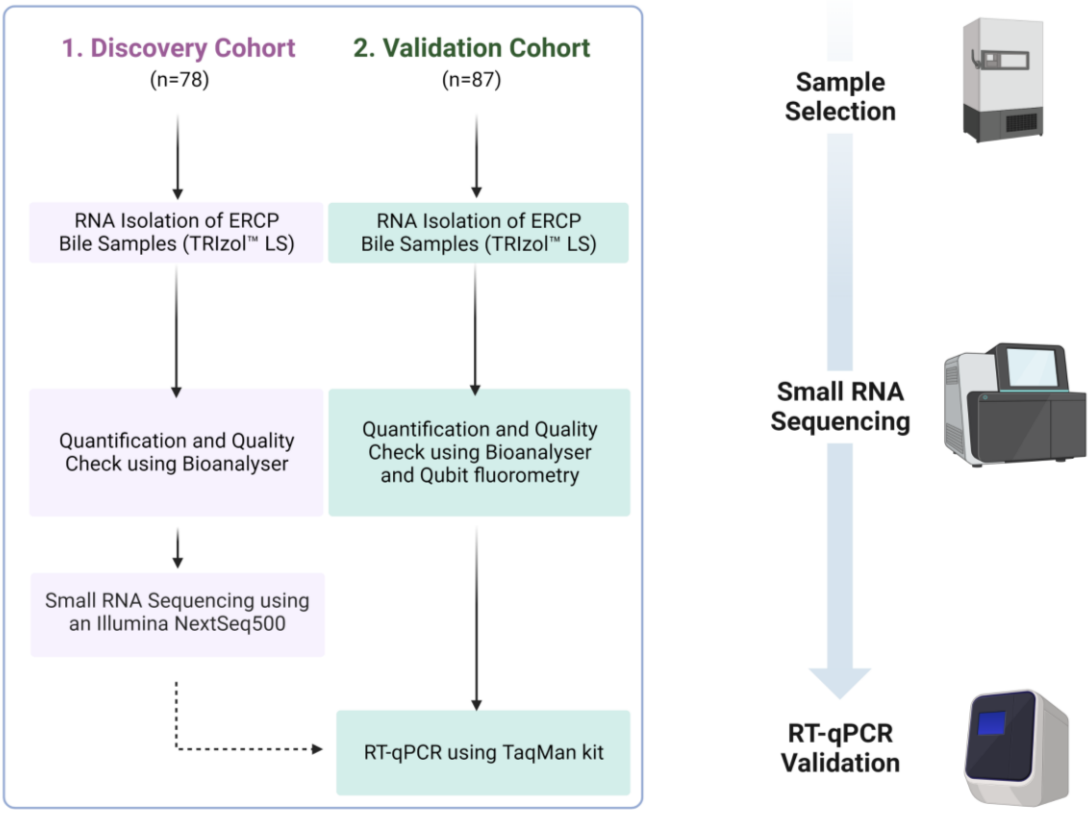


Figure 8: Schematic diagram showing the workflow for microRNA biomarker analysis.
Schematic diagram showing the workflow which involved sample selection and matching of Benign controls. RNA isolation from patient samples was performed using a TRIzol protocol to obtain total RNA which was followed by quantification by Bioanalyzer. Initially samples from the discovery cohort underwent small RNA sequencing followed by differential expression analysis. This then informed the candidates which would be validated on a separate cohort by RT-qPCR using a TaqMan primer-probe system.
RNA –Ribonucleic Acid, RT-qPCR – Reverse transcription quantitative polymerase chain reaction.

Retrospective analysis of the RNA concentrations determined using this analysis is shown (FIGURE 9A and 9B). Analysis of all sequenced samples by Bioanalyzer was undertaken to determine average quantities extracted by our method (FIGURE 9B). Average concentrations for the patients from the AMPULLARY cohort were calculated at 6.296 ng/µl, Benign at 15.08 ng/µl, CCA at 78.98 ng/µl, and PDAC at 19.83 ng/µl. Comparison between each disease by ANOVA showed a clear statistical difference in the bile cfRNA samples as determined by Student t-test. More specifically, CCA concentrations were statistically higher than the Benign cohort ($p= 0.0004$), whilst comparisons of AC vs Benign and PDAC vs Benign did not show any differences.

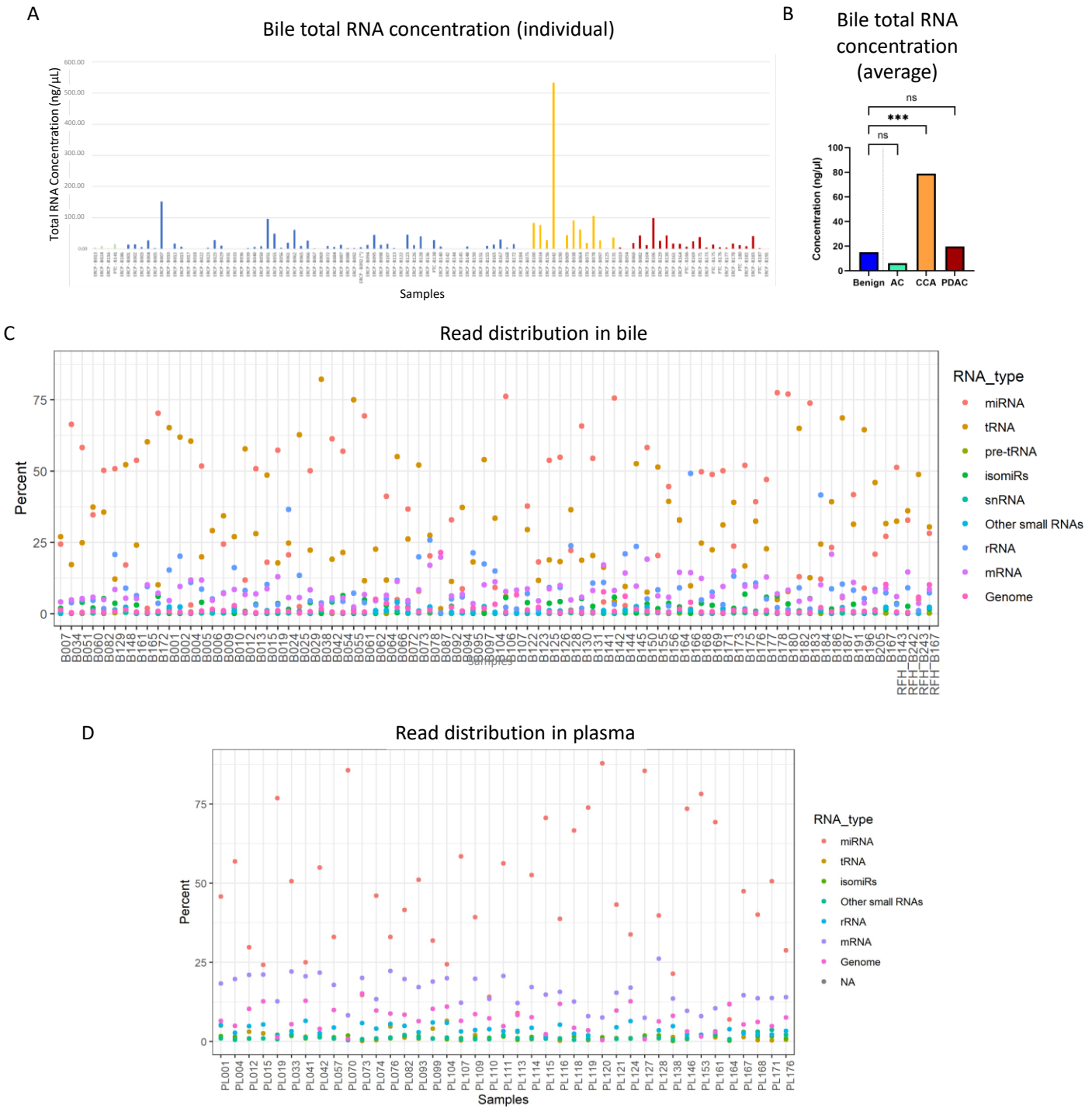


Figure 9: RNA quantification and composition analysis of bile and plasma discovery cohorts.

A. Total RNA concentrations (ng/µL) extracted from bile samples sent for discovery expressed as individual bar plots for each sample demonstrating extraction variation (n=81; 37 Benign, 14 CCA, 27 PDAC, 3 AC). This includes 3 additional sample from a cohort which included Ampullary Cancer later omitted from the analysis.

B. Mean cfRNA concentration (ng/µL) isolated from bile samples (n=105; 60 Benign, 23 PDAC, 14 CCA, 5 AC) according to patient disease.

C. Relative distribution of small RNA reads across different RNA classes for each bile sample from the discovery cohort (n=81; 37 Benign, 14 CCA, 27 PDAC, 3 AC) (shown as percentages).

D. Relative distribution of small RNA reads across different RNA classes for each plasma sample from the discovery cohort (n=41; 10 Benign, 5 CCA, 26 PDAC) (shown as percentages).

AC – Ampullary Cancer, PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma

3.2.5 Discovery of miRNAs present in bile and plasma cfRNA by next generation sequencing is differentially expressed in cancer cohorts.

The discovery cohorts (78 bile samples and 41 plasma samples) underwent small RNA sequencing with FIGURE 9C & 9D showing the small non-coding RNA composition of bile and plasma studies respectively.

Small RNA-sequencing was performed by Dr Morten Venø and the team at Omiics (Aarhus, Denmark) (see Methods 2.2.13). In this case I prepared 150ng of RNA and this was sent to the facility on dry ice where they performed library preparation. 100ng was used as the input amount for each reaction. Library preparation from the RNA samples was undertaken using the QIAseq miRNA Library Kit and sequenced on an Illumina NextSeq500. Single-end 75 base pair small RNA sequencing was performed. After quality filtering and trimming, the average total number of reads measured from bile and plasma were 15657465 and 9946268 respectively. The manufacturer's recommendation is to allocate 5–10 million reads per sample which was achieved in all samples. However, samples varied in their small RNA composition (FIGURE 9C) which demonstrated high variation in both miRNA and tRNA reads (mean percentage 36%, range 0-85%). Interestingly, the number of samples with a proportion of reads comprising <1% miRNA was significantly higher prior to October 2018 (mean proportion of reads constituting miRNA in each sample was 32% rising to 40%), which could either be attributed to differences in specimen collection with a dedicated research fellow or length of time in storage. We therefore felt that low miRNA proportion could be a byproduct of sample degradation. As a result, 6 samples with low miRNA percentage were omitted from further analysis reducing the number to 72 samples taken forward (4 Benign and 2 CCA comprising 7.7% of the samples).

Not all miRNAs were uniformly expressed in our sequencing data, so filtering was undertaken to remove data that could be constituted as "noise". This threshold was based on the data and set to include miRNAs found to have more than 5 sequencing reads in 8 or more samples (present in 10% of our samples and therefore considered well expressed). This resulted in 457 miRNAs taken forward for differential expression analysis in the bile dataset and 375 miRNAs in the plasma dataset based on these conditions (out of a total of 2656 mature miRNAs based on miRBase v22).

Small RNA composition of RNA showed improved consistency in plasma (mean percentage 61%, range 37-90%) in miRNAs with no samples with <1% miRNA content (FIGURE 9D).

3.2.6 Graphical analysis of 457 miRNAs in bile shows differential expression by cancer cohort.

Principal component analysis plot is an exploratory data visualisation method and shows correlation of different samples according to miRNA expression values (FIGURE 10A). CCA and PDAC samples (turquoise and purple respectively) show a broad clustering of the samples with occasional outliers. In contrast widely dispersed expression of benign samples (dark green) appears to indicate heterogeneity of expression consistent with a control population.

Heatmaps were generated to show potential patterns in expression. FIGURE 10B shows an unsupervised heatmap (cohorts shown along the x axis are not grouped) with hierarchical clustering of the miRNAs according to similar expression in samples. Various pairwise comparisons from the clinical data of the samples are shown along the right.

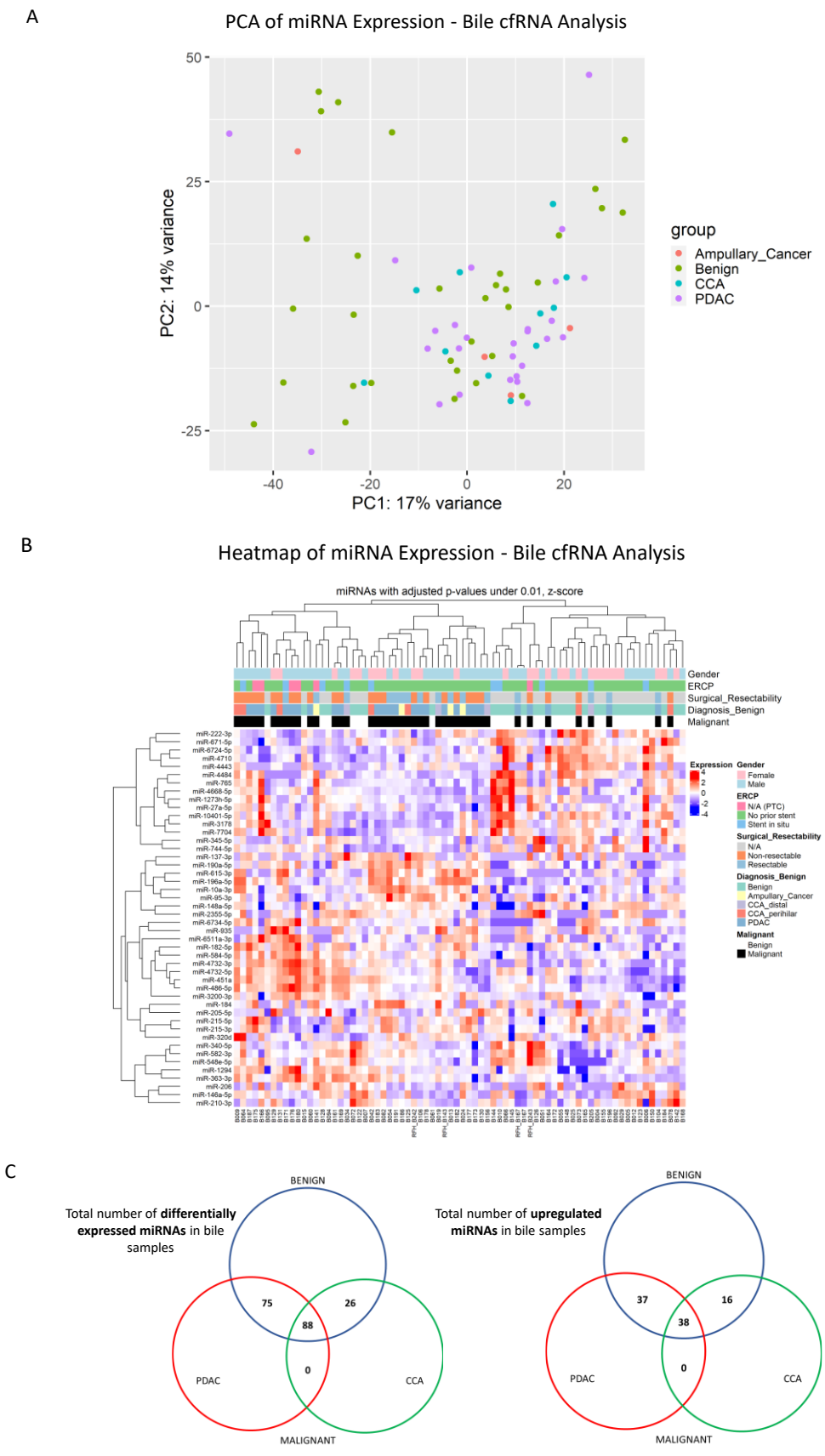


Figure 10: Differential expression of miRNAs present in bile samples (n=81).

The discovery cohort underwent small RNA sequencing to evaluate bile miRNA expression profiles and after filtering we can demonstrate this graphically as follows:

A. A principal component analysis plot showing grouping of the samples according to disease.
B. Heatmap with unsupervised hierarchical clustering showing all miRNAs with adjusted p-values <0.1.
C. Venn diagrams showing the total number of differentially expressed miRNAs (left) and total number of upregulated miRNAs only (right) between PDAC, CCA and benign cohorts with P value (adjusted) <0.05 in bile.

ERCP - Endoscopic retrograde cholangiopancreatography, PDAC – Pancreatic ductal adenocarcinoma, PTC – Percutaneous Transhepatic Cholangiography, CCA – Cholangiocarcinoma

3.2.7 Looking for statistically significant candidates within our differential expression analysis of bile miRNAs.

457 miRNAs in total were available for differential expression analysis from our bile samples. In order to compensate for low p-values and draw conclusions from our differential expression analysis, the Benjamini–Hochberg (299) methodology was used to control the False Discovery Rate (FDR). This reduced the number of candidates further with FIGURE 10C showing the overall number of statistically significant results (after adjustment) for each comparison with only the number upregulated in our cancer samples to the right.

As a result, 88 differentially expressed miRNAs were found statistically different between Malignant vs Benign with 75 candidates in PDAC vs Benign and 26 candidates between CCA vs Benign. Interestingly, no candidates were shown to be significantly differentially expressed between PDAC and CCA, indicating a high level of congruence in the expression profiles although the markers that were differentially expressed were members of the miR-320 family (upregulated in CCA relatively) (SEE FIGURE 11A). When filtered for only those highly expressed in cancer samples, we can see that this was roughly half with 38 miRNAs found between Malignant vs Benign, 37 miRNAs in PDAC vs Benign and 16 miRNAs in CCA vs Benign (FIGURE 9C).

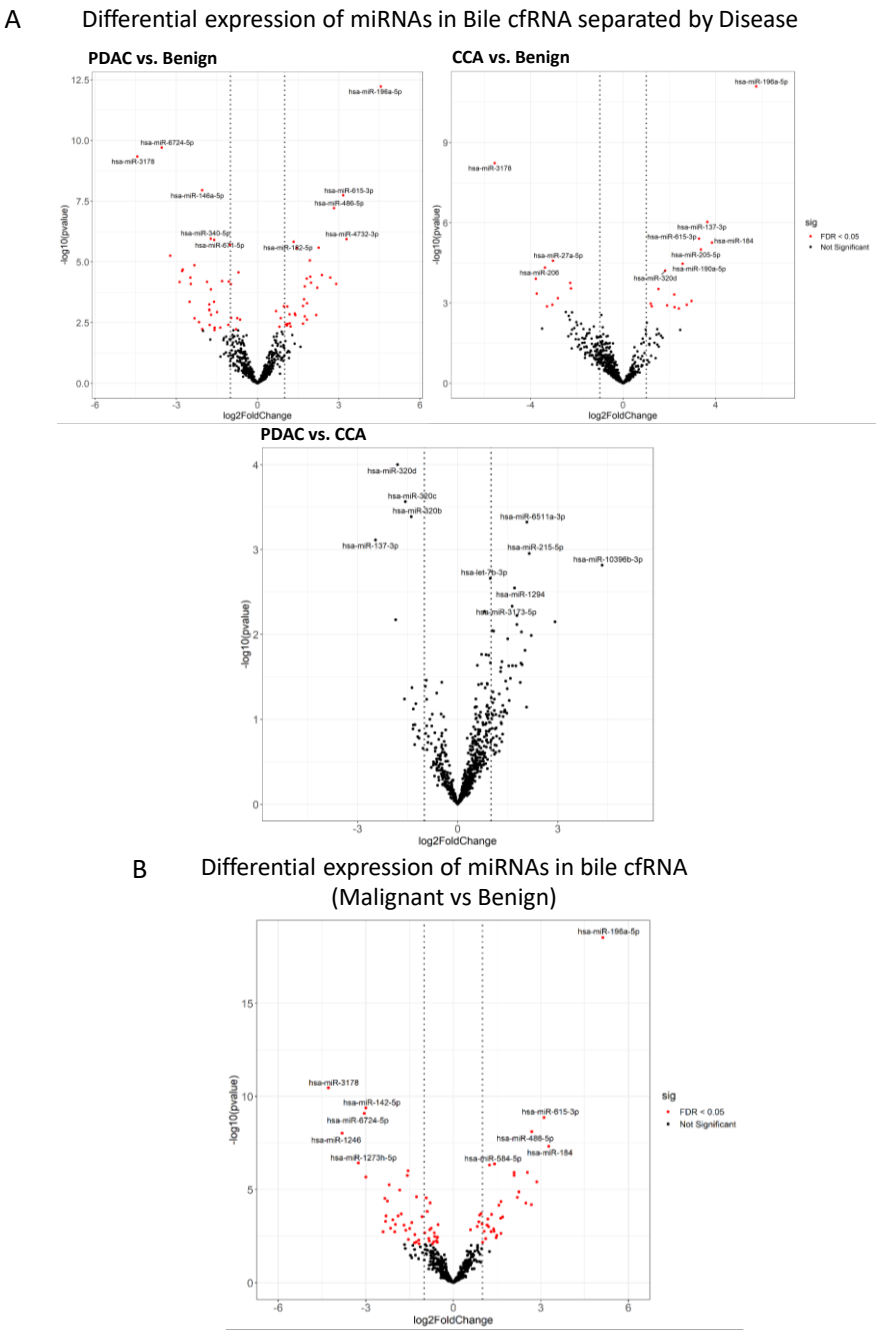


Figure 11: Volcano plots showing significant miRNAs present in bile cfRNA samples (n=78).

A. Volcano plots with top 10 miRNAs found to be significant in the following pairwise comparisons (PDAC vs Benign; CCA vs Benign; PDAC vs CCA) with an FDR < 0.05 (highlighted in red).
B. Volcano plot with top 10 miRNAs (upregulated and downregulated) found to be significant in the pairwise comparison Malignant (PDAC & CCA) vs Benign.

Plotted are candidates with significance (-log10 p-value) vs expression (log2 Fold Change). Vertical lines indicate log2 Fold Change ±1. Only the 10 most significant miRNAs are labelled for clarity. PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, FDR – False Discovery Rate

Volcano plots show differentially expressed miRNAs with significance (adjusted p-value) against fold change on the y and x-axis respectively. Highlighted in red are the statistically significant candidates (adjusted p-value ≤0.05) both up and down-regulated.

FIGURE 11A shows the comparisons PDAC vs Benign, CCA vs Benign and PDAC vs CCA for miRNAs. FIGURE 11B shows the 88 candidates in the comparison Malignant vs Benign with miR-196a-5p demonstrating outstanding significance and log2foldchange. There were no significant differences noted between the PDAC and CCA populations (FIGURE 11A). Members of the miR-320 family (miR-320d, miR-320b and miR-320c) do trend towards significance in the CCA samples (vs PDAC) with adjusted p-values 0.0972, 0.1317, 0.1826 respectively).

To identify potential candidates more easily, the top twenty miRNAs with the greatest difference in expression (i.e. highest expression fold change expressed as a logarithmic value) for the pairwise comparison Malignant vs Benign are shown in TABLE 12.

Table 12: Top Twenty miRNAs identified as expressed in Malignant (PDAC and CCA) vs Benign bile samples.

Top twenty miRNA candidates identified as significantly upregulated by RNA sequencing sorted by descending log fold change values. Expression averaged across all samples with normalization by DESeq2 to allow direct comparison between the samples and expressed as the mean RPM (reads per million mapped reads). Standard errors are of the log2fold change and p-values are adjusted using the Benjamini-Hochberg method. In bold are candidates used in later validation.

MicroRNA	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
miR-196a-5p	**741.15**	**4.9335**	**0.59768**	**1.50E-13**
miR-615-3p	7.1160	3.9384	0.53941	9.31E-11
miR-184	10.464	3.7308	0.56470	7.71E-09
miR-486-5p	31181	3.4113	0.50089	2.38E-09
miR-935	7.9558	3.3745	0.70418	4.90E-05
miR-144-5p	784.77	3.0805	0.56952	4.44E-06
miR-10a-3p	75.876	2.8946	0.46348	5.18E-08
miR-190a-5p	49.429	2.8766	0.48702	3.42E-07
miR-1247-5p	10.290	2.7171	0.66068	0.000479
miR-4732-3p	20.559	2.5272	0.58601	0.00024
miR-4732-5p	39.129	2.5082	0.49298	1.77E-05
miR-215-3p	145.18	2.4818	0.49213	2.07E-05
miR-10a-5p	4503.4	2.4783	0.42679	5.20E-07
miR-214-3p	5.4338	2.4038	0.51115	6.62E-05
miR-1224-5p	8.7578	2.3500	0.46777	2.07E-05
miR-205-5p	85.139	2.3411	0.46684	2.08E-05
miR-20b-5p	366.04	2.2600	0.44928	2.07E-05
miR-182-5p	**4542.0**	**2.2391**	**0.34096**	**8.37E-09**
miR-196b-5p	153.05	2.1340	0.57262	0.001696
miR-484	79.509	2.0607	0.43467	5.93E-05

Additionally, we show the top ten miRNAs with greatest differential expression in PDAC vs Benign in TABLE 13A and the top ten miRNAs with most expression in CCA vs Benign are shown in TABLE 13B. Upregulated miRNAs in PDAC included miR-196a, miR-615-3p, miR-184, miR-486-5p and miR-935. Additional PDAC specific miRNAs included miR-4732-5p whilst CCA samples showed high expression of miR-205-5p in particular.

Table 13: Top Ten miRNAs identified as highly expressed in PDAC vs Benign and CCA vs Benign bile samples separately (n=81).

A. Top ten miRNA candidates identified by RNA sequencing in PDAC vs BENIGN sorted by descending log fold change values.
B. Top ten miRNA candidates identified by RNA sequencing in CCA vs BENIGN sorted by descending log fold change values.

PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, RPM – reads per million mapped reads.

A

MicroRNA	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
miR-196a-5p	741.15	4.5327	0.67938	4.82E-09
miR-615-3p	7.1160	4.0491	0.59852	3.18E-09
miR-4732-5p	72.882	3.9485	0.60274	6.82E-09
miR-486-5p	31181	3.8081	0.56256	3.18E-09
miR-935	7.9558	3.7160	0.78382	3.76E-05
miR-184	10.464	3.6498	0.63774	4.91E-07
miR-215-5p	10393	3.4195	0.50232	3.18E-09
miR-363-3p	223.12	3.1486	0.53517	2.02E-07
miR-451a	18118	3.0953	0.57689	2.75E-06
miR-215-3p	145.18	2.9536	0.54457	2.32E-06

B

MicroRNA	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
miR-196a-5p	741.15	5.3233	0.86403	6.89E-07
miR-615-3p	7.1160	4.0018	0.73382	2.36E-05
miR-184	10.464	3.7428	0.79202	0.000313
miR-205-5p	85.139	3.2897	0.64601	8.44E-05
miR-190a-5p	49.429	3.2131	0.69621	0.000469
miR-1247-5p	10.290	3.0363	0.95004	0.029125
miR-10a-3p	75.876	2.8919	0.67012	0.001381
miR-486-5p	31181	2.6061	0.71658	0.010127
miR-10a-5p	4503.4	2.5484	0.62027	0.002532
miR-4732-5p	72.883	2.3975	0.76496	0.032999

Differential expression analysis was also undertaken for variables including tumour location (distal vs perihilar CCA), prior stent in situ at ERCP (i.e. those with previous biliary endoscopic intervention) and resectability of the tumour in PDAC. Volcano plots shown in FIGURE 12 demonstrate the similarity of cfRNA expression. One miRNA was found to be upregulated in those with no prior stents (miR-205-5p), but small sample sizes (<30) overall meant that the sequencing data was likely underpowered to determine any significant differences in this multivariate analysis.

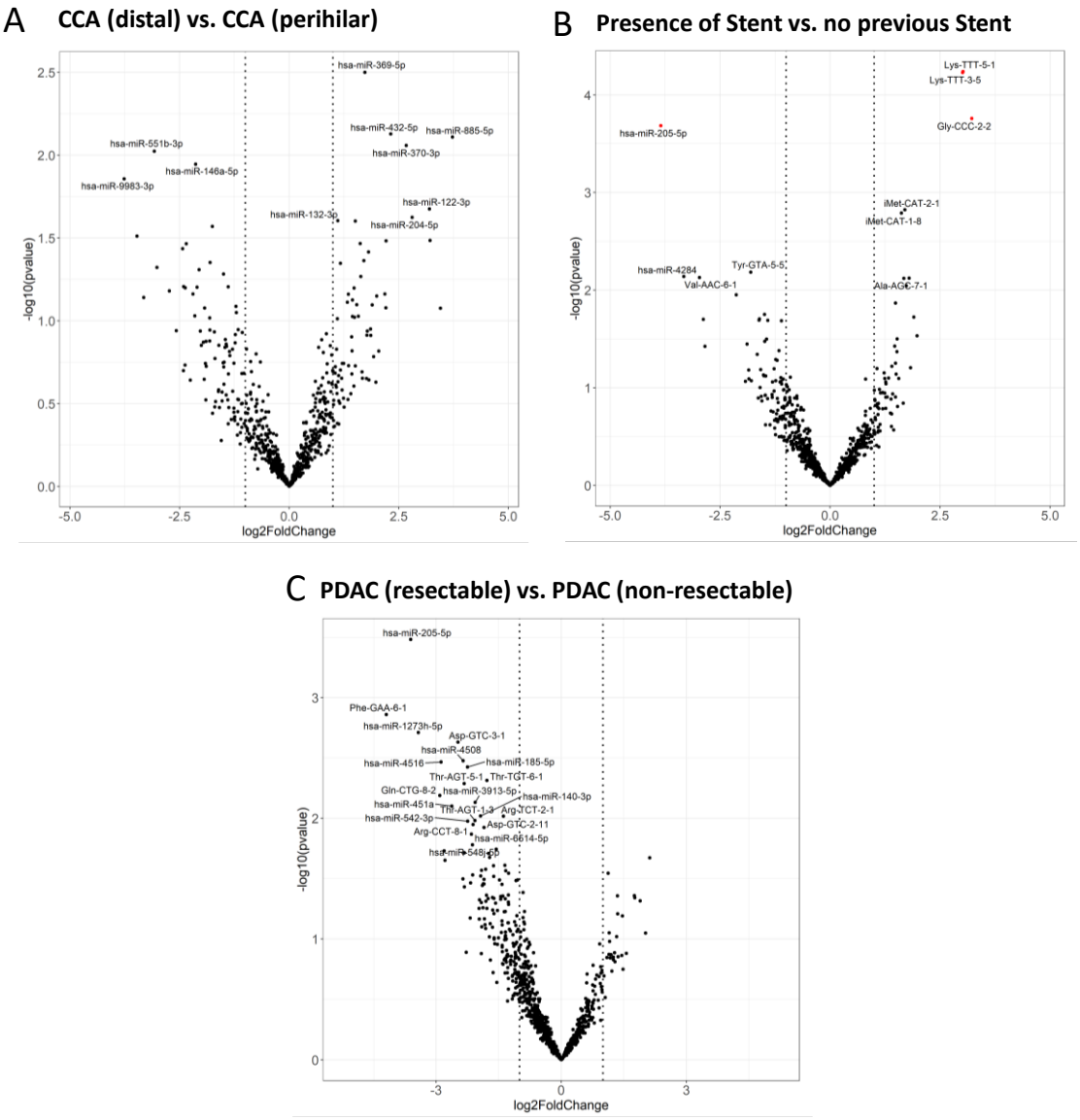


Figure 12: Further multivariate analysis showing non-significant differential expression in bile samples when compared for anatomical location, presence of stent or resection status (n=81).

Volcano plots with top 10 miRNAs found to be significant in the following pairwise comparisons with an FDR < 0.05 (highlighted in red). Plotted are candidates with significance ($-\log^{10}$ p-value) vs expression ($\log^2$ Fold Change). Vertical lines indicate $\log^2$ Fold Change ±1. Only the 10 most significant RNAs are labelled for clarity.

A. Distal CCA vs perihilar CCA – miRNA only comparison.
B. Stent present at the time of ERCP vs No prior stent – miRNA and tRNA comparison.
C. PDAC deemed resectable vs PDAC not resectable – miRNA and tRNA comparison.

PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, FDR – False Discovery Rate

3.2.8 Small RNA sequencing also reveal differential expression of tRNA-derived fragments.

tRNA fragments (tRFs) were identified as highly expressed in our bile cfRNA samples with a mean 37% composition in bile (range 4-84%) which led us to investigate these for

differentially expressed candidates. tRFs are derived from well-conserved 76 nt long tRNAs which are cleaved into either halves (30-35nts) or smaller fragments. These smaller fragments are defined by their relation to the original tRNA, mapping either to the 5' or 3' end, with those remaining termed 'internal tRFs' (i-tRFs). Unassigned tRFs are labelled 'NA'. In comparison to miRNAs, there are a higher number of experimentally validated tRFs found present in the human body (26,531 according to MINTbase v2.0). After filtering for those with adequate expression (expression in more than half of samples), 3176 unique candidates were identified that were analysed in a similar fashion.

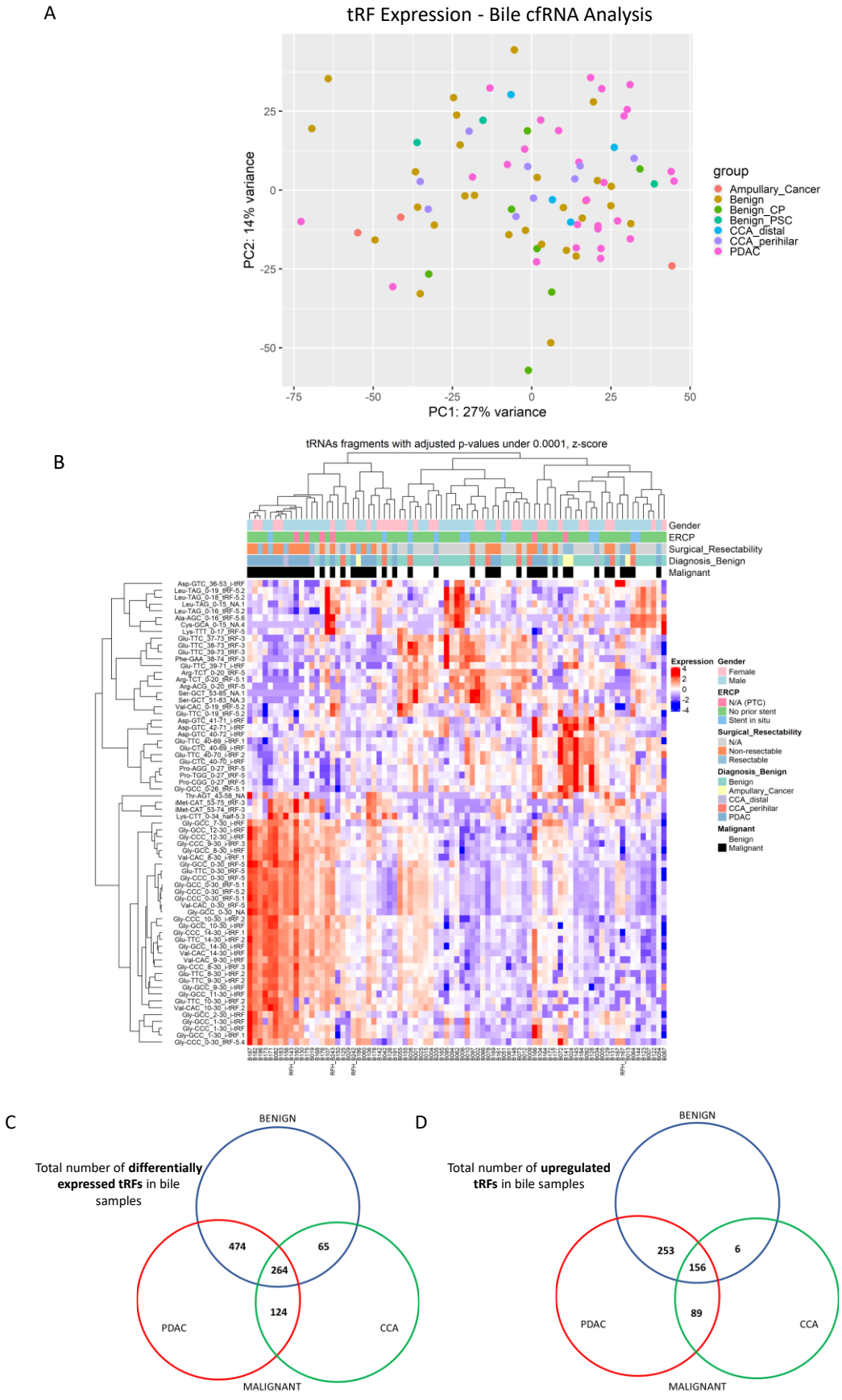


Figure 13: Differential expression of tRFs present in bile cfRNA samples (n=78).

A. Discovery cohort bile tRF expression shown using a principal component analysis plot showing grouping of the samples.
B. Heatmap with unsupervised hierarchical clustering showing all tRFs with adjusted p-values <0.001.
C. Venn diagram showing the total number of differentially expressed tRFs between PDAC, CCA and benign cohorts with P value (adjusted) <0.05 in bile.
D. Venn diagram showing the total number of upregulated tRFs only with P value (adjusted) <0.05 in bile.

PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, tRFs – Transfer RNA fragments

FIGURE 13A shows the PCA for our samples when analysed according to cohort. The high overlap initially indicates that tRFs may have a poor overall diagnostic value. Again a heatmap was generated and is shown in FIGURE 13B. Analysis generated a large number of potential candidates and after correction using FDR we found 264 differentially expressed tRFs between Malignant vs Benign bile samples with 474 candidates in PDAC vs Benign and 65 candidates between CCA vs Benign (FIGURE 13C). FIGURE 13D shows the number of candidates after filtering for only those upregulated in cancer samples which highlights the potential for biomarker research.

Shown below are volcano plots for the pairwise comparisons Malignant vs Benign, PDAC vs Benign and CCA vs Benign (FIGURE 14).

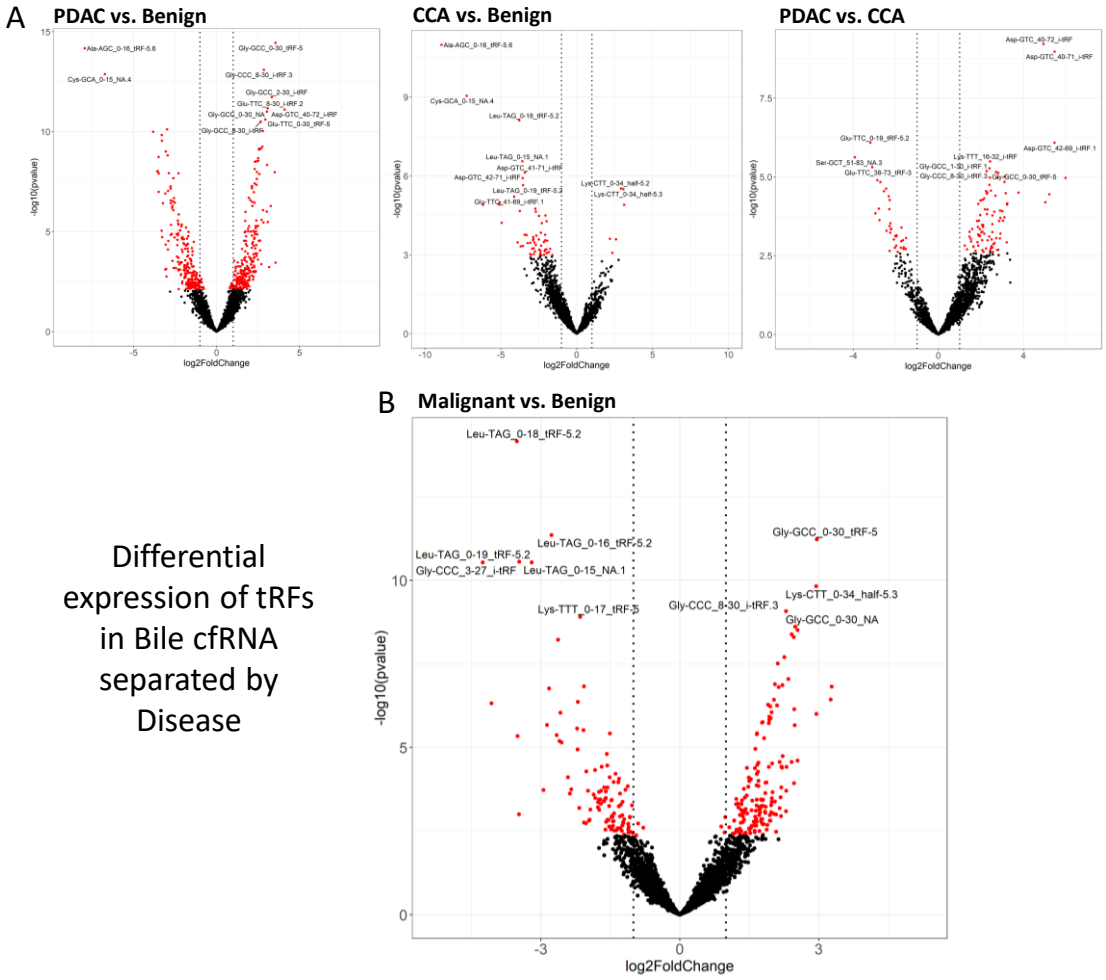


Figure 14: Volcano plots showing significant tRFs present in bile samples (n=78).

A. Volcano plots with top 10 miRNAs found to be significant in the following pairwise comparisons (PDAC vs Benign; CCA vs Benign; PDAC vs CCA) with an FDR < 0.05 (highlighted in red). Plotted are significance (-log10 p-value) vs expression (log2 Fold Change). Vertical lines indicate log2 Fold Change ±1. Only the 10 most significant miRNAs are labelled for clarity.

B. Volcano plot with top 10 miRNAs (upregulated and downregulated) found to be significant in the pairwise comparison Malignant (PDAC & CCA) vs Benign.

PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, FDR – False Discovery Rate, tRFs – Transfer RNA fragments

The comparisons PDAC vs Benign, CCA vs Benign and PDAC vs CCA for tRFs are shown in FIGURE 14A. 264 candidates were compared in the comparison Malignant vs Benign with Gly-GCC_0-30_trF-5 demonstrating outstanding significance (p-adj = 6.04E-09) and log2foldchange (FIGURE 9B).

As a result, the top ten tRFs over-expressed in Malignant vs Benign samples (i.e. potential biomarker candidates) are shown in TABLE 14A and in PDAC samples in TABLE 14B. Only 6 tRFs were found to be statistically over-expressed in the CCA vs Benign comparison and are shown in TABLE 14C.

Table 14: Top 10 tRFs identified as upregulated in the bile cfRNA biomarker analysis (n=78).

A. Top 10 tRF candidates identified as significantly upregulated in the comparison Malignant vs Benign and sorted by descending log fold change values (upregulated in malignant i.e. PDAC and CCA samples).
B. Top 10 tRF candidates identified as significantly upregulated in the comparison PDAC vs Benign and sorted by descending log fold change values (upregulated in PDAC samples).
C. Top tRF candidates identified as significantly upregulated in the comparison CCA vs Benign and sorted by descending log fold change values (upregulated in CCA samples).

Expression averaged across all samples with normalization by DESeq2 to allow direct comparison between the samples. Standard errors are of the log2fold change and p-values adjusted using the Benjamini-Hochberg method. RPM – Reads per million reads.

A

tRNA Fragment	MINTbase Unique ID	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
Gly-GCC_52-71_i-tRF	tRF-19-WE8SPOJU	6.4752	3.2842	0.6253	2.18E-05
Gly-CCC_37-70_i-tRF	tRF-33-WYL1M3WE8S68DM	4.7185	3.2624	0.6417	4.56E-05
Gly-GCC_0-30_tRF-5	tRF-30-P4R8YP9LON4V	124300	2.9616	0.4304	6.04E-09
Thr-AGT_43-58_NA	NA	5.0630	2.9532	0.6036	8.94E-05
Lys-CTT_0-34_half-5.3	tRF-34-PSQP4PW3FJIKE5	112.21	2.9451	0.4599	6.61E-08
Glu-TTC_8-30_i-tRF.2	tRF-22-R918VBY9M	11.166	2.5431	0.4290	8.49E-07
Asp-GTC_28-50_i-tRF	tRF-22-L7S5QKF14	6.0747	2.5421	0.6024	0.001308
Gly-GCC_0-30_NA	NA	548.85	2.4962	0.4186	7.55E-07
Gly-TCC_37-52_NA	NA	2.2686	2.4858	0.5249	0.000158
Val-CAC_0-34_half-5	tRF-34-Q99P9P9NH57S15	113.15	2.4714	0.4987	7.09E-05

B

tRNA Fragment	MINTbase Unique ID	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
Asp-GTC_40-72_i-tRF	tRF-32-R008R959KUMKH	1046.1	4.0967	0.5988	3.43E-09
Asp-GTC_50-72_i-tRF.3	tRF-22-R959KUMKH	32.144	3.5531	0.9949	0.005732
Gly-GCC_0-30_tRF-5	tRF-30-P4R8YP9LON4V	124300	3.5500	0.4511	1.03E-11
iMet-CAT_53-74_tRF-3	tRF-21-WB0Q37Q5D	18.926	3.5049	0.6127	1.02E-06
Gly-GCC_2-30_i-tRF	tRF-28-IK9NJ4S2I7DM	204.66	3.3282	0.4724	1.14E-09
Thr-AGT_43-58_NA	NA	5.0630	3.2695	0.6516	2.59E-05
Val-TAC_41-73_NA.1	NA	18.116	3.1720	0.9229	0.008419
Gly-GCC_52-71_i-tRF	tRF-19-WE8SPOJU	6.4752	3.0837	0.6630	0.000139
Glu-TTC_8-30_i-tRF.2	tRF-22-R918VBY9M	11.166	3.0794	0.4489	3.43E-09
Lys-CTT_0-34_half-5.3	tRF-34-PSQP4PW3FJIKE5	112.21	3.0399	0.5188	5.26E-07

C

tRNA Fragment	MINTbase Unique ID	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
Gly-CCC_36-69_i-tRF	tRF-33-YZQD3KYUYRRHD8	5.8543	3.1252	0.71563	0.002645
Lys-CTT_0-34_half-5.3	tRF-34-PSQP4PW3FJIKE5	112.21	3.0669	0.6583	0.001041
Lys-CTT_0-34_half-5.2	tRF-34-PSQP4PW3FJI0E5	50.783	2.9153	0.6242	0.001041
Val-CAC_0-34_half-5	tRF-34-Q99P9P9NH57S15	113.15	2.5914	0.7091	0.024037
Gly-TCC_37-52_NA	NA	2.2686	2.3497	0.7039	0.044312
Gly-GCC_37-68_NA	NA	13.273	2.1872	0.5966	0.024037

One feature that can be identified is the presence of similar tRFs present in bile cfRNA that vary only by a short number or single nucleotide and yet show high expression

individually in our samples. For example, members of the 5'-tRF Gly-GCC were found to be upregulated in the Malignant vs Benign comparison. Gly-GCC_0-30_tRF-5 and Gly-GCC_0-30_NA vary by a single nucleotide and were both recorded in the top ten upregulated tRFs (TABLE 14A). It can also be noticed that members of the i-tRF Asp-GTC were found to be upregulated in the PDAC vs Benign comparison with Asp-GTC_40-71_i-tRF and Asp-GTC_50-72_i-tRF shorter by 10 nucleotides.

This can be alternatively shown in a coverage plot according to tRNA isotype (i.e. the tRNA from which the fragment is likely to have arisen) which shows that the dominant tRNA isotypes in bile samples appear to be Gly-GCC and Glu-CTC, making up the majority of the sample (FIGURE 15). Although it is unclear whether this reflects the proteome of bile, it is possible that this reflects the overall tRNA profile which has not previously been investigated (and for which currently used sequencing strategies are inadequate (300)). Glutamine is however the most abundant amino acid in the body (301), and glycine is a necessary amino acid for conjugation of bile acids and subsequent excretion (302). The tRFs found in our discovery sequencing was not taken forward into the validation phase (due to lack of expertise and time to develop a robust RT-qPCR protocol) but common short RNA sequences composed of these could be a potential area of further research.

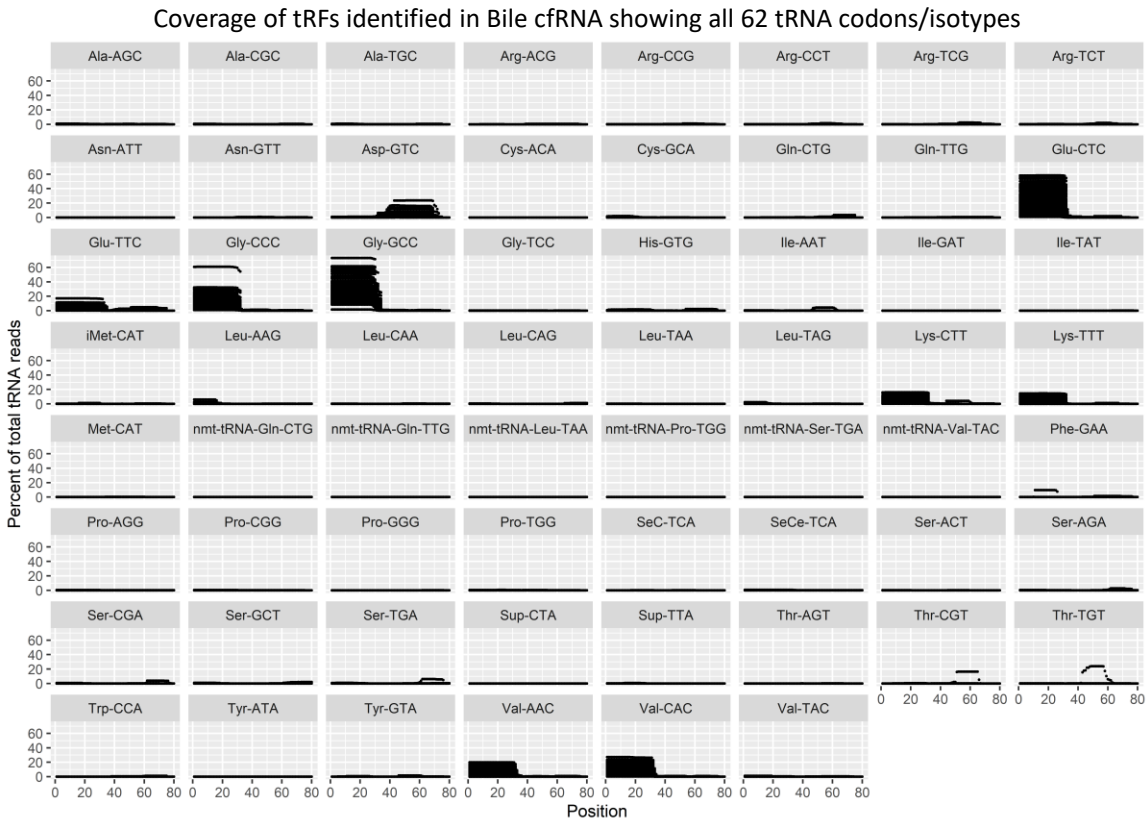


Figure 15: Bile cfRNA tRF expression expressed according to tRNA isotype (n=78).
A coverage plot of the tRF candidates is shown according to tRNA isotype (i.e. the tRNA from which the fragment is likely to have arisen). This shows Gly-GCC and Glu-CTC as dominant tRNA isotypes in bile samples. Expression as a percentage of total tRNA reads with relative nucleotide position within the tRNA expressed along the y-axis.

3.2.9 Individual bile miRNA candidates from RNA Sequencing show AUCs above 0.80.

As mentioned earlier, candidates showing highest positive expression fold change included miR-196a, miR-615-3p, miR-184, miR-486-5p and miR-935, whilst miR-4732-5p appeared specifically in the PDAC vs Benign comparison and miR-205-5p appeared in CCA samples.

Receiver operating characteristic curves (ROC curves) can be used to show the potential diagnostic utility of a candidate by plotting the true positive rate against the false positive rate of the test at various threshold settings. Deviation from the no-discrimination line which is drawn at 45 degrees demonstrates "goodness" of a classifier. Area under the ROC Curve (AUC) is one of the statistics that can be visually understood using the ROC plot and varies between 0 and 1 — with an uninformative classifier yielding 0.5 and a perfect classifier obtaining an AUC of 1. This was undertaken for each of these candidates (those with highest logfoldchange and $p<0.05$) based on identification of malignancy (FIGURE 16A) and PDAC (FIGURE 16B).

ROC curves for miRNA candidates identified as over expressed in bile cfRNA (miR-196a, -615, -184, -486, -935, -4732, -205)

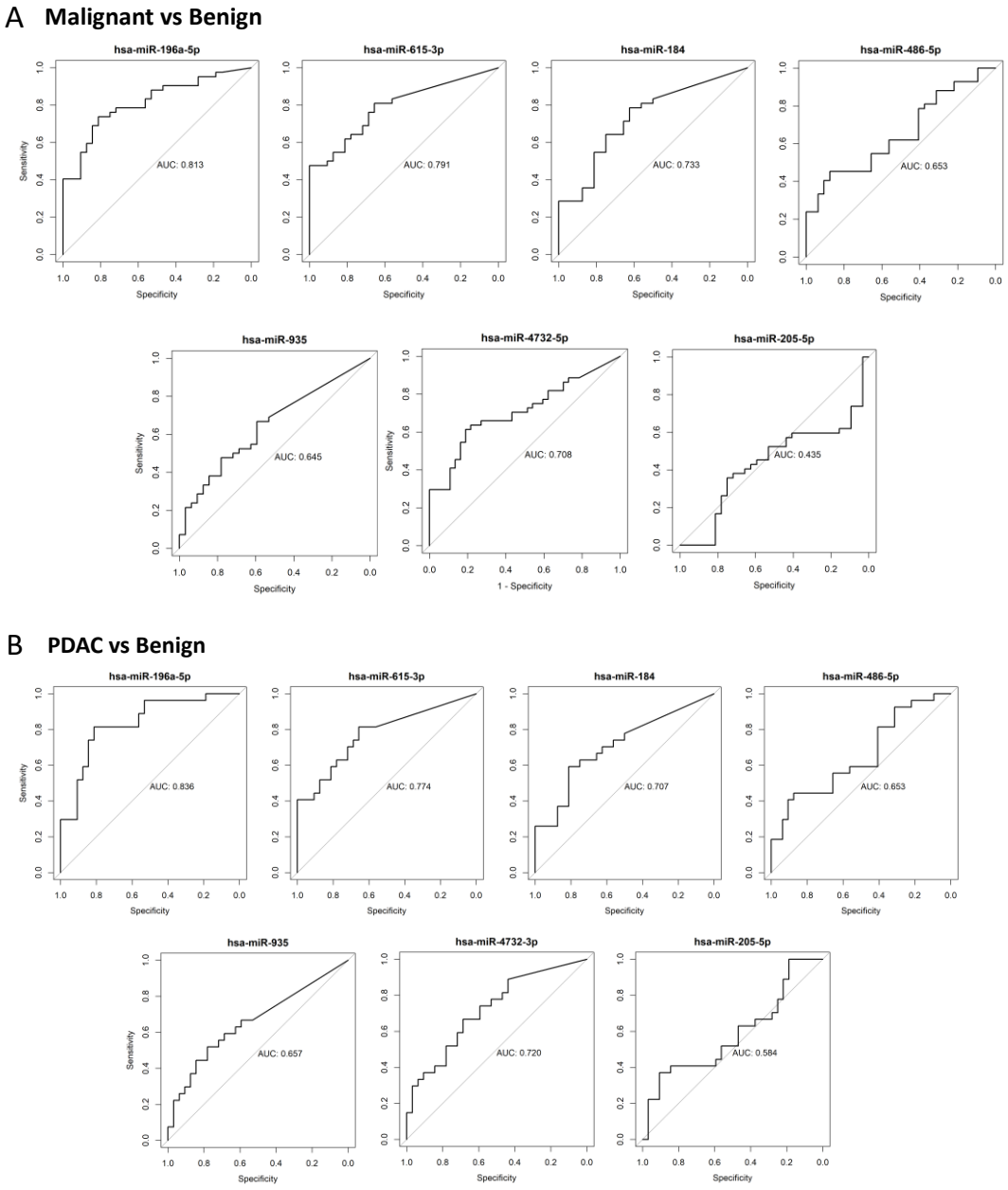


Figure 16: Diagnostic potential of selected individual miRNAs identified as over-expressed in bile cfRNA in Malignant vs Benign and PDAC vs Benign (n=78).

From RNA sequencing data it is possible to estimate the diagnostic potential of selected miRNA candidates which were shown to be over expressed in cancer. Here are:

A. ROC plots for top miRNA candidates (as shown in TABLE 12) for Malignant vs Benign identified by RNA sequencing with AUCs annotated.
B. ROC plots for top miRNA candidates (as shown in TABLE 13A) for PDAC vs Benign identified by RNA sequencing with AUCs annotated.

Due to the expression profiles seen in PDAC and CCA (see FIGURE 10A) the values obtained for each miRNA candidate in either comparison are roughly similar. For these

selected miRNAs, AUCs ranged from 0.435-0.813 for discrimination of Malignant vs Benign and 0.584-0.836 for discrimination of PDAC vs Benign.

TABLE 15-17 show the miRNAs with greatest ability to discriminate as measured by AUC. This was determined using thresholds obtained at the maximum Youden index.

Table 15: Top performing miRNAs by AUC expressed in Malignant samples in bile cfRNA.
Top ten miRNAs in Malignant vs Benign shown according to decreasing area under the curve (AUC) as estimated from RNA sequencing data.

MicroRNA	Threshold	Specificity	Sensitivity	95% CI	AUC	Youden
miR-148a-3p	0.54874	0.78125	0.78571	0.72 - 0.91	0.82	0.56696
miR-196a-5p	0.53376	0.8125	0.73810	0.72 - 0.91	0.81	0.55060
miR-196b-5p	0.62428	0.78125	0.73810	0.69 - 0.9	0.79	0.51935
miR-615-3p	0.77781	1	0.47619	0.69 - 0.89	0.79	0.47619
miR-195-5p	0.52973	0.65625	0.80952	0.64 - 0.87	0.75	0.46577
miR-182-5p	0.68358	0.9375	0.52381	0.64 - 0.86	0.75	0.46131
miR-934	0.69702	0.875	0.52381	0.63 - 0.86	0.75	0.39881
miR-4710	0.51062	0.625	0.83333	0.61 - 0.86	0.74	0.45833
miR-584-5p	0.65122	0.875	0.54762	0.63 - 0.85	0.74	0.42262
miR-190a-5p	0.48974	0.5	0.88095	0.63 - 0.85	0.74	0.38095

Table 16: Top performing miRNAs by AUC expressed in PDAC samples in bile cfRNA.
Top ten miRNAs in PDAC vs Benign shown according to decreasing area under the curve (AUC).

MicroRNA	Threshold	Specificity	Sensitivity	95% CI	AUC	Youden
miR-196a-5p	0.43050	0.8125	0.8148	0.73 - 0.94	0.84	0.62732
miR-1247-5p	0.41128	0.6875	0.8148	0.66 - 0.89	0.78	0.50232
miR-615-3p	0.42993	0.6563	0.8148	0.66 - 0.89	0.77	0.47107
miR-196b-5p	0.53100	0.7813	0.7407	0.64 - 0.89	0.77	0.52199
miR-4710	0.38173	0.5938	0.9259	0.64 - 0.89	0.76	0.51968
miR-190a-5p	0.35147	0.5000	0.9259	0.64 - 0.88	0.76	0.42593
miR-6511a-3p	0.59274	0.9063	0.5185	0.63 - 0.88	0.76	0.42477
miR-195-5p	0.43002	0.6563	0.8519	0.63 - 0.89	0.76	0.50810
miR-345-5p	0.45456	0.7188	0.8148	0.63 - 0.88	0.75	0.53357
miR-584-5p	0.54996	0.8750	0.5926	0.63 - 0.88	0.75	0.46759

Table 17: Top performing miRNAs by AUC expressed in CCA samples in bile cfRNA.
Top ten miRNAs in CCA vs Benign shown according to decreasing area under the curve (AUC). Table headings include the threshold with maximum Youden index (threshold), sensitivity % and specificity %, 95% Confidence interval (CI), Area under the Curve (AUC) and Youden index (calculated as Specificity + Sensitivity -1).

MicroRNA	Threshold	Specificity	Sensitivity	95% CI	AUC	Youden
miR-615-3p	0.56346	1	0.6364	0.7 - 1	0.85	0.63636
miR-196b-5p	0.15569	0.5625	1	0.66 - 0.94	0.80	0.56250
miR-196a-5p	0.59944	1	0.5455	0.59 - 0.96	0.77	0.54546
miR-190a-5p	0.12764	0.3750	1	0.56 - 0.91	0.74	0.37500
miR-137-3p	0.41764	0.9375	0.4545	0.55 - 0.91	0.73	0.39205
miR-184	0.24243	0.6250	0.9091	0.56 - 0.89	0.72	0.53409
miR-10a-3p	0.40130	0.9375	0.4545	0.51 - 0.91	0.71	0.39205
miR-9983-3p	0.23094	0.7188	0.7273	0.53 - 0.88	0.71	0.44602
miR-320d	0.16890	0.3750	1	0.52 - 0.87	0.70	0.37500
miR-548e-5p	0.22182	0.5313	0.9091	0.53 - 0.85	0.69	0.44034

Read counts from the discovery cohort can also be shown as bar plots (FIGURE 16) to demonstrate the logarithmic pattern of expression. Expressed as a linear scale, it can be

shown that for the top miRNA (miR-196a) the expression is only captured in a small proportion of samples. Expressed as a logarithmic scale however, it is easier to determine the diagnostic threshold for this miRNA biomarker.

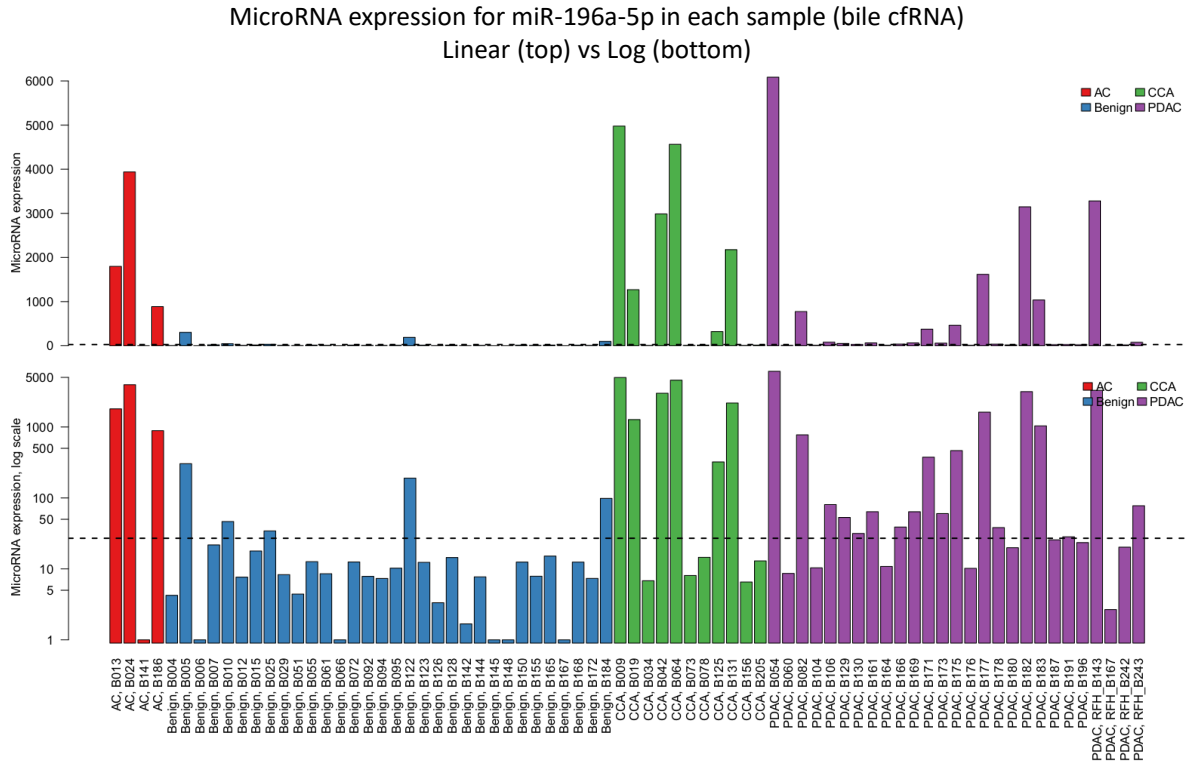


Figure 17: Expression of miR-196a by RNA sequencing in bile cfRNA (n=78).
Expression of the most significant miRNA in PDAC vs Benign and Malignant vs Benign, miR-196a-5p, in all samples sent for RNA sequencing. Colour coded by diagnosis. Linear scaled bar graph above shows large variation and heterogeneity whilst the logarithmic scaled bar graph below allows visualization and threshold determination across many orders of magnitude. PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, FDR – False Discovery Rate

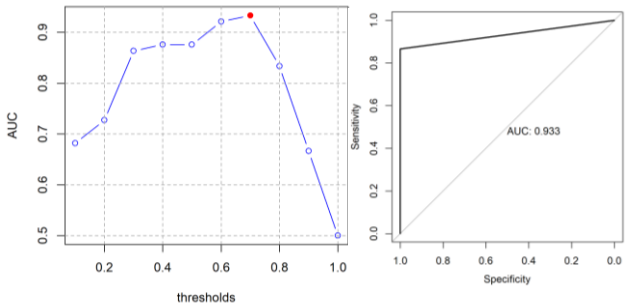
3.2.10 Regression analysis potentially identifies signature in bile with outstanding diagnostic potential.

Least absolute shrinkage and selection operator (LASSO) is a type of regression analysis which can be used to fit models and this was done by performing L1 regularization (303). This allows us to combine potential candidates into a simpler, more parsimonious model reducing other coefficients to zero. This generated a model from a wide number of potential variables leading to fewer candidate biomarkers and which could therefore be clinically suitable for analysis. To do this, the dataset was randomly split into a training and test cohort. 66% of the dataset was randomly sampled as the training set and the remainder was used as test samples.

Using this method, a subset of miRNAs from our bile samples which could detect our Malignant (FIGURE 18A) and PDAC (FIGURE 18B) cohorts with an AUC of 0.933 and 1 respectively. As shown in these figures, different thresholds (also known as regularisation parameters) are shown and their effect on the AUC. To further reduce the variables analysed significant miRNAs were filtered by expression level, requiring miRNAs to be detected with 10,000 or more normalized reads across all bile samples. This analysis resulted in a model which included six upregulated miRNAs (miR-199a-5p, miR-196a-5p, miR-335-5p, miR-182-5p, miR-340-5p, miR-424-5p). Four of these were also shown to be part of a model which generates an AUC of 1 when identifying PDAC vs Benign (miR-199a-5p, miR-335-5p, miR-196a-5p, miR-182-5p).

A

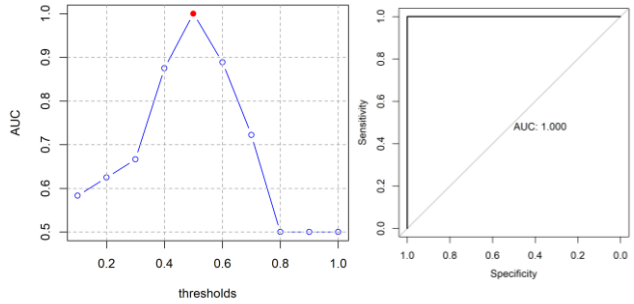
Optimal model generated by LASSO analysis for Malignant vs Benign with AUC at varying thresholds



miRNA	Coefficients	miRNA	Coefficients
miR-199a-5p	**0.000899736**	miR-146a-5p	-6.36E-05
miR-196a-5p	**0.000837807**	miR-3178	-0.000123588
miR-335-5p	**0.000774522**	miR-30e-5p	-0.000210694
miR-182-5p	**0.000134247**	miR-222-3p	-0.000614738
miR-340-5p	**8.70E-05**	miR-769-5p	-0.000723796
miR-424-5p	**7.86E-05**	miR-532-3p	-0.000839583
miR-148a-3p	-1.10E-05	miR-345-5p	-0.00117954

B

Optimal model generated by LASSO analysis for PDAC vs Benign with AUC at varying thresholds



miRNA	Coefficients
miR-199a-5p	0.000398194
miR-335-5p	0.000240827
miR-196a-5p	0.000146821
miR-182-5p	7.29E-05
miR-146a-5p	-4.16E-05
miR-345-5p	-0.000479226
miR-222-3p	-0.000769465

Figure 18: LASSO Regression analysis to determine optimal biomarker model (n=78)
Graph plot demonstrating the LASSO (Least absolute shrinkage and selection operator) regression analysis which splits the dataset into a training set and test set. Using different thresholds (also known as the regularization parameter and ranging from 0 to 1) this was subsequently tested for their effect on the AUC. MiRNAs with 10,000 or more normalized reads across all bile samples only were included. ROC curves were generated using the optimal thresholds (maximum AUC) for that pairwise comparison.

A. Malignant vs Benign with potential AUC = 0.933.
B. PDAC vs Benign with potential AUC = 1.

PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, AUC – Area Under ROC Curve, ROC - Receiver operating characteristic.

A semi-supervised heatmap (FIGURE 19) showed relative expression for the six candidate miRNAs together with the others indicated by the LASSO regression analysis. This was

clustered to show the benign and malignant samples as well as by miRNA to show similar expression patterns.

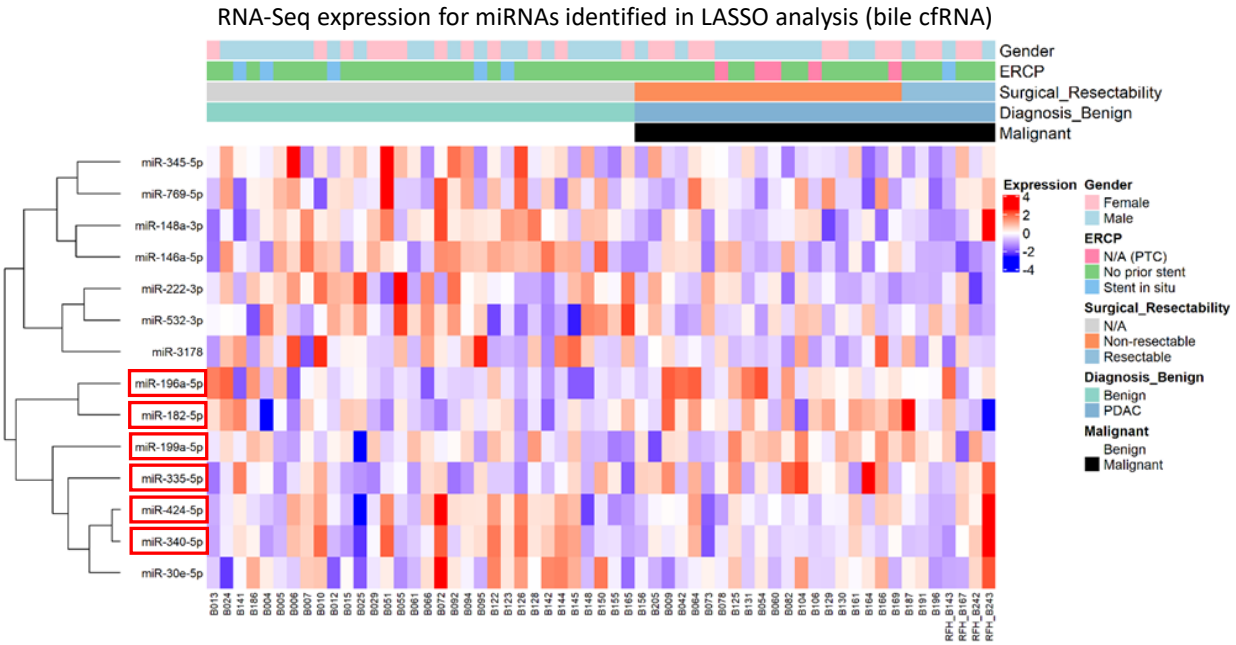
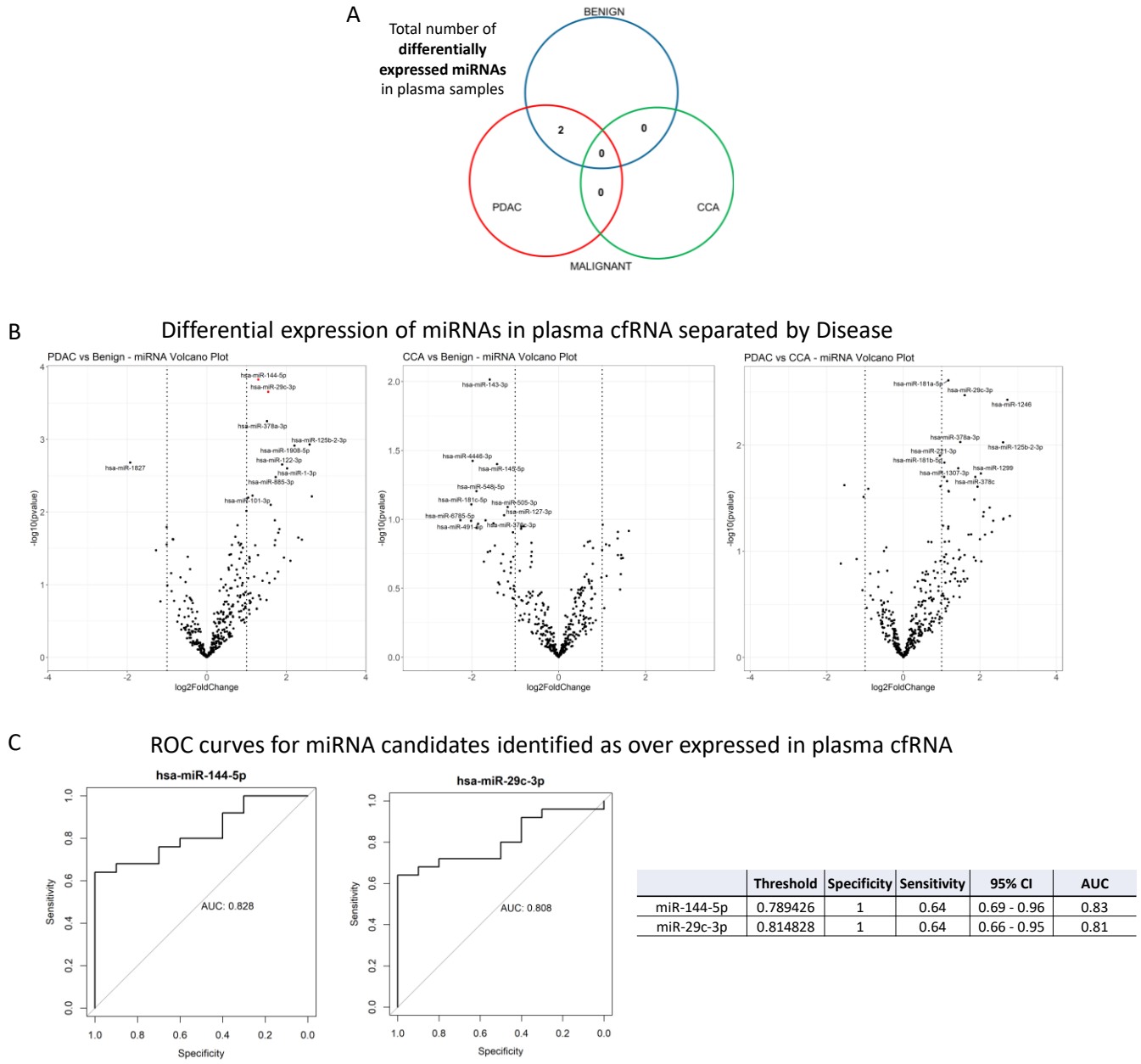


Figure 19: Semi-supervised heatmap showing discovery cohort expression of markers identified by LASSO in bile samples.
Benign and Malignant samples are grouped left and right (white and black) respectively with the six candidate miRNAs chosen for validation shown in red on the left. High relative expression (logfold) in a sample is demonstrated with red whilst blue shows low relative expression.

3.2.11 Graphical analysis of cfRNA sequencing from blood plasma miRNAs fails to show significant differential expression.

RNA sequencing was undertaken in n=41 plasma samples (26 PDAC, 5 CCA, 10 Benign) as part of a discovery cohort into miRNAs differentially expressed between PDAC and Benign patients. After controlling for multiple comparisons only 2 candidates were found to be differentially expressed (upregulated) in the pairwise comparison PDAC vs Benign (FIGURE 20A). These were miR-144-5p and miR-29-3p and can be seen in the volcano plot (FIGURE 20B). No significantly differentially expressed miRNAs were found for the other pairwise comparisons. ROC curves were generated for these two candidates which have an AUC of 0.828 and 0.808. Statistics for the candidates are shown in FIGURE 20C and show that at the optimum threshold, these miRNAs had a specificity of 100% and sensitivity of 64%. Given the low number of candidates found, it was decided not to proceed to validation with these candidates despite their reasonable AUCs and focussed

instead on validation of the bile-based cfRNA candidates given the promising differential expression between our malignant and benign samples.



	Threshold	Specificity	Sensitivity	95% CI	AUC
miR-144-5p	0.789426	1	0.64	0.69 - 0.96	0.83
miR-29c-3p	0.814828	1	0.64	0.66 - 0.95	0.81

Figure 20: Differential expression analysis of miRNAs present in plasma samples (n=41; 26 PDAC, 5 CCA, 10 Benign).

A. Venn diagram showing differentially expressed miRNAs between PDAC, CCA and benign cohorts with P value (adjusted) <0.05 in bile.
B. Volcano plots with top 10 miRNAs found to be significant in the following pairwise comparisons (PDAC vs Benign; CCA vs Benign; PDAC vs CCA) with an FDR < 0.05 (highlighted in red). Plotted are significance ($-\log^{10}$ p-value) vs expression ($\log^2$ Fold Change). Vertical lines indicate $\log^2$ Fold Change ±1. Only the 10 most significant miRNAs are labelled for clarity.
C. ROC Curves for the two differentially expressed (upregulated) markers miR-144-5p and miR-29c-3p in the plasma cfRNA analysis with table showing potential diagnostic statistics.

3.2.12 Bile and plasma cfRNA concentrations and quantity are a limitation to TaqMan RT-qPCR validation.

As many samples as possible were processed in a prospective fashion underwent RNA extraction to validate the candidate microRNAs identified. This was evaluated using fluorometry to detect RNA present in the samples and measured prior to storage for RT-qPCR. 133 patient samples in total underwent an optimised RNA extraction protocol involving dilution, centrifugation and filtration given our power calculation suggested sample size of 126 patients (i.e. 63 in each group). TRIzol LS reagent was used for the extraction of total RNA from bile samples.

FIGURE 21 summarises the RNA concentrations quantified. The average concentration measured was 40.8 ng/µl (range 0-590 ng/µl) whilst 33 samples which comprised 25% of samples extracted contained RNA that was below the limit of detection (<0.2ng/µl). Adequate concentration was deemed at more than 10 ng/µl for the purposes of TaqMan validation (arbitrarily set, although previous studies had shown that our yield should be higher and therefore should be repeated) and this was achieved in 96 samples (69%). 21 samples were repeated after having low RNA and 17 samples were able to demonstrate adequate concentrations of RNA (81%) which indicates that a significant proportion of low RNA quantities was due to extraction inefficiency.

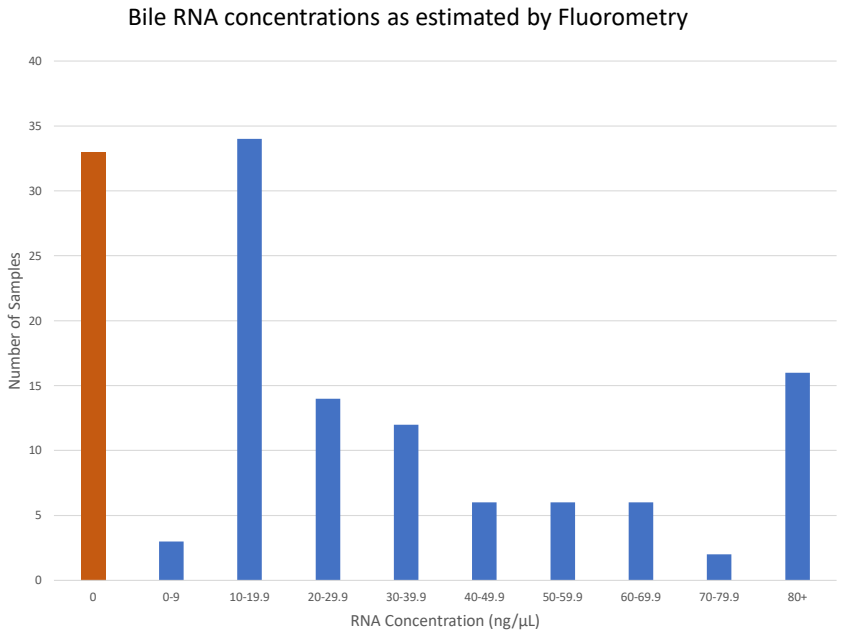


Figure 21: Quantification of RNA extraction from bile by Qubit Fluorometry (n=133).
Number of samples with analysed total RNA concentrations were measured after 500µL of bile underwent extraction with TRIzol LS using an elution volume of 15µL. Zero indicates measurements below the detectable limit (0.25ng/µL).

Plasma extracellular RNA was also measured using a smaller cohort of 6 samples to determine whether the standard TRIzol method could be improved with newer column kit RNA extraction methods. Average measured RNA concentrations were 228 pg/µl (range 10-1096 pg/µl) from both methods with no difference in yields (data not shown) but an improvement in time and standardisation of methodology utilising a kit. As the intention was to undertake validation using a fixed volume rather than RNA quantity for the validation, I decided therefore that there was no lower limit for further sequencing and RT-qPCR of our plasma samples.

3.2.13 RT-qPCR Validation from (n=87) bile samples in our six candidates from the LASSO regression analysis.

RT-qPCR required consideration of our normalisation protocol to control for variations in extraction yield, reverse-transcription yield, and enabling comparisons of microRNA concentrations across different samples. Known reference genes such as U6 usually assumed as internal controls have not been experimentally validated in bile cfRNA so our sequencing data was utilised to identify stably expressed microRNAs. TABLE 18 shows the individual results of our NormFinder (294) algorithm using the discovery cohort data to find potential stable endogenous controls. As multiple references are required, the best pair was found (TABLE 19) to be miR-26a and miR-let-7b.

Table 18: MiRNAs identified as stably expressed for normalization individually.
Top 20 miRNAs individually identified as stably expressed in all samples sorted by descending stability using NormFinder algorithm.

MicroRNA	GroupDif	GroupSD	Stability
miR-339-3p	0.56	0.79	0.14
miR-500a-3p	0.88	0.62	0.16
miR-30e-3p	0.87	0.76	0.17
miR-664a-5p	0.38	0.89	0.17
miR-345-5p	1.1	0.78	0.17
miR-30d-5p	0.85	0.81	0.17
miR-186-5p	0.59	0.72	0.17
miR-532-5p	0.72	0.78	0.18
miR-598-3p	0.22	0.82	0.18
miR-222-3p	0.85	0.64	0.18
let-7d-5p	0.47	0.85	0.2
miR-362-5p	0.65	0.93	0.21
miR-24-3p	0.53	0.58	0.21
miR-28-3p	0.82	0.87	0.21
miR-132-3p	0.51	0.82	0.21
miR-148a-3p	1.35	0.73	0.21
miR-148b-3p	0.44	0.85	0.21
miR-324-5p	0.69	0.93	0.21
let-7b-5p	0.51	0.67	0.21
miR-320a-3p	0.62	0.99	0.21

Table 19: MiRNAs identified as stably expressed for normalization as a pair.
Top 10 pairs of miRNAs identified as stably expressed in all samples sorted by descending stability.
GroupDif is the estimated difference in expression between the groups; GroupSD is a weighted average of the estimated intra-group variation and Stability is the NormFinder expression stability measure with lower values indicating stable gene expression. See Methods 2.2.19 for more information.

MicroRNA 1	MicroRNA 2	Stability
let-7a-5p	miR-26a-5p	0.1
miR-26a-5p	miR-92a-3p	0.11
miR-26a-5p	miR-101-3p	0.11
let-7a-5p	miR-93-5p	0.12
let-7b-5p	miR-26a-5p	0.12
miR-26a-5p	miR-26b-5p	0.12
let-7f-5p	miR-26a-5p	0.13
let-7b-5p	miR-93-5p	0.13
miR-26a-5p	miR-103a-3p	0.13
let-7f-5p	miR-93-5p	0.14

From the LASSO methodology, six candidates were identified in our model and taken forward for validation. These were miR-199a-5p, miR-196a-5p, miR-335-5p, miR-182-5p, miR-340-5p and miR-424-5p.

During validation, miR-199a-5p and miR-424-5p did not show adequate expression in RT-qPCR (>50% of samples no expression detected) for comparison and so were excluded. FIGURE 22 shows the results of RT-qPCR validation of the remaining 4 markers: miR-196a-5p, miR-182-5p, miR-335-5p and miR-340-5p.

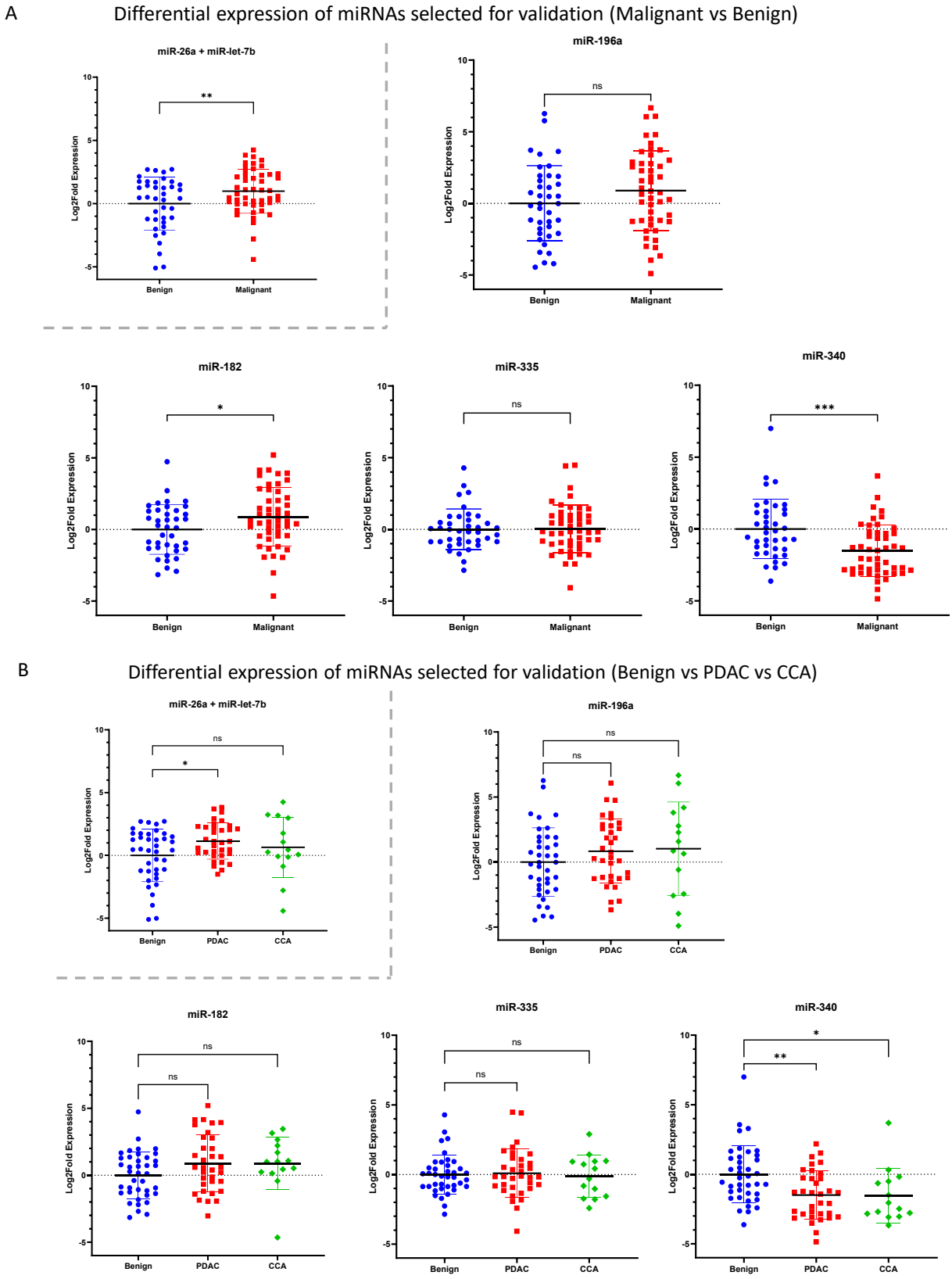


Figure 22: RT-qPCR Validation of 4 marker panel identified by LASSO in bile samples (n=87).

Shown are the results of the RT-qPCR validation for the four microRNAs: miR-196a, miR-182, miR-335 and miR-340 showing expression values for each sample analysed. Bar and whisker plots show mean expressions and standard deviation. Endogenous normalisers miR-26a and let-7b were used (MiR-26a + miR-let-7b indicates the geometric mean of the two expression values) and log fold expression was measured

A. Benign (blue) vs Malignant (red) (48 Malignant, 39 Benign) showing statistical significance in miR-340 and to a lesser extent miR-182.
B. Same 4 marker panel shown by cohort Benign (blue) vs PDAC (red) vs CCA (green) (34 PDAC, 14 CCA, 39 Benign) showing that the difference is greater for PDAC vs Benign.

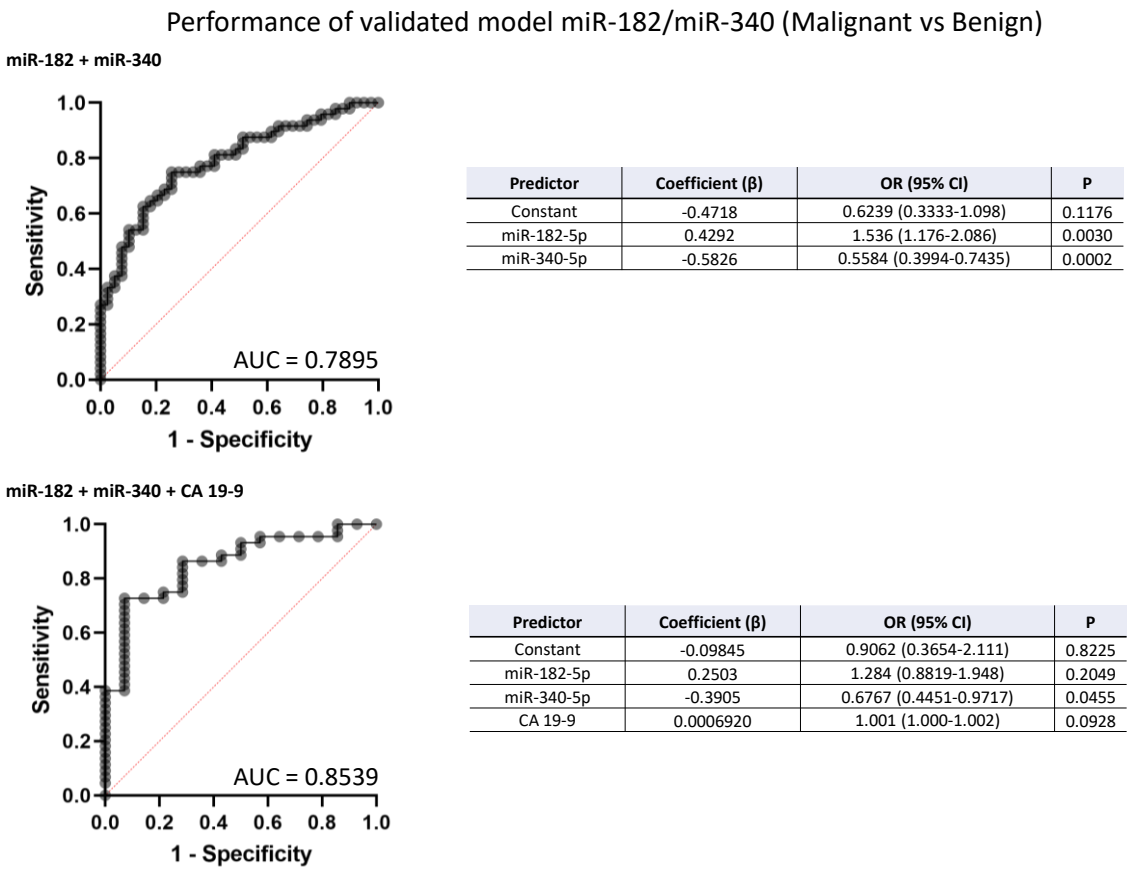
All reactions were performed in triplicate and the mean value used. Asterisks show statistical significance using Student's t-test. * P <0.05, ** P<0.005, *** P<0.0005.

Geometric means of the endogenous controls miR-26a and miR-let-7b were compared between Malignant and Benign populations with increased expression in the Malignant samples (p=0.0093; FIGURE 22A). This was not expected although low stability does not mean that there is no inter-group difference, only that it would be minimised using these. We identified a significant increase in miR-182-5p in Malignant samples (p=0.0179) whilst conversely miR-340-5p was shown to be significantly downregulated in Malignant samples (p=0.0002). miR-196a-5p trended towards an increase in Malignant samples (p=0.0671) whilst miR-335-5p showed no significant difference (p=0.4607) overall. No differences were detected when PDAC and CCA samples were treated separately (FIGURE 22B). However, PDAC and CCA remained significantly differentially expressed in terms of miR-340-5p when compared on their own to the control population (p=0.0275 and p=0.0007, respectively).

To combine these two markers that appear to be differentially expressed in cancer, a multiple logistic regression was undertaken to analyse the fold change in expression of miR-182-5p and miR-340-5p against endogenous controls against the likelihood of malignancy within our validation cohort. This generated a model which can be expressed as:

$$\log(\text{Odds of Malignancy}) = -0.4718 + 0.4292(\text{miR-182-5p}) - 0.5826(\text{miR-340-5p})$$

A



Predictor	Coefficient (β)	OR (95% CI)	P
Constant	-0.4718	0.6239 (0.3333-1.098)	0.1176
miR-182-5p	0.4292	1.536 (1.176-2.086)	0.0030
miR-340-5p	-0.5826	0.5584 (0.3994-0.7435)	0.0002

Predictor	Coefficient (β)	OR (95% CI)	P
Constant	-0.09845	0.9062 (0.3654-2.111)	0.8225
miR-182-5p	0.2503	1.284 (0.8819-1.948)	0.2049
miR-340-5p	-0.3905	0.6767 (0.4451-0.9717)	0.0455
CA 19-9	0.0006920	1.001 (1.000-1.002)	0.0928

Figure 23: Validated ROC Curves for miR-182 and miR-340 in bile cfRNA (Malignant vs Benign; n=87).

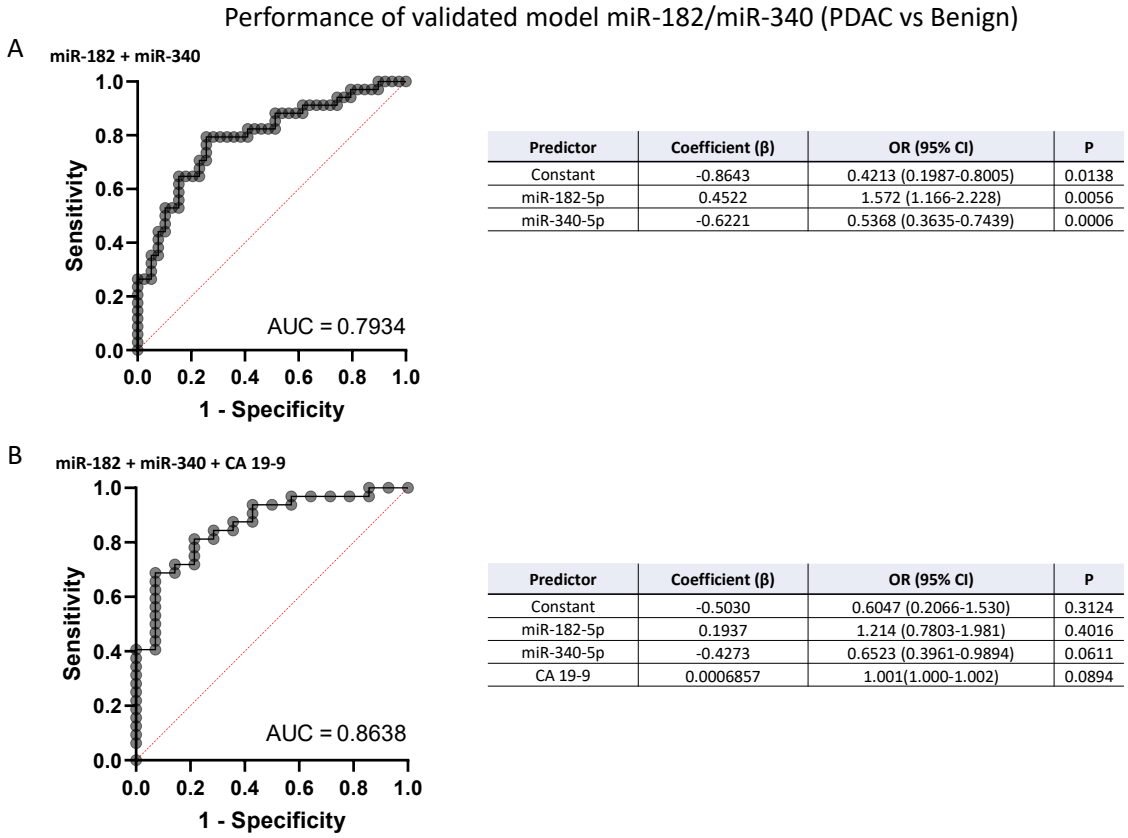
A. Receiver Operator Characteristic analysis for miR-182 & miR-340 using the validated expression values from RT-qPCR show an AUC of 0.79 together with reported results of the multivariate logistic regression used to predict Malignant (PDAC and CCA). Positive values for ß imply a positive association whilst negative values imply a negative association.

B. Receiver Operator Characteristic combined analysis for miR-182, miR-340 and CA 19-9 showing a superior AUC of 0.85 together with results of the logistic regression.

OR – Odds ratio.

Predicted probabilities were generated for this model and were expressed as a ROC curve (FIGURE 23) which had an AUC value of 0.79 (95% Confidence Interval (CI) 0.70 – 0.88, p<0.0001) and at the optimum cut-off (>0.6275) determined by maximum Youden Index classified the Malignant samples with a sensitivity 64.6%, specificity 82.1%, negative predictive value (NPV) of 65.3% and positive predictive value (PPV) of 81.6%. Combining this model with the clinically validated marker CA 19-9 improved the diagnostic potential yielding a curve which had an AUC value of 0.85 (95% CI 0.75 – 0.96, p<0.0001). This model was better at the optimum cut-off (>0.6637) with a diagnostic accuracy of 82.7%, sensitivity of 86.4%, specificity of 71.4%, NPV 62.5% and PPV of 90.5% in the detection of malignancy. Further improvements could be yielded if the model (with CA 19-9) was

limited to the determination of PDAC only (AUC 0.86, 95% CI 0.75 – 0.97, diagnostic accuracy 80.4%, NPV 64.7%, PPV 89.7%: FIGURE 24).



Predictor	Coefficient (β)	OR (95% CI)	P
Constant	-0.8643	0.4213 (0.1987-0.8005)	0.0138
miR-182-5p	0.4522	1.572 (1.166-2.228)	0.0056
miR-340-5p	-0.6221	0.5368 (0.3635-0.7439)	0.0006

Predictor	Coefficient (β)	OR (95% CI)	P
Constant	-0.5030	0.6047 (0.2066-1.530)	0.3124
miR-182-5p	0.1937	1.214 (0.7803-1.981)	0.4016
miR-340-5p	-0.4273	0.6523 (0.3961-0.9894)	0.0611
CA 19-9	0.0006857	1.001(1.000-1.002)	0.0894

Figure 24: Validated ROC Curves for miR-182 and miR-340 in bile cfRNA (PDAC vs Benign; n=87).

A. Receiver Operator Characteristic analysis for miR-182 & miR-340 showing an AUC of 0.79. NPV 80.0%, PPV 71.0% together with reported results of the logistic regression used to predict PDAC only. Positive values for β imply a positive association whilst negative values imply a negative association.

B. Receiver Operator Characteristic combined analysis for miR-182, miR-340 and CA 19-9 showing a superior AUC of 0.86. NPV 69.2%, PPV 84.9% together with reported results of the logistic regression used to predict PDAC only.

OR – Odds ratio.

3.2.14 Correlation analysis of miRNA expression

Expression for these four markers was correlated using Spearman's rank method with standard laboratory markers of biliary obstruction, inflammation, and cancer (bilirubin, CRP and CA 19-9, FIGURE 25). This generates a correlation coefficient (abbreviated here to r_s) which ranges from 1 to -1 where 1 is perfect positive correlation and -1 is perfect negative correlation to determine presence of co-linearity. Overall expression of the 4 miRNAs show low correlation with each other (between -0.26 and 0.50) overall which indicates suitability for use as independent markers. Bilirubin was positively correlated

with CA 19-9 (r_s=0.51, P<0.001) and the most positively correlated miRNA and laboratory marker was miR-182 with CA 19-9 (r_s=0.32, P=0.014).

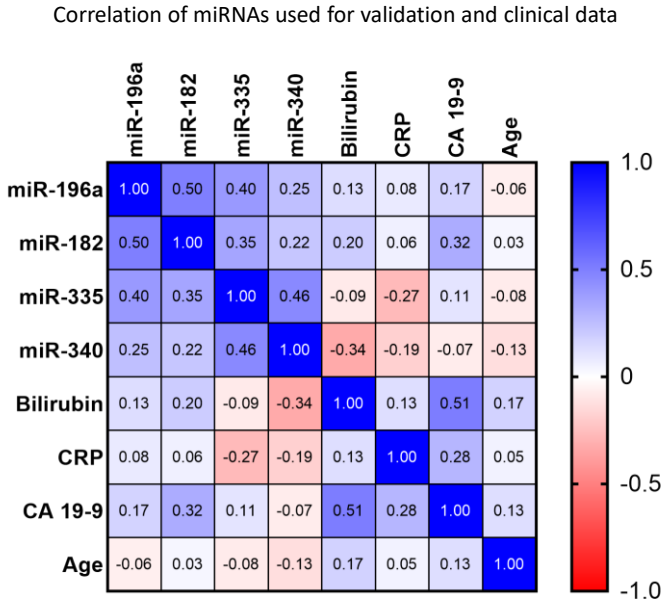


Figure 25: Correlation matrix for miRNAs of interest, laboratory parameters and age.
Heatmap showing the four microRNAs of interest and their expression (miR-196a, miR-182, miR-335 and miR-340) together with a measure of inflammation (CRP), known biomarkers (CA 19-9) and age to determine if there is any association between these. Spearman's rank correlation coefficient where 1.0 is a perfect positive correlation and -1.0 is perfectly inversely correlated. Colour coded with blue demonstrating positive correlation and red, an inverse correlation to assist interpretation.

3.2.15 Functional Analysis of identified miRNAs.

MiEAA (304), an online web-based tool utilising the GeneTrail toolkit, was used to undertake Gene Set Enrichment Analysis (GSEA) of the six candidates. This is a widely used methodology used computationally to determine whether a given gene set (or in this case miRNA set) is significantly enriched in a list ranked by correlation with different phenotypes or processes (305) and can be used to determine whether known biological functions or processes are statistically over-represented within our experimentally derived miRNA list. This is integrated with predicted target databases specifically mirWalk which hosts potential binding sites for miRNAs and precursors of interest. P-values were calculated and adjusted for multiple comparisons. We identified 40 functional pathways for all 6 candidates (including the two which were not validated by RT-qPCR) with an adjusted p-value <0.05 and the top ten are shown in TABLE 20.

This analysis highlighted pathways including Bone morphogenetic proteins (BMP) signalling, Toll-like Receptor (TLR) pathways and involvement of the endosome. BMP signalling is a group of proteins which act as growth factors and belong to the TGF-β superfamily. Although several are described, these proteins are believed to be active and present in pancreatic cancer, leading to increased invasiveness in PANC-1 cells via increased Smad1 signalling (306). In particular, BMP 2 and 4, (and to a lesser extent BMP7) were potent mediators of epithelial to mesenchymal transition (EMT) but have also been found to positively regulate glucose sensing and insulin secretion in normal pancreatic cells (307). Toll like receptors are another class of proteins that recognise structural patterns derived from microbes in order to activate immune responses and therefore are found on various immune cells (e.g. dendritic, macrophages, T-cell, and B cells) as well as epithelial cells and fibroblasts in humans. Interestingly, certain miRNAs (such as miR-21 and miR-29a) activate TLR, specifically TLR8 *in vitro* and have been shown to induce a pro-metastatic inflammatory response (TNF-α and IL-6 secretion) in mice exposed to exosomes containing these miRNAs (308). TLR 1, 3, 5, 7 and 9 are also found expressed in PDAC tissue (309) so it may be possible that these miRNAs act on these to activate the immune system and mediate tumour progression.

Table 20: MiRNA Gene Set Enrichment Analysis.

Results of a functional pathways analysis undertaken using MiEAA using the miRNAs (miR-199a-5p, miR-196a-5p, miR-335-5p, miR-182-5p, miR-340-5p and miR-424-5p) and adjusted p-values (Benjamini-Hochberg) generated using hypergeometric distribution. MiRWalk was used as the database for Functional target interactions and top ten pathways are shown for simplicity.

Gene Ontology			
Subcategory	**Enrichment**	**P-value**	**P-adjusted**
GO0030509 BMP Signalling Pathway	over-represented	1.47e-5	0.021
GO0007595 Lactation	over-represented	6.94e-5	0.021
GO0009791 Post Embryonic Development	over-represented	8.89e-5	0.021
GO0005768 Endosome	over-represented	1.30e-4	0.021
GO0003705 RNA Polymerase II Distal Enhancer Sequence Specific DNA Binding Transcription Factor Activity	over-represented	1.48e-4	0.021
GO0002224 Toll Like Receptor Signaling Pathway	over-represented	1.55e-4	0.021
GO0034130 Toll Like Receptor 1 Signaling Pathway	over-represented	1.55e-4	0.021
GO0034134 Toll Like Receptor 2 Signaling Pathway	over-represented	1.55e-4	0.021
GO0005635 Nuclear Envelope	over-represented	1.69e-4	0.021
GO0008063 Toll Signalling Pathway	over-represented	1.69e-4	0.021

3.3 Discussion

3.3.1 Previous literature on bile cfRNAs shows heterogenous results.

Only 5 human studies have been published utilising bile fluid as a source of extracellular microRNAs for the identification of PDAC or CCA (TABLE 21). No consistent biomarker has been found in any of these studies although different methodologies for discovering markers was used in each. Ours is the largest study to date to prospectively identify miRNAs from bile using next generation RNA sequencing and validation of them by RT-qPCR. Despite this, our study did not appear to correlate with any marker previously investigated.

Table 21: Studies identified to-date involving small RNA markers in bile.
Literature search involved keywords including RNA, miRNA, pancreatic cancer and bile taken from the Early Stage Report submitted Friday 20th September 2019 and updated 1st November 2022.
AUC – Area Under the receiver operator characteristic Curve, EV – Extracellular Vesicle, PDAC – Pancreatic ductal adenocarcinoma, PSC – Primary Sclerosing Cholangitis, CCA – Cholangiocarcinoma.

MicroRNA studies in Bile						
First Author	**Year**	**No. of Patients**	**Biofluid**	**EV Study**	**Bile Biomarker**	**Significant comments**
Shigehara et al.	2011	CCA (n=9), Controls (n=9)	Human bile	No	miR-9, miR-145*, miR-105, miR-147b, let-7f-2*, let-7i*, miR-302c*, miR-199a-3p, miR-222* and miR-942	Showed stability of endogenous miRNA in bile at room temperature up to 480 minutes later. AUC for miR-9 was 0.975.
Cote et al.	2014	PDAC (n=77), Benign (n=67), Controls (n=71)	Human plasma and bile	No	miRNA-10b, -30c, -106b, -155, and -212	AUC >0.90 for this 5-miRNA panel.
Li et al.	2014	CCA (n=46), Controls (n=50)	Human bile	Yes	A panel of 5 microRNAs: miR-191, miR-486-3p, miR-1274b, miR-16, miR-484	Showed stability of miR expression at room temperature and across at least three freeze-thaw cycles. Comparison of different mathematical modelling was used and focussed on specificity. 5-miR panel displayed a sensitivity of 67% and specificity of 96%.
Voigtländer et al.	2015	CCA (n=19), PSC (n=52), PSC/CCA (n=12), Controls (n=12)	Human bile and serum	No	MiR-412, miR-640, miR-1537 and miR-3189	Array-based screening was used to identify markers. Combination of miR-1537 and CA 19–9 in a logistic regression model resulted in AUC of 0.91.
Han et al.	2019	CCA (n=48), Benign (n=58)	Human bile	No	miR-30d-5p and miR-92a-3p	AUC for the combination of two miRNAs and serum CA19-9 was 0.689.

Shigehara *et al.* (310) in 2011 investigated n=9 patients with CCA and compared for the first-time different fractions of bile obtained by centrifugation for total RNA microRNAs with 9 patients with choledocholithiasis. They identified the presence of miR-21, with let-7b being the most common miRNA found overall. 335 out of 667 miRNAs were found to be differentially expressed with a $p<0.05$ whilst 10 miRNAs were found to be differentially expressed with four miRNAs (miR-9, miR-302c*, miR-199a-3p, and miR-222*) having a sensitivity level of 88.9% with a specificity of 100% for the detection of CCA ($p<0.0005$). They were able to show that endogenous miRNAs miR-21, let-7c, and miR-9 were relatively stable whilst synthetic exogenous miRNA cel-miR-39 showed degradation (in minutes) when added to bile, which could be due to their presence in EVs or bound to proteins such as Ago (192).

Cote et al (311) is the largest study to date to investigate bile miRNA markers, using a pre-selected panel of 10 markers (miR-10b, -21, -30c, -106b, -132, --155, -181a, -181b, -196a, and -212) with prior putative evidence for PDAC associated overexpression. Focussing only on their bile study, they used a training cohort consisting of 25 patients with PDAC, 23 patients with chronic pancreatitis and 22 controls. Our study discovery cohort in comparison investigated a similar size (27 patients with PDAC and 37 controls).

In total, 110 unique patient samples were used in this study and corresponded to 38 patients in the PDAC cohort, 19 patients in the CCA cohort and 54 patients in the Benign cohort. Although this did not achieve the numbers determined by our *a priori* power calculation (limited in part by restrictions caused by COVID-19), we were able to achieve comparable numbers to the Cote study and additionally is the largest cohort of bile samples to be sequenced for the purposes of PDAC and CCA biomarker discovery.

3.3.2 A bile cfRNA biomarker has been shown to work in a suitable population at risk for PDAC/CCA

Our large cohort of benign patients allowed for age and sex to be matched within the discovery phase which is a strength of our study in comparison with other biomarker studies. These are often biased by volunteer populations used as controls which are not representative of a screening population and unlikely to yield diagnostic value when the markers are taken forward into a clinical environment. Patients with choledocholithiasis present at an earlier age, so selecting older patients from this is likely to yield additional diagnostic information.

It is important to note that our study evaluated patients with some degree of biliary obstruction (serum jaundice levels ranging from 42.1-253.6) and which affected the measured CA 19-9 in these patients. In benign patients, CA 19-9 is correlated with jaundice levels (312), which means that in this population CA 19-9 has a reduced diagnostic value in detecting cancer (see Figure 7), which makes the benefit of a miRNA diagnostic test in these patients more significant.

Initial analysis of the cfRNA concentrations identified significant differences in quantity obtained from patients with cancer compared with those with benign diagnoses. Average concentrations for the patients from the Ampullary population were calculated at 6.296 ng/µl, Benign at 15.08 ng/µl, CCA at 78.98 ng/µl, and PDAC at 19.83 ng/µl. This was

significantly raised in CCA and indicates that there is a sufficiently ample source of RNA to be found in the bile of patients with malignant biliary strictures.

3.3.3 Biomarker discovery using sequencing and RT-qPCR validation generates models with good predictive ability.

Next generation sequencing is the preferred standard for miRNA biomarker discovery without *a priori* selection bias(313), and all 2656 mature known human miRNAs (miRBase v22) and all 26,531 known tRNA fragments (MINTbase v2.0) have been investigated. This gave us a profile of 457 miRNAs and 3176 tRNA fragments were identified with sufficient expression for analysis. After multiple hypotheses correction, 88 differentially expressed miRNAs were found between Malignant vs Benign, 75 miRNAs between PDAC vs Benign and 26 candidates between CCA vs Benign.

LASSO regression analysis was used to generate a sparse model which included six upregulated miRNAs (miR-199a-5p, miR-196a-5p, miR-335-5p, miR-182-5p, miR-340-5p, miR-424-5p). RT-qPCR was a necessary step if we were to generate a clinically useful biomarker which could be prospectively validated in bile samples. We demonstrated that miR-182-5p showed a significant increase in Malignant samples ($p=0.0179$) whilst conversely miR-340-5p was shown to be significantly downregulated in Malignant samples ($p=0.0002$). miR-196a-5p trended towards an increase in Malignant samples ($p=0.0671$) whilst miR-335-5p showed no significant difference ($p=0.4607$) overall. miR-199a-5p and miR-424-5p were excluded due to their low detection (in <50% of samples measured).

From this miR-182-5p and miR-340-5p were combined to generate a model based on RT-qPCR that predicted Malignancy with an AUC value of 0.79 (95% CI 0.70 – 0.88, $p<0.0001$) and at the optimum cut-off (>0.6275) sensitivity 64.6%, specificity 82.1%, NPV of 65.3% and PPV of 81.6%. Combining with the clinically validated marker CA 19-9 improved the diagnostic potential to an AUC value of 0.85 with further improvement if the regression analysis was limited to the determination of PDAC only (AUC 0.8539, diagnostic accuracy 80.4%, NPV 64.7%, PPV 89.7%). This has the potential to increase the sensitivity of Ca 19-9 in the presence of biliary obstruction and jaundice, allowing us to distinguish between a malignant rise in CA 19-9 and benign disease.

3.3.4 Novel small RNA biomarkers include tRNA fragments (tRFs) in bile and may be a source of further investigation.

This was the first study of its kind to analyse the presence of tRFs in bile for significant candidate biomarkers. I found that this species was identified as highly expressed with a mean 37% composition in bile (range 4-84%) compared to miRNAs (36%, range 0-85%) and 3176 unique candidates were found to be significantly expressed and investigated for differential expression. After multiple hypothesis correction, 264 differentially expressed tRFs were found between Malignant vs Benign with 474 potential candidates in PDAC vs Benign and 65 potential candidates between CCA vs Benign.

The tRFs found in our discovery sequencing was not taken forward into the validation phase but could be a source of future biomarkers, in isolation or better still in combination. This is further supported by our identification that a number of bile samples (6 out of 78 samples, 7.7%) had less than 1% miRNA content and were not available for miRNA analysis (FIGURE 8C).

A brief analysis of tRFs found similar fragments varying only by a short number or single nucleotide within our differential expression analysis. For example, members of the 5'-tRF Gly-GCC were found to be upregulated with Gly-GCC_0-30_tRF-5 and Gly-GCC_0-30_NA varying by a single nucleotide. It was also noticed that members of the i-tRF Asp-GTC were found to be upregulated in the PDAC vs Benign comparison with Asp-GTC_40-71_i-tRF and Asp-GTC_40-72_i-tRF varying by a nucleotide in length. Various methods have been proposed for classification of these RNA molecules derived from tRNAs, including tRNA-derived fragments, tsRNAs, tiRNAs, SHOT-RNAs (Sex HOrmone-dependent tRNAs) and tRNA halves (314). It may be that these molecules share a functional target with overlapping sequence homology which is yet to be elucidated.

Our study contained several limitations. First, our analysis featured samples from two centres with a predominance of samples from our primary study undertaken at the Hammersmith Hospital. Independent validation in a separate cohort would provide stronger evidence for our model and would strengthen implementation in clinical practice. We investigated a pre-diagnostic cohort of patients presenting with biliary obstruction which meant that the clinical demographics of the PDAC population included predominately advanced disease. This prevented us from identifying markers that demonstrate a positive correlation with disease progression. Ideally, a marker in

bile cfRNA would be utilised for the screening of early PDAC which would be potentially curative.

3.3.5 Limitations and future strategies

Cell-free RNA as a source of markers was shown to have some drawbacks clinically. Heterogeneity of small RNA species was identified with proportions of miRNAs varying from 0 to 85%. Plasma cfRNA showed more consistent miRNA expression (mean miRNA composition 61% and range 37-90%) which indicated that plasma might be more suitable for miRNA analysis, but we were only able to identify two significant candidates for further evaluation (miR-144-5p and miR-29-3p). In combination with the number of failed ERCPs (samples either not collected or insufficient material for analysis) which was 11.7% (FIGURE 5C), this leads to potential further problems with the clinical applicability of cfRNA, at least with current RNA isolation techniques.

In particular, TaqMan RT-qPCR remained reliant on RNA concentration and quantification. 33 samples which comprised 25% of samples extracted contained RNA that below the limit of detection (<0.2ng/μl). Repeating RNA extraction in 21 samples showed that 81% (17 samples) contained subsequently higher concentrations of RNA implying that low RNA was sometimes due to extraction inefficiency. Alteration of RNA extraction methodology may lead to improved yields and ultimately diagnostic potential. Further alternative RT-qPCR methods have been subsequently developed with improved binding strength and specificity which are discussed in the next chapter.

Our use of LASSO regression as a method of biomarker identification led to six candidate miRNAs which resulted in a two-marker model after RT-qPCR validation. This is an automated method to generate a sparse model when there are many differentially expressed candidates. It is unclear if these candidates are functionally related, but we were able to show that this set of miRNAs were statistically overrepresented in pathways associated with cancer progression. Alternatively, we could use a functional analysis to identify candidates of interest. We also could consider adjusting the filter threshold to evaluate some of the more lowly expressed markers.

NormFinder was used to generate a pair of miRNAs and we chose to use miR-26a and miR-let-7b as they had a low stability. However, the results of our RT-qPCR suggest that these in fact are differentially expressed in Malignant compared to Benign samples which may introduce some Type II bias into the results (i.e. reducing the overall power

of the study). Subsequent experiments in the next chapter removed this issue by repeating several endogenous candidate normalisers by RT-qPCR in order to confirm their stability with the different methodology.

In summary, we have shown that there is differential expression of cfRNAs (miRNAs and tRFs) which are suitable for the diagnosis of PDAC and CCA. RNA sequencing can be used for the discovery of these molecules and identified candidates for RT-qPCR validation. Our final model (miR-182-5p and miR-340-5p) generated an AUC value of 0.85 in the detection of malignancy, rising to 0.86 when used for the prediction of PDAC only. These two miRNAs were furthermore shown to target BMP signalling suggesting that they could be associated with EMT. Although this has not appeared to corroborate with any previous historical studies investigating bile cfRNA, we have demonstrated that bile cell-free miRNAs can indeed be a source of biomarkers for the detection of PDAC and biliary tract malignancy.

Chapter Four – Small RNA markers for the detection of PDAC and CCA in Bile Extracellular Vesicles

4.1 Background

Extracellular vesicles (EVs) are lipid bilayered structures which are produced by cells and can be classified according to their biogenesis. The term "extracellular vesicle" can be defined as particles that are "naturally released from the cell … delimited by a lipid bilayer and cannot replicate, i.e. do not contain a functional nucleus." based on the MISEV (Minimal Information for Studies of Extracellular Vesicles) position statement by the International Society for Extracellular Vesicles published in 2014. What constitutes EVs make up a diverse category that include large membrane bound structures such as apoptotic bodies and smaller structures such as microvesicles. Whilst all vesicles are included in this umbrella term, many types have been characterised and shown to have distinct cellular biogenesis, as well as differing size distribution and protein markers. A popular term is "exosome" and is generally used for vesicles that are 30-150nm in size and specifically derived from late endocytic compartments, known as multivesicular endosomes (MVEs) which fuse with the membrane of the cell. It is known that part of the extracellular RNA found in biofluids are within this compartment and are a potential source of specific marker, excreted by cells and resistant to degradation by enzymes.

The aims of the second part of this project were to investigate small non-coding RNAs (miRNAs in particular) present in EVs as biomarkers that could aid diagnosis in patients with a biliary stricture. This occurred in three stages. First step was (1) the isolation and characterisation of EVs from patients with benign and malignant biliary strictures (i.e. pancreaticobiliary cancers) through samples obtained from surgery. This was then optimised and used on bile obtained prospectively through the MIRABILE study to (2) identify differentially expressed EV-miRNAs in plasma and bile suitable for detecting pancreaticobiliary malignancy. The final stage of this process was to (3) validate these for clinical use and if possible, determine the functional relevance of putative biomarkers in relation to pancreatic cancer.

4.2 Results

4.2.1 Surgical samples used for initial optimisation of EV isolation.

Surgical bile was selected for our initial optimisation of EV isolation (see 2.1.1 Sample Collection from Surgery). This was obtained from patients consented for pancreaticoduodenectomy or cholecystectomy and obtained by aspiration from the common hepatic duct (during pancreaticoduodenectomy) or gallbladder using a sterile syringe and large-gauge needle. After collection bile samples were kept at -80°C until further use. Samples had been collected between 17th May 2017 and 29th May 2018.

4.2.2 Initial use of ultracentrifugation as a method of EV isolation.

Our initial study was based on the use of ultracentrifugation (UC) as a method for the isolation of EVs from our bile and plasma samples. This was considered the "gold standard" for EV isolation and remains the most popular method in the literature at the time we began the project (2018). In order to develop a protocol that would isolate EVs from bile, a literature search was undertaken to identify papers used to determine the features common to each UC protocol for isolation from bile. Studies were narrowed down to 3: Masyuk *et al.* (253) published in 2010, Li *et al.* published in 2014 (254) and Severino *et al.* (223) in 2017. These were compared to each other, and a schematic of their methods can be seen below in FIGURE 26.

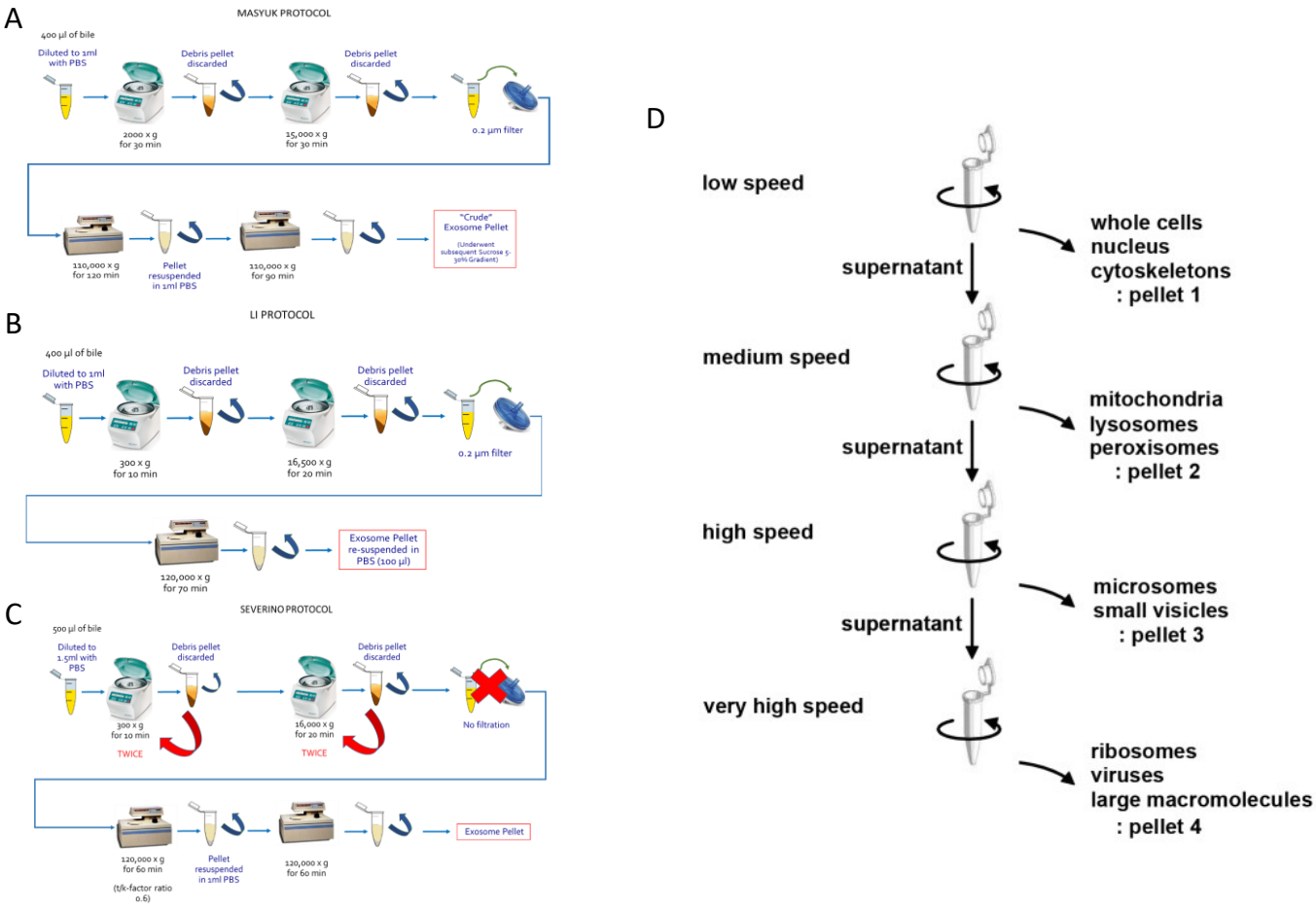


Figure 26: Differential ultracentrifugation protocols used in prior bile EV studies.

Previous studies investigating the role of biliary EVs in disease were evaluated for their method of EV isolation and this was used to inform the ultracentrifugation protocol used in my initial studies.

A. Schematic for the protocol used in the 2010 paper by Masyuk *et al.* (253). This involved a dilution of bile to 1mL before a 200 x g spin for 30 minutes. This was followed by a 15,000 x g spin for 30 minutes before filtration and two 110,000 x g ultracentrifugation spins. The final pellets were resuspended in PBS and further purified using a 5-30% sucrose gradient.
B. Schematic for the protocol used in the 2014 paper by Li *et al.* (254). This also involved a dilution of bile to 1 mL before a 300 x g for 10 minutes. This was followed by a 16,500 x g spin for 20 minutes before filtration and a single 120,000 x g ultracentrifugation spin.
C. Schematic for the protocol used in the 2017 paper by Severino *et al.* (223). This involved a dilution of bile to 1.5 mL before a spin at 300 x g for 10 minutes twice, followed by a 16,000 x g spin for 20 minutes, twice. A 120,000 x g ultracentrifugation spin for 60 minutes was then performed twice with the remaining pellet.
D. Schematic drawn from the 2011 paper by Shigehara *et al.* (310) explaining the rationale for a four-step protocol.

PBS – Phosphate Buffered Saline

The earliest paper identified was Masyuk *et al.* (253) who obtained bile from the cannulated common bile duct (CBD) of male Sprague-Dawley rats and were isolated through a differential ultracentrifugation method previously devised for human urine (FIGURE 26A). Their method consisted of 400 µL of bile diluted to 1mL before a low-speed spin of 200 x g for 30 minutes. This was followed by a 15,000 x g spin for 30 minutes before filtration through a 0.2 µm filter and two 110,000 x g ultracentrifugation spins (at 120 minutes and then at 90 minutes). The final pellet was called "crude exosome pellets".

Crude pellets were resuspended in Phosphate-buffered saline (PBS) and further purified using a 5-30% sucrose gradient, obtaining exosomes with a density of 1.12–1.21 g/ml. They were able to demonstrate interactions with cholangiocyte cilia *in vitro* and that this led to increased expression of miR-15A, an miRNA associated with cholangiocyte proliferation and with roles in the ERK signalling pathway.

Li et al (254) investigated a total of 96 human bile samples (CCA n=46, Controls n=50) obtained at ERCP or PTC and used a more simplified UC protocol (FIGURE 26B) to obtain EVs suitable for RNA microarray discovery and validation by RT-qPCR. They used a sample volume that ranged from between 400-1000 μL of bile sample which was diluted to 1 mL and processed with a low-speed spin at 300 x g for 10 minutes. This was followed by a 16,500 x g spin for 20 minutes and passed through a 0.2 μm filter. They undertook a single 120,000 x g ultracentrifugation spin and resuspended the pellet in 100 μL of PBS.

The most recent paper was one by Severino *et al.* (223) and which drew bile samples from 50 patients with PDAC (n=20), CCA (n=5), non-malignant/chronic pancreatitis (n=5) and compared this to control patients with biliary stones (n=10). 500 μL samples were taken at ERCP and diluted to 1.5 mL. A low-speed spin at 300 x g for 10 minutes was undertaken twice, followed by a 16,000 x g spin for 20 minutes, twice. A 120,000 x g ultracentrifugation spin for 60 minutes was repeated which resulted in a final exosome pellet. From this they determined that concentrations using Nanoparticle tracking analysis (NTA) could at particular thresholds diagnose PDAC and CCA in bile and serum.

Whilst all three protocols varied in protocol, the general principles of each could be understood through the explanation given by Shigehara *et al.* (FIGURE 25C) in a paper from 2011 (310). This group had undertaken bile sample analysis and compared the pellets obtained at various stages of the centrifugation process, describing a low-speed spin to remove cells, nuclei, and cellular cytoskeletons, followed by an intermediate spin to obtain mitochondria, lysosomes, and peroxisomes. Microsomes and small vesicles were described as pelleting at high speeds (e.g. 80,000 x g for 1 hour) followed by the smallest components (e.g., ribosomes, viruses, and large macromolecules). They concluded from these results that, in fact, the vast majority of miRNAs circulating in bile were to be found in the first pellet i.e. cells and nuclei. From this, it was therefore decided *a priori* to combine the various features of each protocol i.e. a three-stage spin: low, medium and high, together with initial dilution, subsequent filtration and a double high-speed spin to ensure high EV purity.

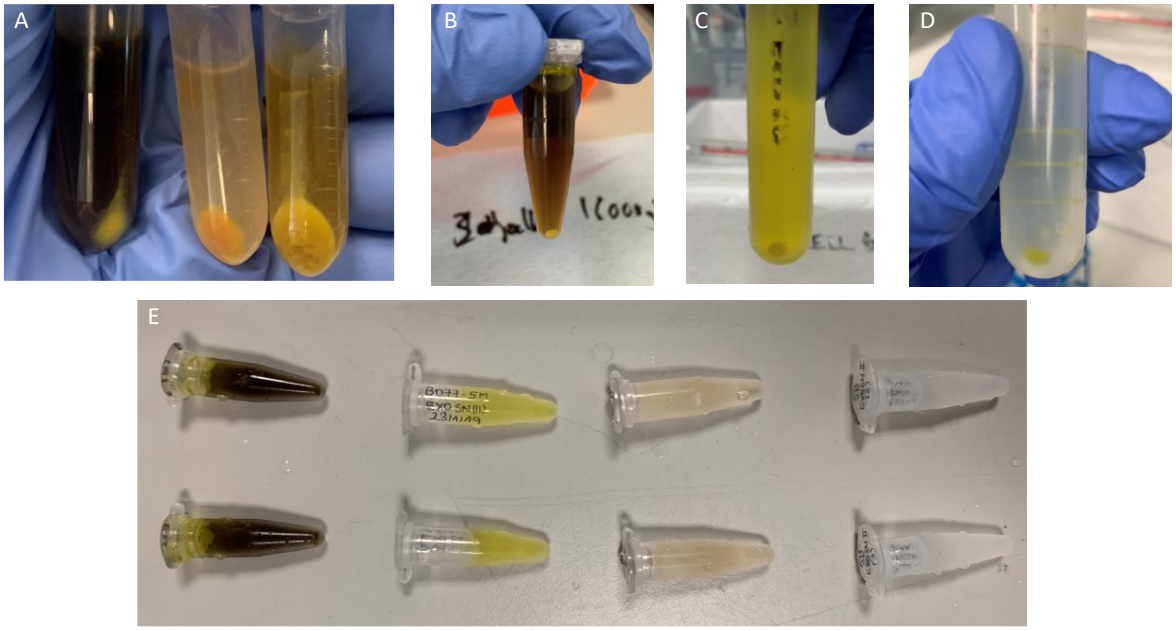


Figure 27: Photos showing initial optimization using differential ultracentrifugation as a method for EV isolation.

Photographs showing the results during and after the UC protocol (explained in detail below). Samples underwent low speed centrifugation at 300 x g for 10 minutes at 4°C occurred twice, with the pellets discarded (A). The supernatant was then centrifuged at 16,000 x g for 20 minutes at 4°C twice and the pellet stored for later protein analysis (B). Samples for UC were added to 10mL polypropylene oak ridge bottles and diluted with 9 mL PBS, closed using a non-sealing cap assembly and placed in a fixed angle rotor and spun at 120,000 x g for 70 minutes at 4°C in a Sorvall Discovery 100S. Pellets were retained (C), diluted with 9 mL PBS and re-spun at 120,000g for 70 minutes at 4°C with final pellet prior to removal of supernatant shown (D).

A. Large pellets after initial 300 x g spin which is subsequently discarded.
B. Pellet after 16,000 x g spin which is subsequently discarded.
C. EV Pellet after 120,000 x g spin which is retained.
D. EV Pellet after second 120,000 x g spin which is retained and stored for further analysis.
E. Photo highlighting the variation in bile samples obtained by colour and viscosity. Shown here are the supernatants remaining from the second centrifugation spin (16,000 x g).

FIGURE 27 shows pictures of pellets obtained using our protocol on surgical bile samples. Our protocol involved a preparation step whereby 500 µL bile samples were thawed on ice and diluted with 500 µL PBS. Low speed centrifugation at 300 x g for 10 minutes at 4°C occurred twice, with the pellets (FIGURE 27A) discarded. The supernatant was then centrifuged at 16,000 x g for 20 minutes at 4°C twice and the pellet stored for later protein analysis (labelled 16k; FIGURE 27B). Samples for UC were added to 10mL polypropylene oak ridge bottles and diluted with 9 mL PBS, closed using a non-sealing cap assembly and placed in a fixed angle rotor and spun at 120,000 x g for 70 minutes at 4°C in a Sorvall Discovery 100S. Pellets were retained (FIGURE 27C), diluted with 9 mL PBS and re-spun at 120,000g for 70 minutes at 4°C. The final pellets (FIGURE 27D) were resuspended in

100 µL PBS and stored for analysis. FIGURE 27E shows the variability in bile samples obtained from surgery with different compositions ranging from dark, inspissated and extremely viscous fluids, to watery pink samples almost devoid of protein.

Difficulties arose with UC at this stage of the EV isolation process. Specific polypropylene tubes suitable for use with our fixed angle T-1270 rotor took several months to order and obtain and were limited in number to 50. Re-use with sterilisation might be possible between patients, but re-suspension of pellets led to visible damage to the tubes themselves and was furthermore a time-consuming process. There were additional failures due to the UC machines used which required high speeds under vacuum and which would intermittently fail, leading to an incomplete isolation of the samples. After each spin, it was also required to carefully pipette away the supernatant which is a user-dependent process with a small but significant loss of material during each stage. With a maximum of 12 samples run over 2 hours and additional time required to process the EVs (resuspension in PBS and time taken to sterilise equipment) it was determined not to be a feasible process for batch processing of clinical samples to obtain EVs.

It was therefore decided to abandon UC as a method of isolation, and we instead looked for other potential methods. After attending an EV workshop in May 2019 run by EMBO in Heidelberg, other methods such as density-gradient centrifugation, size-exclusion chromatography (SEC) and proprietary kits were all investigated and compared. We eventually settled on SEC as a method which was only beginning to be utilised, due to its rapid turn-around time, low user-dependence and quality of EV isolate (212, 315).

4.2.3 Size-exclusion as a method of EV isolation improves workflow.

FIGURE 28A shows the initial process of size exclusion chromatography (SEC) using Izon qEVoriginal columns. These are disposable columns containing a proprietary inert material which was stored at 4˚C until required. The schematic diagram below drawn for the review paper (Liu *et al.* (215)) shows how a stationary phase (in this case a gel) is held in a column which allows a liquid mobile phase to pass through. This contains the analyte in aqueous buffer solution, and which leaves the column at a rate proportional to its size/hydrodynamic volume. Different gels have an optimal molecular weight range which

will allow good linear separation (FIGURE 28B) and the Izon columns used are optimised for isolation of EVs between 70-1000 nm. Shown are 500µL fractions which change colour as bile pigments (bilirubin and biliverdin plus others) are eluted with continuous passage of PBS. Additionally, the speed of the SEC process can be further optimised with automation if desired.

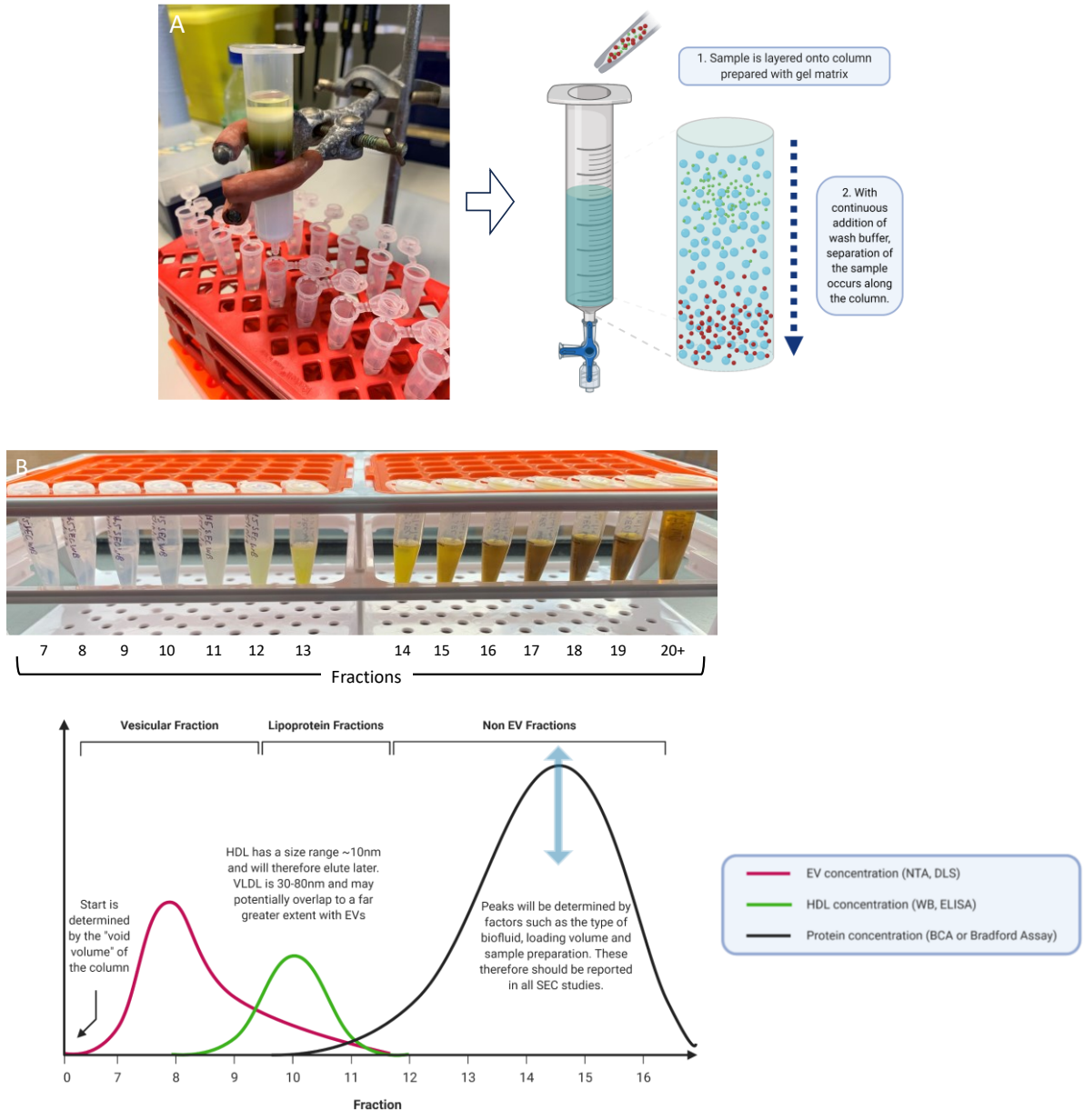


Figure 28: Photos and diagrams showing optimization using SEC as a method for EV isolation.

A. Clamp stand with Izon 70nm size exclusion column containing bile sample and eluting into prepared and labelled 1.5mL eppendorf tubes. Below is a schematic drawn from Liu *et al.* (215) explaining the mechanism of size exclusion.
B. Eppendorfs containing eluted labelled fractions (500µL) showing the change in colour as increasing non-vesicular protein passes through the column. Below is a schematic showing the theoretical performance expected in terms of vesicular and non-vesicular protein measured.

Our workflow for the processing of bile samples for EV analysis is shown in FIGURE 29. Bile samples required initial processing to remove cellular debris and microlithiasis which involved 500 µL bile samples thawed on ice, diluted with 500 µL PBS and centrifuged at 300 x g for 10 minutes at 4°C twice, taking the supernatant and discarding the pellets. The supernatant was then centrifuged at 16,000 x g for 20 minutes at 4°C twice before it was ready for SEC isolation.

Sample selection was undertaken based on patient age and sex and according to retrospectively confirmed diagnosis and EVs isolated from a fixed volume (500µL). Fractions from benign (n=3) and PDAC (n=3) were individually analysed using Nanoparticle Tracking Analysis, Transmission Electron Microscopy and Protein analysis (Western blotting and quantification) to determine EV and non-EV components (FIGURE 29B). Based on these results, pooled EV fractions (subsequently determined as fractions 7-12) were to be used to determine EV concentrations of n=46 samples (to verify the findings by Severino *et al.)* and a selection of samples (n=18) would go on for RNA sequencing after extraction of RNA from the EVs. As I will describe below, this selection process was based on choosing 6 samples from each disease cohort based on a maximum EV-protein purity ratio i.e., those with a high number of seen particles by NTA and minimum protein concentration determined by BCA.

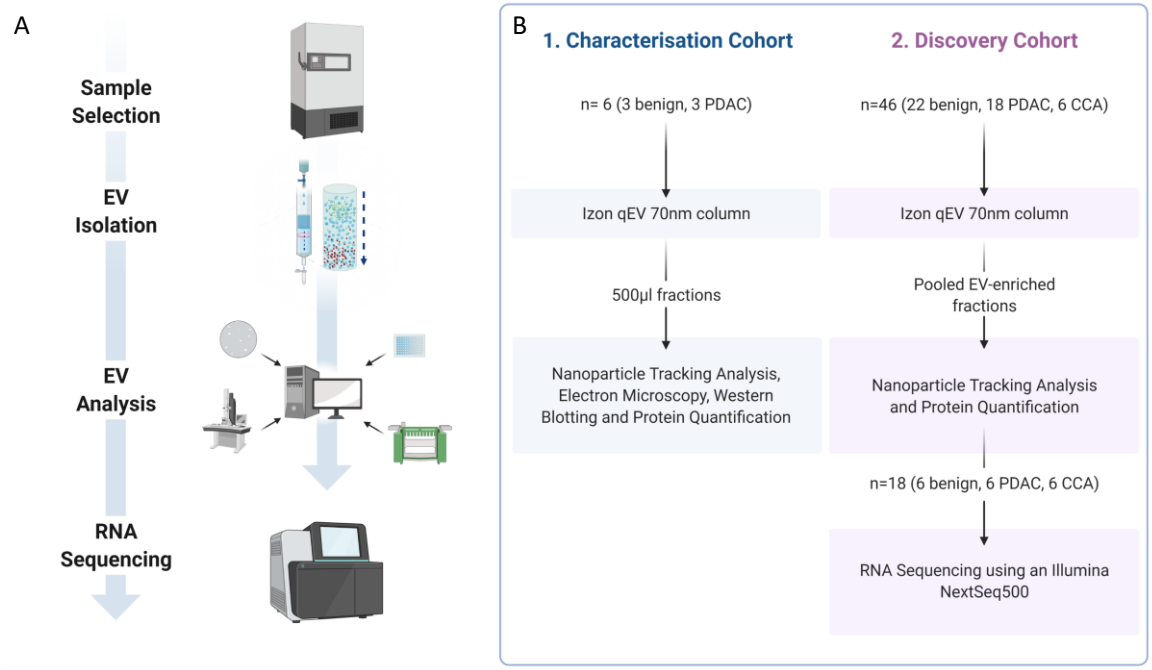


Figure 29: Study Design for the evaluation of EVs from surgical bile samples.

A. Contrasted with Chapter 3 Figure 3 the EV workflow includes additional steps which involve EV isolation (either by UC or as later explained, by SEC), EV analysis and RNA extraction (using a TRIzol protocol) of these samples.

B. Initially samples underwent determination of EV characteristics according to MISEV guidelines using Nanoparticle Tracking Analysis, Transmission Electron Microscopy and Protein analysis (Western blotting and quantification). This was followed by further characterisation of a discovery cohort of 46 samples with sequencing of 18 samples comprising of PDAC, CCA and Benign cohorts. These were selected on the basis of EV purity (protein to particle ratio). Small RNA sequencing data could then be followed by differential expression analysis of small RNA markers. This then informed the candidates which would be validated on a separate cohort by RT-qPCR using a TaqMan primer-probe system.

PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, EV – Extracellular Vesicle, MISEV – Minimal Information for the Study of Extracellular Vesicles

4.2.4 Surgical Bile EV Characterisation by NTA and TEM

Individual 500 µL fractions were eluted (up to 25 fractions for each sample) and analysed for the presence of individual vesicles using TEM. FIGURE 30A showed the presence of ~100nm bilayered structures with negative staining using uranyl acetate. These are described as typically "cup-shaped" due to the desiccation process that takes place when the grids are prepared. Qualitatively, it could also be seen in the wide field views (4000x magnification) the presence of other extracellular protein structures which appeared at fraction 14 and continued into fraction 20. Immunogold labelling of CD63, a tetraspanin molecule present on the surface of EVs was done for a single Benign and PDAC bile sample. FIGURE 30B shows the binding of 10nm gold conjugated secondary antibody to EVs.

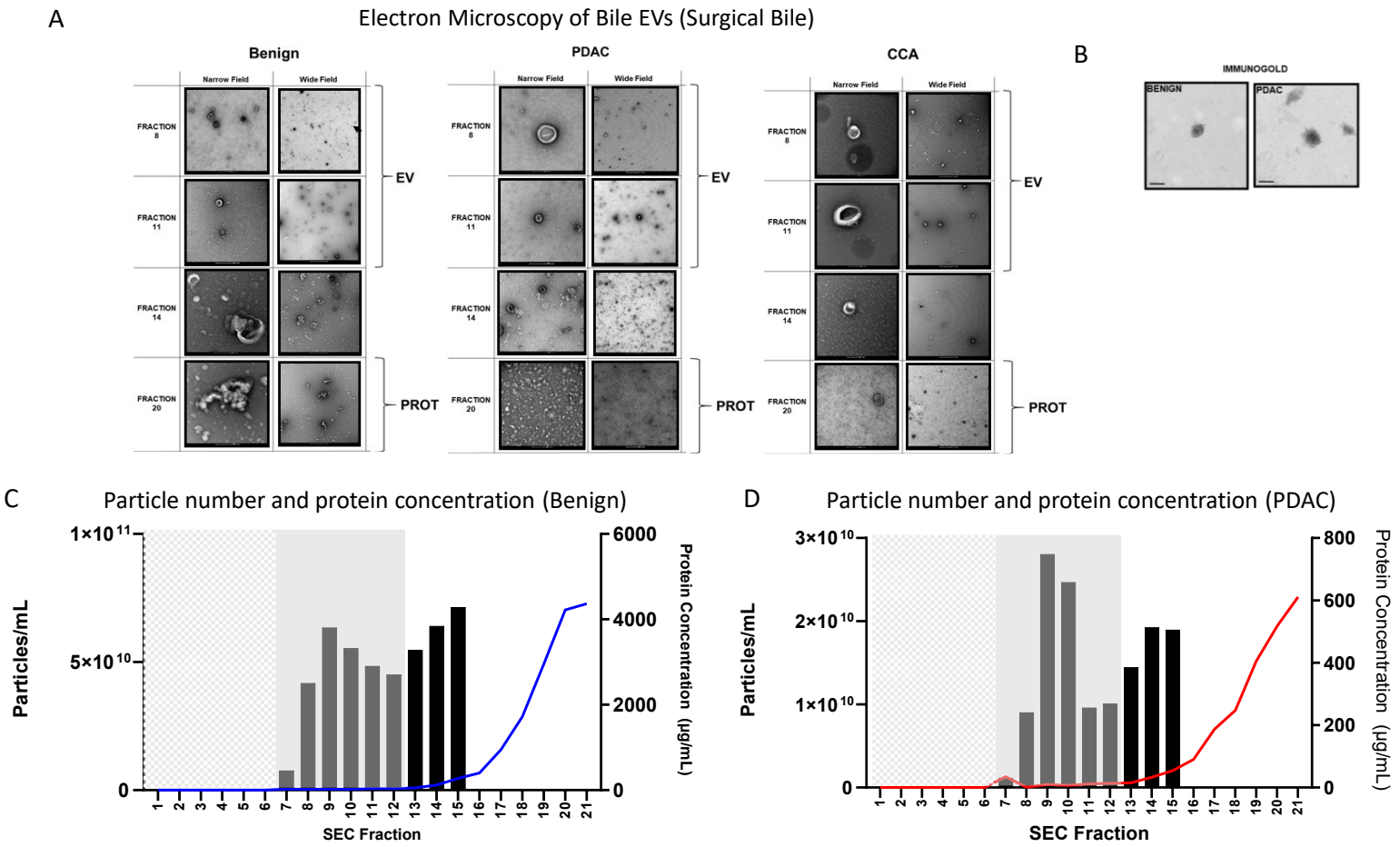


Figure 30: Characterisation of individual fractions show EVs from surgical bile samples (n=6).

A. The shape and structure of bile EVs isolated by SEC are shown here using narrow field (72000x) and wide field (4200x) electron microscopy. 8 µL samples were deposited onto 200-300 mesh glow-discharged copper grids and negatively stained with 2% uranyl acetate. The grids were observed using a Tecnai T12 TWIN transmission electron microscope at magnification ranges 4,000-120,000x. As the fraction number increases so does the protein content shown in the background until EVs are poorly resolved at Fraction 20.

B. Immunogold labelling performed by Dr Padmini Sarathchandra at the Magdi Yacoub Institute Heart Science Centre was done on two samples show gold secondary antibody binding to CD63 present on EVs (72000x). In brief samples were incubated with 4% Paraformaldehyde (PFA) and 5 µL deposited on 200 mesh copper grids. Samples were washed and blocked with 1% BSA before incubation with primary antibody (anti-CD63). Grids were then washed and labelled with secondary goat anti mouse IgG antibody conjugated with 10 nm gold particles before negative staining with 2% uranyl acetate.

C. NTA results showing average particle concentration in each fraction eluted in Benign samples (n=3; plotted as bars) against the protein concentration quantified by BCA and plotted as a line.

D. NTA results showing average particle concentration in each fraction eluted in PDAC samples (n=3; plotted as bars) against the protein concentration quantified by BCA and plotted as a line.

BCA - Bicinchoninic acid assay, PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, EV – Extracellular Vesicle, MISEV – Minimal Information for the Study of Extracellular Vesicles

EVs were estimated using NTA and recorded for each fraction as particles/ mL and protein quantities were measured. Averages (FIGURE 29C+D) for Benign samples (n=3) on the left and malignant samples (n=3) on the right were measured. This showed undetectable/low protein concentrations (<20µg/mL) until fraction 12 when a steep increase can be seen. The minimum concentration of particles measured was 7.8 x10^9 and the maximum was 7 x 10^{10} per mL which gave a total particle yield of 2.3x10^{10}-2.1x10^{12} particles per 500µL of bile. This was consistent with previous studies (Li et al. reported a range of 2.5x10^{10} – 3x10^{11} EVs/mL of bile (254)).

Unexpectedly NTA continued to obtain results at higher fractions which indicated that vesicles continue to elute into the higher fractions (13+). This was confirmed by TEM results which did show vesicular like structures present. Reproducibility studies were also undertaken to assess the NTA results in two samples. SEC was undertaken twice and measured separately (FIGURE 31). This showed good reproducibility at fractions up to 12 in both concentration and particle size. However the variation increased from 13 onwards.

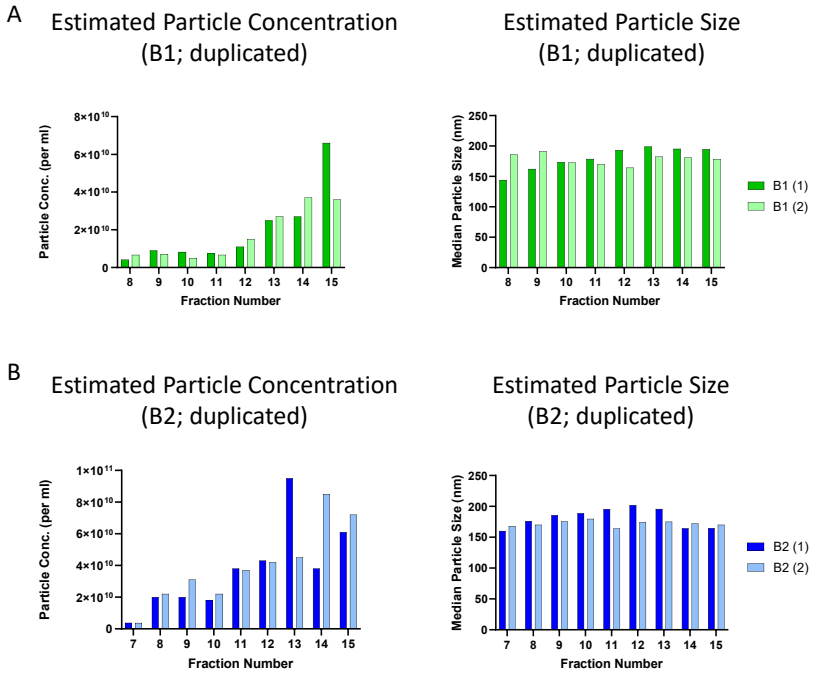


Figure 31: NTA of individual fractions show SEC reproducibility in surgical bile samples (n=2).

A. NTA results in sample B1 (Benign) showing fractions eluted with particle concentration (per ml) on the left and median particle size (nm) on the right.
B. NTA results in sample B2 (Benign) showing fractions eluted with particle concentration (per ml) on the left and median particle size (nm) on the right.

NTA – Nanoparticle tracking Analysis

It was determined that this was due to protein aggregates that are visible on TEM (FIGURE 31A) and decrease the sensitivity of particles measured by NTA. Thus, NTA can only support the presence of EVs but cannot be used solely as a surrogate for EV concentration, particularly when high protein concentrations are also present in the sample i.e. below fraction 13.

4.2.5 Surgical Bile EVs obtained by SEC analysed by Western Blotting.

To confirm the presence of EVs in our samples we analysed our bile EVs for the presence of EV proteins reported in the literature. The previously reported bile EV study by Severino *et al.* (223) compared bile pellets to unextracted bile and found the presence of TSG101, CD9, ALIX and HSP70.1. These showed high expression in malignant samples compared with benign. Initially, we performed UC to isolate EVs and FIGURE 32A showed 3 samples (one excluded due to low protein quantity) obtained. Final 120k pellets on the left (presumed to contain EVs) underwent SDS-PAGE (under reducing conditions except for CD63) and were analysed for expression of EV markers. This was compared with 16k pellets obtained from the intermediate spin and likely to contain subcellular structures such as mitochondria and lysosomes. Expression of EV markers were found in both pellets, but also widespread expression of GAPDH and Calnexin, which was surprising.

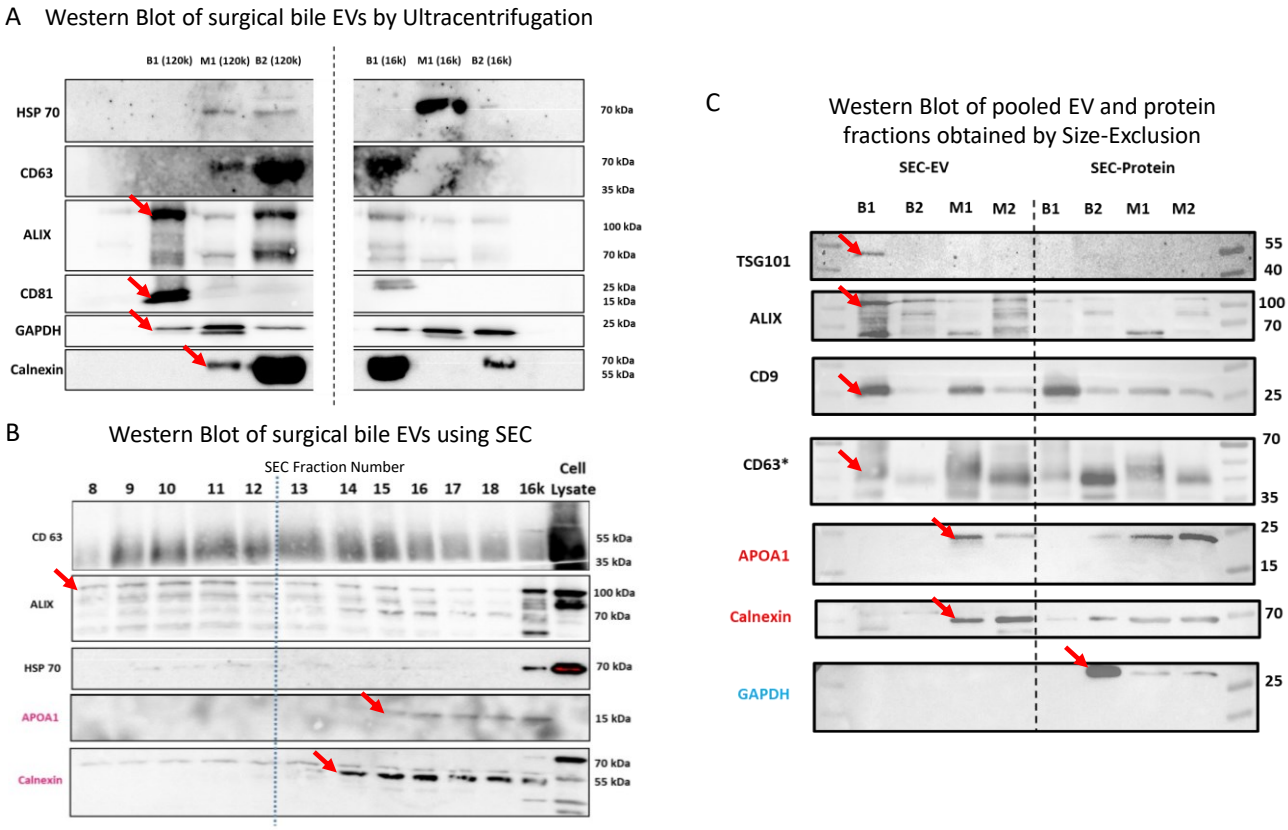


Figure 32: Characterisation by western blot of EVs in surgical bile samples (n=6).

A. Initial western blotting using UC to isolate EVs looking for expression of HSP70, TSG101, ALIX, CD63, CD9 and negative markers APOA1, Calnexin and GAPDH. 20 µg of protein was loaded into each well and compared between intermediate (16k) and high speed (120k) pellets. Dashed line represents unfilled well separating 120k from 16k samples. Red arrows show relevant protein bands.

B. SEC fractions (7 µg loaded) from 8 to 18 using a benign surgical bile sample, as well as the cell pellet from the 16,000 x g spin and SUIT-2 cell lysate as a positive control. A blue dotted line marks the fractions from which the expression of Calnexin appears. Red arrows show relevant protein bands with APOA1 and Calnexin highlighted as proteins not expected in EVs.

C. Pooled EV fractions (7-12) and Protein fractions (22-24) isolated from surgical bile samples are compared and show differential expression. For pooled EV samples 10 µg of protein was loaded into each well. Red arrows show relevant protein bands with APOA1 and Calnexin highlighted as proteins not expected in EVs.

APOA1 – Apolipoprotein A1, CD – Cluster of Differentiation, EV – Extracellular Vesicle, HSP70 – Heat Shock Protein 70, SEC – Size Exclusion Chromatography, TSG101 – Tumour susceptibility gene 101, GAPDH – Glyceraldehyde 3-phosphate dehydrogenase.

In comparison, I then performed SEC of our surgical bile samples and Western blotting of consecutive fractions. Examining one sample from fraction 8 to 18 (fraction 7 excluded due to low protein loading) shows expression of EV related protein (FIGURE 31B) and was compared with the cell pellet from the 16,000 x g spin and SUIT-2 cell lysate as a positive control (cells derived from a metastatic liver tumor of human PDAC (316)). Unexpectedly, CD63 and ALIX showed expression into fraction 18 which together with the NTA and TEM data supported the presence of vesicles not just in the early fractions but also much later. Another alternative explanation could be that these were lysed EV proteins present in solution. However, the presence of Calnexin from fraction 13 onwards indicated that these would definitely be unsuitable for vesicle isolation. It was therefore

deduced that SEC might not able to completely separate EVs from extracellular protein but was able to isolate EVs in the early fractions in a pure form that is not associated with extracellular protein present in the samples. The alternative explanation could be that the EVs are undergoing lysis and releasing free ALIX and CD63 into the later SEC fractions. In either case, I decided that there would undoubtedly be a proportion of EVs potentially lost in the EV extraction process using this method and could not be fully evaluated (no process yet has 100% efficacy of EV isolation from biofluids). As Western blotting appeared to demonstrate that UC generated pellets with more cellular contamination than SEC, I felt reasonable comfortable that SEC preserved EVs due to the low speeds used and TEM evidence of well-defined EVs. Although it could be argued that this isolated a particular sub-population of EVs I decided to move on and take this into account when evaluating the differential expression of small non coding RNAs present.

Four EV samples (B1, B2 – Benign samples; M1, M2 – PDAC samples) obtained from the pooled lysis of SEC fractions 7-12 (labelled SEC-EV) were analysed for EV protein expression alongside negative markers and compared with the later fractions 22-24 (labelled SEC-Protein) containing a mixture of EV and non-EV associated proteins. Here we note a singular sample (B1) expressing TSG101 in the EVs and relative over-expression of ALIX and CD63. Calnexin and APOA1 (present in non-EV cellular material) were found in later fractions but also found in EV samples M1 and M2. GAPDH is notably absent in our EV samples but present in later fractions.

In summary, the results were taken together to show that SEC could be used to obtain EVs from bile better than UC. EVs were found present in all fractions from 7 onwards as shown by the presence of vesicular shaped structures on TEM and EV-related protein by western blot up to fraction 18. In order to obtain EV samples relatively depleted of extracellular protein which may contain RNA not associated with EVs (for example bound to Argonaute (AGO) proteins), the cut-off indicated by both TEM and western blot was at fraction 12-13 with an upper threshold at 13-14 where TEM began to identify protein aggregates. When fractions 7-12 were pooled and analysed, these showed overexpression of EV-related proteins such as TSG 101 and ALIX compared with a higher fraction from the same samples, which confirmed that EVs proteins are over-expressed in our EV fractions and therefore any RNA extracted would be EV-associated. Furthermore, the NTA results from these fractions (7-12) appeared to give particle concentrations that were

reproducible and suitable for concentration comparisons and analysis. We therefore termed these pooled fractions going forward simply as EVs. In a similar fashion, later fractions (20+) whilst not entirely devoid of EVs were termed "Protein" fractions.

4.2.6 Surgical Bile EVs obtained by SEC show no difference in particle size or EV concentration.

Based on the Severino *et al.* paper it was hypothesised that the concentration of EVs found in the bile of those with cancer would be higher than those found in Benign disease. The median EV concentration from all samples was 2.7×10^{10} particles/ml (Range: 4.7×10^{7} – 10.0×10^{11}) which gave an estimated average yield of 4.05×10^{10} EVs per mL of bile.

EV concentrations were compared for 46 samples divided into three cohorts, PDAC (n=18), CCA (n=6) and Benign (n=22). EV concentrations ranged from 6.0×10^{9} to 1.00×10^{11} Particles/mL. Median particle sizes ranged from 134.5nm to 207.2nm. FIGURE 33A shows the three cohorts. An ANOVA was run to determine statistical significance and the smallest p-value obtained was 0.0843 showing a trend towards significance in the pairwise comparison PDAC vs Benign. The pairwise comparison Malignant (PDAC + CCA) vs Benign did not however show any significance (p=0.3323).

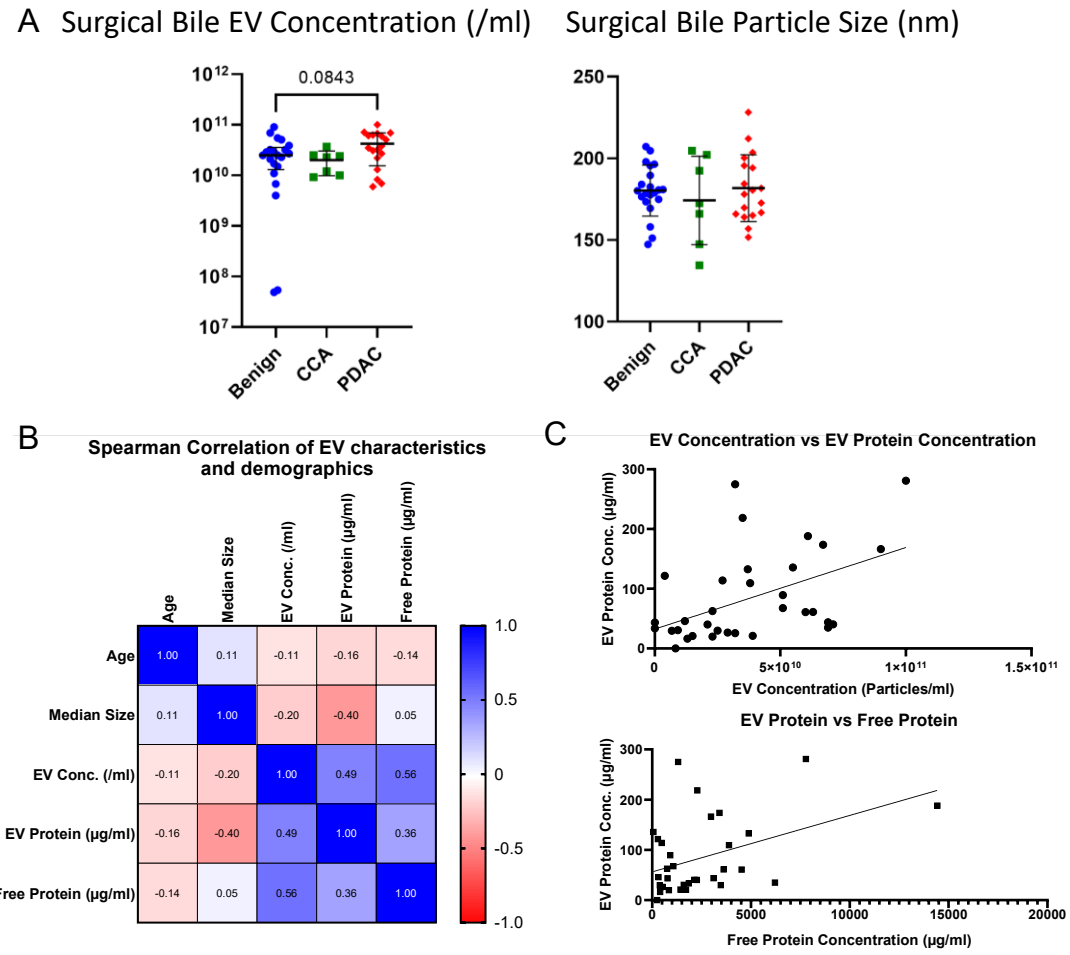


Figure 33: Analysis of particle and protein concentration in surgical bile EVs obtained by SEC (n=46).

A. EV Concentrations and median particle sizes were analysed and showed no differences in surgical bile.
B. Rank correlation was calculated for the continuous variables identified by NTA (Median Particle Size and EV Concentration) and protein quantification of EV and PROT fractions using a BCA Assay. Spearman's rank correlation coefficient was calculated and a heatmap with positive correlation in blue and negative correlation in red.
C. Graphs plotted of EV protein against EV concentration (log scale) show a positive correlation. Best fit line had a slope of 1.36 x 10-9 and a y-intercept of 32.47µg. EV Protein against Free Protein also showed a weak positive correlation (R – 0.36) which indicated a small degree of protein contamination in our EV samples.

NTA – Nanoparticle Tracking Analysis. BCA – Bicinchoninic Acid

Spearman rank correlation was computed for the other quantifiable variables obtained from the SEC-EV fractions which included variables obtained by NTA (Median Particle Size and EV Concentration) and protein quantification from both the EV and protein (PROT) fractions using a BCA Assay (after lysis with RIPA). Measured protein from the EV fraction was called EV Protein, whilst measured protein from the PROT fraction was termed "Free protein." These showed that median particle size was negatively correlated with measured protein concentrations in EV fractions (r_s=-0.40, p=0.019) whilst EV concentration was positively correlated with EV protein measured (r_s=0.49, p=0.004).

These are shown in a heatmap (FIGURE 33B) and can also be plotted as a scatter dot diagram (FIGURE 33C). Additionally measured protein in the EV and PROT fractions were also positively correlated but to a lesser degree (r=0.36, p= 0.04).

In summary, it appeared that EV protein and the concentration of EVs were correlated, but not to the degree that protein concentrations could be a surrogate for NTA. The confounding factor appears to be the presence of "Free Protein" or extracellular cell-free protein in samples which also is somewhat correlated with both EV concentration and protein likely due to co-isolation by SEC.

4.2.7 Surgical Bile EVs undergo RNA extraction and can be sequenced for differential expression.

In total 18 samples (6 samples were chosen from each disease cohort) were chosen for analysis based on the EV-protein purity ratio described by Webber et al (317) which was used as a surrogate for EV purity in order to minimise non-EV related RNA. Clinical data for the samples are shown in TABLE 21. Small RNA-sequencing was In this case I prepared 150ng of RNA and this was sent to the facility on dry ice where they performed library preparation These were sent for next-generation small RNA sequencing performed by Dr Morten Venø and the team at Omiics (Aarhus, Denmark) (NGS; single-end 75 base pair, see Methods 2.2.13 for more details) to discover the presence of small RNAs within our samples. As previous, this was sent to the facility on dry ice where they performed library preparation. Library preparation from the RNA samples was undertaken using the QIAseq miRNA Library Kit and sequenced on an Illumina NextSeq500. Single-end 75 base pair small RNA sequencing was performed.

Table 22: Patients selected for characterization and sequencing in the surgical bile EV microRNA biomarker analysis.

Overall demographic characteristics and recorded pathology results were collected retrospectively and analysed by Fisher's exact test for each cohort used for surgical bile EV microRNA biomarker analysis.
PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, SD – Standard Deviation

Variables	Subcategory	PDAC	CCA	Benign*
All cases		**18**	**7**	**21**
Age, *y*	<65	15	4	20
	>65	3	3	1
	Mean (SD)	58 (8.1)	67 (9.1)	49 (12.6)
	Median	55	62	47
Sex	Male	11	6	10
	Female	7	1	11
Tumour stage	I	0	0	-
	II	1	0	-
	III	9	6	-
	IV	2	0	-
	Unknown	6	1	-
Sequencing Cohort		**6**	**6**	**6**
Age, *y*	<65	4	3	3
	>65	2	3	3
	Mean (SD)	61 (11.3)	68 (8.8)	61 (14.7)
	Median	57	67	63
Sex	Male	3	5	3
	Female	3	1	3
Tumour stage	I	0	0	-
	II	1	0	-
	III	4	6	-
	IV	1	0	-
	Unknown	0	0	-

After quality filtering and trimming, the average filtered reads after trimming from the RNA sequencing libraries were 9,824,699 reads. Among these, 100% were mapped to known RNA species and the mean as well as individual read distributions are shown in FIGURE 34. Reads were mapped into the following small non-coding RNAs: transfer RNAs (tRNAs) followed by miRNAs (allowing addition of A and T nucleotides at the 3' end). Reads not mapped to tRNAs or miRNAs were then mapped to other relevant transcriptomes (snRNA, snoRNA, yRNA, vaultRNA and RNAs from the Rfam non-coding RNA database). Finally, the remainder were mapped to Human mRNAs followed by the human (hg19) genome.

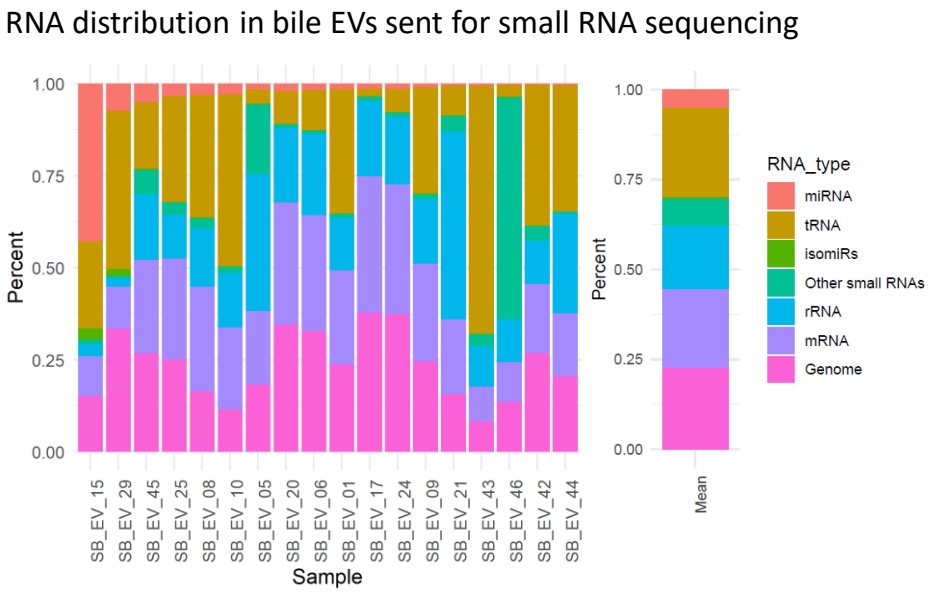


Figure 34: Relative distribution of small RNA reads for each bile sample from the discovery cohort (n=18).

Surgical bile EV samples were sent for single end 75 base pair small RNA sequencing by Dr Morten Venø and the team at Omiics (Aarhus, Denmark). Shown are the mapped RNA reads and relative proportions in each sample making up each RNA subtype (colour legend to the right). Stacked bar plots are shown ordered by sample number with results in percentages. Mean proportions across all samples are represented in the solitary bar plot to the right.

TABLE 23 which shows the top expressed miRNAs present revealed that there was a predominance of miRNAs related to the let-7 family (namely let-7b, let-7a, let-7f, let-7i, let-7c) and miR-200 family (namely miR-141, miR-200a, miR-200c, miR-200b) as well as miR-21 and miR-16. Other miRNAs shown are also genomically associated with these families e.g. miR-125b clustered with let-7a.

Table 23: Top twenty microRNAs present in bile EVs (n=18; Discovery cohort 6 PDAC, 6 CCA, 6 Benign).

EVs from 18 bile samples underwent RNA extraction and miRNAs are shown sorted by descending overall expression expressed as the average normalised reads.
RPM – Reads per Million reads.

Top miRNAs by expression	Mean RPM
miR-122-5p	2105.06808
let-7b-5p	1614.77947
let-7a-5p	996.847207
miR-21-5p	893.466237
let-7f-5p	682.219234
miR-192-5p	500.556672
miR-141-3p	404.421139
miR-16-5p	401.492863
miR-26a-5p	400.088416
let-7i-5p	281.236501
miR-200a-3p	274.055807
miR-125a-5p	272.036701
miR-200c-3p	260.373573
miR-223-3p	198.036668
miR-125b-5p	178.365169
miR-194-5p	157.862644
miR-26b-5p	150.647281
let-7c-5p	143.114158
miR-30d-5p	113.703257
miR-320a-3p	108.008406

4.2.8 Differential expression analysis undertaken for EV-miRNA identified from surgical bile samples.

Analysis of the mapped reads revealed a high variability in RNAs with tRNAs being the most common, accounting for 25.0% of all mappable RNAs, followed by 22.5% genomic RNA, 21.8% mRNAs, 17.8% rRNAs, and 5.1% miRNAs (FIGURE 34). Global differential expression profiling for miRNAs was undertaken between cohorts (PDAC, CCA and benign, as well as Malignant (PDAC + CCA) vs benign) with an adjusted $p<0.05$ and log2fold change > 1 as a cut-off for determining significance.

Filtering of miRNAs to exclude those with no expression in more than 50% of the samples was undertaken to reduce noise and compared with unfiltered results which showed fewer significant candidates (results not shown). Overall, we found 2, 6, and 8 deregulated miRNAs between PDAC vs Benign, CCA vs Benign, and Malignant vs Benign, respectively (FIGURE 35A).

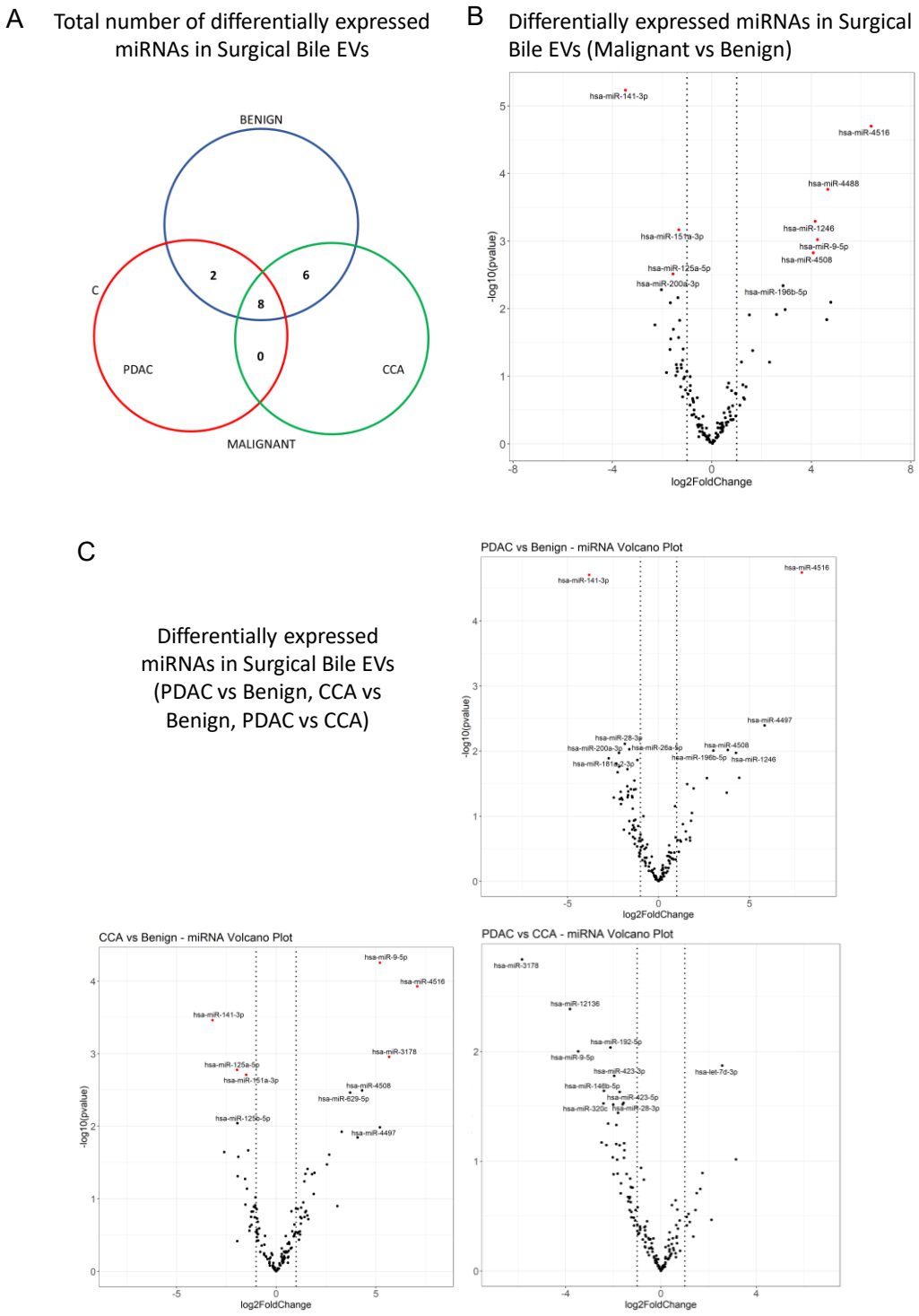


Figure 35: Differential expression using filtered RNA sequencing data for surgical bile EVs (n=18).

A. Venn diagram showing the total number of differentially expressed miRNAs between PDAC, CCA and benign cohorts with P value (adjusted) <0.05 in bile.
B. Volcano plot with top 10 miRNAs labelled. Candidates shown were found to be significant in the following pairwise comparisons (Malignant vs Benign) with an FDR < 0.05 (highlighted in red).
C. Pairwise comparisons PDAC vs Benign; CCA vs Benign; PDAC vs CCA shown for reference.

Plotted are significance (-log10 p-value) vs expression (log2 Fold Change). Vertical lines indicate log2 Fold Change ±1. Only the 10 most significant miRNAs are labelled for clarity.

PDAC – pancreatic ductal adenocarcinoma, CCA – cholangiocarcinoma.

Table 24: Significant miRNAs (p<0.05) common from each pairwise comparison alongside adjusted p-value and log2fold change.

Differentially expressed miRNA candidates from the small RNA sequencing data were identified as statistically significant after adjustment for FDR and included below with normalised expression and log fold change. Candidates are ordered in descending adjusted p-values. Highlighted in bold are miRNAs common to all comparisons and which remained significant when filtering was removed. For the pairwise comparison, PDAC vs CCA, no significant hits were identified. Values rounded to 5 sig. fig.
PDAC – pancreatic ductal adenocarcinoma, CCA – cholangiocarcinoma.

Comparison	MicroRNA	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
CCA vs Benign	miR-9-5p	2.7665	5.2047	1.2912	0.0083
	miR-4516	342.88	7.0748	1.8376	0.0089
	miR-141-3p	145.05	-3.1909	0.8916	0.0173
	miR-3178	10.786	5.6639	1.7364	0.0415
	miR-125a-5p	50.336	-1.9534	0.6216	0.0489
	miR-151a-3p	23.609	-1.5032	0.4854	0.0489
PDAC vs Benign	**miR-4516**	342.88	7.8788	1.8373	0.0015
	miR-141-3p	145.05	-3.8227	0.8954	0.0015
PDAC vs CCA	-	-	-	-	-
Malignant vs Benign	**miR-141-3p**	145.05	-3.4740	0.7667	0.0008
	miR-4516	157.24	6.4087	1.5029	0.0013
	miR-4488	2.7271	4.6691	1.2428	0.0074
	miR-1246	14.212	4.1580	1.1966	0.0165
	miR-151a-3p	23.609	-1.3296	0.3914	0.0176
	miR-9-5p	2.7665	4.2518	1.2868	0.0205
	miR-4508	4.6514	4.0782	1.2839	0.0275
	miR-125a-5p	50.336	-1.5653	0.5288	0.0496

TABLE 24 shows that 2 candidate bile EV miRNAs for diagnosis featured in all three analyses. MiR-141-3p was significantly downregulated (log2foldchange -3.4740; p=0.0008) in the malignant vs benign comparison and miR-4516 significantly upregulated (log2foldchange 6.4087; p=0.0013). These 2 miRNAs were also consistently differentially expressed in the PDAC vs Benign and CCA vs Benign comparisons indicating they may be a marker of hepatobiliary adenocarcinoma. They remained statistically significant when filtering was removed. No miRNAs appeared to be significantly differentially expressed when the PDAC cohort was compared with the CCA cohort. To assess diagnostic value of these two candidate's receiver operating characteristic (ROC) curve analysis was performed for each individually (see FIGURE 36) and the AUC for miR-4516 was 0.625 and AUC for miR-141-3p was 0.792. The combined AUC from both candidates from the sequencing data was calculated as 0.736.

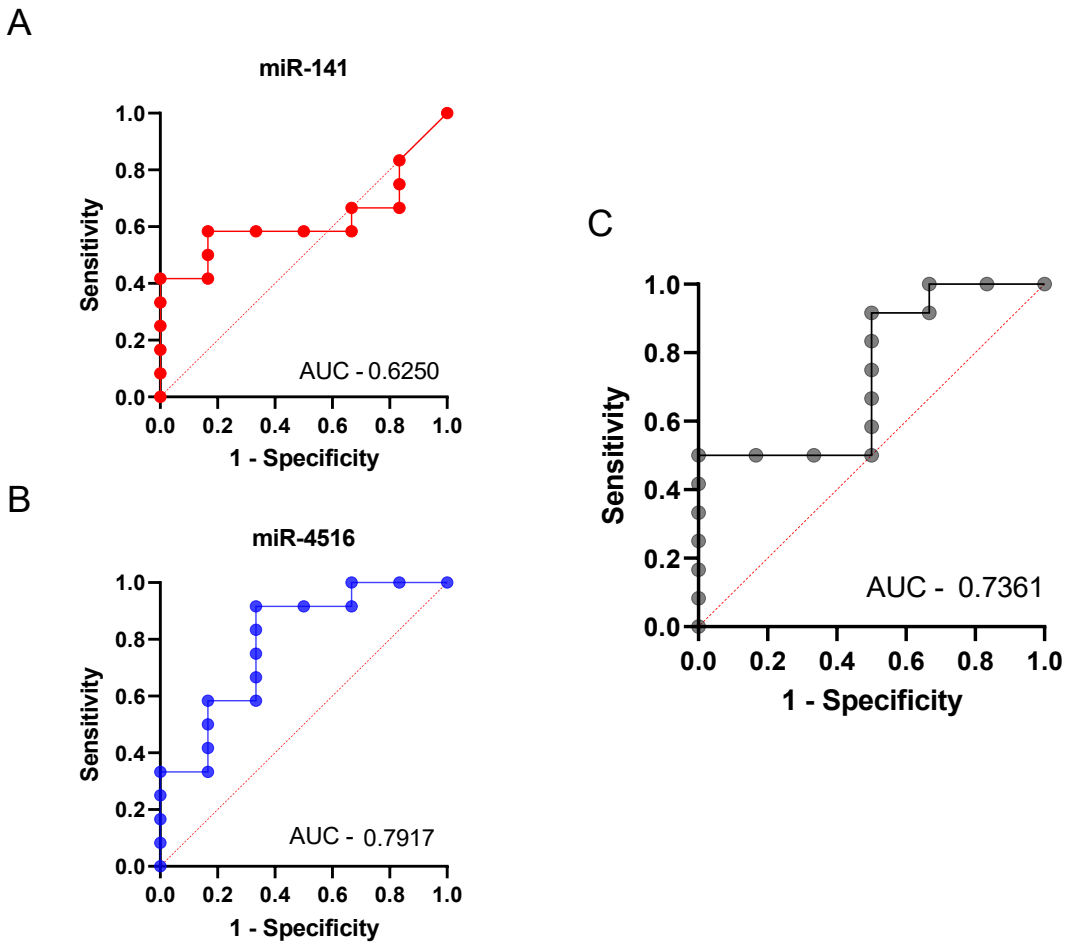


Figure 36: RNA sequencing identifies a two miRNA signature in surgical bile EVs (n=46).
From small RNA sequencing data it is possible to estimate the diagnostic potential of selected miRNA candidates which were shown to be over expressed in cancer (Malignant vs Benign). Here are:

A. & B. Individual receiver operating characteristic (ROC) curve for miR-4516 and miR-141-3p respectively. AUC for miR-4516 (upregulated) was 0.625 and AUC for miR-141-3p (downregulated) was 0.792.
C. Combined ROC curve for the two miRNAs using the pairwise comparison Malignant vs Benign.

4.2.9 Differential expression analysis undertaken for EV-tRNAs identified from surgical bile samples.

From the RNA sequencing data, tRNA fragments (tRFs) were again identified as highly expressed in our bile samples with a mean 25% composition in bile EVs which led us to hypothesise that these might be a source of differentially expressed candidates.

Analysis in a similar fashion to that demonstrated for our bile cfRNA analysis was undertaken by Dr Venø and generated a larger number of potential tRF candidates. After correction using FDR, we found 92 differentially expressed tRFs between Malignant vs Benign with 69 candidates in PDAC vs Benign and 163 candidates between CCA vs Benign (FIGURE 37A). FIGURE 37B shows the number of candidates upregulated in cancer samples.

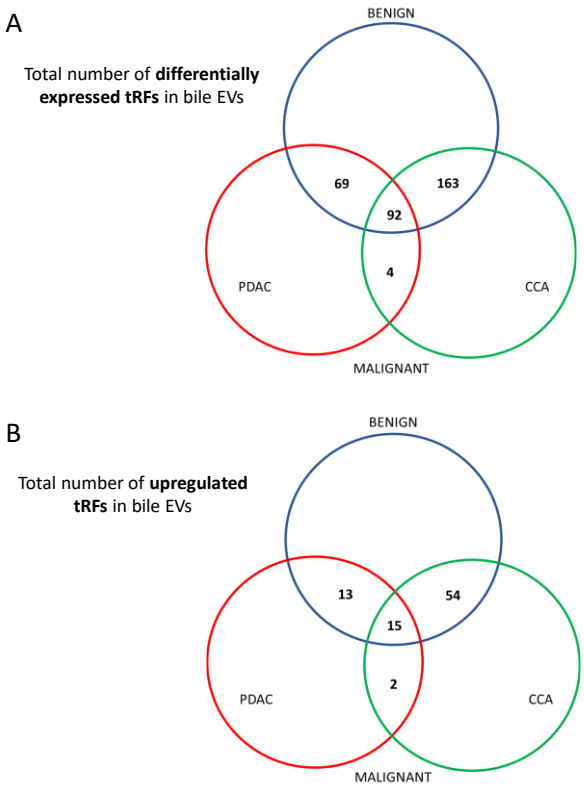
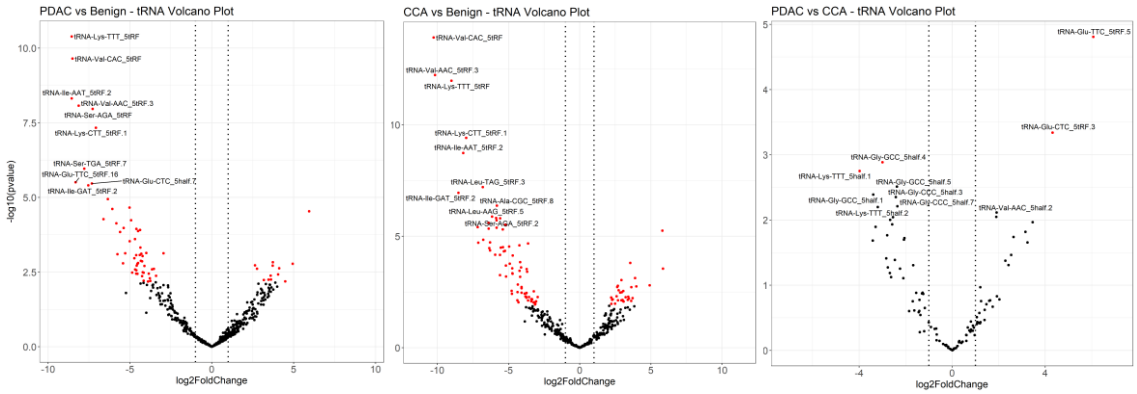


Figure 37: Venn diagrams showing the total number of differentially expressed tRFs in surgical bile EVs (n=18).

A. Total number of differentially expressed tRFs between PDAC, CCA and benign cohorts.
B. Total number of upregulated tRFs only (P value (adjusted) <0.05).

PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, tRFs – Transfer RNA fragments.

A Differentially expressed tRFs in Surgical Bile EVs (PDAC vs Benign; CCA vs Benign; PDAC vs CCA)



B Differentially expressed tRFs in Surgical Bile EVs (Malignant (PDAC+CCA) vs Benign)

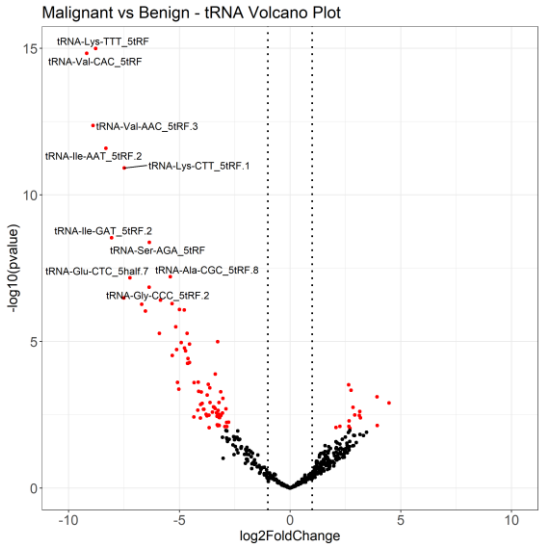


Figure 38: Volcano plots showing significant tRFs identified by small RNA sequencing as present in surgical bile EV samples (n=18).

From small RNA sequencing data analysis of the discovery cohort each diagnostic potential of selected trF candidates were investigated. Plotted are significance (-log10 p-value) vs expression (log2 Fold Change). Vertical lines indicate log2 Fold Change ±1. Only the 10 most significant tRFs are labelled for clarity. Here are:

A. Volcano plots with top 10 tRFs found to be significant in the following pairwise comparisons (PDAC vs Benign; CCA vs Benign; PDAC vs CCA) with an FDR < 0.05 (highlighted in red).
B. Volcano plot with top 10 tRFs (upregulated and downregulated) found to be significant in the pairwise comparison Malignant (PDAC & CCA) vs Benign.

PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, FDR – False Discovery Rate, tRFs – Transfer RNA fragments

Volcano plots for the pairwise comparisons Malignant vs Benign, PDAC vs Benign and CCA vs Benign are shown in FIGURE 38 with top miRNAs labelled. tRFs with the highest significance and differential expression appeared to skew towards downregulation in our cancer samples. This could indicate a relative paucity of candidate biomarkers in cancer bile EVs, so we focussed on only those that appeared to be upregulated in Malignancy.

TABLE 25 showed the top ten tRFs expressed in Malignant samples, PDAC samples and CCA samples respectively. Comparison with the cfRNA analysis (TABLE 14) reveal the recurrence of upregulation of Gly-GCC 5' tRF (tRF-28-P4R8YP9LOND5 present in EVs whilst tRF-30-P4R8YP9LON4V which is identical except for additional 2 GC nucleotides; highlighted in red) and Asp-GTC-itRF (tRF-17-8HM2OS4 present in EVs in PDAC and tRF-22-L7S5QKF14 present in cfRNA in PDAC albeit different in size and start point) among the top ten tRFs identified. Additionally, it can be seen that two tRFs, Lys-TTT_5tRF.20 and Gly-GCC_5tRF.9 are present in both populations indicating a potential for shared markers present in Malignancy (i.e. PDAC and CCA).

Table 25: Top 10 tRFs identified as upregulated in the surgical bile EV analysis (n=18).

A. Top 10 tRF candidates identified as significantly upregulated in the comparison PDAC vs Benign and sorted by descending log fold change values (upregulated in PDAC samples).
B. Top 10 tRF candidates identified as significantly upregulated in the comparison CCA vs Benign and sorted by descending log fold change values (upregulated in CCA samples).
C. Top 10 tRF candidates identified as significantly upregulated in the comparison Malignant vs Benign and sorted by descending log fold change values (upregulated in malignant i.e. PDAC and CCA samples).

Expression averaged across all samples with normalization by DESeq2 to allow direct comparison between the samples. Standard errors are of the log2fold change and p-values are adjusted using the Benjamini-Hochberg method. Candidates in bold are common to both pairwise analyses, whilst those in red are common/similar to tRFs identified in the earlier cfRNA analysis. RPM – Reads per million reads.

A

tRNA Fragment	MINTbase Unique ID	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
Glu-TTC_5tRF.5	tRF-28-86J8WPMN1E0J	105.02	5.9491	1.4229	0.0011
Lys-TTT_5tRF.20	**tRF-29-PS5P4PW3FJF2**	**12.341**	**4.9382**	**1.5710**	**0.0214**
Pro-AGG_5tRF.3	tRF-27-6978WPRLXN3	4.0217	4.4861	1.6472	0.0498
Gly-GCC_5tRF.9	**tRF-28-P4R8YP9LOND5**	**643.30**	**4.1164**	**1.3541**	**0.0271**
His-GTG_5tRF.2	tRF-28-MUWLV47PU90M	1.6983	4.0497	1.3991	0.0344
Gly-GCC_5tRF.10	tRF-28-PNR8YP9LOND5	1293.3	3.7322	1.1742	0.0210
Gly-CCC_5tRF.7	tRF-28-Q1Q89P9L84DF	322.08	3.7236	1.2004	0.0230
Asp-GTC_itRF.40	**tRF-17-8HM2OS4**	**3.9106**	**3.6134**	**1.3056**	**0.0468**
Glu-CTC_5tRF.4	tRF-29-87R8WP9N1EJ4	1519.1	3.5944	1.2550	0.0372
Pro-TGG_itRF.39	tRF-16-8U53K80	1.9749	3.4663	1.2539	0.0468

B

tRNA Fragment	MINTbase Unique ID	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
Pro-TGG_itRF.6	tRF-17-R3IEWSN	10.963	5.8546	1.6126	0.00296
Lys-TTT_5half.1	tRF-31-PS5P4PW3FJHPB	255.63	5.8075	1.2792	0.00012
Pro-TGG_itRF.7	tRF-16-RLXN48E	5.4257	4.9089	1.5538	0.01286
Lys-TTT_5half	tRF-30-PS5P4PW3FJHP	25.754	3.9780	1.2725	0.01388
Ala-TGC_5half.1	tRF-32-R29P4P9L5HJVQ	8.8179	3.8921	1.1551	0.00648
Lys-TTT_5tRF.20	**tRF-29-PS5P4PW3FJF2**	**12.341**	**3.8404**	**1.5626**	**0.05631**
Gly-GCC_5tRF.9	**tRF-28-P4R8YP9LOND5**	**643.30**	**3.7196**	**1.3537**	**0.03214**
Lys-TTT_5half.2	tRF-32-PS5P4PW3FJHP1	902.85	3.6308	1.1695	0.01418
Pro-TGG_5half.2	tRF-31-6978WPRLXN48E	44.211	3.6195	1.3645	0.03652
His-GTG_5half.3	tRF-32-PW5SVP9N15WVN	31.543	3.5497	0.9392	0.00187

C

tRNA Fragment	MINTbase Unique ID	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
Lys-TTT_5tRF.20	tRF-29-PS5P4PW3FJF2	12.341	4.4671	1.3846	0.0013
Pro-AGG_5tRF.3	tRF-27-6978WPRLXN3	4.0217	3.9351	1.4683	0.0074
Gly-GCC_5tRF.9	tRF-27-PNR8YP9LON3	643.30	3.9314	1.1681	0.0008
Ala-TGC_5half.1	tRF-32-R29P4P9L5HJVQ	8.8179	3.1747	1.1028	0.0040
Gly-GCC_5tRF.10	**tRF-28-PNR8YP9LOND5**	**1293.3**	**3.1439**	**1.0363**	**0.0024**
Gly-CCC_5tRF.7	tRF-28-Q1Q89P9L84DF	322.08	3.1250	1.0631	0.0033
Gly-GCC_5tRF.11	tRF-29-P4R8YP9LONHK	105.43	2.9197	0.9905	0.0032
His-GTG_5half.3	tRF-32-PW5SVP9N15WVN	31.543	2.8418	0.9078	0.0017
Gly-CCC_5tRF.10	tRF-29-Q1Q89P9L84FS	238.24	2.7486	0.7846	0.0005
Lys-TTT_5tRF.20	tRF-29-PS5P4PW3FJF2	58.907	2.7018	1.0352	0.0091

These two candidates can be seen as representing a high number of the tRF expression as mapped according to tRNA isotype (FIGURE 39). 5' tRFs can be seen as predominating in a very similar expression pattern to the whole bile cfRNA picture from the previous chapter.

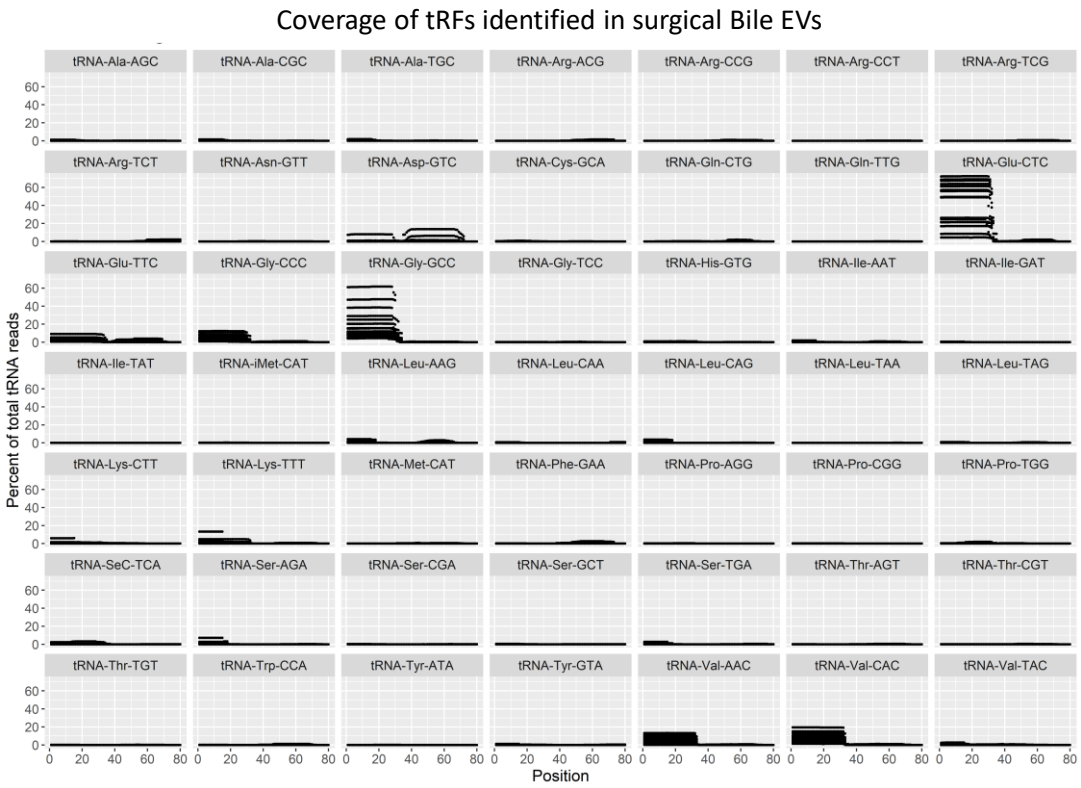


Figure 39: Surgical Bile EV tRF expression expressed according to tRNA isotype (n=18).
A coverage plot of the tRF candidates is shown according to tRNA isotype (i.e. the tRNA from which the fragment is likely to have arisen). This also shows Gly-GCC and Glu-CTC as dominant tRNA isotypes in bile EV samples as well as cfRNA. Normalised read distribution of tRFs in bile according to tRNA isotypes (tRNA of origin). Expression as a percentage of total tRNA reads with relative nucleotide position within the tRNA expressed along the y-axis.

4.2.10 Characterisation of EVs present in endoscopic bile samples obtained prospectively (MIRABILE).

The promising findings in our surgical samples allowed me to take forward the protocol for isolating EVs using SEC in bile samples forward into preparation of EVs from bile obtained from the MIRABILE study (see 3.2.1 for sample collection details). Patient data was available to determine three samples which were from the benign, PDAC and CCA populations and characterised by TEM, NTA and western blotting.

TEM allowed visualisation of individual vesicles and FIGURE 40A shows the presence of ~100nm bilayered structures with negative staining using uranyl acetate. Once again "cup-shaped" particles were seen and can be compared with FIGURE 30. Qualitatively, it was noticed that there were fewer other extracellular protein structures seen in the wide field views (4000x magnification) of MIRABILE samples in comparison to the surgical samples. It was also possible to obtain clearer vesicular structures with a more circular morphology, particularly in the early fractions.

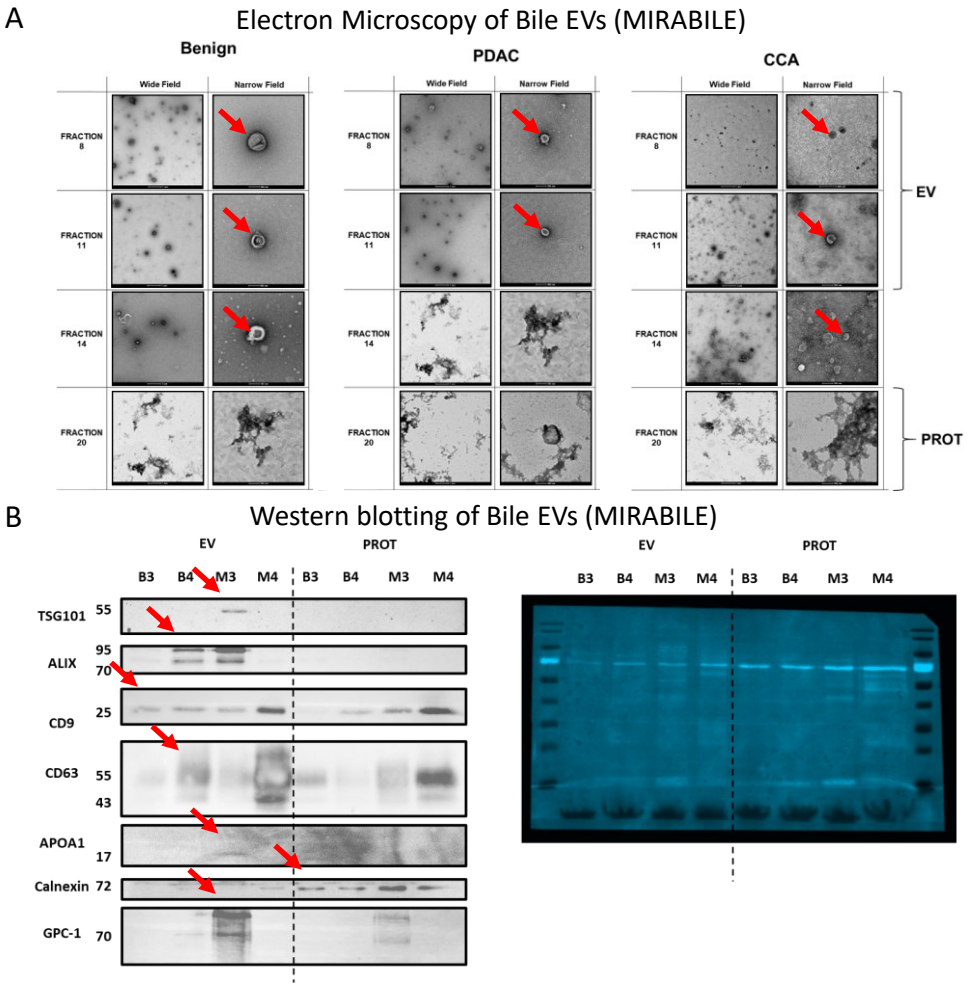


Figure 40: Characterisation of individual EV fractions using TEM and Western blotting of the MIRABILE bile samples (n=6).

A. The shape and structure of bile EVs isolated by SEC and shown here using narrow field (72000x) and wide field (4200x) electron microscopy. 8 µL samples were deposited onto 200-300 mesh glow-discharged copper grids and negatively stained with 2% uranyl acetate. The grids were observed using a Tecnai T12 TWIN transmission electron microscope at magnification ranges 4,000-120,000x. As in previous studies, with increasing fraction number so does the protein content shown in the background until EVs are not apparent in Fraction 20. Red arrows to show EVs with lipid bilayer.

B. Western blot of EV (fractions 7-12) compared to Protein (22-24) fractions looking for expression of TSG101, ALIX, CD9 CD63, and negative markers APOA1, Calnexin. GPC-1 was also assessed. Red arrows show relevant protein bands. Also shown are total membrane protein levels for normalization purposes (No-Stain Protein Labelling Reagent, Invitrogen).

APOA1 – Apolipoprotein A1, CCA – Cholangiocarcinoma, CD – Cluster of Differentiation, EV – Extracellular Vesicle, PDAC – Pancreatic ductal adenocarcinoma, SEC – Size Exclusion Chromatography, TSG101 – Tumour susceptibility gene 101, GPC-1 – Glypican-1.

Western blotting of four EV samples (2 benign, 2 PDAC) revealed differential expression of ALIX and TSG101 compared with PROT (fractions 22-24). Total protein was used as a normaliser (shown in blue) with 10 µg of protein loaded in each well. Glypican-1 was also identified in one of our PDAC samples (M3) which had not previously been shown as expressed in bile EVs.

NTA was undertaken for each fraction initially alongside quantification of protein concentrations for one patient sample with the diagnosis of benign, PDAC and CCA. This showed a clear peak particle measurement between fractions 7-12 (FIGURE 40A). Fractions 7-12 were therefore pooled as EV samples and these three samples (1 PDAC, 1 CCA and 1 Benign) were analysed by NTA generating a size distribution graph shown in FIGURE 41B.

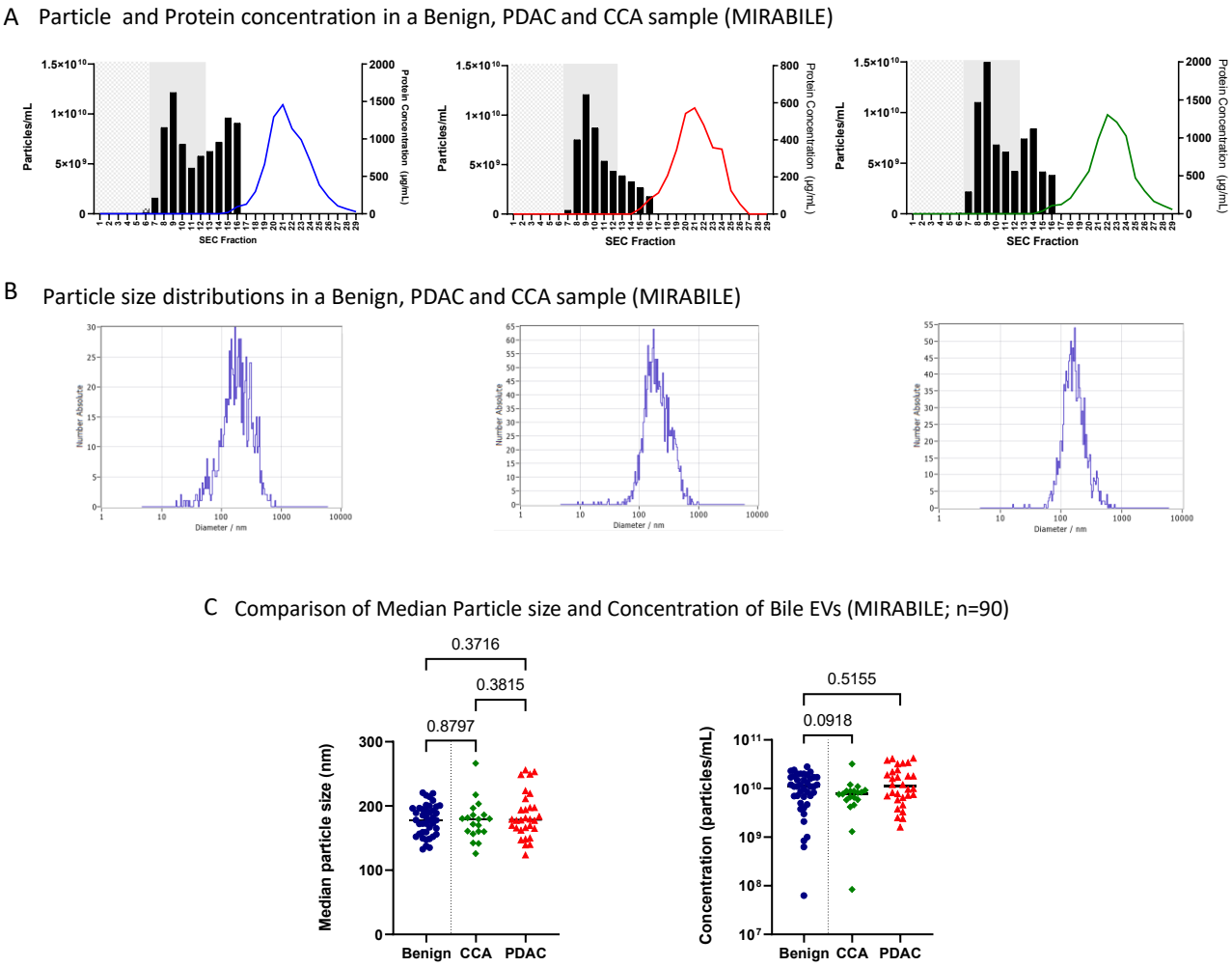


Figure 41: NTA and protein concentration analysis in MIRABILE bile samples (n=9; n=90).

A. Average particle and protein concentrations measured for each fraction in each cohort (3 benign, 3 PDAC, 3 CCA from left to right). Shaded region shows fractions 7 to 12 which were pooled for subsequent RNA extraction and analysis.
B. Individual size distribution plots from the same pooled EV samples analysed during characterisation show EV particle size (concentration measured at 1:100).
C. Median particle size and concentrations measured show each cohort and statistical significance determined by Kruskal-Wallis (non-parametric concentrations) in all samples (n=90). P-values show significance using uncorrected Dunn's test.

Once the protocol for analysis of EVs was finalised, all samples (n=90) were retrospectively analysed for EV median particle size and concentration. This showed

benign patients had an average particle size of 178.4nm (Interquartile range/IQR – 135.8 to 196.5) and yielded EVs with a concentration of 1.1 x 10^{10} particles/mL (IQR – 5.9 x 10^{9} to 1.7 x 10^{10}). PDAC samples had an average particle size of 184.8nm (IQR – 164.1 to 200.8) and concentration of 1.1 x 10^{10} particles/mL (IQR – 6.3 x 10^{9} to 2.3 x 10^{10}), whilst CCA samples contained an average particle size of 177.2nm (IQR – 160.0 to 186.1) and a slightly lower concentration of EVs with an average of 7.7 x 10^{9} particles/mL (IQR – 5.8 x 10^{9} to 8.9 x 10^{10}; p=0.0918). EV concentrations for the pairwise comparison Malignant (PDAC + CCA) vs Benign was also compared which did not show any difference (p=0.6882). This contradicts the Severino *et al.* study which had found an increase in median concentration of EVs in malignant CBD stenoses.

4.2.11 Clinical characteristics for the patients comprising the MIRABILE EVs analysis.

After EV characterisation was complete, it was decided to separate the bile samples into two cohorts for the purpose of discovery and validation of potential RNA biomarkers. The details of the patient samples selected for each cohort for the analysis of RNA are summarised in TABLE 26. A total of 90 unique patient samples were used for this study with a cohort of 30 taken forward for RNA sequencing, corresponding to 10 from each cohort, with the aim of minimising differences in age and sex. The validation cohort comprised of 81 samples (27 PDAC, 17 CCA, 37 Benign) and underwent RT-qPCR validation.

Table 26: Patients selected for the discovery and validation cohorts used in the bile EV biomarker analysis.

Overall demographic characteristics and recorded pathology results were collected retrospectively and analysed by Fisher's exact test for the discovery and validation cohort as well as overall. PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, PSC – Primary sclerosing cholangitis

Table. Demographic and Pathological Characteristics of the Patient Cohorts comprising the Bile EV study

Patient Characteristics		Total (n=90)				Discovery (n=30)				Validation (n=81)			
		PDAC	CCA	Benign	P value	PDAC	CCA	Benign	P value	PDAC	CCA	Benign	P value
All cases (n, %)		**30**	**19**	**41**		**10**	**10**	**10**		**27**	**17**	**37**	
Age, *y*	<60	6 (20)	4 (21)	19 (46)		1 (10)	2 (20)	2 (20)		5	4	18	
	≥60	24 (80)	15 (79)	22 (54)	**0.0125**	9 (90)	8 (80)	8 (80)	1.0000	22	13	19	**0.0096**
	Mean (SD)	69 (11)	68 (9)	62 (16)	**0.0207**	70 (12)	71 (7.9)	66 (11)	0.2369	69.4 (11)	67.8 (9)	61 (17)	**0.0149**
Gender (%)	Female	8 (27)	8 (42)	19 (46)		2 (20)	3 (30)	3 (30)		6	7	18	
	Male	22 (73)	11 (58)	22 (54)	0.2005	8 (80)	7 (70)	7 (70)	1.0000	21	10	19	0.1085
History of PSC (%)		0 (0)	1 (5)	2 (5)	0.5896	0 (0)	1 (10)	0 (0)	1.0000	0 (0)	1 (6)	2 (5)	0.5898
Chronic Pancreatitis (%)		2 (7)	0 (0)	8 (20)	**0.0390**	0 (0)	0 (0)	0 (0)	1.0000	2 (7)	0 (0)	8 (22)	**0.0378**
Pre-diagnostic blood parameters*	CA 19-9 (U/mL)	10444	6465.0	285.6	0.0671	2627.3	12250.4	14	*0.3422*	11145.84	3427.5	285.6	0.0734
	Bilirubin (ng/mL)	246.6	168.8	49.0	**<0.0001**	167.2	257.1	42.8	***0.0117***	256.1	170.9	49.0	**<0.0001**
	CRP (µg/mL)	55.3	45.1	64.2	0.2907	51.8	35.8	133.3	*0.1341*	49.1	47.8	45.7	0.6595
Anatomical location	Perihilar CCA	-	11 (58)	-		-	4 (40)	-		-	9 (53)	-	
	Distal CCA	-	5 (26)	-		-	3 (30)	-		-	5 (29)	-	
	Intrahepatic CCA	-	3 (16)	-		-	3 (30)	-		-	3 (18)	-	
	Head of Pancreas PDAC	23 (77)	-	-		9 (90)	-	-		21 (78)	-	-	
	Uncinate PDAC	3 (10)	-	-		0 (0)	-	-		3 (11)	-	-	
	Mixed PDAC (HOP/Uncinate)	4 (13)	-	-		1 (10)	-	-		3 (11)	-	-	
Tumour stage (%)	T1	0 (0)	0 (0)	-		0 (0)	0 (0)	-		0 (0)	0 (0)	-	
	T2	4 (13)	2 (11)	-		1 (10)	1 (10)	-		3 (11)	2 (12)	-	
	T3	7 (23)	9 (47)	-		2 (20)	4 (40)	-		7 (26)	8 (47)	-	
	T4	19 (63)	8 (42)	-		7 (70)	5 (50)	-		17 (63)	7 (41)	-	
Nodal disease (%)		18 (60)	8 (42)	-		5 (50)	1 (10)	-		17 (63)	5 (29)	-	
Metastatic disease (%)		10 (33)	2 (11)	-		3 (30)	5 (50)	-		9 (53)	1 (6)	-	

Statistical significance was determined using Fisher's exact test for categorical variables (Sex, Age >60) whilst an unpaired two-tailed Student t-test or Mann Whitney U test using Malignant (PDAC + CCA) vs Benign as the pairwise comparison. P-values <0.05 highlighted in bold. Key: HOP, Head of Pancreas; PDAC, pancreatic ductal adenocarcinoma; CA 19-9, Carbohydrate antigen 19-9; CRP, C-Reactive Protein.

*Blood results where available from patient electronic hospital records.

Malignant vs Benign comparison of the demographic data showed no statistical difference for age and sex in the discovery cohort as expected due to sample selection. Within the validation cohort there did appear to be a significant difference in mean age (P value 0.0149) but no difference in the sex ratio. This accounts for the overall statistically significant difference in the ages between our disease and control cohorts with an older population making up patients who present with PDAC and CCA (P 0.0207).

Looking at the overall study, the PDAC cohort included n=30 patients of which 27% were female and 73% were male. The mean age was 69 years with a standard deviation of 11 years. The CCA cohort included n=19 patients of which 42% were female and 58% male. The mean age was 68 years with a standard deviation of 9 years. The Benign control cohort included n=41 patients of which 46% were female and 54% were male. The mean age was 62 years with a standard deviation of 16 years.

Comparatively there was no difference to the sampled population for the cfRNA (See Table 10) analysis. The mean ages in PDAC and CCA were identical (69 and 68 respectively) and the sex ratios were 34% female for PDAC and 32% female for CCA (slightly lower than in this population). The benign cohort had the greatest tendency for deviation as these populations did not overlap, but the mean age in the cfRNA was still the same at 62 (SD – 16) but with a sex ratio of 39% female which was lower than in this population as well. In summary, the populations sampled were of similar make-up with a slightly higher number of females in the Benign cohort this time around.

The mean CA 19-9 in the PDAC cohort was very high, at >10,000 U/mL, in the CCA cohort was lower at 6465U/mL and in the Benign cohort was 285.6U/mL. The mean serum bilirubin level in the PDAC cohort was 246.6ng/mL, in the CCA cohort was 168.8ng/mL and in the Benign cohort was 49.0ng/mL. The mean CRP in the PDAC cohort was 55.3µg/mL. in the CCA cohort was slightly lower at 45.1µg/mL and in the Benign the mean CRP was 64.2µg/mL.

Differences in the discovery cohort were analysed but not deemed statistically applicable as there were a high (>50%) number of missing data in this cohort. Differences were identified in the pre-therapeutic bilirubin levels (P <0.0001) with them significantly raised in the Malignant samples within all cohorts whilst CRP levels were found to be no different across all patients. CA19-9 interestingly did appear to trend towards significance

in both the overall and validation cohorts but was in fact below the level of significance (P = 0.0671 overall) in the Malignant population.

Examining the cancer phenotypes of the overall population analysed showed that most cancers causing biliary obstruction anatomically were located in the head of pancreas (77%) with the remainders Uncinate (10%) or a mixture of the two radiologically (13%). CCA included mainly perihilar but also distal and intrahepatic. Tumour staging was predominantly T4 (63% in PDAC and 42% CCA) as previously noted, with a small proportion T2 (11-13%; PDAC and CCA). A high degree of nodal disease was present in all cohorts (60% in PDAC, 42% in CCA) and metastatic disease was present in 33% of PDAC patients, 11% of CCA patients.

4.2.12 Workflow for the biomarker analysis of EV-RNA from the MIRABILE samples.

The workflow for the EV study can be broken down into three main steps. Initial characterisation in n=9 samples followed MISEV guidelines to ensure that the SEC protocol followed results in EVs that are suitable for analysis. This required nanoparticle tracking analysis, transmission electron microscopy, protein quantification and Western blotting. The results of this are shown above (4.2.10).

The discovery cohort of samples underwent next generation sequencing (NGS) and differential expression analysis for biomarker discovery of small RNAs including miRNAs and tRNA-derived fragments, which we had previously shown as present in surgical bile EVs. Finally selected candidates were validated by RT-qPCR for identification of potential markers for PDAC and/or CCA (FIGURE 42)

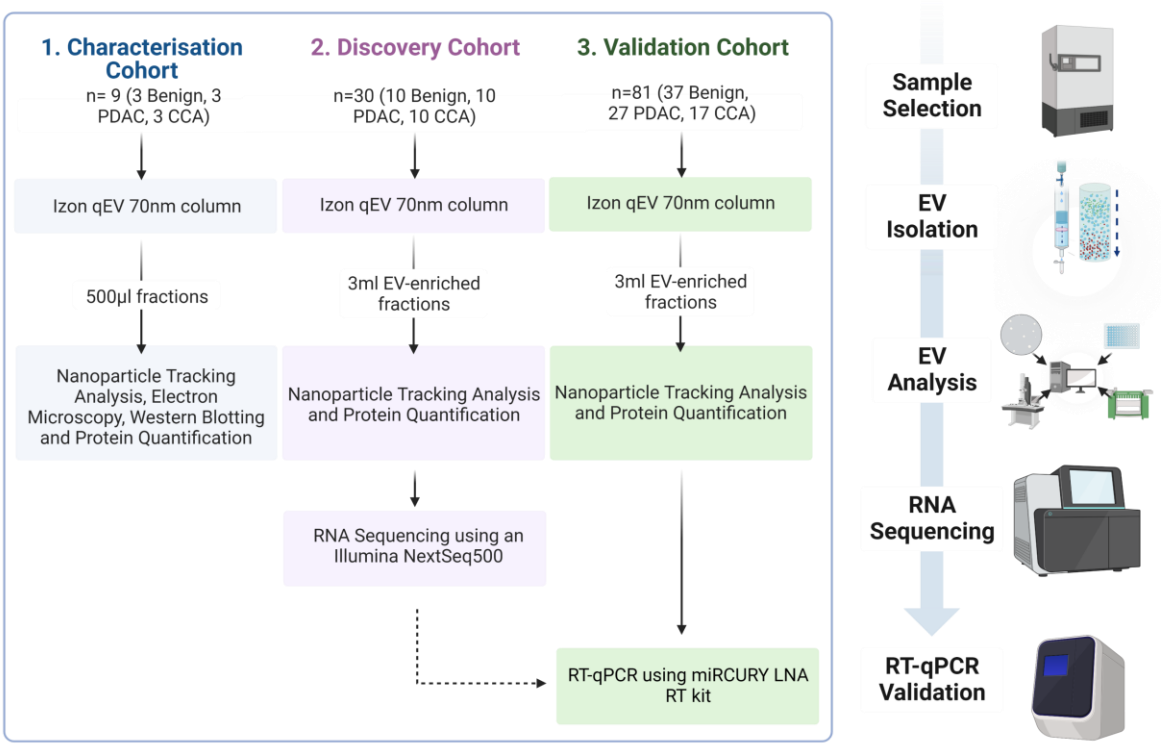


Figure 42: Study Design for the evaluation of EVs from MIRABILE bile samples.
Compared to Figure 28 the EV workflow now included additional validation of the EV-RNA biomarkers by RT-qPCR. The study design includes establishing EV characteristics according to MISEV guidelines using Nanoparticle Tracking Analysis, Transmission Electron Microscopy and Protein analysis (Western blotting and quantification). Discovery in n=30 samples inform the larger validation cohort.
PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, EV – Extracellular Vesicle, MISEV – Minimal Information for the Study of Extracellular Vesicles, RT-qPCR – Reverse transcription quantitative polymerase chain reaction.

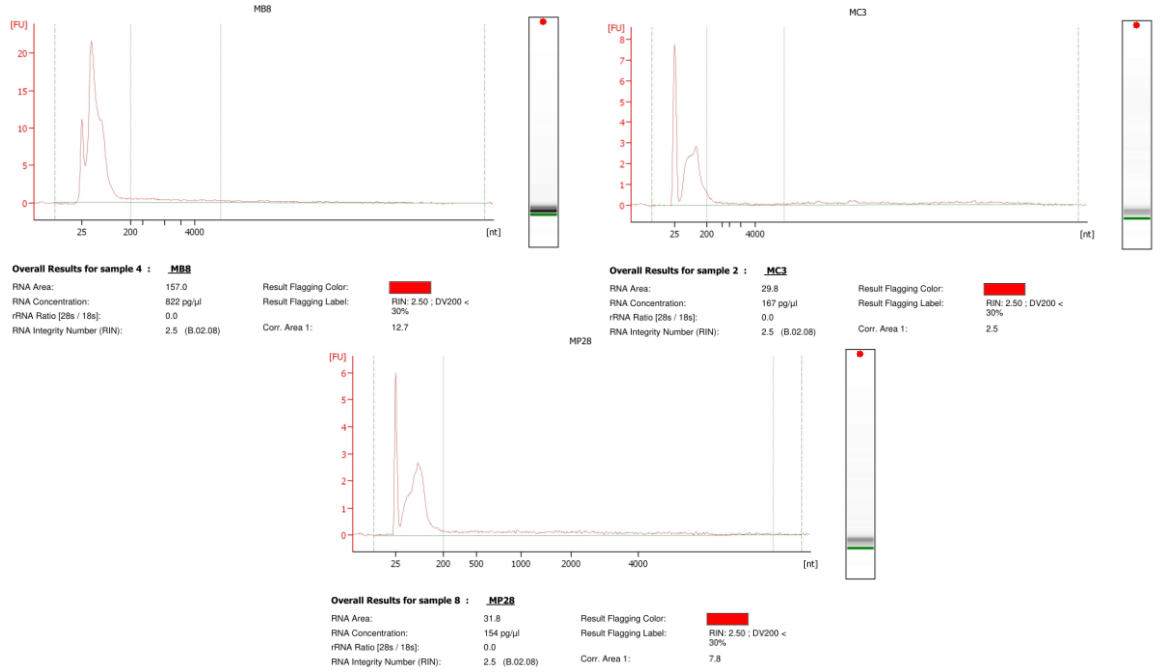
4.2.13 RNA quantification in bile-derived EVs.

RNA quantification was performed using automated electrophoresis using an Agilent Bioanalyzer to determine the presence of RNAs. A sample from each is shown in FIGURE 43A. Plots show a predominance of small RNAs <200-nts with concentrations ~100pg/µL.

RNA distribution by stacked bar plots are shown in FIGURE 43B. Two samples were found to have low (<1%) miRNA composition (1 Benign, 1 CCA) and as outliers were excluded from the analysis. An average of 749,754 total reads (range 222,570-1,672,643) were found in the remaining samples and after mapping (see 2.2.12 for method) 28 samples had an average of 462,392 mapped reads (range 123,259 – 972,205) available for differential expression analysis.

The composition of small RNAs were heterogenous among the samples and miRNAs comprised an average of 8.525% of the mapped reads (range 0.237%-51.252%). tRNA reads were another high proportion, comprising 10.416% on average (range 0.745%-62.729%) and the remaining reads were constituted by ribosomal RNA (21.347%, range 6.955%-35.240%), mRNA (29.6%, range 16.036%-41.876%) and genomic reads (27.943%, range 5.684%-36.885%).

A RNA electrophoresis showing RNA size distribution in 3 bile EV samples (benign above left, PDAC above right, CCA below)



B Relative distribution of mapped RNA reads in Bile EVs according to RNA type

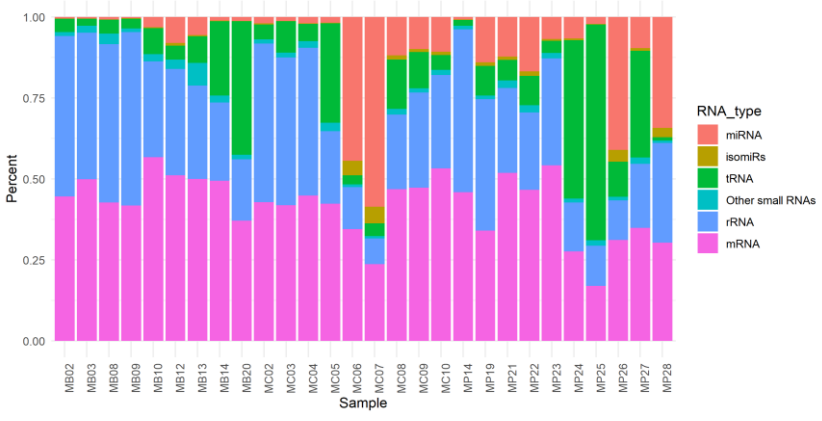


Figure 43: RNA Composition of MIRABILE Bile EV-RNA (n=30).

A. RNA electrophoresis was performed to determine RNA quantity and quality with plots shown for three samples (Benign, PDAC and CCA).

B. Bile EV samples were sent for single end 75 base pair small RNA sequencing by Dr Morten Venø and the team at Omiics (Aarhus, Denmark). Shown are the mapped RNA reads and relative proportions in each sample making up each RNA subtype (colour legend to the right). Stacked bar plots are shown ordered by sample number with results in percentages.

PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma.

4.2.14 Differential expression analysis of bile EV-miRNAs

After filtering, the small RNA sequencing data for bile EVs included expression of 766 miRNAs out of a total of 2656 mature miRNAs based on miRbase v22. Principal component analysis was undertaken to investigate correlation of different samples according to miRNA expression values and showed good separation of our cohorts by expression (FIGURE 44A). CCA and PDAC samples (green and blue respectively) were shown as one colour below and appeared to group together.

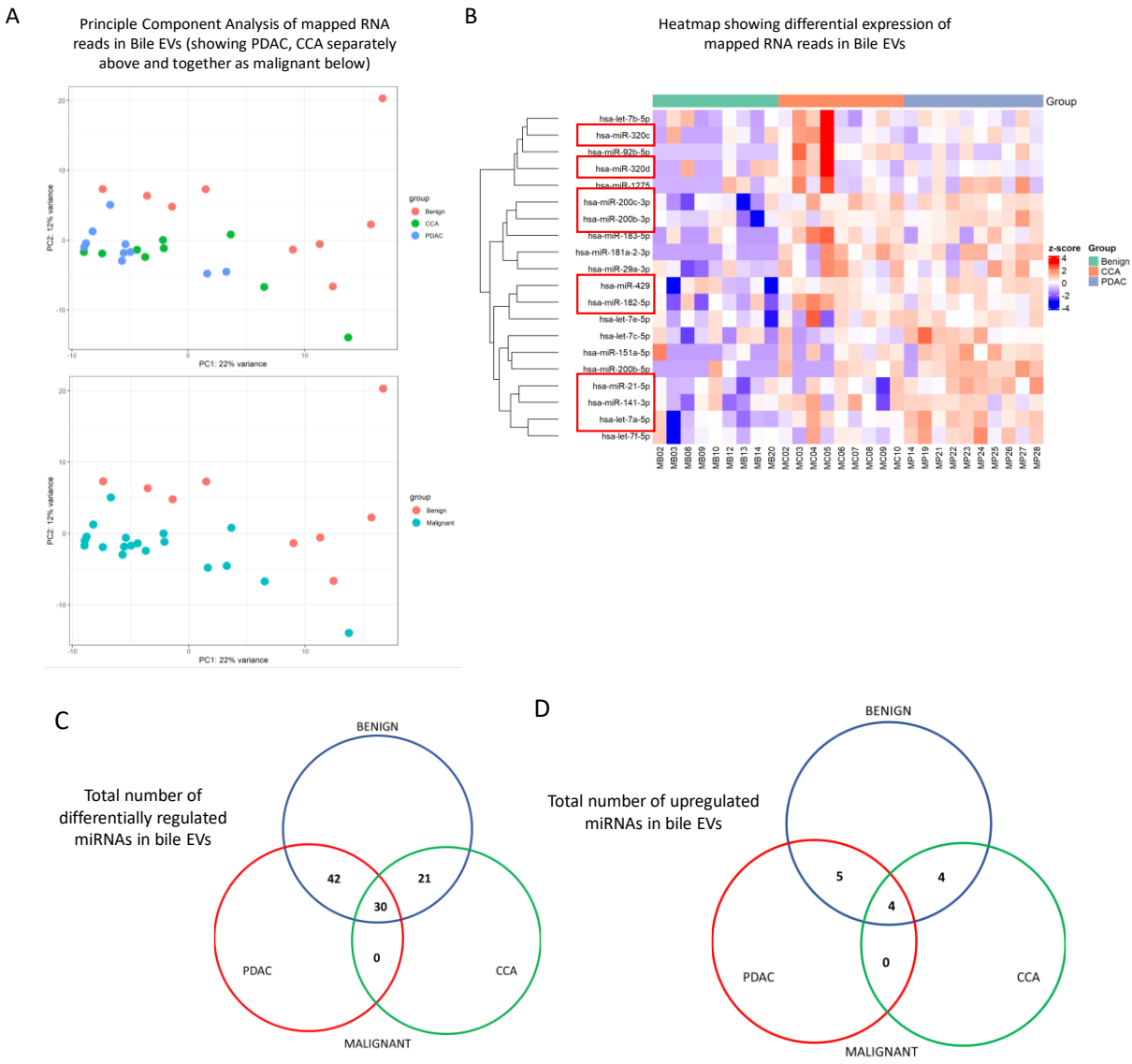


Figure 44: Differential expression analysis of miRNAs identified from small RNA sequencing in bile EVs (n=28).

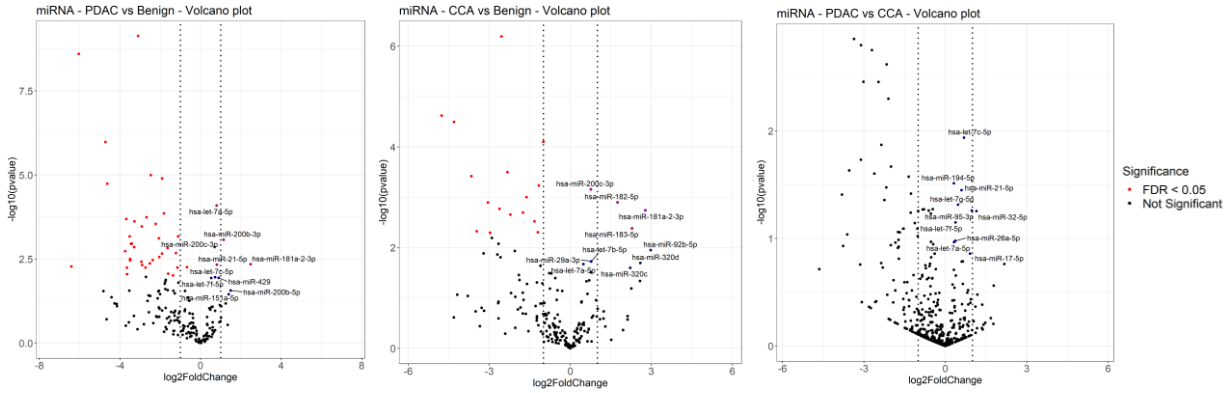
A. Principal component analysis plot showing broad grouping of the samples by cohort according to miRNA expression (PDAC vs CCA vs Benign above and Malignant (PDAC+CCA) vs Benign below).
B. Supervised heatmap with hierarchical clustering showing all miRNAs with adjusted p-values <0.1 in the pairwise comparisons shown below. Highlighted in a red box are the candidates taken forward for subsequent validation. Red means high relative expression with blue signifying low relative expression.

PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, FDR – False Discovery Rate

Heatmaps were generated to show potential patterns in expression. A partially supervised heatmap (cohorts grouped) and hierarchical clustering of the miRNAs is shown in FIGURE 44B with grouping according to PDAC, CCA or Benign. One group appeared to distinctly be downregulated in Benign and upregulated in CCA compromising of miR-320c/miR-320d/let-7b/miR-92b/miR-1275. After adjustment using Benjamini–Hochberg methodology to control the False Discovery Rate (FDR), we identified 30 differentially expressed miRNAs between Malignant vs Benign. Broken down by each cancer, there were 42 potential candidates in PDAC vs Benign and 21 candidates between CCA vs Benign. FIGURE 44C shows the overall number of statistically significant results (after adjustment) and the number upregulated in our cancer samples. Once again, no candidates were shown to be significantly differentially expressed between PDAC and CCA which was the same result found in our bile cfRNA study and in our surgical bile EVs. A much smaller number of candidates (compared with bile cfRNA) were found upregulated in cancer samples with 4 miRNAs found between Malignant vs Benign, 5 miRNAs in PDAC vs Benign and 4 miRNAs in CCA vs Benign (FIGURE 44D). It was therefore decided to focus my attention on these upregulated miRNAs found in bile EV cancer samples as a potential biomarker for disease. As RT-qPCR can be feasibly performed with 10-12 candidates depending on time and technical expertise, we investigated these markers rather than using the LASSO regression method to determine an optimal model.

Volcano plots (FIGURE 45) show differentially expressed miRNAs with significance (adjusted p-value) against fold change on the y and x-axis respectively. Highlighted in red are the statistically significant candidates (adjusted p-value ≤ 0.05) and top ten miRNAs upregulated are labelled. These show a skew towards downregulation of miRNAs in bile EVs in Malignancy.

A Differentially expressed miRNAs in MIRABILE Bile EVs (PDAC vs Benign; CCA vs Benign; PDAC vs CCA)



B Differentially expressed miRNAs in MIRABILE Bile EVs (Malignant (PDAC+CCA) vs Benign)

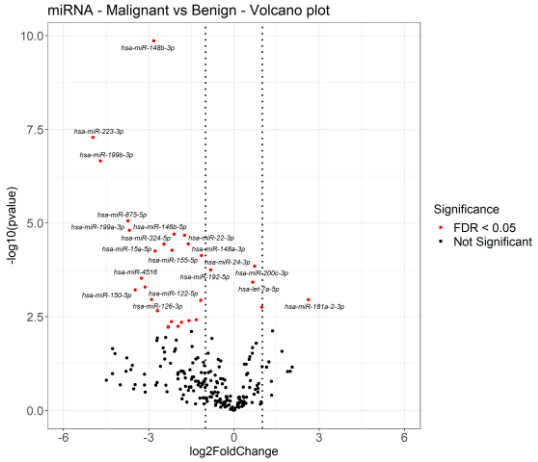


Figure 45: Volcano plots showing significant miRNAs present in bile EV (MIRABILE) samples (n=28).

A. Volcano plots with miRNAs found to be significant highlighted in red in the following pairwise comparisons (PDAC vs Benign; CCA vs Benign; PDAC vs CCA) with an FDR < 0.05.
B. Volcano plot with top miRNAs found to be significant in the pairwise comparison Malignant vs Benign.

Plotted are significance ($-\log^{10}$ p-value) vs expression ($\log^2$ Fold Change). Vertical lines indicate $\log^2$ Fold Change ±1. Only the 10 most significant miRNAs are labelled for clarity.
PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, FDR – False Discovery Rate

TABLE 27 shows the 5 miRNA candidates statistically upregulated in PDAC vs Benign, 4 candidates in CCA vs Benign and 4 candidates in Malignant vs Benign. We focussed only on those upregulated in cancer which reduced our candidates significantly.

Table 27: Significant miRNAs (p<0.05) common from each pairwise comparison alongside adjusted p-value and log2fold change in MIRABILE bile EVs (n=28).

Highlighted in bold are miRNAs found in all comparisons. Those highlighted in red are members of the miR-200 family which are subsequently of interest in Chapter Five (Plasma EV analysis). No candidates were found which were common to this analysis and the cfRNA or surgical EV samples. For the pairwise comparison, PDAC vs CCA, no significant hits were identified. Values rounded to 5 sig. fig.
PDAC – pancreatic ductal adenocarcinoma, CCA – cholangiocarcinoma.

Comparison	MicroRNA	Expression	Log2FoldChange	Standard Error	Adjusted P-value
PDAC vs Benign	**miR-181a-2-3p**	4.9274	2.4858	0.8758	0.0283
	miR-200b-3p	79.609	1.1437	0.3427	0.0100
	miR-21-5p	893.47	0.8100	0.2863	0.0283
	let-7a-5p	996.85	0.8007	0.2032	0.0023
	miR-200c-3p	260.37	0.6999	0.2188	0.0135
CCA vs Benign	**miR-200c-3p**	260.37	0.7607	0.2244	0.0175
	miR-182-5p	14.655	1.7455	0.5413	0.0233
	miR-181a-2-3p	4.9274	2.7754	0.8902	0.0282
	miR-183-5p	6.8059	2.2837	0.7966	0.0490
PDAC vs CCA	-	-	-	-	-
Malignant vs Benign	**miR-181a-2-3p**	4.9274	2.6196	0.8034	0.0127
	miR-200b-3p	79.609	0.9783	0.3130	0.0186
	miR-200c-3p	260.37	0.7289	0.1916	0.0025
	let-7a-5p	996.85	0.6581	0.1852	0.0055

Of these, 2 candidate bile EV miRNAs for diagnosis featured in all three analyses. MiR-181a-2-3p (log2foldchange 2.6196; p-adj=0.0127) and miR-200c-3p (log2foldchange 0.7289; p=0.0025) was significantly upregulated in the Malignant vs Benign comparison. These miRNAs were also consistently differentially expressed in the PDAC vs Benign and CCA vs Benign comparisons indicating they may be a marker of hepatobiliary adenocarcinoma. Other candidates include let-7a-5p, miR-21-5p and miR-200b-3p which were upregulated in PDAC vs Benign, and miR-182-5p, miR-183-5p which were upregulated in CCA vs Benign.

Additionally, we show the top twenty miRNAs with most expression in PDAC vs Benign in TABLE 28, the top twenty miRNAs with most expression in CCA vs Benign in TABLE 29 and the top twenty in Malignant vs Benign in TABLE 30. Upregulated but non-significant miRNAs in PDAC included miR-200b-5p (miR-200b*), miR-151-5p, miR-1275 and miR-95-3p. Upregulated but non-significant miRNAs in CCA also included miR-92b-5p, miR-320d, and again miR-1275.

Table 28: Top Twenty miRNAs identified as expressed in PDAC samples in the MIRABILE bile EV biomarker analysis (n=28).

Top twenty miRNA candidates identified by RNA sequencing as differentially expressed in PDAC vs Benign sorted by descending log fold change values with candidates in bold those with a significant adjusted p value (<0.05). Expression averaged across all samples with normalization by DESeq2 to allow direct comparison between the samples. Standard errors are of the log2fold change and p-values are adjusted using the Benjamini-Hochberg method.

MicroRNA	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
miR-181a-2-3p	**4.9274**	**2.4858**	**0.8758**	**0.02825**
miR-200b-5p	2.7370	1.4906	0.6746	0.09917
miR-151a-5p	2.4736	1.3938	0.6627	0.11496
miR-1275	21.152	1.3476	1.2635	0.46018
miR-95-3p	4.0840	1.2440	0.6775	0.18775
miR-200b-3p	**79.609**	**1.1437**	**0.3427**	**0.01002**
miR-135b-5p	4.2227	1.0240	0.5853	0.21489
miR-454-3p	2.2477	1.0166	0.7512	0.33679
miR-3065-5p	3.4728	0.9928	0.8209	0.40299
miR-182-5p	14.655	0.9570	0.5354	0.20343
miR-183-5p	6.8059	0.9357	0.7931	0.41929
miR-429	33.815	0.8938	0.3536	0.05064
miR-21-5p	**893.47**	**0.8100**	**0.2863**	**0.02825**
let-7a-5p	**996.85**	**0.8007**	**0.2032**	**0.00233**
miR-224-5p	13.708	0.7231	0.4514	0.25512
let-7c-5p	143.11	0.7229	0.2843	0.05026
miR-3065-3p	1.4673	0.7200	0.8797	0.58071
miR-200c-3p	**260.37**	**0.6999**	**0.2188**	**0.01346**
miR-215-5p	6.9384	0.6580	0.8965	0.61628
miR-20a-5p	5.6362	0.6026	0.5152	0.41929

Table 29: Top Twenty miRNAs identified as expressed in CCA samples in the MIRABILE bile EV biomarker analysis (n=28).

Top twenty miRNA candidates identified by RNA sequencing as differentially expressed in CCA vs Benign sorted by descending log fold change values with candidates in bold those with a significant adjusted p value (<0.05).

MicroRNA	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
miR-92b-5p	2.8711	2.9705	1.1713	0.0939
miR-181a-2-3p	**4.9274**	**2.7754**	**0.8902**	**0.0282**
miR-320d	6.6861	2.5908	1.1152	0.1308
miR-1275	21.152	2.5765	1.2911	0.2194
miR-183-5p	**6.8059**	**2.2837**	**0.7966**	**0.0490**
miR-10a-5p	9.7473	2.2553	1.2226	0.2738
miR-320c	6.3525	2.2129	0.9888	0.1491
miR-432-5p	1.8086	2.1216	1.9201	0.5695
miR-196a-5p	2.2682	2.1207	1.7633	0.5469
miR-196b-5p	3.4712	1.9506	2.4994	0.7111
miR-182-5p	**14.655**	**1.7455**	**0.5413**	**0.0233**
miR-345-3p	1.0210	1.5125	3.6694	0.8532
miR-671-5p	2.6095	1.2941	1.4471	0.6707
miR-320b	9.7312	1.2359	0.9331	0.5034
miR-574-5p	30.285	1.2289	1.0856	0.5621
miR-455-5p	1.1230	1.0954	0.9818	0.5656
miR-92b-3p	1.6893	1.0100	0.7269	0.4729
miR-1307-3p	9.2791	0.9803	0.8236	0.5469
miR-135b-5p	4.2227	0.9524	0.5969	0.3767
miR-200b-5p	2.7370	0.8438	0.6977	0.5469

Table 30: Top twenty miRNAs identified as expressed in Malignant samples in the MIRABILE bile EV biomarker analysis (n=28).

Top twenty miRNA candidates identified by RNA sequencing as differentially expressed in PDAC vs Benign sorted by descending log fold change values with candidates in bold those with a significant adjusted p value (<0.05).

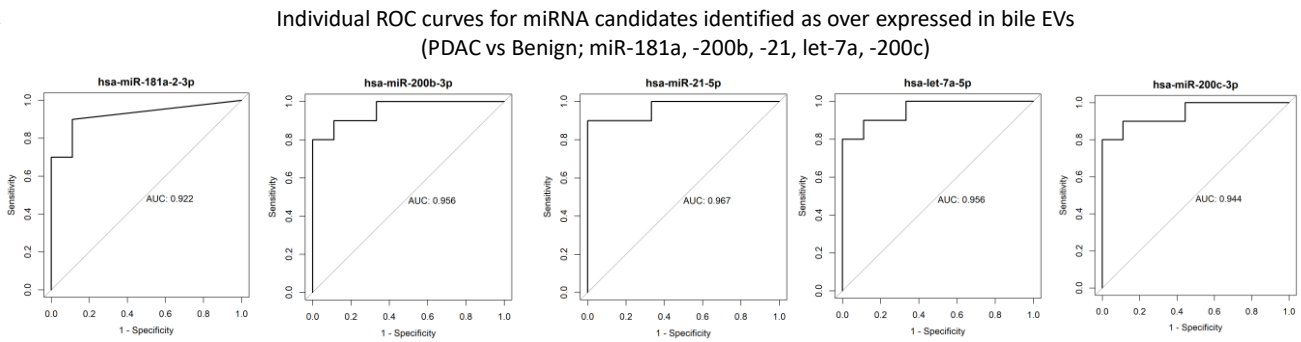
MicroRNA	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
miR-181a-2-3p	**4.9274**	**2.6196**	**0.8034**	**0.0127**
miR-1275	21.1519	2.0482	1.1300	0.2325
miR-92b-5p	2.8711	1.9779	1.1719	0.2697
miR-10a-5p	11.616	1.8842	1.1191	0.2697
miR-183-5p	6.8059	1.6888	0.7609	0.1361
miR-182-5p	14.655	1.3548	0.5071	0.0562
miR-196a-5p	2.2682	1.3251	1.5494	0.6004
miR-320c	6.3525	1.3176	0.9584	0.3686
miR-200b-5p	2.7370	1.2312	0.6315	0.2032
miR-151a-5p	2.4736	1.1491	0.6352	0.2325
miR-135b-5p	4.2227	0.9923	0.5446	0.2325
miR-200b-3p	**79.609**	**0.9783**	**0.3130**	**0.0186**
miR-432-5p	1.8086	0.9547	1.8122	0.7552
miR-196b-5p	3.4712	0.9304	2.2611	0.7862
miR-95-3p	4.0840	0.8701	0.6658	0.3910
miR-3065-5p	3.4728	0.8428	0.7480	0.4446
miR-455-5p	1.1230	0.8426	0.8595	0.5208
miR-92b-3p	1.6893	0.8162	0.6574	0.3994
miR-429	33.815	0.7791	0.3238	0.0980
miR-454-3p	2.2477	0.7506	0.6896	0.4593

4.2.15 Candidates chosen for RT-qPCR Validation show outstanding AUC in the discovery cohort.

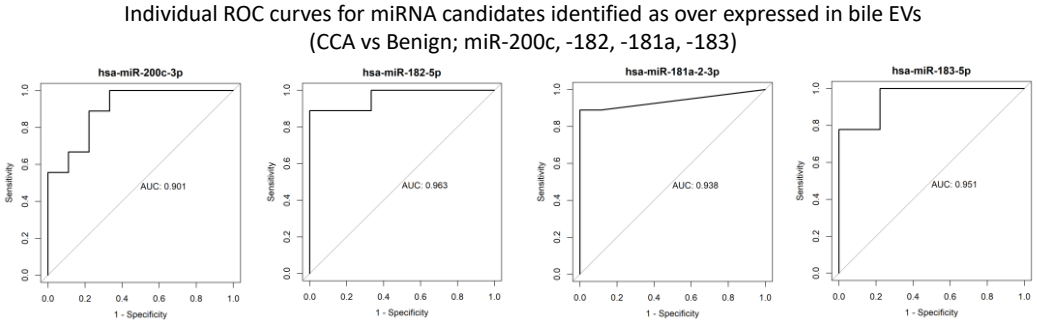
From the sequencing data results we chose 12 candidates for further validation. This was based on selecting those significantly upregulated in the comparison Malignant vs Benign (let-7a-5p, miR-200b-3p, miR-200c-3p), those that had specificity for PDAC (miR-21-5p) and those with specificity for CCA (miR-181a-2-3p, miR-182-5p). We also wanted to investigate the other members of the miR-200 family, having identified them in our plasma EV study (miR-200a-3p, miR-429). Finally, there was a group of non-significant but interesting cluster of miR-320 which could have CCA specificity, for which we undertook an exploratory study (miR-320a-3p, miR-320b, miR-320c, miR-320d). miR-183-5p was significantly upregulated in CCA vs Benign but we decided not to include this for validation due to experimental constraints.

FIGURE 46 shows the ROC plots for these based on PDAC and CCA pairwise comparisons. They showed the potential for outstanding AUCs of >0.9.

A



B



C

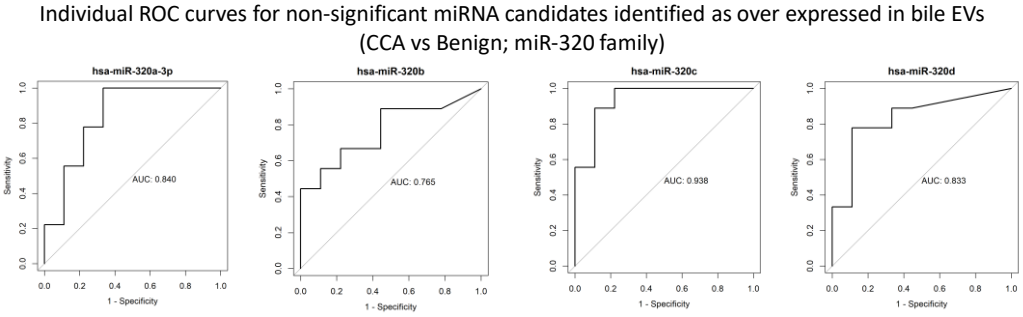


Figure 46: Receiver-Operator Curve plots for selected miRNAs identified as differentially expressed in PDAC and CCA samples in the bile EV biomarker analysis.

A. Significant miRNA candidates (as shown in TABLE 27) for PDAC vs Benign identified by RNA sequencing with AUCs as annotated.
B. Significant miRNA candidates (as shown in TABLE 28) for CCA vs Benign identified by RNA sequencing with AUCs as annotated.
C. Non-significant miRNA candidates chosen from miR-320 family for CCA vs benign pairwise comparison identified by RNA sequencing.

Whilst previously a LASSO technique was necessary to reduce the number of candidates suitable for a predictive model, we did not feel that shrinkage of the existing model was required. We decided also to look only at upregulated candidates despite the relative downregulation in bile EVs seen (see Figure 45). These included miRNAs such as miR-148b-3p, miR-223-3p and miR-199b-3p. By focussing on only those deemed statistically significant we had 6 possible candidates (let-7a-5p, miR-200b-3p, miR-200c-3p, miR-21-5p, miR-181a-2-3p, miR-182-5p). We additionally included an exploratory evaluation of the other two miR-200 family members (miR-200a and miR-429) as well as the promising miR-320 family (miR-320a, miR-320b, miR-320c, miR-320d) which showed a trend

towards expression in CCA. With the exception of miR-320b, these all showed a predicted AUC of greater than 0.80 when based on RNA sequencing reads only.

4.2.16 A note on RT-qPCR Normalisation.

Using the sequencing data, we were able to run the NormFinder algorithm to identify suitable stably expressed miRNAs for use as a normaliser. Both individual miRNAs (TABLE 31) and pairs (TABLE 32) are shown. We chose miR-125a-5p and miR-194-5p as paired normalisers with the best stability. Additionally, it was decided that to compensate for any inter-plate variation, we would use an exogenous synthetic spike-in (UniSp6) and added a fixed amount (0.0375 fmol) to each sample. This was added to the protocol because the miRCURY method did not require a fixed amount of RNA (as was required for TaqMan, the RT-qPCR kit used in Chapter Three). By adding in an exogenous spike-in at the beginning of the reverse transcription stage, we ensured technical efficacy of the validation stage could be checked and standardised across our samples. This was also recommended by a previous study for further evaluation of EV RNA in bile (254).

The geometric mean, using principles by Vandesompele et al (318) was calculated and an average (EndoMean) was used to normalise the data (see Section 2.2.17 for a detailed explanation).

Table 31 & 32: MiRNAs identified as stably expressed for normalization in bile EVs.
Top twenty miRNAs (left) identified as stably expressed in all samples sorted by descending stability using the NormFinder algorithm. Top ten pairs (right) of miRNAs identified as stably expressed in all samples sorted by descending stability. GroupDif is the estimated difference in expression between the groups; GroupSD is a weighted average of the estimated intra-group variation and Stability is the NormFinder expression stability measure with lower values indicating stable gene expression.

MicroRNA	GroupDif	GroupSD	Stability
miR-194-5p	0.49	0.34	0.29
miR-125a-5p	0.17	0.62	0.29
miR-151a-3p	0.35	0.46	0.3
miR-93-5p	0.3	0.66	0.35
miR-26b-5p	0.37	0.67	0.37
miR-27b-3p	0.44	0.67	0.39
miR-30d-5p	0.51	0.64	0.39
miR-16-5p	0.53	0.71	0.4
miR-30e-5p	0.34	0.89	0.41
miR-125b-5p	0.45	0.73	0.41
let-7b-5p	0.71	0.58	0.42
let-7g-5p	0.68	0.57	0.42
let-7i-5p	0.53	0.69	0.42
miR-29c-3p	0.49	0.75	0.43
miR-92a-3p	0.5	0.81	0.43
miR-26a-5p	0.73	0.58	0.44
miR-191-5p	0.7	0.64	0.44
let-7f-5p	0.81	0.5	0.45
miR-192-5p	0.89	0.61	0.47
miR-200a-3p	1.03	0.63	0.5

MicroRNA 1	MicroRNA 2	Stability
miR-194-5p	miR-151a-3p	0.16
miR-194-5p	miR-125a-5p	0.2
miR-194-5p	miR-93-5p	0.21
miR-125a-5p	miR-151a-3p	0.24
miR-93-5p	miR-125a-5p	0.25
miR-93-5p	miR-151a-3p	0.25

4.2.17 Validation of our 12 candidates shows diagnostic potential in EV-miRNA markers for PDAC and CCA.

FIGURE 46 showed the results of RT-qPCR validation of 10 miRNAs compared in the PDAC samples. These were let-7a-5p, miR-21-5p, miR-200a-3p, miR-200b-3p, miR-200c-3p, miR-429, miR-320a-3p, miR-320b, miR-320c and miR-320d (grouped by family).

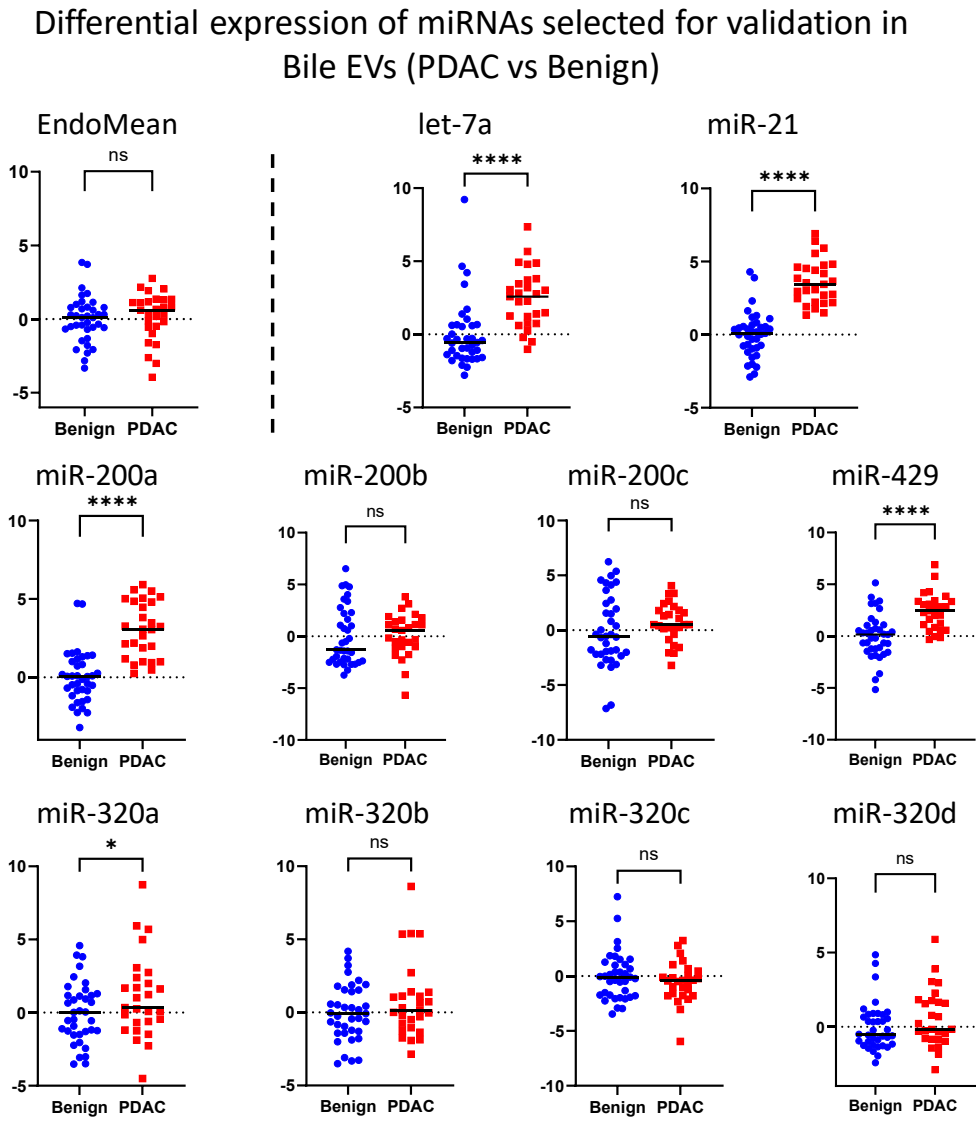


Figure 47: RT-qPCR Validation of marker panel in bile EV RNA samples (PDAC vs Benign; n=64).

The 10-marker panel let-7a-5p, miR-21-5p, miR-200a-3p, miR-200b-3p, miR-200c-3p, miR-429, miR-320a-3p, miR-320b, miR-320c and miR-320d (grouped by family) were evaluated using RT-qPCR and their expression calculated using the comparative (ΔΔCT) method. Differences in mean log fold expression were evaluated by Student t-test with asterisks to demonstrate significance. All reactions were performed in triplicate and the mean value used. Endogenous normalisers miR-125a-5p, miR-194-5p and UniSp6 were used (shown as EndoMean).
* P <0.05, ** P<0.005, *** P<0.0005, **** P<0.0001.

No difference was detected as expected in the EndoMean (miR-125a-5p, miR-194-5p and UniSp6) whilst let-7a, miR-21, miR-200a and miR-429 were upregulated in PDAC samples ($p<0.0001$). miR-320a was also investigated, and although this was thought to be a CCA specific marker was found also to be differentially expressed in PDAC vs Benign ($p=0.0424$). Statistical significance was not achieved for miR-200b ($p=0.4182$) and miR-200c ($p=0.1679$) although a trend towards upregulation in PDAC can be seen in the scatter plots. miR-181-a-2-3p and miR-182-5p did not show expression in ~50% of our samples and so was not able to be evaluated (results not shown).

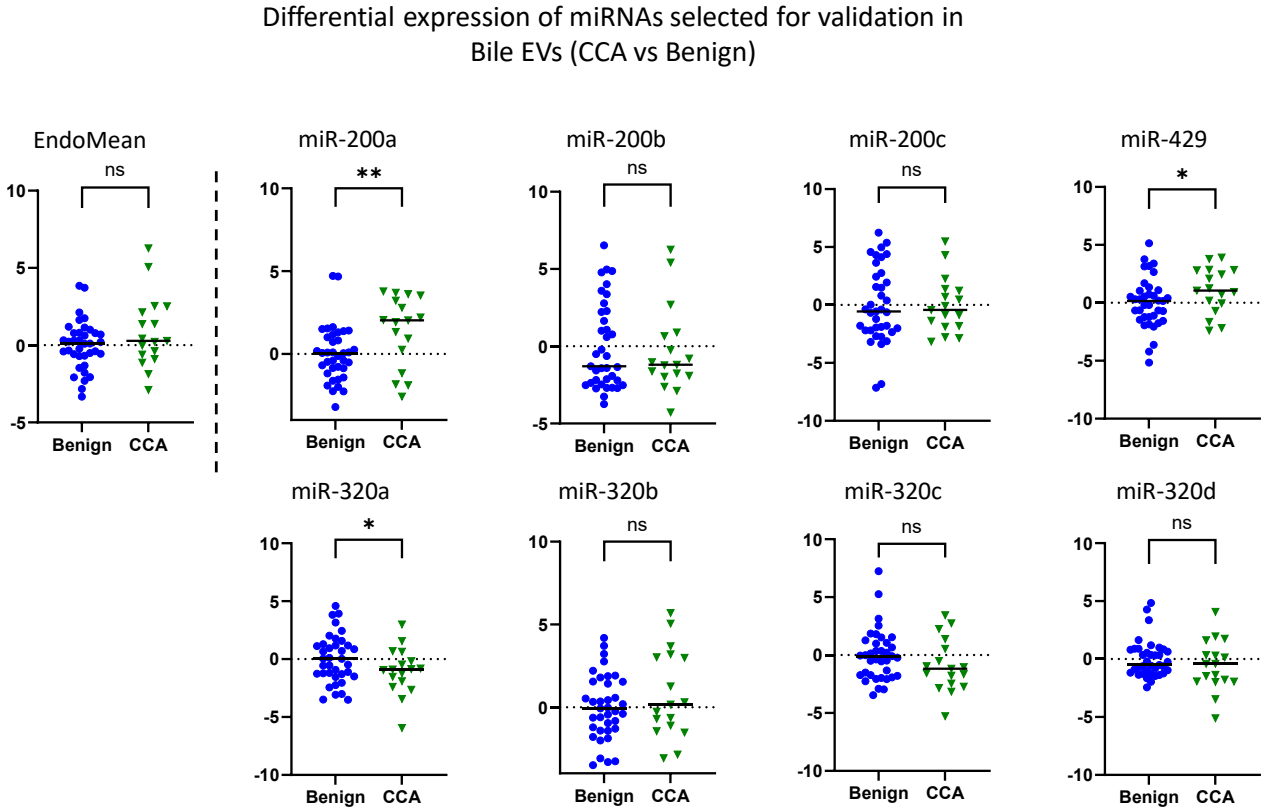


Figure 48: RT-qPCR Validation of marker panel in bile EV RNA samples (CCA vs Benign; n=54).

A 10-marker panel let-7a-5p, miR-21-5p, miR-200a-3p, miR-200b-3p, miR-200c-3p, miR-429, miR-320a-3p, miR-320b, miR-320c and miR-320d (grouped by family) were evaluated using RT-qPCR (mean of triplicate shown) and their expression calculated using the comparative (ΔΔCT) method. Differences in mean log fold expression were evaluated by Student t-test with asterisks to demonstrate significance. All reactions were performed in triplicate and the mean value used. Endogenous normalisers miR-125a-5p, miR-194-5p and UniSp6 were used (shown as EndoMean).
* P <0.05, ** P<0.005, *** P<0.0005, **** P<0.0001.

FIGURE 47 shows the results of RT-qPCR validation of 8 miRNAs compared in the CCA samples. It was decided to omit let-7a and miR-21 as these did not show significance in the CCA vs Benign pairwise comparison and due to experimental constraints. After analysis, miR-181-a-2-3p and miR-182-5p did not show expression (C_T >40 cycles) in ~50% of our samples so the results are not shown. Again, EndoMean is shown to be no different

in either population highlighting their appropriate stability and therefore use as normalisers. miR-200a again shows a statistically significant upregulation (p=0.0057) in CCA vs Benign as does miR-429 (p=0.0468) and miR-320a (p= 0.0491).

In order to ensure RT-qPCR primers were suitable, i.e. binding and amplifying the target miRNAs with equal efficacy across a broad range of RNA input (especially important as EV samples can have wide variations in RNA concentration) a standard curve was performed for each miRNA primer used. Comparative standard curves to demonstrate equivalent primer efficiencies were undertaken in a single highly expressed sample, triplicated, and is shown in FIGURE 49. Slope values calculated for a relative ΔC_T experiment should approach 0, i.e. the relative efficiency of the overall expression calculated is invariant with respect to input RNA. From these graphs, it can be seen that for miR-200b, miR-200c, let-7a, miR-21, miR-429, the slope value was close to zero. With the miR-320 family and miR-200a there appeared to be an upward trend with increasing RNA input which could explain the lack of significant findings particularly in miR-320. A positive slope would indicate that at higher RNA input there might be higher differential expression compared to with low input and therefore increase the risk of Type I error i.e. masking any differential expression that might occur. It may also lead to a Type II error if RNA input is variable within cohorts (high in PDAC and low in Benign for example).

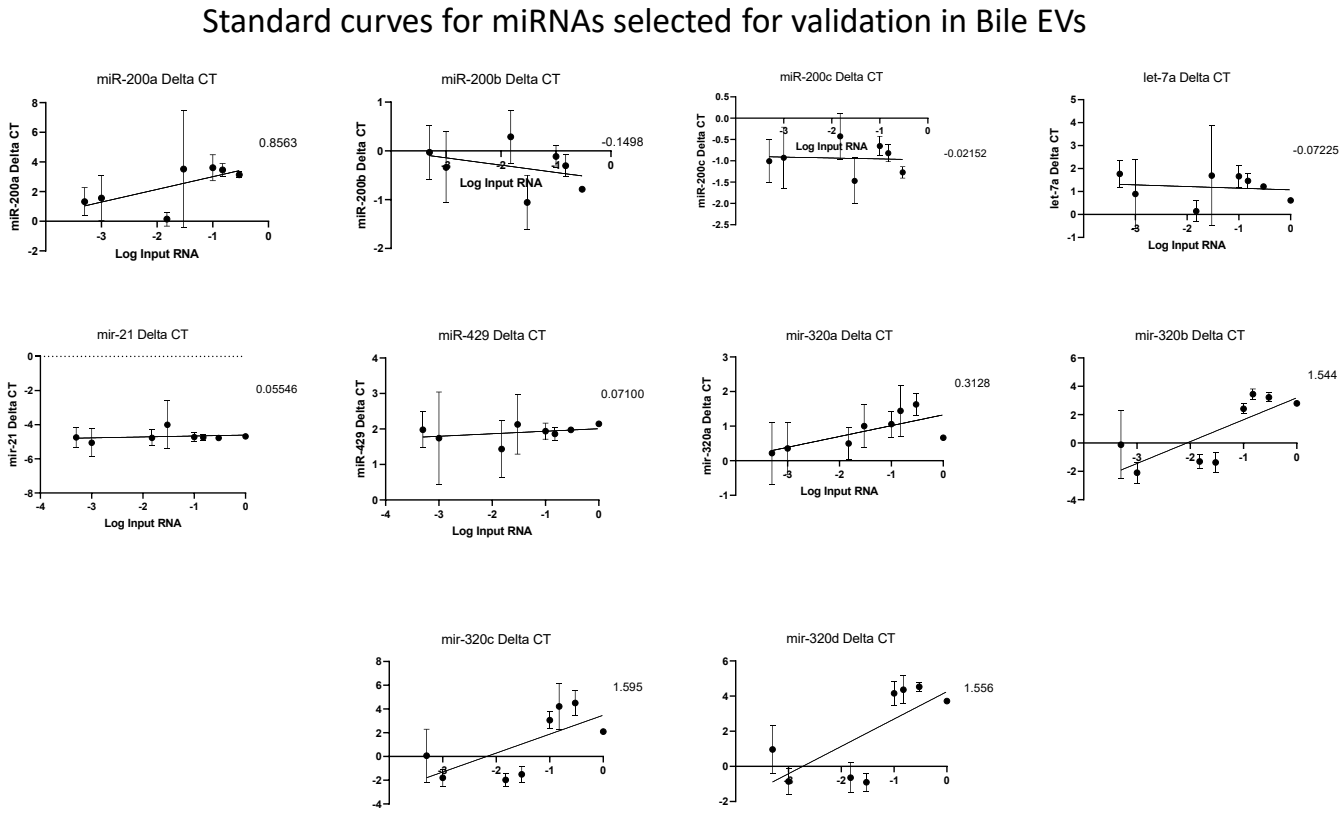


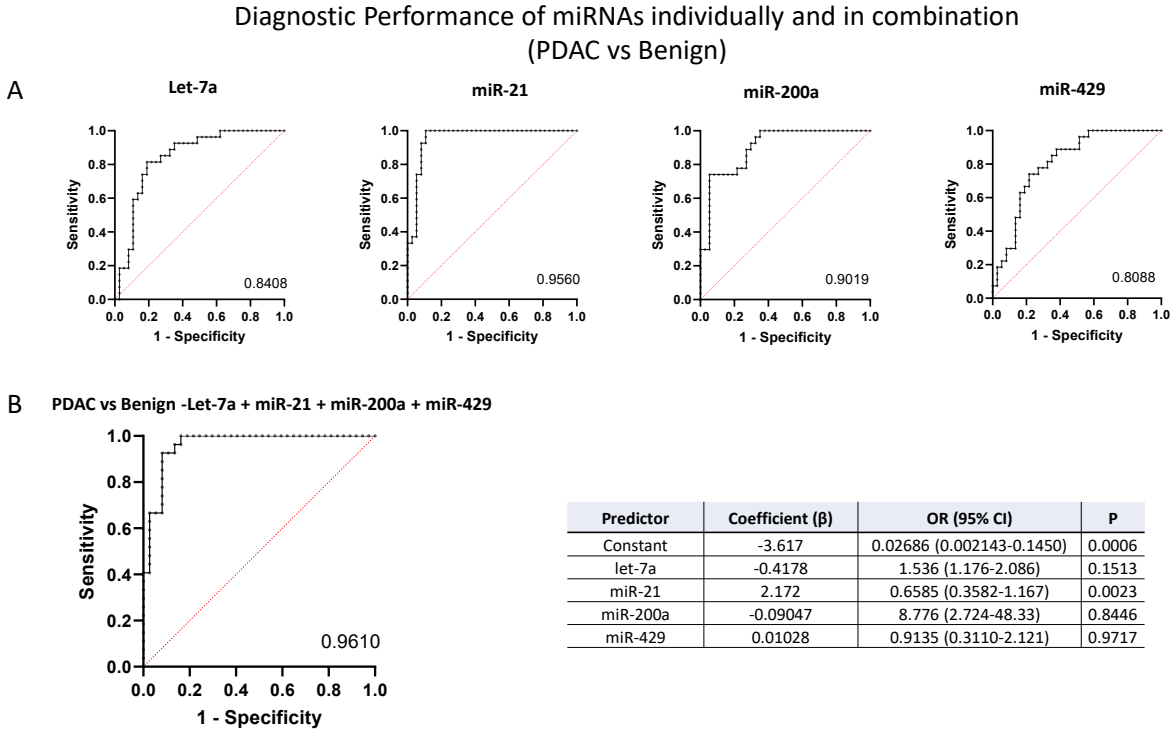
Figure 49: RT-qPCR standard curves for each miRNA candidate validated (n=1 in triplicate).

In order to ensure RT-qPCR primers bind and amplify with equal efficacy across a broad range of RNA input, standard curves were undertaken in a sample with high miRNA expression previously determined. The 10-marker panel let-7a-5p, miR-21-5p, miR-200a-3p, miR-200b-3p, miR-200c-3p, miR-429, miR-320a-3p, miR-320b, miR-320c and miR-320d (grouped by family) was measured with EV-RNA input diluted as shown. Trendlines show the slope values with the ideal gradient as <0.1. Endogenous normalisers miR-125a-5p, miR-194-5p and UniSp6 were used.

4.2.18 Combination of candidates using logistic regression improves diagnostic sensitivity.

The four candidates which showed most promising upregulation in our cancer samples were let-7a-5p, miR-21-5p (in PDAC vs Benign) miR-200a-3p, and miR-429 (in both PDAC and CCA vs Benign). To demonstrate their diagnostic potential we analysed them individually by logistic regression and predicted probabilities were used to generate ROC curves for each (FIGURE 50A). For the identification of PDAC, let-7a had an estimated AUC of 0.8408 (95% CI 0.7417 - 0.9400, p<0.0001), whilst miR-21-5p exhibited the highest AUC 0.9560 (95% CI 0.9049 - 1.000, p<0.0001). miR-200a-3p demonstrated an AUC of 0.9019 (95% CI 0.8291 - 0.9747, p<0.0001) and miR-429 demonstrated a lower AUC of 0.8088 (95% CI 0.7037 - 0.9140, p<0.0001). Linear combination into a 4-marker panel subset of EV-miRNAs which could detect PDAC gives an outstanding AUC of 0.9610 (95% CI 0.9179 - 1.000, p<0.0001, FIGURE 50A). Using this combined panel, the optimum cut-off was determined (>0.4397) which resulted in a sensitivity of 70%, and specificity 87.1%.

Statistics for each marker in the panel are shown in FIGURE 50B with miR-200a achieving significance (p=0.0023).



Predictor	Coefficient (β)	OR (95% CI)	P
Constant	-3.617	0.02686 (0.002143-0.1450)	0.0006
let-7a	-0.4178	1.536 (1.176-2.086)	0.1513
miR-21	2.172	0.6585 (0.3582-1.167)	0.0023
miR-200a	-0.09047	8.776 (2.724-48.33)	0.8446
miR-429	0.01028	0.9135 (0.3110-2.121)	0.9717

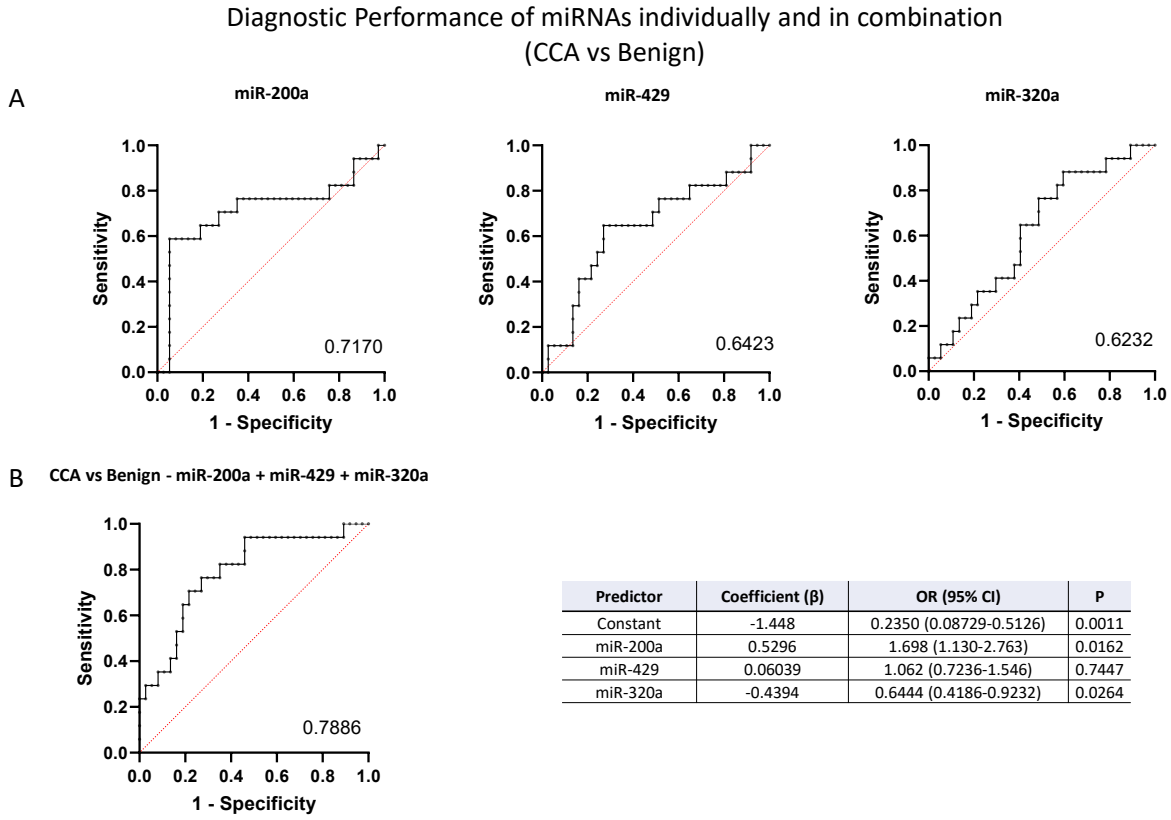
Figure 50: Regression analyses to validate bile EV-miRNA biomarker model (PDAC; n=64).

ROC curves were generated using RT-qPCR validation expression data and analysed by multiple logistic regression for each of the 4 candidates overexpressed in the pairwise comparison PDAC vs Benign.

A. Each individual miRNA (let-7a-5p, miR-21-5p, miR-200a-3p, and miR-429).
B. Combined 4-marker panel with validated AUC = 0.961 (95% CI 0.9179 – 1.000, p<0.0001).

PDAC – Pancreatic Ductal Adenocarcinoma, AUC – Area Under ROC Curve, ROC - Receiver operating characteristic.

Three markers were statistically differentially expressed in our CCA samples and so were also analysed for their diagnostic potential in CCA. This was less successful (FIGURE 51) with the model (miR-200a, miR-429 and miR-320a) showing a validated AUC of 0.7886 (95% CI 0.6573 - 0.9198, p=0.0007). Individually, miR-200a-3p demonstrated a moderately useful AUC of 0.7170 (95% CI 0.5440 - 0.8900, p=0.0110), miR-429 showed an AUC of 0.6423 (95% CI 0.4772 - 0.8074, p=0.0955) and miR-429 demonstrated a lower AUC of 0.6232 (95% CI 0.4686 - 0.7778, p=0.1489).



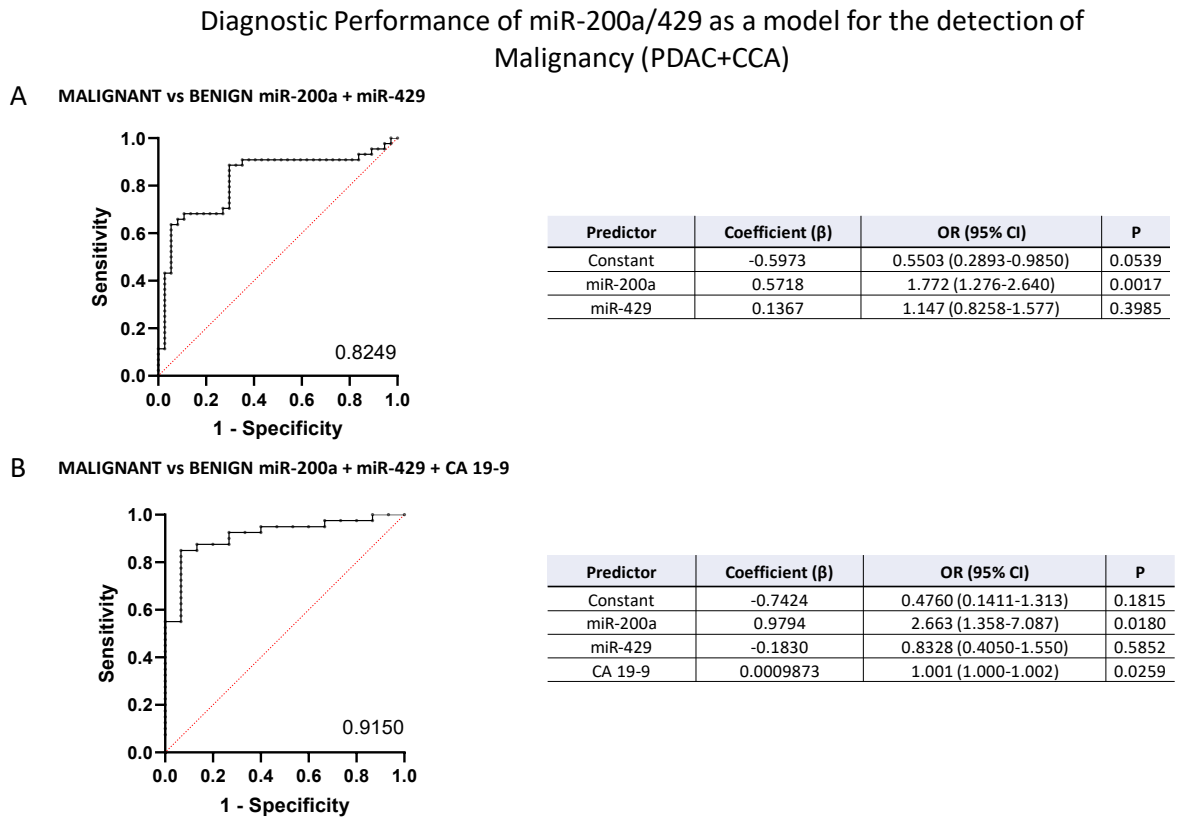
Predictor	Coefficient (β)	OR (95% CI)	P
Constant	-1.448	0.2350 (0.08729-0.5126)	0.0011
miR-200a	0.5296	1.698 (1.130-2.763)	0.0162
miR-429	0.06039	1.062 (0.7236-1.546)	0.7447
miR-320a	-0.4394	0.6444 (0.4186-0.9232)	0.0264

Figure 51: Regression analyses to validate bile EV-miRNA biomarker model (CCA; n=54).
ROC curves were generated using RT-qPCR validation expression data and analysed by multiple logistic regression for each of the 3 candidates statistically overexpressed in the pairwise comparison CCA vs Benign.

A. Each individual miRNA (miR-200a-3p, miR-429 and miR-320a).
B. Combined 3-marker panel with validated AUC = 0.789 (95% CI 0.6573 - 0.9198, p=0.0007).

CCA – Cholangiocarcinoma, AUC – Area Under ROC Curve, ROC - Receiver operating characteristic.

Finally, we combined our results, taking the 2-markers that were cross-validated in both pairwise comparisons panel (miR-200a, miR-429) to evaluate its value in detecting Malignant vs Benign. The ROC curve can be seen in FIGURE 52A and has a validated AUC of 0.8249 (95% CI 0.7297 - 0.9202, p<0.0001). This was combined with the CA 19-9 (FIGURE 52B) to give an AUC of 0.9150 (95% CI 0.8362 - 0.9938, p<0.0001). In both analyses, miR-200a expression retains its significance (p<0.05). At the threshold determined by the maximum Youden Index (>0.6599) the combined model predicts cancer with a sensitivity of 78.3%, a specificity of 93.3% and a PPV of 97.14%.



Predictor	Coefficient (β)	OR (95% CI)	P
Constant	-0.5973	0.5503 (0.2893-0.9850)	0.0539
miR-200a	0.5718	1.772 (1.276-2.640)	0.0017
miR-429	0.1367	1.147 (0.8258-1.577)	0.3985

Predictor	Coefficient (β)	OR (95% CI)	P
Constant	-0.7424	0.4760 (0.1411-1.313)	0.1815
miR-200a	0.9794	2.663 (1.358-7.087)	0.0180
miR-429	-0.1830	0.8328 (0.4050-1.550)	0.5852
CA 19-9	0.0009873	1.001 (1.000-1.002)	0.0259

Figure 52: Evaluation of a 2 marker model (miR-200a and miR-429) by regression analysis from bile EVs (Malignant vs Benign; n=81).

ROC curves were generated utilizing the RT-qPCR validation data for 2 EV-miRNAs found to be upregulated (miR-200a-3p and miR-429) using predetermined thresholds at maximum Youden Index in the pairwise comparison Malignant vs Benign.

A. 2 marker panel alone with validated AUC = 0.8249.
B. 2 marker panel in combination with CA 19-9 with validated AUC = 0.9150.

AUC – Area Under ROC Curve, ROC - Receiver operating characteristic.

4.2.19 tRNAs in bile EVs as a potential future source of biomarkers.

As tRFs comprised a read percentage of 25% (FIGURE 34) it was deemed worth investigating for candidates suitable for future biomarker analysis to determine malignancy. This was performed as part of the discovery cohort that went for sequencing and principal component analysis was undertaken to investigate correlation of different samples according expression values, (FIGURE 53A+B) and shows grouping of our cohorts. CCA and PDAC samples (green and blue respectively) are shown as one colour to the right and can be seen to roughly group together.

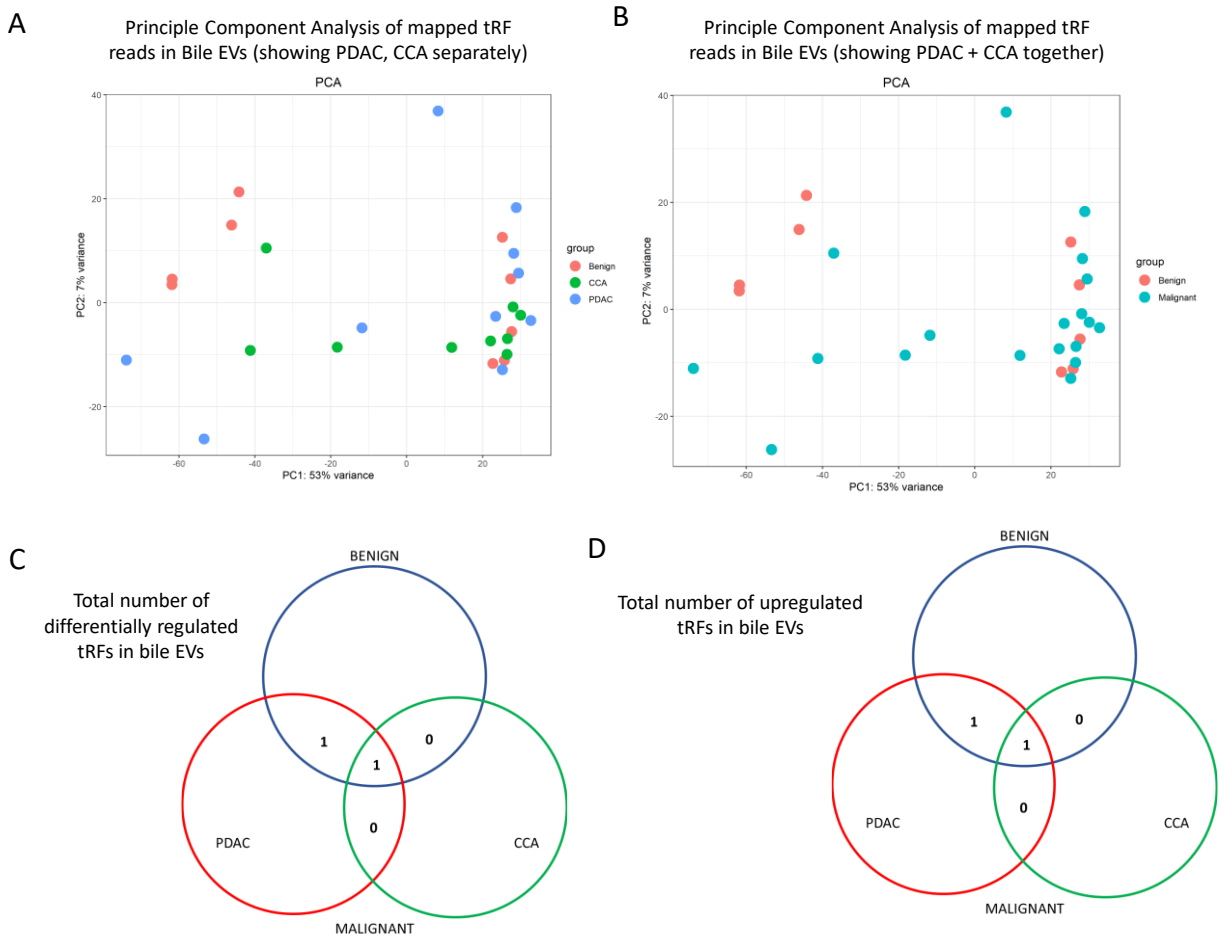


Figure 53: Differential expression analysis for tRFs in MIRABILE bile EVs from RNA sequencing data (n=30; 10 PDAC, 10 CCA, 10 Benign).

A. Principal component analysis (PCA) plot showing broad grouping of the samples by cohort according to tRF expression (PDAC vs CCA vs Benign)
B. PCA with grouping of Malignant (PDAC+CCA) vs Benign according to tRF expression.
C. Venn diagram showing the total number of differentially expressed tRFs between PDAC, CCA and benign cohorts with P value (adjusted) <0.05 in bile.
D. Venn diagram showing the total number of upregulated tRFs only with P value (adjusted) <0.05 in bile.

PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, tRFs – Transfer RNA fragment

After filtering, the small RNA sequencing data for bile EVs included expression of 10,962 tRFs but after adjustment using Benjamini–Hochberg methodology to control the False Discovery Rate (FDR), disappointingly we identified only one differentially expressed tRF between Malignant vs Benign. This single candidate was also found to be differentially expressed in the PDAC vs Benign comparison. FIGURE 52C shows the overall number of statistically significant results (after adjustment) and FIGURE 52D shows the number upregulated in our cancer samples. Once again, no candidates were shown to be significantly differentially expressed between PDAC and CCA and this was also the case for CCA vs Benign.

FIGURE 54 shows volcano plots for the pairwise comparisons. FIGURE 54A shows that tRNA-Leu-TAG_5tRF.7 (tRF-18-RP9830D4) was significantly expressed but was found to be upregulated in our Benign samples. As previously mentioned, although this could be a marker that might have functional importance, the focus was on upregulated markers that would be suitable as a candidate for the diagnosis of cancer.

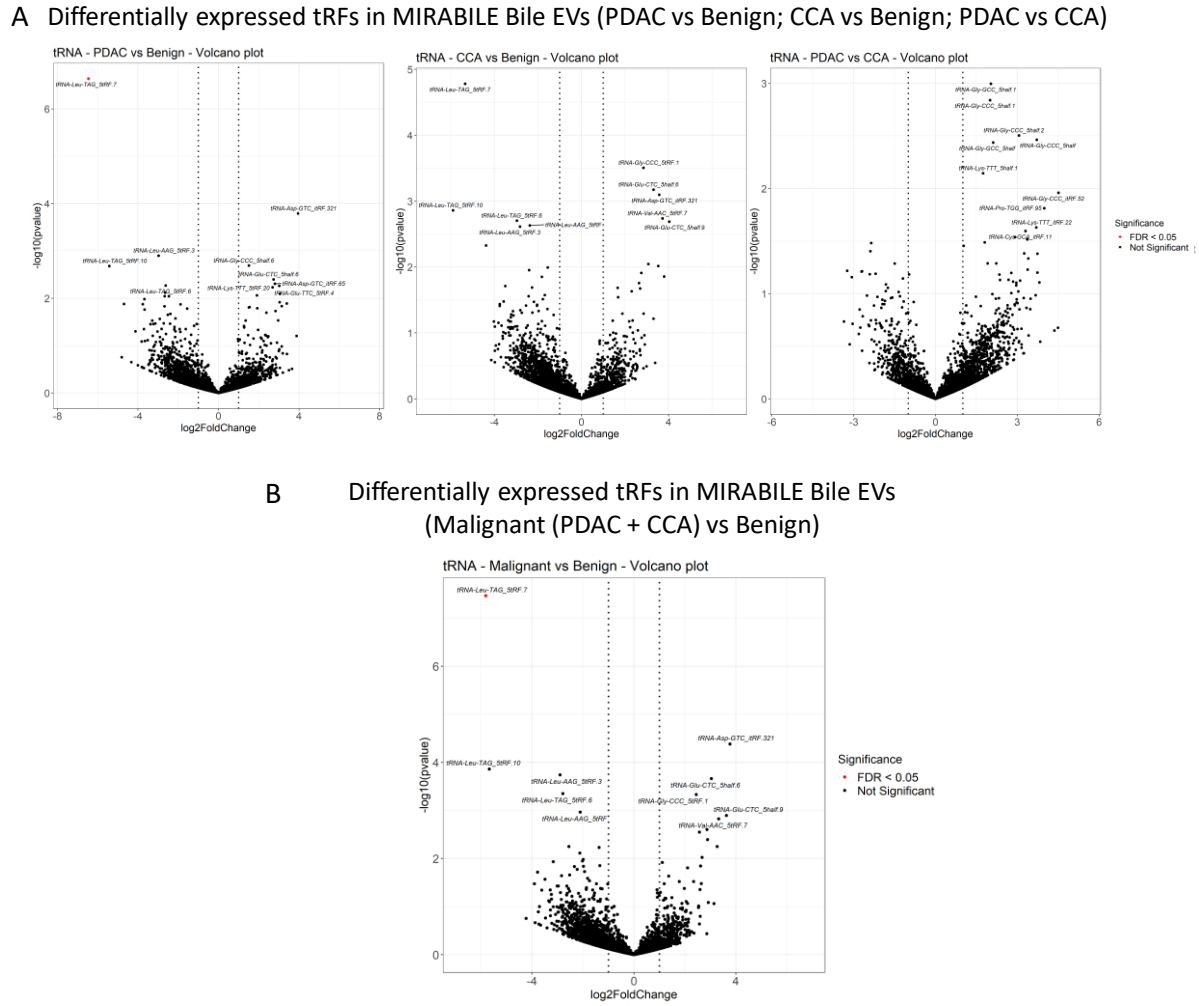


Figure 54: Volcano plots showing significant tRFs present in bile EV (MIRABILE) samples (n=30).

A. Volcano plots with miRNAs in the following pairwise comparisons (PDAC vs Benign; CCA vs Benign; PDAC vs CCA).
B. Volcano plot with top miRNAs in the pairwise comparison Malignant vs Benign.

Plotted are significance ($-\log^{10}$ p-value) vs expression ($\log^2$ Fold Change) with those candidates found to be significant highlighted in red with an FDR < 0.05. Vertical lines indicate $\log^2$ Fold Change ±1. Only the 10 most significant miRNAs are labelled for clarity.

PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, FDR – False Discovery Rate, tRF – Transfer RNA fragment

TABLE 33-35 showed the top ten tRF candidates sorted by descending log fold change in the comparisons PDAC vs Benign, CCA vs Benign and Malignant vs Benign respectively. These were not deemed significant, but four candidates were represented in both pairwise comparisons and could be potential markers for future validation. These were tRNA-Glu-

CTC_5half.9 (tRF-36-87R8WP9N1EWJQ7D), tRNA-Val-AAC_5tRF.7 (tRF-17-79MP9PP), tRNA-Glu-CTC_5half.12 (tRF-34-87R8WP9N1EWJI5) and tRNA-Glu-CTC_5half.6 (tRF-35-87R8WP9N1EWJQ7). There were no matched candidates between this and the surgical bile EV analysis (Table 24).

Table 33: Top ten tRNA fragments identified as upregulated in bile EV biomarker analysis (PDAC vs Benign).

Top ten tRF candidates identified in the comparison PDAC vs Benign (top) in bile EVs, sorted by descending log fold change values (upregulated in PDAC samples). Expression averaged across all samples with normalization by DESeq2 to allow direct comparison between the samples. Standard errors are of the log2fold change and p-values are adjusted using the Benjamini-Hochberg method. tRFs in bold are represented in both PDAC vs Benign and CCA vs Benign analyses. PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, tRFs – Transfer RNA fragment.

tRNA Fragment	MINTbase Unique ID	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
Asp-GTC_itRF.321	tRF-32-L85DMKYUYRLH4	10.331	3.7666	0.9191	0.2279
Glu-CTC_5half.9	**tRF-36-87R8WP9N1EWJQ7D**	**10.721**	**3.6288**	**1.1259**	**0.9988**
Val-AAC_5tRF.7	**tRF-17-79MP9PP**	**3.1090**	**3.3247**	**1.0473**	**0.9988**
Glu-CTC_5half.12	**tRF-34-87R8WP9N1EWJI5**	**2.7686**	**3.2650**	**1.1798**	**0.9988**
Gly-CCC_itRF.35	tRF-21-62K876IRD	1.7450	3.1426	1.8359	0.9988
Glu-CTC_5half.6	**tRF-35-87R8WP9N1EWJQ7**	**48.942**	**3.0345**	**0.8206**	**0.4765**
Glu-CTC_itRF.145	tRF-33-0VL8K87SIRMMY	1.2237	2.9187	1.6813	0.9988
Val-CAC_itRF.93	tRF-16-QB1MK8E	4.8851	2.8878	1.0044	0.9988
Glu-TTC_5tRF.4	tRF-23-86J8WPMNY	4.6130	2.8650	0.9476	0.9988
Asp-GTC_3half.28	tRF-37-8HM2OSRNLNKSEKL	1.2856	2.8595	3.1755	0.9988

Table 34: Top ten tRNA fragments identified as upregulated in bile EV biomarker analysis (CCA vs Benign).

Top ten tRF candidates identified in the comparison CCA vs Benign (middle) in bile EVs, sorted by descending log fold change values (upregulated in CCA samples).

tRNA Fragment	MINTbase Unique ID	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
Glu-CTC_5half.9	**tRF-36-87R8WP9N1EWJQ7D**	**10.721**	**4.0265**	**1.3066**	**1**
Lys-CTT_itRF.241	NA	1.5516	3.7968	1.5441	1
Val-AAC_5tRF.7	**tRF-17-79MP9PP**	**3.1090**	**3.7232**	**1.1945**	**1**
Asp-GTC_itRF.321	tRF-32-L85DMKYUYRLH4	10.331	3.5722	1.0655	1
Glu-CTC_5half.12	**tRF-34-87R8WP9N1EWJI5**	**2.7686**	**3.5235**	**1.3615**	**1**
Asp-GTC_itRF.534	tRF-34-8HM2OSRN2NKSE4	0.6670	3.3798	3.1577	1
Val-CAC_5tRF.19	tRF-17-Q99P9PP	1.2885	3.3115	1.7666	1
Glu-CTC_5half.6	**tRF-35-87R8WP9N1EWJQ7**	**48.942**	**3.3058**	**0.9718**	**1**
Val-CAC_5tRF.2	tRF-17-79MP9PP	2.8408	3.0713	1.1764	1
Asp-GTC_3tRF.17	tRF-24-3K7SIR3DIY	0.5633	2.9873	3.0194	1

Table 35: Top ten tRNA fragments identified as upregulated in bile EV biomarker analysis (Malignant vs Benign).

Top ten tRF candidates identified in the comparison Malignant vs Benign (bottom) in bile EVs, sorted by descending log fold change values (upregulated in both PDAC and CCA samples).

tRNA Fragment	MINTbase Unique ID	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
Asp-GTC_itRF.321	tRF-32-L85DMKYUYRLH4	10.331	3.9516	1.0480	0.8916
Gly-CCC_itRF.35	tRF-21-62K876IRD	1.7450	3.8791	2.0832	1
Tyr-GTA_itRF.171	NA	1.0406	3.6530	3.5950	1
Gly-CCC_itRF.6	tRF-17-EH62K8N	1.4279	3.4943	3.5879	1
Phe-GAA_itRF.53	tRF-16-R95RM3E	0.9761	3.3992	2.3627	1
Gly-GCC_5half.15	tRF-30-PLR8YP9LONNV	1.3113	3.3894	1.3643	1
Ile-AAT_itRF.61	tRF-16-VNJNRX0	0.9889	3.2001	3.5960	1
Cys-GCA_5tRF.63	tRF-20-R2VP4PRL	0.9143	3.1878	2.1937	1
Glu-CTC_5half.9	tRF-36-87R8WP9N1EWJQ7D	10.721	3.1415	1.2882	1
Gly-CCC_itRF.3	tRF-17-2VR00S4	2.1599	3.1403	1.9348	1

4.3 Discussion

4.3.1 Bile-derived EVs are an under investigated source of small RNAs which are differentially expressed in PDAC.

Vesicles present in bile have been identified as far back as 1985 by Sömjen et al (250) who identified these as having a potential role in cholesterol transport . Since 2011, only 5 studies have previously investigated miRNAs present in bile (See 3.3.1). It is only in the last 10 years, that the RNA cargo present in these bile-derived EVs have been investigated as a source of biomarker for cancer. To date, four studies (223, 254, 319, 320) detailed in TABLE 36 have investigated their role in PDAC and CCA looking at miRNAs, concentrations, long non-coding RNAs (lncRNAs) and most recently circular RNAs (circRNAs).

Table 36: Studies identified to-date investigating bile EVs.

Literature search involved keywords including pancreatic cancer, EVs, exosomes and bile taken from the Early Stage Report submitted Friday 20th September 2019 and updated 1st November 2022.
CircRNA – circular RNA, lncRNA – long non-coding RNA, EV – extracellular vesicle, PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma

Bile EV studies						
First Author	**Year**	**No. of Patients**	**Biofluid**	**EV Study**	**Bile Biomarker**	**Significant comments**
Li et al.	2014	Controls (n=50), CCA (n=46)	Human bile	Yes	A panel of 5 microRNAs: miR-191, miR-486-3p, miR-1274b, miR-16, miR-484	Showed stability of miR expression at room temperature and across at least three freeze-thaw cycles. Comparison of different mathematical modelling was used and focussed on specificity. 5-miR panel displayed a sensitivity of 67% and specificity of 96%.
Severino et al.	2017	PDAC (n=20), CCA (n=5), Controls (CP, n=15)	Human bile	Yes	Median concentration of EVs	A threshold value of 9.46×10^{14} nanoparticles/L correctly classified all malignant samples and outperformed both serum CA19-9 and bile CEAM6.
Ge et al.	2017	Controls (n=56), CCA (n=35)	Human bile	Yes	lncRNAs - ENST00000588480.1 and ENST00000517758.1	Found upregulated in bile exosomes with greater sensitivity than CA19-9 (82.9% vs 74.3%), significant increasing trend in advanced disease and worse overall survival.
Xu et al.	2020	Controls (n=5), perihilar CCA (pCCA; n = 40), and distal CCA (dCCA; n = 40)	Human serum and bile	Yes	Hsa_circRNA_102064	Bile-resident exosomal circ-CCAC1 (synthesized from exons 23-26 of ERBB2) functioned better as a diagnostic biomarker than serum-derived EVs from patients with CCA.

Common to these studies are the size of their sample populations (~90 patients) with the predominant source of EVs in each case ultracentrifugation. The use of SEC to isolate EVs from biofluids rapidly makes the results of this study unique. As mentioned in the introduction, we demonstrate that EVs contain ~100pg/µL of RNA which from the cfRNA work done in the previous chapter (average 24.49 ng/µL) results in EV concentrations which are roughly 0.5% of the overall RNA content found in bile samples. This correlates what has been found by other studies (191) but we believe that the compartment of EVs

contains small RNAs which are both differentially expressed in cancer and may have a functional role in tumourigenesis.

4.3.2 Optimisation on Bile-derived EVs from surgery led to SEC as a novel isolation method with high throughput and reproducibility.

This study investigated SEC as a novel EV isolation method which is quick and clinically applicable for multiple samples. We investigated each 500 μL fraction eluted from a small cohort of samples (3 per cohort) and demonstrated the presence of EV associated protein, cup-shaped vesicular structures by EM and reproducible particle concentrations using NTA. The estimated turn-around time of a single sample was ~20 minutes which could be performed in batches and potentially automated. Early EV containing fractions could additionally be compared to later fractions as controls for the presence of non-EV associated protein. At present, UC remains the most popular choice for the isolation of EVs (321) but our results question the assumption that EVs can be separated efficiently into two compartments with either SEC or UC methodologies.

However, our previous work in bile cfRNA had identified that there was some diagnostic value in non-EV associated RNA with 38 upregulated candidates found (Figure 10). Further evaluation and comparison between the two compartments are needed, particularly to identify why differences in the small RNA population might be occurring and ultimately which candidates might be more clinically useful.

4.3.3 Comparison of EV characterisation between surgical and endoscopic samples

Our western blots undertaken in surgical samples (FIGURE 32C) showed that EV associated proteins ALIX were found in both EV and PROT fractions, whilst there was a much clearer separation between EV and PROT fractions in the MIRABILE EVs (comparison with FIGURE 40B showing absence of ALIX). This together with qualitative TEM indicated that the newer endoscopic bile samples contained well-preserved EVs which were potentially not lysed or during storage (and/or freeze/thaw cycles). The literature suggests that although RNA content is preserved in EVs, some loss due to freeze/thaw cycles has been reported (225). This may have contributed to the different miRNAs and tRFs identified by differential expression analysis in the two different studies. Surgical samples can be used as a source of bile EVs, but a possible confounder

for why these did not appear to reveal similar expression includes the differences in age and storage conditions.

From the surgical bile EV-miRNAs two candidates featured in all three analyses (PDAC vs Benign, CCA vs Benign and Malignant vs Benign) with miR-141-3p significantly downregulated (log2foldchange -3.4740; p=0.0008) and miR-4516 significantly upregulated (log2foldchange 6.4087; p=0.0013; TABLE 23) in the Malignant vs Benign comparison. A combined AUC from both candidates from the sequencing data was calculated as 0.736 which is reasonable but no better than CA 19-9. This was not validated by RT-qPCR but paved the way for a model to be identified from the MIRABILE Bile EV analysis.

A comparison of particle concentrations using NTA failed to find a statistical difference in concentration between Malignant and Benign bile EV samples in either the surgical bile (n=46, p=0.3323) or the MIRABILE study (n=90, p=0.6882) thus contradicting the findings of Severino et al (223). Our larger sample size compared to that original study appears to refute this as a potential marker that can be utilised for the diagnosis of malignant biliary strictures.

4.3.4 Let-7a, miR-21 and miR-200 family has been detected in Bile EVs for the first time.

Li *et al.* was the largest study to-date investigating miRNAs derived from bile EVs (254). They chose RT-qPCR miRNA microarrays using TaqMan Low-density Array cards performed on a total of 6 specimens, and from 137 expressed miRNAs, chose the top 11 miRNA species by fold expression for further analysis. Our method on the other hand, involved NGS with no *a priori* assumptions about miRNA species and analysed all differentially expressed small RNAs.

As a result for the first time, we have undertaken RNA-Seq of bile EVs in PDAC and detected the presence of differentially expressed miRNAs. The skew towards downregulation of miRNAs in Malignancy (PDAC + CCA) led to a smaller number of potential cancer biomarkers (reducing our number of candidates from 32 to 4; Figure 44). This was a decision based on only assessing for upregulated markers present in cancer. Although assessing for downregulation would be equally likely to identify cancer-associated RNA markers, we tuned our focus to highly expressed and differentiated small

RNAs that could be found by RT-qPCR in bile EVs. This identified let-7a and miR-21, as upregulated in the bile EVs of patients with PDAC ($p<0.0001$). These are well recognised miRNAs associated with PDAC and found to be upregulated in bulk PDAC tumours as well as in tumour-associated fibroblasts present in the stroma of patients with PDAC (166). MiR-21 was also highlighted by Li *et al.* but not identified as a differentially expressed candidate for diagnostic purposes. It is therefore of significant interest that we have identified potentially oncogenic miRNAs in the EVs of bile obtained from ERCP.

We proceeded to show validated differential expression of miR-200a and miR-429 ($p<0.0001$), two related miRNAs both from the miR-200 family and co-located on chromosome 1, in bile EVs from patients with PDAC and CCA. Interestingly, miR-141 which was identified in our surgical bile EVs was significantly downregulated in our cancer samples and is also a member of the miR-200 family. The miR-200 family has been recognised as playing an important role in the epithelial-to-mesenchymal transition (EMT) and has been associated with PDAC and CCA in multiple *in vitro* studies (322, 323). Szafranska *et al.* identified the miR-200 family as upregulated in macro dissected PDAC tissue compared with normal tissue in a small cohort (n=25, (169)) whilst in 2009, Zhang *et al.* further characterised 95 miRNAs in tissue and also found miR-200b as significantly upregulated in PDAC (324). A study by Kent *et al.* found that the most highly overexpressed miRNA in 21 PDAC cell lines was miR-200a and identified that enforced expression of the miR-200b cluster, miR-200a, and the miR-17-92 cluster together resulted in significantly accelerated proliferation but not miR-200b or miR-200a alone (322). This family of miRNAs is believed to be a tumour suppressor in cancer tissue, leading to inhibition of the EMT pathway is via a negative feedback loop on transcription factors ZEB1 and ZEB2 (325). Interestingly, expression of the miR-200 family is found to be increased in a number of solid cancers (breast, prostate, oesophageal, gastric, ovarian, and metastatic colorectal cancers) and associated with changes in overall survival (OS), being negatively correlated with survival when present in blood plasma or serum (326) (further discussion in Chapter 6).

4.3.5 Diagnostic potential of these as a panel and in combination with CA 19-9

Four candidates (let-7a-5p, miR-21-5p, miR-200a-3p, and miR-429) were combined in our PDAC vs Benign validation to form a subset of EV-miRNAs which could detect PDAC. The ROC curve shown in FIGURE 50B demonstrates that this four-marker panel resulted

in a validated AUC of 0.961. miR-21-5p demonstrated an individual AUC of 0.9560 (95% CI 0.9049 - 1.000, p<0.0001) whilst miR-200a-3p demonstrated an AUC of 0.9019 (95% CI 0.8291 - 0.9747, p<0.0001). Using this combined panel, the optimum cut-off was determined (>0.4397) which resulted in a sensitivity 70%, and specificity 87.1%. Our model for CCA was less successful with the model (miR-200a, miR-429, miR-320a) showing a validated AUC of 0.7886. Once again miR-200a-3p demonstrated the highest individual statistic with an AUC of 0.7170 (95% CI 0.5440 - 0.8900, p=0.0110).

We finally combined our results into a 2-marker panel (miR-200a, miR-429) to evaluate its value in detecting hepatobiliary malignancy. The ROC curve can be seen in FIGURE 52 which has a validated AUC of 0.8249. This was combined with the CA 19-9 to give an AUC of 0.9150. In both analyses, miR-200a expression retains its significance (p<0.05) indicating that it is the most relevant miRNA in this model. At the threshold determined by the maximum Youden Index (>0.6599) the combined model predicts cancer with a sensitivity of 78.3%, a specificity of 93.3% and a PPV of 97.14%. Our aim was to identify a biomarker with an ideal test sensitivity of 95% for a bile-based miRNA signature and although this was not quite achieved, we were able to show additional improvement when used in conjunction with currently existing markers (CA 19-9). The next step would be to prospectively measure expression of these markers (i.e. bile EV-let-7a-5p/miR-21-5p/miR-200a/miR-429) in patients with diagnostic uncertainty. By performing a study which then monitors survival we would hopefully be able to indirectly assess whether these EV miRNAs are correlated with tumour burden.

In a real-world scenario however, it is uncertain how useful a bile-based marker will be, unless it is extremely specific or sensitive, there is an element of morbidity associated with the test. This test may be most beneficial to patients who by necessity have an endoscopic procedure, for example those with indeterminate biliary strictures and jaundice, and those with PSC or an obstructing pancreatic mass for whom inflammation may be a differential diagnosis. Since our cohort includes patient with chronic pancreatitis (CP), this marker panel should be applicable in this cohort of difficult to manage patients as a potential indicator for further surgery.

4.3.6 Novel tRFs may be a future promising source of biomarkers and are found present in bile EVs for the first time.

Finally, it is worth noting that this was the first study to have analysed for differential expression the presence of tRFs within bile-derived EVs. A smaller study involving the RNA-Seq of bile exosomes was conducted by Ge *et al.* and comprised 82 patients. investigating the difference in exosomal RNA between patients with malignant and benign biliary obstruction (319). They focussed on long non-coding RNAs present in the exosomes of patients with PDAC, CCA, and Ampullary carcinoma patients and interestingly also found that tRNAs comprised the most mapped small non-coding RNAs with 27% compared with 24% for miRNAs. In our samples we found that tRFs comprised 10.416% of overall reads compared with an average of 8.525% for miRNAs in our discovery sequencing.

The limitations of this study include an issue with the heterogeneity of bile-derived EV-miRNA composition. This precluded all the discovery samples from being analysed and 2 samples were excluded on the basis of low RNA input and thus poor discrimination from the rest of the samples when differentially analysed. With the improvements in sequencing technology and speed, it is hoped that small input will not be a barrier to future biomarker analysis. Our clinical data suggests that there remain significant differences between our cancer and non-cancer cohorts including degree of biliary obstruction and age, which remain a problem throughout the study (and a major feature of others). There remains also the question of the functional significance of these biomarkers and their overall role in PDAC/CCA which was not able to be addressed in this current project. It is hoped that with further evidence for these as upregulated, future *in vitro* work may be able to determine the role of miR-200a/miR-429 in pancreatic cancer.

Chapter Five – Small RNA markers for the detection of PDAC in Plasma Extracellular Vesicles

5.1 Background

Plasma is an extracellular component of the circulating volume and absent of blood cells and contains proteins, lipids, simple molecules as well as nucleic acids. It is also rich in extracellular vesicles (EVs) which are released into the circulation presenting a source of biomarkers for analysis. Bile-derived EV-miRNAs have been shown to provide good diagnostic potential in the detection of PDAC and CCA but the collection of this by endoscopy is invasive and therefore limited in availability. A blood-based biomarker allows for easier testing as well as repeatability leading to potential longitudinal analysis during oncological treatment of a patient with PDAC or CCA. The "MIcroRNAs as BILE-based biomarkers in Pancreaticobiliary Cancers" (MIRABILE) study collected plasma samples contemporaneously with their ERCP procedure which allowed paired samples of bile and blood for additional analysis of patients at a unique timepoint prior to intervention.

The aims of the third part of this project were to investigate EV-miRNAs from plasma in patients with a biliary stricture. Once again this took place in three broad stages. First was (1) the isolation and characterisation of EVs from the blood plasma of patients with benign and malignant biliary strictures (i.e. pancreaticobiliary cancers) followed by optimisation of the protocol. The second part (2) identified differentially expressed EV-miRNAs (and transfer RNA derived fragments or tRFs) in plasma suitable for detecting pancreaticobiliary malignancy. The final stage of this process involved further longitudinal analysis to (3) determine potential prognostics in relation to pancreatic cancer in samples obtained pre- and post-surgery.

5.2 Results

5.2.1 Samples used for plasma EV isolation.

All patients (see 2.1.2 Sample Collection for the MIRABILE study) attended the endoscopy unit at Imperial College NHS Trust (Hammersmith Hospital, London, UK) between 2nd November 2017 and 11th March 2020. Additional samples for longitudinal analysis were obtained from patients at the Royal Surrey NHS Foundation Trust (Guildford, UK). The fasting samples were obtained either via a freshly inserted 18-22G peripheral cannula or 21-23G butterfly syringe into BD K2 Ethylenediaminetetraacetic acid (EDTA) lined tubes and processed within 4 hours. Blood was centrifuged at 2,500 x g for 10 minutes at room temperature to remove cells and the resulting platelet-free plasma was aliquoted and samples used immediately or frozen at -80°C until required.

5.2.2 Study design for the analysis of plasma EVs

Optimisation of EV isolation using the SEC protocol took place using 6 samples (3 PDAC and 3 Benign). This used a similar protocol to that used for bile with the exception that there was no requirement to process the samples by centrifuge before EV isolation (to remove cellular debris). 1mL samples of plasma were thawed on ice, then brought to room temperature and loaded onto a qEVoriginal 70nm SEC column, followed by elution with PBS. The eluate was initially collected in 30 sequential fractions of 0.5 mL (labelled from 1 to 30), which were used immediately for EV characterisation and RNA extraction or frozen at -80°C until required. Subsequent sequencing and validation took place with pooled EV samples. The overall workflow can be seen in FIGURE 55.

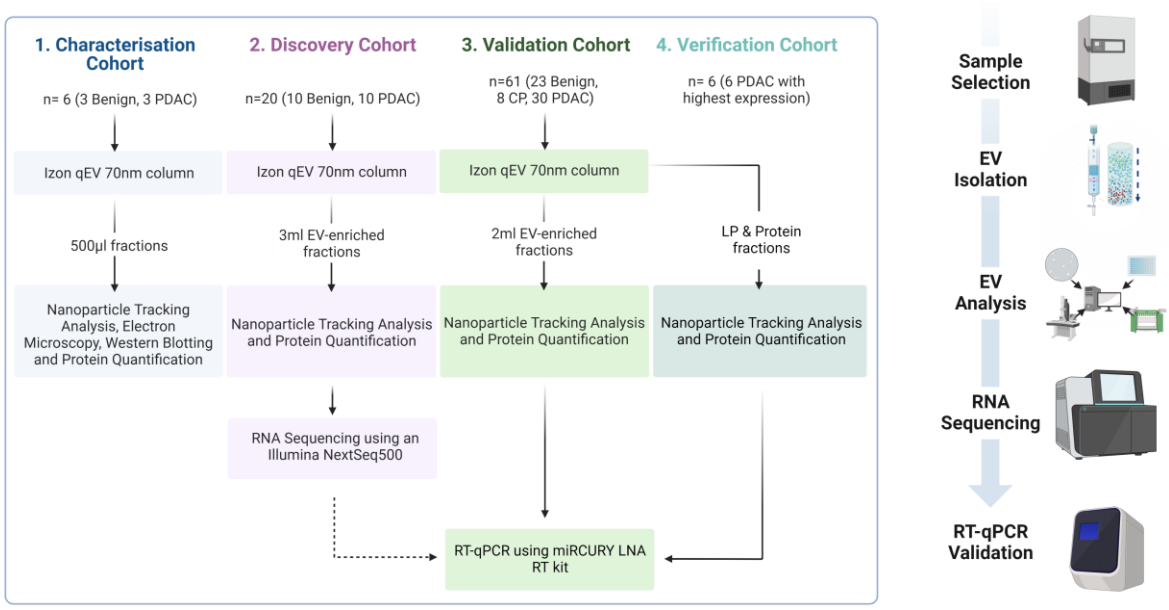


Figure 55: Study Design for the evaluation of EVs from plasma samples.
In contrast with Chapter 4 Figure 42, the EV workflow now includes additional steps which involved a final EV verification step after validation of samples. The study design for optimization involved establishing EV characteristics according to MISEV guidelines using Nanoparticle Tracking Analysis, Transmission Electron Microscopy and Protein analysis (Western blotting and quantification) in 6 samples followed by sequencing of 20 samples comprising of PDAC and Benign cohorts only.
PDAC – Pancreatic ductal adenocarcinoma, EV – Extracellular Vesicle, MISEV – Minimal Information for the Study of Extracellular Vesicles.

5.2.3 Characterisation of EVs present in plasma samples obtained prospectively (MIRABILE)

Transmission electron microscopy (TEM) was undertaken to enable visualisation of vesicles and FIGURE 56A shows the presence of ~100nm bilayered structures with negative staining using uranyl acetate in a PDAC plasma sample which underwent SEC isolation. As previously, four fractions were compared for the presence of "cup-shaped" particles and extracellular protein. Qualitatively, it could be noticed that there were significantly higher amounts of "extra-vesicular" structures seen in the wide field views (5000x magnification) compared to that in bile samples. These structures appear to have a single outer layer and are particularly present from fraction 11 onwards. It was considered likely that these were lipoproteins which are present in blood plasma in high concentrations. The same protein aggregates noticed in our bile samples were also noted at fractions 13 onwards which obscured detection of EVs.

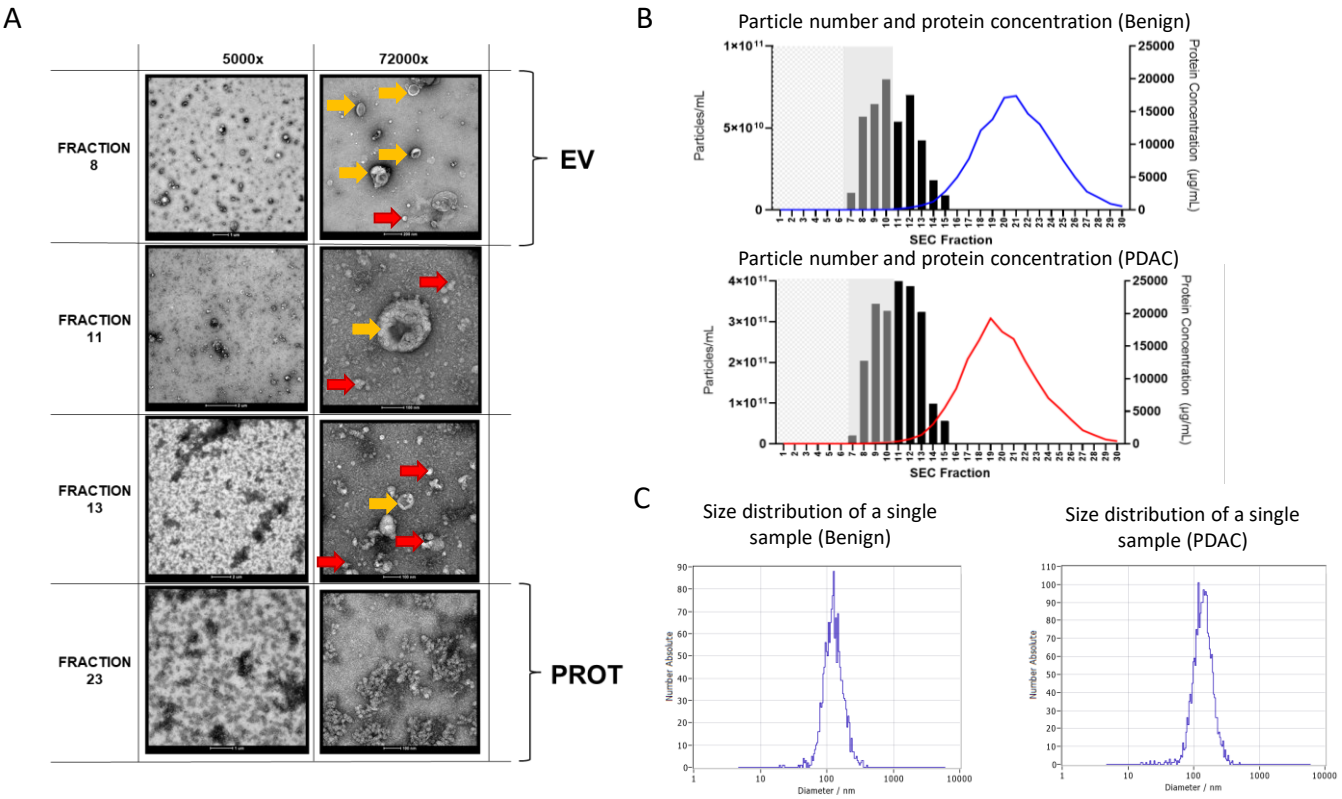


Figure 56: Characterisation of individual and pooled EV fractions using TEM and NTA of the MIRABILE plasma samples (n=6).

A. Shape and structure of plasma EVs isolated in a PDAC sample by SEC are shown here using narrow field (72000x) and wide field (5000x) electron microscopy. 8 µL samples were deposited onto 200-300 mesh glow-discharged copper grids and negatively stained with 2% uranyl acetate. The grids were observed using a Tecnai T12 TWIN transmission electron microscope at magnification ranges 4,000-120,000x. Highlighted in yellow are likely EVs with a lipid bilayer. In red are smaller structures likely extravesicular contamination e.g. protein or lipoprotein. As the fraction number increases so does the protein content shown in the background until EVs are poorly resolved at Fraction 20.

B. Average particle and protein concentrations measured by NTA and BCA for each fraction in each cohort (3 benign - above, 3 PDAC - below). Shaded region shows fractions 7 to 10 which were pooled for subsequent RNA extraction and analysis.

C. Individual size distribution plots from the EV pooled fractions sample analysed during characterisation by NTA show EV particle size (concentration measured at 1:2000).

BCA – Bicinchoninic acid assay, EVs – extracellular vesicles , PDAC – Pancreatic ductal adenocarcinoma, NTA – Nanoparticle tracking analysis.

NTA was undertaken for each fraction alongside protein quantification and average results for 3 benign samples (above) and the average of 3 PDAC samples (below) can be seen in FIGURE 56B. This showed that there was a possible bimodal distribution with a peak at 9 and again at 12 which could also be attributed to the presence of lipoprotein particles with a similar size. Size distribution using NTA of an individual Benign sample (above) and PDAC sample (below) can be seen in FIGURE 56C demonstrating a peak size of just over 100nm.

5.2.4 Western blotting of EVs present in plasma samples obtained prospectively (MIRABILE) highlights specific differences between bile and plasma derived EVs.

Western blotting was undertaken to determine the protein content of each individual 500 μL fraction and was significantly easier than for bile-derived EVs due to the increased amounts of protein found (approximately 10-fold higher concentrations). Initial blotting was undertaken by Flora Upton (MRes student) (FIGURE 57A) was done for fractions 11 through to 22 due to difficulties in obtaining sufficient protein for analysis from the pure EV fractions (7-10). This showed similar early expression of CD63, CD9, TSG 101 and HSP 70, EV-associated proteins, which unexpectedly persisted into the later fractions. APOA-1 which was used as a marker for lipoproteins appeared in fraction 12 onwards, whilst Calnexin appeared from 13 onwards (a component of the endoplasmic reticulum). The blue marked line indicated the original defined end of the EV fraction as determined previously with bile.

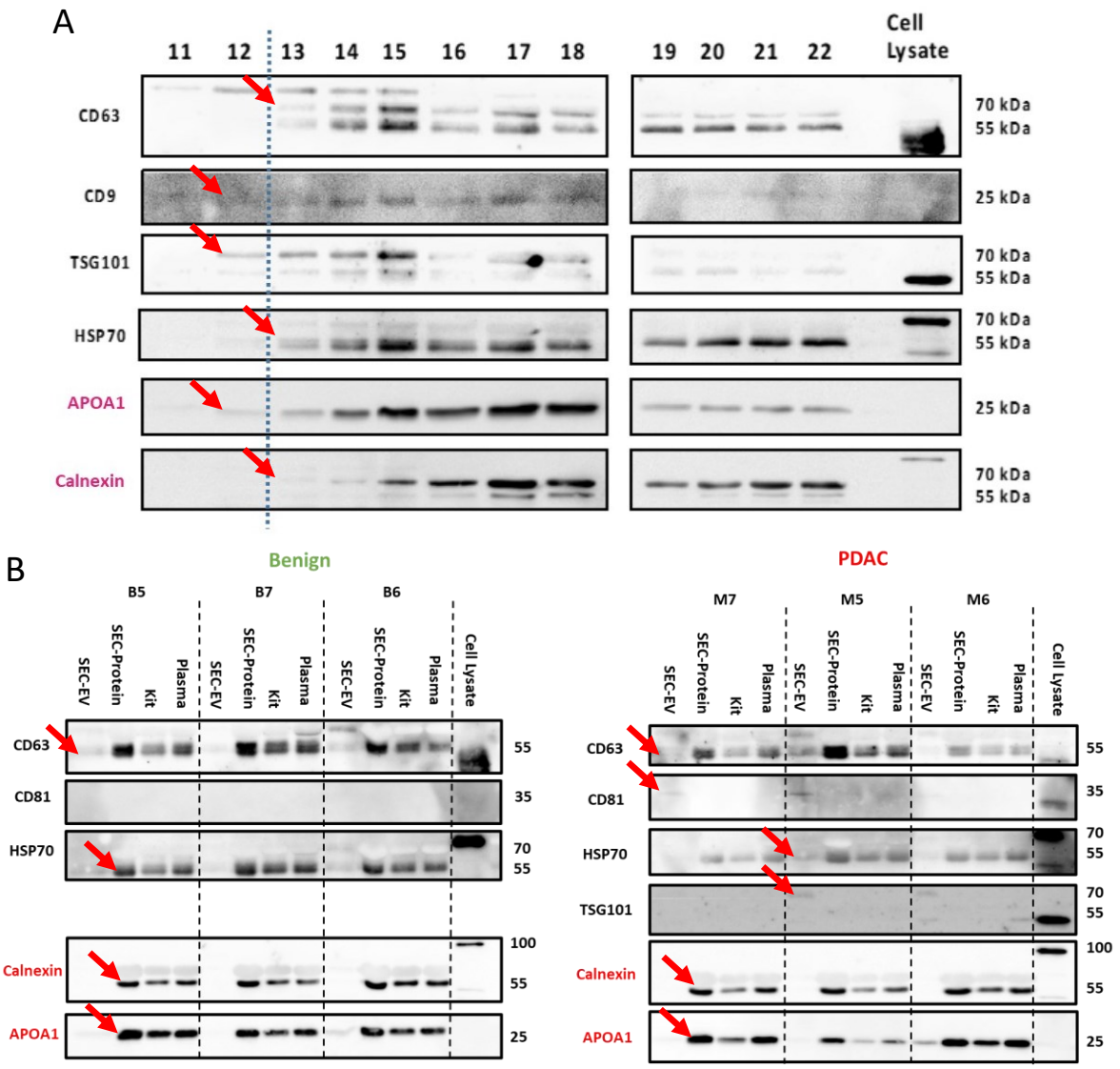


Figure 57: Characterisation of individual and pooled EV fractions using western blotting of the MIRABILE plasma samples (n=6).

A. Western blot of individual fractions (11 to 22) analysed for expression of TSG101, CD9 CD63, HSP 70 and negative markers APOA1, Calnexin. 20 µg of protein was loaded into each well and compared (all analyses here performed by Flora Upton (MRes). Red arrows show relevant protein bands with APOA1 and Calnexin highlighted as proteins not expected in EVs.

B. Western blot of pooled samples to compare EVs isolated from kits and unprocessed plasma in benign (B5, B6, B7) and PDAC (M5, M6, M7) samples. SEC-EV included fractions 7-12 whilst SEC-Protein was fractions 20-22. Red arrows show relevant protein bands with APOA1 and Calnexin highlighted as proteins not expected in EVs. Cell lysate was obtained from lysis of cultured HPNE.

APOA1 – Apolipoprotein A1, CCA – Cholangiocarcinoma, CD – Cluster of Differentiation, EV – Extracellular Vesicle, HPNE – Human pancreatic epithelial cells, HSP 70 – Heat Shock Protein 70, PDAC – Pancreatic ductal adenocarcinoma, SEC – Size Exclusion Chromatography, TSG101 – Tumour susceptibility gene 101

Additional work was undertaken to try to definitively prove EV-associated protein expression and we also investigated the use of a kit isolation system, the Plasma/Serum Exosomes Purification and RNA Isolation Mini Kit from Norgen BioTek (Ontario, Canada). This utilised a precipitation method which contained a proprietary slurry that pulls down EVs directly from plasma samples. FIGURE 57B shows the results of three benign samples (B5, B6, B7) on the left and three PDAC samples (M5, M6, M7) on the

right, alongside expression o. Comparison was made between these (labelled “kit”) alongside fractions 7-12 (SEC-EV), fractions 20-22 (SEC-Protein) and unprocessed plasma. As can be seen, whilst SEC-Protein can be seen as most similar in expression to unprocessed plasma, the kit EV samples did not show any differential expression and was deemed a poor EV isolation method as it co-purified extracellular protein and lipoprotein alike.

FIGURE 58 shows the same six samples undergoing comparative analysis to determine the presence of ALIX, CD9, and Glypican-1, a marker that has previously been shown to be upregulated in PDAC plasma EVs. FIGURE 58A is a blot of 3 PDAC samples that showed expression of ALIX, CD9 and GPC-1, EV-associated proteins. However, it also showed increased expression of APOA-1 which could indicate lipoprotein co-precipitation. By decreasing the EV fraction to only 7-10, we were able to demonstrate in three further samples (B1, B2 and B3) that EVs containing ALIX and TSG 101 were only present in this smaller fraction (FIGURE 58B) and that this could demonstrate separation of lipoproteins which was increasingly expressed in fractions 11-12 and onwards.

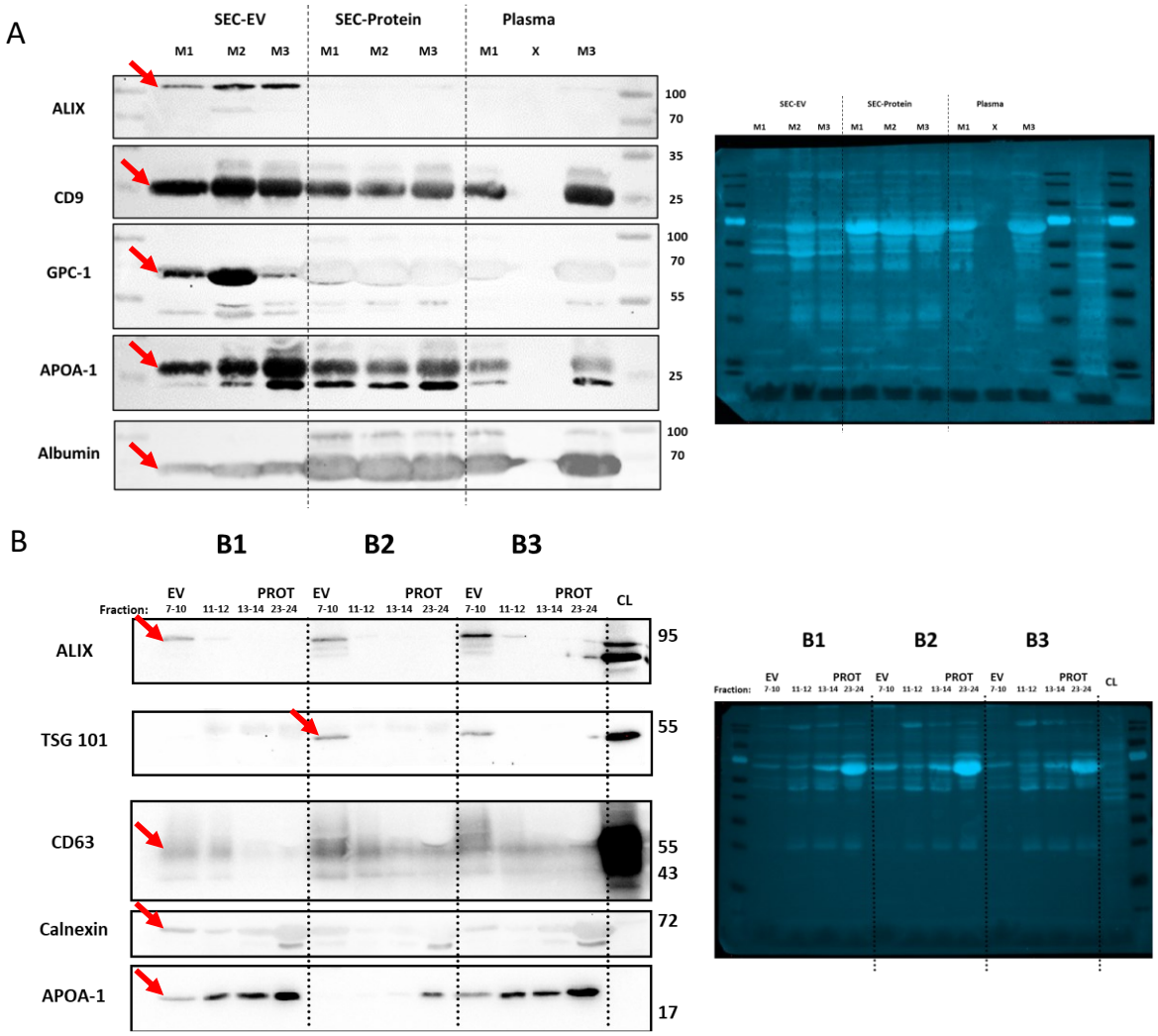


Figure 58: Additional western blotting of SEC EV pooled fractions in plasma samples (n=6).

A. 15 µg of protein from pooled EV fractions (7-12) were compared against pooled Protein fractions (20-22) and unprocessed plasma in 3 PDAC samples. The dotted line marks the different fraction types. Artefacts are present due to incomplete primary antibody stripping and show over exposure due to large variations in protein levels from individual EV samples.

B. Smaller pooled EV fractions 7-10 are compared against fractions 11-12 and "intermediate" fractions (13-14) as well as protein fractions (23-22) isolated from 3 Benign plasma samples. CL was obtained by lysis of HPNE cells.

Also shown are total membrane protein levels for normalization purposes (No-Stain Protein Labelling Reagent, Invitrogen). APOA1 – Apolipoprotein A1, CD – Cluster of Differentiation, EV – Extracellular Vesicle, HPNE – Human pancreatic normal epithelial cell line, HSP70 – Heat Shock Protein 70, SEC – Size Exclusion Chromatography, TSG101 – Tumour susceptibility gene 101.

In summary, western blotting of plasma EVs was able to demonstrate the co-precipitation of extracellular contaminants and failed to demonstrate expression of important EV-associated proteins when derived from a Norgen proprietary EV isolation kit. SEC EV fractions 7-10 however did contain EV-associated proteins, notably ALIX, TSG101 and GPC-1 which is a PDAC specific marker. The results highlighted that inclusion of fractions 11 onwards raised the risk of co-precipitating lipoproteins which could be identified with antibodies against APOA-1, and therefore necessitated an alteration in

the EV isolation protocol for SEC EV isolation in plasma which was taken forward into the study design.

5.2.5 Clinical characteristics for the patients comprising the MIRABILE plasma EVs analysis.

After EV characterisation was complete, we divided the samples into two cohorts for the purpose of discovery and validation of potential RNA biomarkers. The details of the patient samples selected for each cohort for the analysis of RNA are summarised in TABLE 37. A total of 67 unique patient samples were used for this study with 20 samples taken forward for RNA sequencing, split evenly between PDAC and Benign. The validation cohort comprised of 61 samples in total (30 PDAC, 31 Benign) and underwent RT-qPCR. It was decided not to undertake any analysis of patients with CCA for the plasma EVs to simplify analysis and focus on markers that would be relevant to PDAC as we had failed in previous studies to find significant differences in expression between the two diseases.

Table 37: Patients selected for the discovery and validation cohorts used in the plasma EV biomarker analysis.

Overall demographic characteristics and recorded pathology results were collected retrospectively and analysed by Fisher's exact test for the discovery and validation cohorts as well as overall. PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, PSC – Primary sclerosing cholangitis

Table. Demographic and Pathological Characteristics of the Patient Cohorts comprising the Plasma EV study

Patient Characteristics		Total (n=67)				Discovery (n=20)				Validation (n=61)			
		PDAC	CCA	Benign	P value	PDAC	CCA	Benign	P value	PDAC	CCA	Benign	P value
All cases (n, %)		**31**	**0**	**36**		**10**	**0**	**10**		**30**	**0**	**31**	
Age, *y*	<60	6 (19)	-	13 (36)		1 (10)	-	5 (50)		6 (20)	-	10 (33)	
	≥60	25 (81)	-	23 (64)	0.1763	9 (90)	-	5 (50)	0.1409	24 (80)	-	21 (67)	0.3844
	Mean (SD)	69 (11)	-	63 (16)	0.1194	74 (12)	-	52 (16)	**0.0029**	69 (12)	-	66 (14)	0.4643
Gender (%)	Female	9 (29)	-	12 (33)		3 (30)	-	6 (60)		9 (30)	-	9 (29)	
	Male	22 (71)	-	24 (67)	0.7946	7 (70)	-	4 (40)	0.3698	21 (70)	-	22 (71)	1.0000
History of PSC (%)		0 (0)	-	1 (3)	1.0000	0 (0)	-	1 (10)	1.0000	0 (0)	-	0 (0)	1.0000
Chronic Pancreatitis (%)		2 (6)	-	8 (22)	0.0927	0 (0)	-	0 (0)	1.0000	2(7)	-	8 (26)	0.0807
Pre-diagnostic blood parameters*	CA 19-9 (U/mL)	7128.1	-	285.1	**<0.0001**	4090.1	-	18.5	**0.0040**	7371.3	-	303.2	**0.0256**
	Bilirubin (ng/mL)	226.4	-	46.6	**<0.0001**	177.2	-	24.0	**<0.0001**	229.4	-	51.4	**<0.0001**
	CRP (µg/mL)	55.2	-	26.3	0.0561	51.6	-	14.6	0.2370	56.8	-	28.1	0.1479
Anatomical location	Perihilar CCA	-	-	-		-	-	-		-	-	-	
	Distal CCA	-	-	-		-	-	-		-	-	-	
	Intrahepatic CCA	-	-	-		-	-	-		-	-	-	
	Head of Pancreas PDAC	25 (81)	-	-		7 (70)	-	-		24 (80)	-	-	
	Uncinate PDAC	3 (10)	-	-		1 (10)	-	-		3 (10)	-	-	
	Mixed PDAC (HOP/Uncinate)	3 (10)	-	-		2 (20)	-	-		3 (10)	-	-	
Tumour stage (%)	T1	0 (0)	-	-		0 (0)	-	-		0 (0)	-	-	
	T2	6 (19)	-	-		0 (0)	-	-		6 (20)	-	-	
	T3	4 (13)	-	-		0 (0)	-	-		4 (13)	-	-	
	T4	21 (68)	-	-		10 (100)	-	-		20 (67)	-	-	
Nodal disease (%)		20 (65)	-	-		5 (50)	-	-		20 (67)	-	-	
Metastatic disease (%)		11 (35)	-	-		6 (60)	-	-		11 (37)	-	-	

Statistical significance was determined using Fisher's exact test for categorical variables (Sex, Age >60) whilst an unpaired two-tailed Student t-test or Mann Whitney U test using Malignant (PDAC + CCA) vs Benign as the pairwise comparison. P-values <0.05 highlighted in bold. Key: HOP, Head of Pancreas; PDAC, pancreatic ductal adenocarcinoma; CA 19-9, Carbohydrate antigen 19-9; CRP, C-Reactive Protein.

*Blood results where available from patient electronic hospital records.

Comparison of the demographic data showed that there were no significant differences between the PDAC and Benign populations in the Validation cohort and overall. Within the discovery cohort however, there did appear to be a statistical difference for mean age (p=0.0029) but not for sex and this could be due to the relatively small sample size. In general, there remained an older population making up patients who present with PDAC compared with the Benign population.

Looking at the overall study, the PDAC cohort included n=31 patients of which 29% were female and 71% were male. The mean age was 69 years with a standard deviation of 11 years. The Benign controls included n=36 patients of which 33% were female and 67% were male. The mean age was 63 years with a standard deviation of 16 years.
The mean CA 19-9 in the PDAC cohort was 7128 U/mL, whilst in the Benign cohort was only 285.1U/mL. The mean serum bilirubin level in the PDAC cohort was 226.4ng/mL, and in the Benign cohort was 46.6ng/mL. The mean CRP in the PDAC cohort was 55.2µg/mL whilst in the Benign the mean CRP was 26.3µg/mL.

Differences were identified in the pre-therapeutic bilirubin levels (P value <0.0001) with them significantly raised in the PDAC samples within all cohorts whilst CRP levels were found to be no different across all patients. CA19-9 was also significantly raised in all cohorts. Examining the cancer phenotypes of the overall population analysed showed that most cancers causing biliary obstruction anatomically were located in the head of pancreas (81%) with the remainders Uncinate (10%) or a mixture of the two radiologically (10%). Tumour staging was predominantly T4 (63%) as previously noted, with a small proportion T2 (19%). A high degree of nodal disease was present in all cohorts (65%) and metastatic disease was present in 35% of patients.

5.2.6 RNA isolation from plasma EVs shows differential expression of miR-200 family.

RNA quantification was done after extraction from pooled EV fractions using automated electrophoresis using an Agilent Bioanalyzer to determine the presence of RNAs suitable for sequencing. A sample from each cohort (Benign and PDAC) is shown in FIGURE 59A. Plots again show a predominance of small RNAs <200nts with concentrations ~100pg/µL. RNA distribution by stacked bar plots are shown in FIGURE 59B. Two samples were found to have low miRNA composition (2 PDAC) and additionally were seen as outliers

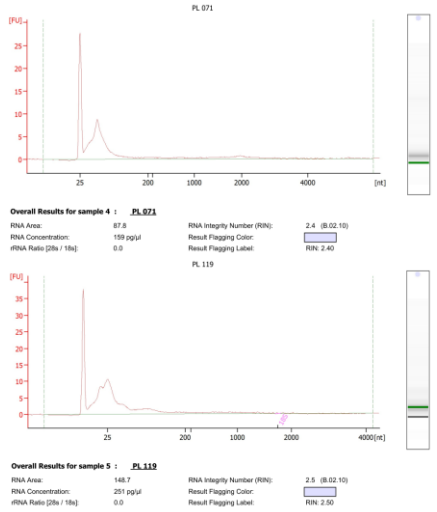
using principal component analysis. These were excluded from the analysis. An average of 13,004,616 total reads (range 8,402,014 - 22,524,635) were found in the remaining samples and available for differential expression analysis.

Analysis of the mapped reads revealed a high variability in RNAs with a much smaller proportion making up tRNAs. miRNAs made up the most common type, accounting for 37.2% of all mappable RNAs, followed by 27.2% mRNAs, 18.6% genomic RNA, 9.6% rRNAs, and 3.4% tRNAs.

Global profiling for miRNAs was undertaken between cohorts for the pairwise comparison PDAC vs Benign with an FDR<0.05 and log2fold change > 1 as cut-off for determining significance. After filtering was undertaken to reduce noise (miRNAs with no expression in more than 50% of samples were removed), we compared a total 283 miRNAs and found 15 significantly differentially expressed miRNAs between PDAC vs Benign. Their relative expression can be seen in a heatmap (FIGURE 59D) to highlight their similarity in expression profiles and as a volcano plot (FIGURE 60) to compare log fold change against significance.

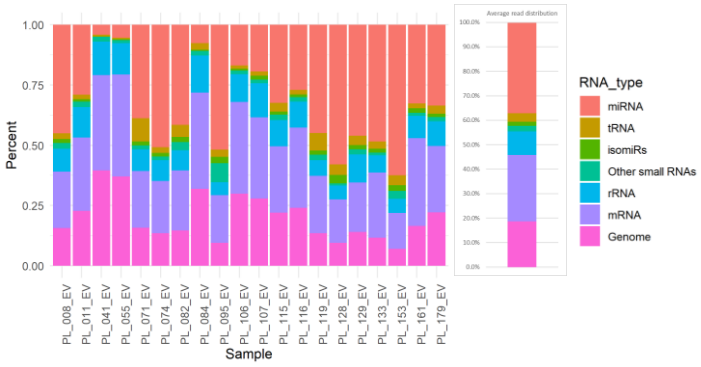
A

RNA electrophoresis showing size distribution in 2 samples (benign above, PDAC below)



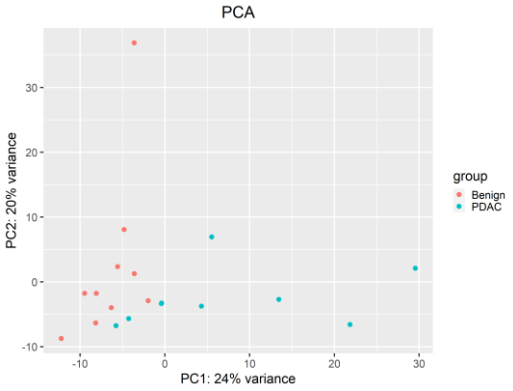
B

Relative distribution of mapped RNA reads in Plasma EVs according to RNA type with average



C

Principle Component Analysis of mapped RNA reads in Plasma EVs



D

Heatmap showing differential expression of mapped RNA reads in Plasma EVs

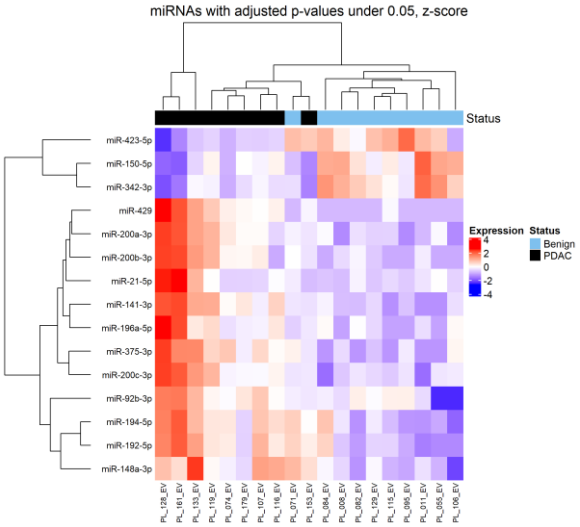


Figure 59: RNA quantification and initial differential expression analysis of miRNAs present in plasma EV samples (n=20).

A. RNA electrophoresis to determine RNA quantity and quality was performed by Bioanalyzer for two samples (Benign above), PDAC below).

B. EV-RNA from plasma samples were subsequently sent for single end 75 base pair small RNA sequencing by Dr Morten Venø and the team at Omiics (Aarhus, Denmark). Shown are the mapped RNA reads and relative proportions in each sample making up each RNA subtype (colour legend to the right). Stacked bar plots are shown ordered by sample number with results in percentages. Mean proportions across all samples are represented in the solitary bar plot to the right. Relative distribution of RNA reads according to RNA type from small RNA sequencing is shown as stacked bar plots with results in percentages.

C. Differential expression of mapped miRNA reads was undertaken and 15 miRNA candidates were found to be statistically over- or under-expressed in PDAC vs Benign. A principal component analysis was performed showing broad grouping of the samples by cohort according to miRNA expression (PDAC vs Benign).

D. Unsupervised heatmap of the 15 miRNA candidates with hierarchical clustering showing all miRNAs with adjusted p-values <0.1 in the pairwise comparisons shown below. Red means high relative expression with blue signifying low relative expression.

PDAC – Pancreatic Ductal Adenocarcinoma

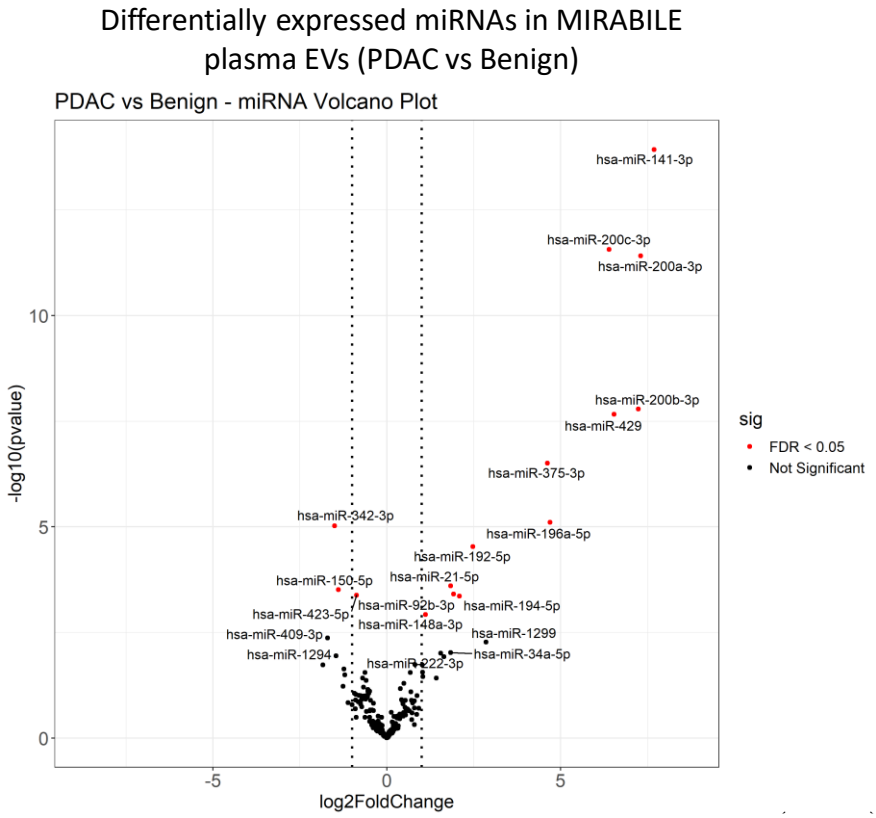


Figure 60: Differential expression of miRNAs present in plasma EV samples (n=18).
From small RNA sequencing data analysis of the discovery cohort (n=18, 8 PDAC, 10 Benign) each diagnostic potential of selected miRNA candidates were investigated. A volcano plot with significant miRNAs found in the following pairwise comparisons (PDAC vs Benign) with an FDR < 0.05 (highlighted in red). Plotted are significance ($-\log^{10}$ p-value) vs expression ($\log^2$ Fold Change). Vertical lines indicate $\log^2$ Fold Change ±1. Top twenty miRNAs are labelled for clarity.
PDAC – Pancreatic Ductal Adenocarcinoma, FDR – False Discovery Rate

TABLE 38 shows the 12 upregulated miRNA candidates with members of the miR-200 family in bold showing highest log fold change. These are miR-141-3p, miR-200a-3p, miR-200b-3p, miR-200c-3p and miR-429. Additional miRNAs detected included miR-196a-5p, miR-375-3p, miR-192-5p, miR-194-5p, miR-92b-3p, miR-21-5p, miR-148a-3p, miR-423-5p, miR-150-5p and miR-342-3p.

Table 38: Significant miRNAs identified as upregulated in PDAC samples from the plasma EV biomarker analysis (n=18).

Significant miRNA candidates identified by RNA sequencing sorted by descending log fold change values. Expression averaged across all samples with normalization by DESeq2 to allow direct comparison between the samples. Standard errors are of the log2fold change and p-values are adjusted using the Benjamini-Hochberg method. In bold are candidates used in later validation.

MicroRNA	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
miR-141-3p	**197.09**	**7.6890**	**0.9961**	**2.22E-12**
miR-200a-3p	**251.64**	**7.3058**	**1.0524**	**2.44E-10**
miR-200b-3p	**88.782**	**7.2339**	**1.2811**	**7.77E-07**
miR-429	**13.077**	**6.5386**	**1.1678**	**8.18E-07**
miR-200c-3p	**235.39**	**6.3926**	**0.9142**	**2.44E-10**
miR-196a-5p	13.382	4.6913	1.0493	0.0002
miR-375-3p	13.756	4.6179	0.9026	0.0000
miR-192-5p	305.94	2.4751	0.5925	0.0006
miR-194-5p	111.04	2.0845	0.5925	0.0059
miR-92b-3p	16.172	1.9158	0.5403	0.0059
miR-21-5p	2359.4	1.8383	0.5021	0.0048
miR-148a-3p	415.54	1.1132	0.3436	0.0151

Comparison can be made with the circulating plasma miRNA differential expression analysis (See 3.2.11) which only identified 2 candidates miR-144-5p and miR-29c-3p as upregulated in PDAC from a total of 375 differentially expressed miRNAs. By examining the EV compartment of circulating plasma and adding in a single EV isolation step (see FIGURE 60) we were able to detect a greater number of potential biomarkers. Comparison with the expression of surgical bile EV-RNA (Section 4.2.8), miR-141-3p appears to be a common candidate, being expressed in cancer samples from both analyses.

Additionally, it can be seen that members of the miR-200 family (miR-200a, miR-200b, miR-200c) as well as the previously evaluated oncogenic miR-21 are upregulated in plasma EVs as well as bile EVs (comparison with Figure 45). This time however there are a predominance of upregulated candidates, which we could then take on to validation with RT-qPCR.

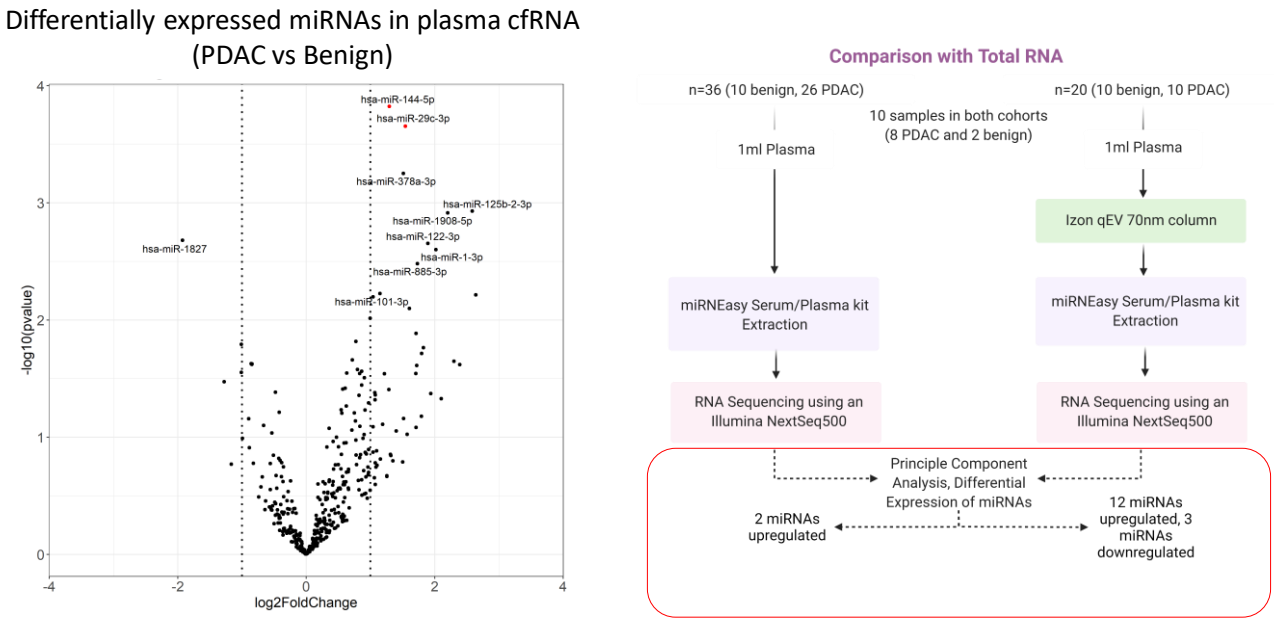


Figure 61: Comparison between circulating plasma and plasma EV differential analysis highlights improved resolution of miRNA candidates.

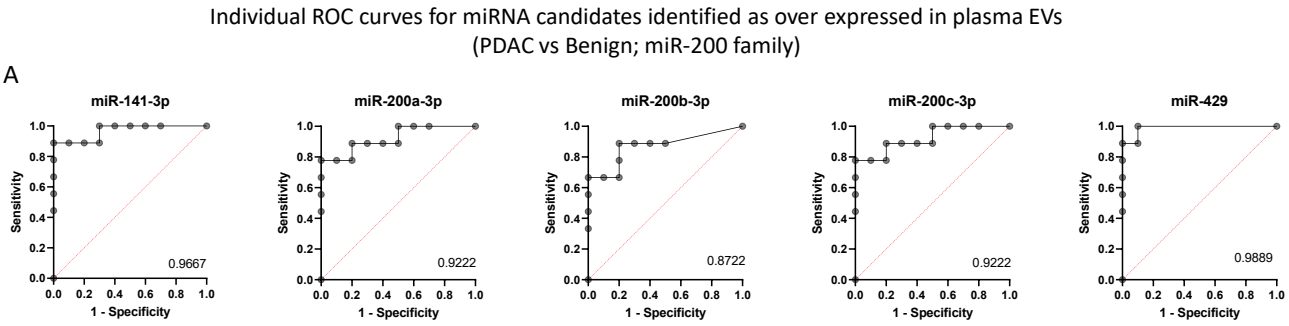
A volcano plot of significant miRNAs found in plasma cfRNA can be compared with Figure 60 and shows PDAC vs Benign with an FDR < 0.05 (highlighted in red). Only 2 significant miRNAs are shown as upregulated in PDAC in the circulating plasma analysis are labelled whilst in the plasma EV RNA analysis we have 12 miRNAs shown as differentially expressed following small RNA sequencing. On the right is shown the number of samples used for comparison and subsequent workflow for each.

PDAC – Pancreatic Ductal Adenocarcinoma, FDR – False Discovery Rate

5.2.7 ROC curves for each miR-200 family candidate demonstrates high diagnostic potential.

The five candidates from the miR-200 family (miR-141-3p, miR-200a-3p miR-200b-3p, miR-200c-3p, and miR-429) were analysed individually by logistic regression and predicted probabilities were used to generate ROC curves for each (FIGURE 62A). These were predicted as they were obtained using RNA sequencing read data which would still need to be validated. FIGURE 62B shows the statistical results of that analysis. They show that potentially expression of these plasma EV-miRNAs could predict PDAC with outstanding AUCs of above 0.9.

A



B

Comparison	MicroRNA	Threshold	Coefficient	95% CI	P-value	AUC
PDAC vs Benign	miR-141-3p	>0.5821	0.2052	0.8913 - 1.000	0.0006	0.9667
	miR-200a-3p	>0.6611	0.1105	0.7970 - 1.000	0.0019	0.9222
	miR-200b-3p	>0.6712	0.1965	0.6968 - 1.000	0.0062	0.8722
	miR-200c-3p	>0.6316	0.09541	0.7970 - 1.000	0.0019	0.9222
	miR-429	>0.7359	1.178	0.9538 - 1.000	0.0003	0.9889

Figure 62: Predicted ROC Curves for selected miRNAs identified as expressed in PDAC samples from the plasma EV discovery (n=18).

A. Candidates from the miR-200 family (as shown in TABLE 37) for PDAC vs benign pairwise comparison identified by RNA sequencing with AUCs as annotated.
B. Statistics for the individual miRNAs with threshold value (cut-off with maximum Youden Index), coefficient and 95% CI (Confidence Interval) with p-values and AUC (Area under the ROC curve).

ROC – Receiver operator characteristic.

5.2.8 miRNAs for normalisation

Using the sequencing data, we ran the NormFinder algorithm to identify suitable stably expressed miRNAs for use to normalise our RT-qPCR validation data. Both individual miRNAs (TABLE 39) and pairs (TABLE 40) are shown. Interestingly, although let-7a was identified as a potential candidate in our biliary EVs, we found that it was stably expressed in plasma EVs. We therefore chose 4 of these: let-7a, miR-23a-3p, miR-26a-5p and miR-126, for further validation by RT-qPCR. Validation of these endogenous miRNAs by RT-qPCR revealed miR-23a-3p and miR-26a-5p as the most stably expressed in four samples (results not shown) so these were used as a pair of endogenous normalisers. Additionally, it was decided that to compensate for any inter-plate variation, we would use an exogenous synthetic spike-in (UniSp6) and added a fixed amount (0.0375 fmol) to each sample. This was the same method used for the bile EV-RNA analysis in order to ensure standardisation of methods. The geometric mean, using principles by Vandesompele et al (318) was calculated and an average (EndoMean) was used to normalise the data (see Section 2.2.18 for a detailed explanation).

Table 39: MiRNAs identified as stably expressed individually for normalization in plasma EVs (n=18).

Top twenty miRNAs identified as stably expressed in samples sent for small RNA sequencing sorted by descending stability using NormFinder algorithm. GroupDif is the estimated difference in expression between the groups; GroupSD is a weighted average of the estimated intra-group variation and Stability is the NormFinder expression stability measure with lower values indicating stable gene expression.

MicroRNA	GroupDif	GroupSD	Stability
let-7b-5p	0	0.29	0.13
miR-26a-5p	0.1	0.19	0.14
miR-26b-5p	0.04	0.27	0.14
let-7d-5p	0.02	0.36	0.17
let-7a-5p	0.08	0.31	0.18
miR-29a-3p	0.02	0.42	0.19
miR-92a-3p	0.03	0.4	0.19
miR-186-5p	0.07	0.37	0.19
let-7f-5p	0.17	0.26	0.2
miR-93-5p	0.09	0.38	0.2
let-7i-5p	0.11	0.34	0.2
miR-101-3p	0.03	0.48	0.22
miR-103a-3p	0.04	0.47	0.23
miR-30e-3p	0.01	0.55	0.23
miR-425-5p	0.05	0.5	0.23
miR-584-5p	0.03	0.53	0.24
miR-23b-3p	0	0.55	0.24
miR-25-3p	0.18	0.36	0.24
miR-27a-3p	0.01	0.57	0.25
miR-23a-3p	0.16	0.44	0.26

Table 40: MiRNA pairs identified as stably expressed for normalization in plasma EVs (n=18).

Top ten pairs of miRNAs identified as stably expressed in samples sent for small RNA sequencing sorted by descending stability.

MicroRNA 1	MicroRNA 2	Stability
let-7a-5p	miR-26a-5p	0.09
miR-26b-5p	miR-26a-5p	0.09
miR-26a-5p	let-7f-5p	0.09
let-7b-5p	miR-26b-5p	0.1
let-7b-5p	miR-26a-5p	0.1
miR-25-3p	let-7f-5p	0.1
let-7d-5p	let-7b-5p	0.11
let-7d-5p	miR-26a-5p	0.11
let-7b-5p	miR-29a-3p	0.11
let-7b-5p	let-7a-5p	0.11

5.2.9 Validation of our miR-200 family 5 miRNA panel shows diagnostic potential for EV-miRNA plasma markers for PDAC.

FIGURE 63 shows the results of RT-qPCR validation of our 5 miRNAs compared in the PDAC samples and performed by myself together with Catherine Blaker (MRes). As expected, no difference was detected when the geometric mean of the normalisers was analysed and is shown in FIGURE 63A labelled "EndoMean" (miR-23a-3p, miR-26a-5p and UniSp6). All 5 miRNAs showed upregulation in our PDAC EV samples after normalisation. miR-200a-3p ($p<0.0001$), miR-141-3p ($p<0.0001$) miR-200b-3p ($p=0.0007$), miR-200c-3p ($p=0.0011$) and miR-429 ($p=0.0112$) showed increased expression with statistical significance in our validation cohort (n=61).

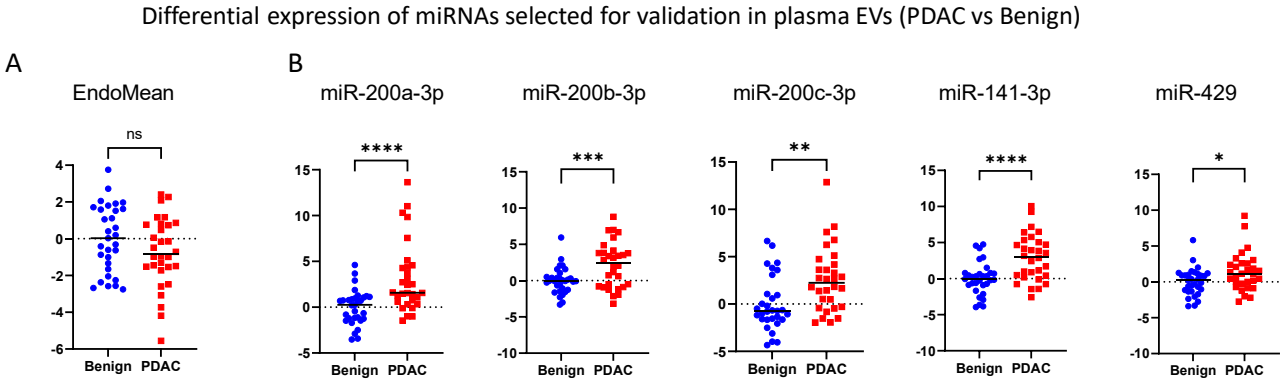


Figure 63: RT-qPCR Validation of marker panel identified in plasma EV RNA samples (n=61).

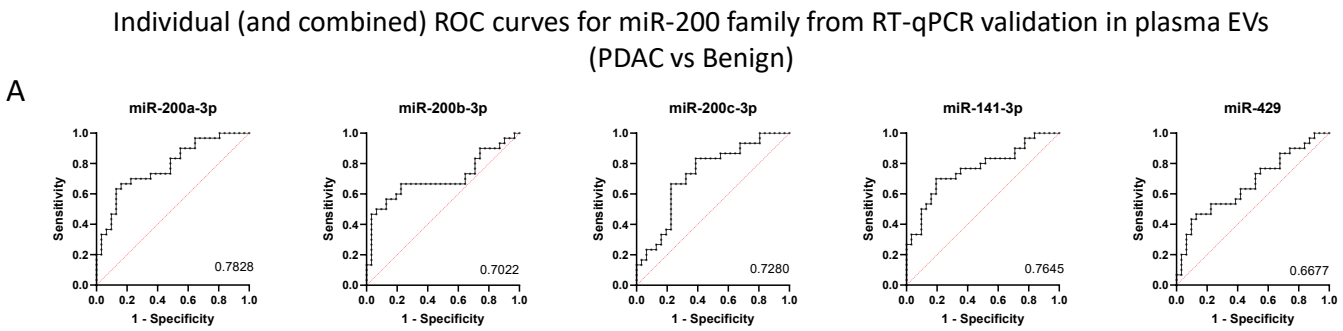
61 samples (31 PDAC and 30 Benign) underwent RT-qPCR validation of the miR-200 family and shown here are the results (performed in triplicate and mean result shown). Horizontal lines show the median and statistical analysis was performed using an unpaired Student t-test.

A. Endogenous normalisers miR-23a, miR-26a and UniSp6 were used and the geometric mean labelled as EndoMean. No significant difference was noted in the comparison PDAC vs Benign.
B. Five panel marker miR-200a, miR-200b, miR-200c, miR-141 and miR-429 showing differences between PDAC vs Benign in log fold expression with asterisks to demonstrate significance.

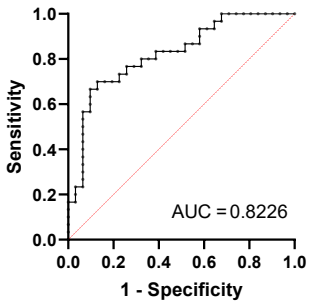
* P <0.05, ** P<0.005, *** P<0.0005, **** P<0.0001

Individual and combined analysis was undertaken of our RT-qPCR expression data and logistic regression used to generate ROC curves for our 5-miRNA panel in detecting PDAC vs Benign. Individually, miR-200a-3p demonstrated an AUC of 0.7828 (95% CI 0.6683 - 0.8973, p=0.0001), miR-200b-3p demonstrated an AUC of 0.7022 (95% CI 0.5642 - 0.8401, p=0.0067), miR-200c-3p demonstrated an AUC of 0.7280 (95% CI 0.5996 - 0.8563, p=0.0022), miR-141-3p demonstrated an AUC of 0.7645 (95% CI 0.6439 - 0.8851, p=0.0004 and miR-429 demonstrated a lower AUC of 0.6232 (95% CI 0.5311 - 0.8044, p=0.0244). The ROC curves can be seen in FIGURE 64A and these 5 were combined to exhibit a validated AUC of 0.8226 (95% CI 0.7172 - 0.9279, p<0.0001, FIGURE 64B). This was

combined with the CA 19-9 (FIGURE 64C) to give an impressive AUC of 0.9617 (95% CI 0.9081 - 1.000, p<0.0001). At the threshold determined by the maximum Youden Index (>0.7075) the combined (with CA 19-9) final model predicts cancer with a sensitivity of 89.29%, a specificity of 100% and a PPV of 100%. Without CA 19-9, the 5-miRNA model alone (threshold >0.4397) predicts cancer with a sensitivity of 70.0%, a specificity of 87.1% and a PPV of 84%. In both analyses however, no individual miRNA achieves significance after logistic regression.

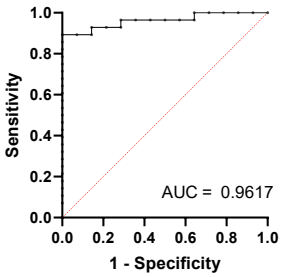


B **PDAC vs Benign – 5 miRNA panel**



Predictor	Coefficient (β)	OR (95% CI)	P
Constant	-0.7101	0.4916 (0.2433 - 0.9242)	0.0349
miR-200a	0.3444	1.411 (0.8944 - 2.526)	0.1794
miR-200b	0.07571	1.079 (0.7420 - 1.548)	0.6786
miR-200c	0.07221	1.075 (0.8291 - 1.387)	0.5710
miR-141	0.3049	1.356 (0.8540 - 2.285)	0.2111
miR-429	-0.4012	0.6695 (0.3649 - 1.099)	0.1378

C **PDAC vs Benign – 5 miRNA panel + CA 19-9**



Predictor	Coefficient (β)	OR (95% CI)	P
Constant	-2.178	0.1132 (0.005828 - 0.6235)	0.0462
miR-200a	-0.9032	0.4053 (0.07255 - 1.640)	0.2050
miR-200b	-0.3249	0.7226 (0.3219 - 1.562)	0.3680
miR-200c	0.2764	1.318 (0.6318 - 3.326)	0.4720
miR-141	1.882	6.567 (1.252 - 100.5)	0.0675
miR-429	-0.1716	0.8423 (0.3464 - 1.974)	0.6745
Ca19-9	0.001495	1.001 (1.001 - 1.003)	0.0156

Figure 64: Regression analyses to validate plasma EV-miRNA biomarker model (PDAC vs Benign; n=61).

After RT-qPCR validation, expression values for each sample can be matched to their original diagnosis and a diagnostic test generated either individually (A) or as a model (B+C) which generates a ROC curve depending on the threshold used.

A. ROC curves for each individual miRNA from the miR-200 family (miR-200a-3p, miR-200b-3p, miR-200c-3p, miR-141-3p and miR-429) for the pairwise comparison PDAC vs Benign.
B. ROC curve for the combined 5-miRNA marker panel for the pairwise comparison PDAC vs Benign with validated AUC = 0.8226.
C. ROC curve for the combined 5-miRNA marker panel and additional CA 19-9 where available for the pairwise comparison PDAC vs Benign with validated AUC = 0.9617.

PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, AUC – Area Under ROC Curve, ROC - Receiver operating characteristic.

5.2.10 EV parameters (Concentration and particle size) improve diagnostic sensitivity.

Interestingly, EV parameters as measured by NTA appeared to show a difference between PDAC and Benign populations. The median EV concentration in the Benign population was 2.3 x 10^{10} particles/mL whilst the mean EV concentration in the PDAC population was measured at 7.8 x 10^{10} particles/mL with PDAC samples containing more EVs ($p=0.0012$). Median particle size was also larger in PDAC samples with Benign samples having an average particle size of 149.4nm whilst PDAC had an average size of 163.0nm ($p<0.0001$).

As a potential marker for PDAC we undertook individual and combined analysis by logistic regression to predict probabilities for PDAC and these were used to generate ROC curves for these specific parameters (FIGURE 65A). Individually, EV concentration demonstrated an AUC of 0.7742 (95% CI 0.6504 - 0.8980, $p=0.0002$) whilst median particle size demonstrated an AUC of 0.8349 (95% CI 0.7307 - 0.9391, $p<0.0001$). Together (FIGURE 65B) they improved the AUC to 0.8484 (95% CI 0.7496 - 0.9472, $p<0.0001$). Finally, if combined with the miR-200 family (FIGURE 65C), we were able to predict PDAC with an outstanding AUC 0.9118 (95% CI 0.8387 - 0.9849, $p<0.0001$).

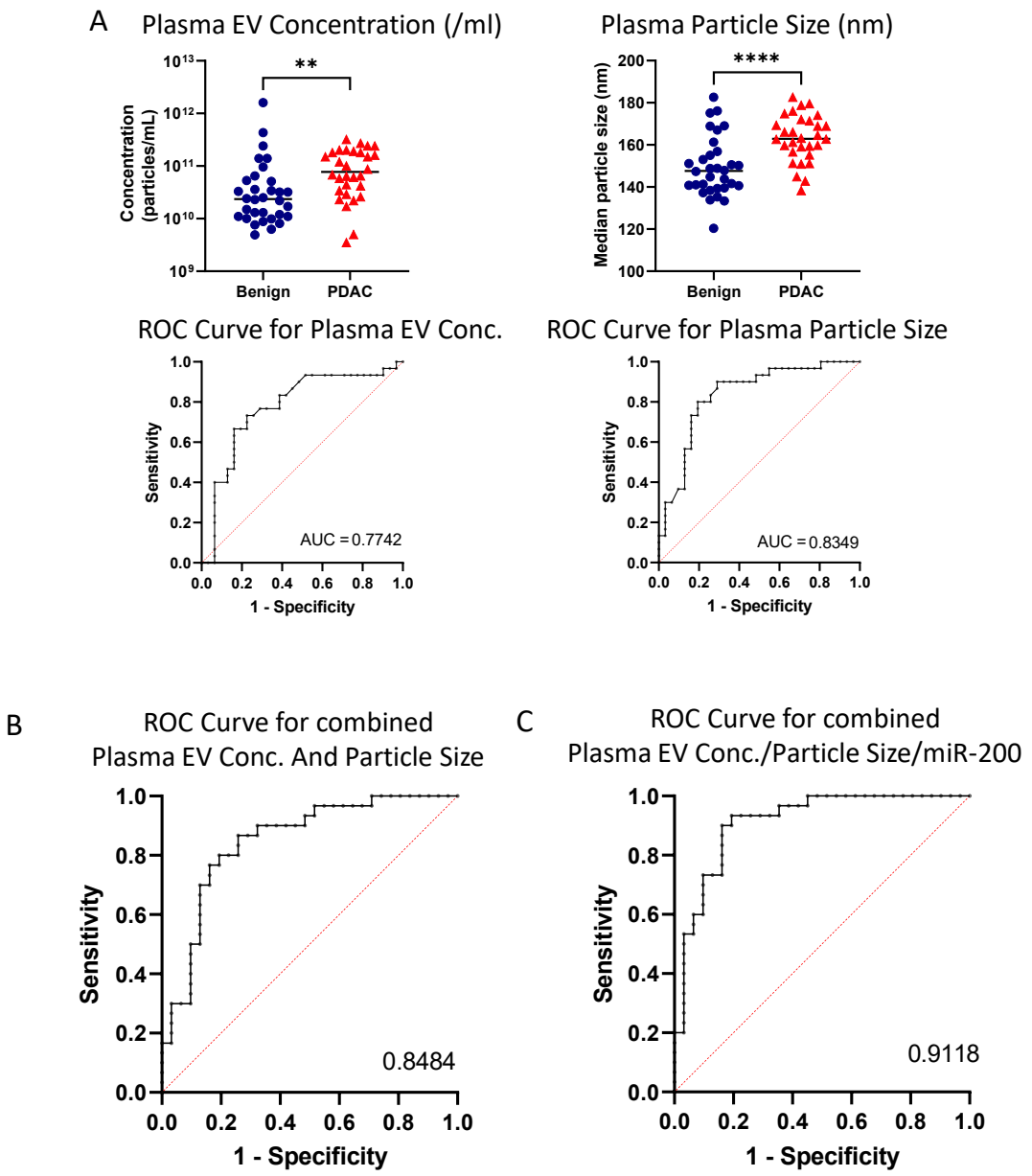


Figure 65: Regression analyses of plasma EV parameters (concentration and size) as an alternative model (PDAC vs Benign) (n=61).

Characteristics of plasma EV samples were obtained using NTA to determine concentration (particles/ml) and median particle size (nm). These are shown with median values labelled as a black line. Statistical analysis was performed using a Mann-Whitney U-test for concentration and a Student t-test for particle size. ROC curves were then generated using logistic regression to determine diagnostic value in identifying PDAC vs Benign and is shown with an AUC.

A. Comparison of EV concentrations and median particle sizes by NTA in all plasma samples showed differences between cohorts with asterisks to demonstrate statistical significance. ROC curves for each parameter (EV concentration and median particle size) for the pairwise comparison PDAC vs Benign.
B. ROC curve for the combined parameters for the pairwise comparison PDAC vs Benign with an AUC = 0.8484.
C. ROC curve for the combined parameters together with 5 miRNA marker panel for the pairwise comparison PDAC vs Benign with an overall AUC = 0.9118.

PDAC – Pancreatic Ductal Adenocarcinoma, CCA – Cholangiocarcinoma, AUC – Area Under ROC Curve, CP – Chronic Pancreatitis, ROC - Receiver operating characteristic. * P <0.05, ** P<0.005, *** P<0.0005, **** P<0.0001

In summary, our 5 miR-200 family panel predicts PDAC with an AUC of 0.8226 ($p<0.001$) and set to the optimum threshold (>0.4397) predicts cancer with a sensitivity of 70.0%, a specificity of 87.1% and a PPV of 84%. Whilst not perfect, this can be improved with current biomarkers such as CA 19-9 and/or with EV analysis using NTA. PDAC samples furthermore appear to have a higher EV concentration ($p=0.0012$) and size ($p<0.0001$) which can be detected using NTA. By combining this with the EV miRNA panel improves the AUC to 0.9118 ($p<0.0001$).

5.2.11 Verification of EV-miRNAs and miR-200 against lipoprotein fractions

Fractions 11-14 from the EV isolation were also collected for each sample, and these showed high expression of APOA-1 (FIGURE 57B). Therefore, these were considered as enriched in lipoprotein (high-density termed HDL or low-density termed LDL) and labelled "LP". Additionally, we collected fractions 22-24 from each sample and these were considered enriched in extracellular protein, therefore labelled "PROT". A small group of PDAC samples (n=6) had RNA isolated from their LP and PROT fractions and compared with the EV fraction for expression of our 5 miR-200 family panel (FIGURE 66) to determine the presence or absence of these markers in each compartment.

FIGURE 66A shows that the endogenous normalisers used (miR-23a and miR-26a) were significantly under expressed in our LP and PROT fractions, which could be due to their presence in EVs. As a result, this was not suitable as a normaliser for our subsequent measurements of each miRNA and the spike-in UniSp6 was substituted for analysis. FIGURE 66B showed a significant decrease in expression for all 5 miRNAs which confirms the relative absence of these markers from the extracellular and lipoprotein compartments of blood plasma in our samples.

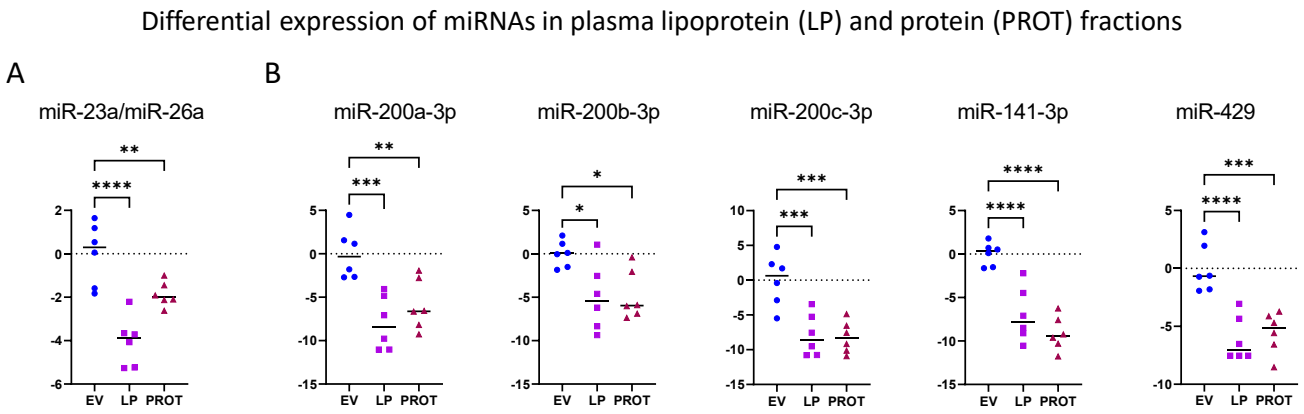


Figure 66: RT-qPCR verification of marker panel in plasma EVs compared with lipoprotein and protein fractions (n=6).

Verification that miR-200 as well as the normalisers (miR-23a and miR-26a) were present in EVs and not found in other SEC fractions was performed by RNA isolation of later fractions. Fractions 11-14 were deemed lipoprotein (LP) fractions constituted and fractions 22-24 were deemed protein-rich (PROT). RNA underwent RT-qPCR and expression of miRNAs was analysed by the ΔΔCT method. Student t-tests were performed to determine statistical differences between these and the EV fractions (fractions 7-9) from earlier. UniSp6 was used as a spiked-in normaliser due to differences in endogenous miR-23a and miR-26a shown. All reactions were performed in triplicate and the mean value used.

A. Endogenous normalisers miR-23a and miR-26a in EV, LP and PROT fractions is shown. A significant difference was noted.
B. Five markers miR-200a, miR-200b, miR-200c, miR-141 and miR-429 showing differences in log fold expression with asterisks to demonstrate significance between the fractions.

* P <0.05, ** P<0.005, *** P<0.0005, **** P<0.0001

Additional standard curves to demonstrate comparative primer efficiencies (and hence validity of our relative $\Delta\Delta C_T$ method) were undertaken in a single highly expressed sample, triplicated, and is shown in FIGURE 67. Once again, the optimal slope value should be <0.1 but due to a high error in some of the calculated expression values, it is difficult to determine if this was a significant finding.

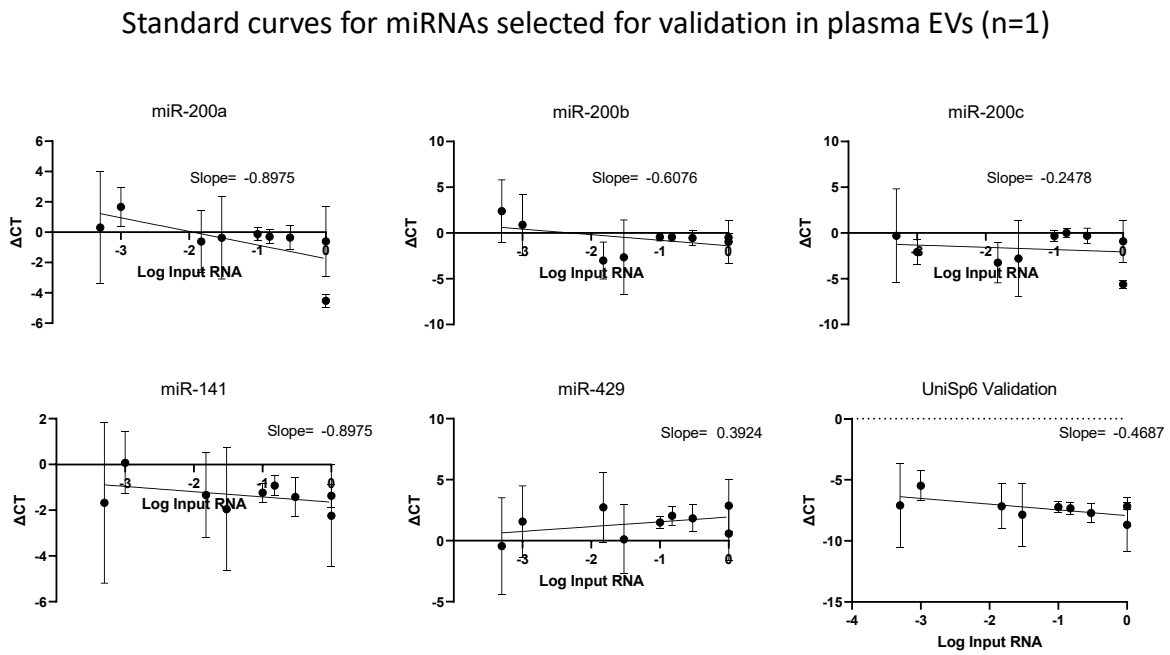


Figure 67: RT-qPCR standard curves for each miRNA (n=1; triplicate).
In order to ensure RT-qPCR primers bind and amplify with equal efficacy across a broad range of RNA input, standard curves were undertaken in a sample with high miRNA expression previously determined. The ΔC_T for each of the 5-marker panel comprising miR-200a, miR-200b, miR-200c, miR-141 and miR-429 was measured with EV-RNA input diluted as shown. Trendlines show the slope values adjacent to each graph with the ideal gradient as <0.1. Endogenous normalisers miR-125a-5p, miR-194-5p and UniSp6 were used.

5.2.12 tRNAs in plasma EVs as a potential future source of biomarkers

tRFs comprised a read percentage of 3.4% (FIGURE 58B) which was a significantly smaller amount than in bile EVs. However we undertook a similar investigation by RNA sequencing for tRFs that might show candidates suitable for future biomarker analysis. TABLE 39 shows the top ten tRF candidates sorted by descending log fold change in the comparisons PDAC vs Benign. These were not deemed significant ($p>0.05$), but overall expression by coverage showed a similar pattern to bile EVs whereby specific sets of tRNAs gave rise to the majority of tRNA reads (tRNA-Glu-CTC 5' halves and tRNA-Gly-GCC 5' halves). Comparison between plasma and bile EV tRFs (Table 33) show the presence of Glu-CTC 5' half which is upregulated in both EVs (tRF-38-87R8WP9N1EWJQ7D1 differing only by one nucleotide). Additionally Glu-TTC 5' tRF is found in both EV studies but as a much shorter fragment (tRF-23-86J8WPMNY) in bile EVs. These have been highlighted in red.

Table 41: Top ten tRNA fragments identified as upregulated in the plasma EV biomarker discovery analysis (n=18).

Top ten tRF candidates identified in the comparison PDAC vs BENIGN in bile EVs, sorted by descending log fold change values (upregulated in PDAC samples). Highlighted in red are the two tRFs found common to plasma and bile EVs. Expression averaged across all samples with normalization by DESeq2 to allow direct comparison between the samples. Standard errors are of the log2fold change and p-values are adjusted using the Benjamini-Hochberg method.

tRNA Fragment	MINTbase Unique ID	Expression (RPM)	Log2FoldChange	Standard Error	Adjusted P-value
Ala-TGC_5tRF.9	tRF-23-RK9P4P9LDS	2.5294	1.9977	1.1277	0.5353
Leu-AAG_5tRF.1	tRF-16-RPM830D	1.2684	1.9251	0.9127	0.4319
Ser-TGA_5tRF.15	tRF-24-PJ58309N0Y	2.1409	1.7943	1.0389	0.5676
Lys-CTT_5half	tRF-31-PSQP4PW3FJI0B	1.7654	1.7530	0.8187	0.4274
Glu-CTC_5half.8	tRF-38-87R8WP9N1EWJQ7D1	2.9201	1.6599	0.8373	0.4724
Ser-CGA_5tRF.3	tRF-24-QW58309N0Y	4.0226	1.4952	1.1754	0.7396
Ser-TGA_5tRF	NA	4.3696	1.4469	0.8562	0.5823
Glu-TTC_5tRF.13	tRF-26-86V8WPMN1EE	2.2060	1.4319	1.0558	0.6957
Lys-TTT_5half.8	tRF-32-PS5P4PW3FJHP1	5.4908	1.3418	0.7124	0.5251
Pro-TGG_5half.3	tRF-31-6978WPRLXN4VE	0.9911	1.3243	0.9682	0.6936

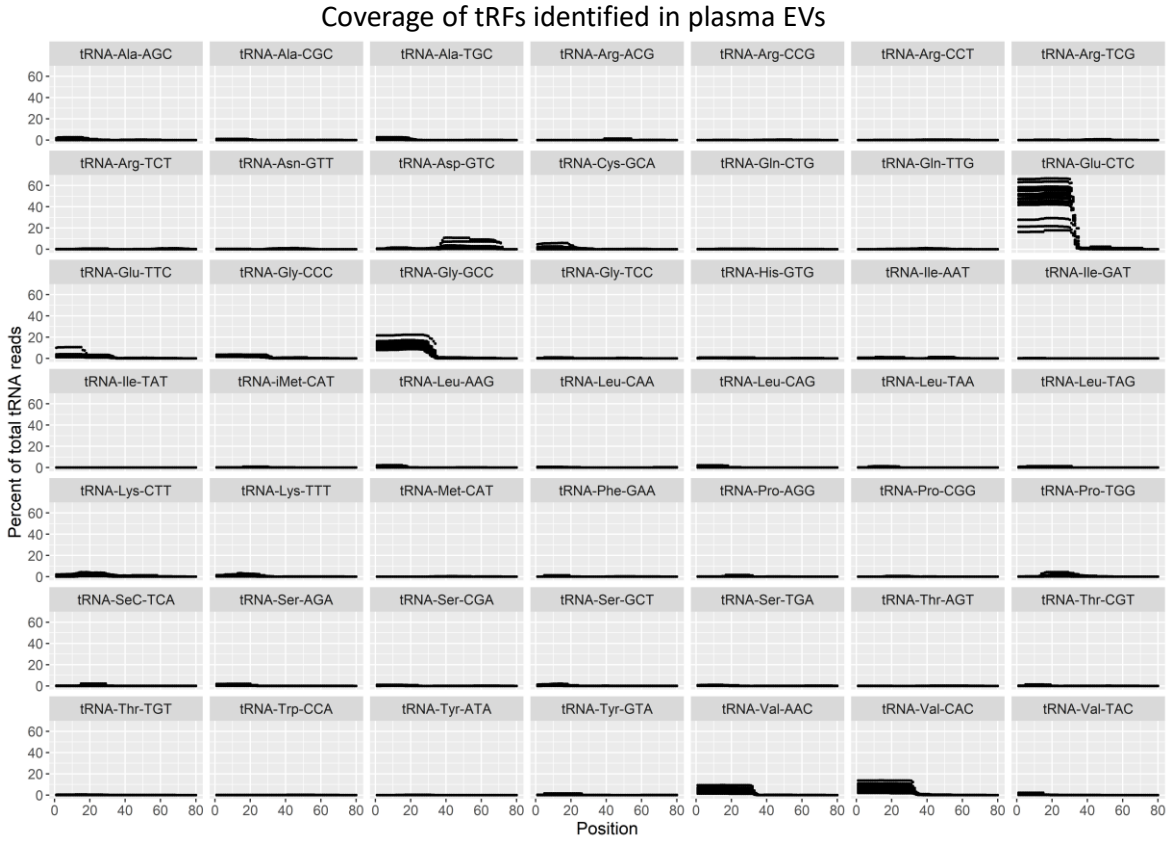


Figure 68: Plasma EV tRF expression expressed according to tRNA isotype (n=18).

A coverage plot of the tRF candidates is shown according to tRNA isotype (i.e. the tRNA from which the fragment is likely to have arisen). This also shows Glu-CTC (but not Gly-GCC) as a dominant tRNA isotype in plasma EV samples. Normalised read distribution of tRFs in bile according to tRNA isotypes (tRNA of origin). Expression as a percentage of total tRNA reads with relative nucleotide position within the tRNA expressed along the y-axis.

5.2.13 Prospective investigation in longitudinal post-operative patient samples shows potential change after surgery.

Pre- and post-operative blood samples were obtained from patients (n=10) undergoing surgery for PDAC (FIGURE 69). EVs were isolated and underwent our workflow to

determine if EV-miRNAs changed after complete resection. Five patients who underwent pylorus preserving pancreaticoduodenectomy (PPPD) for PDAC were compared with controls, made up of patients who had a laparotomy and trial dissection followed by bypass (no resection of tumour; n=2) and patients who underwent PPPD for benign disease (n=3). After EV isolation and RNA extraction using the same methodology as previously, a statistically significant decrease was detected in miR-200b-3p (p=0022) with no differences detected in the control population. Statistical analysis of the overall EV miR-200 change in expression (measured as the expression before minus the expression after surgery) showed no difference was detected between PDAC and controls.

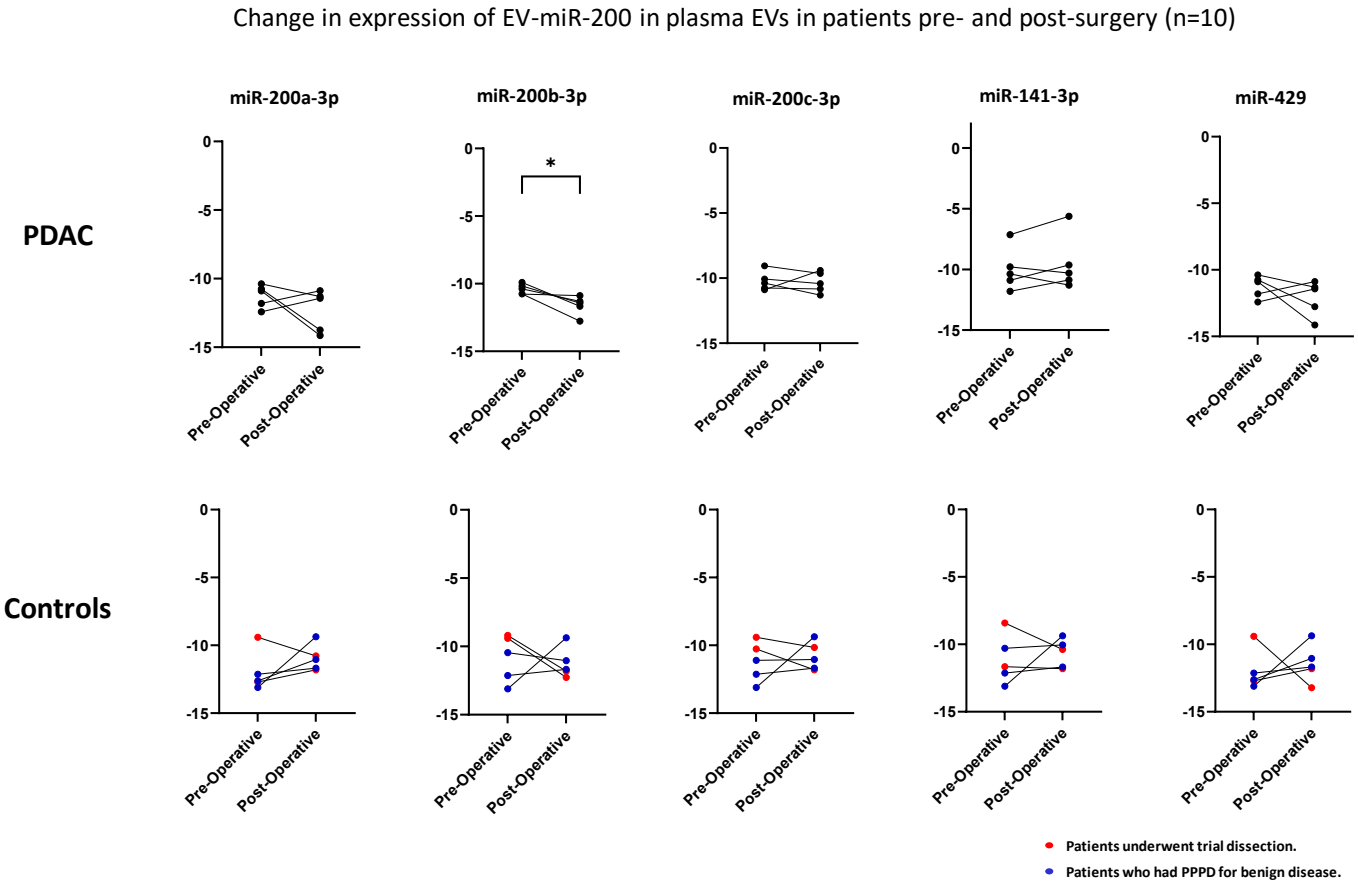


Figure 69: Analysis of expression of miR-200 in blood plasma EVs from patients undergoing surgery for PDAC show potential to predict PDAC tumour burden (n=10).

Pre- and post-operative blood samples were analysed for our 5 EV-miRNA markers (miR-200a-3p, miR-200b-3p, miR-200c-3p, miR-141-3p and miR-429). Red dots indicate control patients who underwent surgery for PDAC but no resection was performed, and blue indicates those who underwent surgery for benign disease. Endogenous normalisers miR-23a, miR-26a and UniSp6 were used, and a two-tailed paired t-test was used to determine statistical significance with asterisks where noted. * P <0.05.

PDAC – Pancreatic ductal adenocarcinoma

5.3 Discussion

5.3.1 Plasma-derived EVs are a source of small RNAs which are differentially expressed in PDAC.

Blood-derived EVs have been a newly recognised source of small RNAs which have been associated with many disease processes in the human body. This is an area with a large number of studies (see APPENDIX Table 1) having been published investigating this with cargos including protein, RNA and DNA from serum and plasma. It is clear from the wide-ranging literature that EVs show differential expression in nucleic acid cargo and continues to be a heavily researched area as a potential source for biomarkers. We identified 12 statistically significant EV miRNAs in plasma which include miR-196a and miR-21 which have been previously described (275, 276, 327, 328). By undertaking a sequencing strategy to discover candidate miRNAs we have been able to evaluate all known small RNA species and undertake composition analysis of these molecules associated with our EV samples.

5.3.2 miR-200 family has been previously identified as overexpressed in previous studies investigating the plasma EVs of patients with PDAC.

A review of the literature highlights a wealth of studies investigating the differential expression of plasma EV miRNAs (See Appendix Table 2). With the recent advances in high-throughput RNA sequencing, studies have been able to undertake large scale differential expression analyses of EV samples obtained by different methods.

The first evidence that these miRNAs might play a diagnostic role in PDAC was a paper published by Reese *et al.* (329) who found overexpression of miR-200b and miR-200c in serum exosomes of PDAC patients compared to healthy controls and patients with chronic pancreatitis. This was undertaken in a larger study of 89 patients. They identified an AUC of 0.79 in distinguishing PDAC from healthy patients and an AUC of 0.77 in distinguishing from CP.

One very recent large study by Nakamura et al (published in November 2022) also supports our results. They used a larger discovery cohort (n=101) of samples from 2009-2017 to identify a total of 13 circulating miRNAs and 17 EV miRNAs that were significantly upregulated in patients with early stage PDAC (I/II by AJCC which correlates to T3N1M0 or less). They used a kit-based method and their final validated

model consisting of 13 miRNAs (5 cf- and 8 exo-miRNAs), of which 5 of the EV miRNAs identified were members of the miR-200 family. They found that their EV-miRNA signature AUC value alone was 0.89 which was improved to 0.93 with the addition of circulating miRNAs when applied to an independent validation cohort.

From our results, individually miR-200a-3p demonstrated an AUC of 0.7828, miR-200b-3p demonstrated an AUC of 0.7022, miR-200c-3p demonstrated an AUC of 0.7280, miR-141-3p demonstrated an AUC of 0.7645 and miR-429 demonstrated a lower AUC of 0.6232. Together these 5 were combined to exhibit a validated AUC of 0.8226 (95% CI 0.7172 - 0.9279, $p<0.0001$). This was combined with the CA 19-9 (FIGURE 64C) to give an impressive AUC of 0.9617 (95% CI 0.9081 - 1.000, $p<0.0001$) which is in line with the results by the other studies. Despite the large number of other studies in this area however, it is surprising that there have been so few consistently identified candidates to-date.

5.3.3 SEC as an isolation method for plasma-EVs

One of the reasons for the heterogeneity of EV biomarker studies is the variation in EV isolation methodology. Ours is the only study identified from the literature which has utilised SEC as a method for the isolation of EVs to detect PDAC and CCA. We have used the MISEV guidelines to approach the characterisation of our samples in order to ensure that RNA is obtained from the EV compartment. With TEM showing cup-shaped structures, NTA showing a size distribution with an average of 149.4-163.0nm (median particle sizes) and western blotting showing expression of ALIX, CD63, TSG 101 and relative downregulation of Calnexin and APOA-1, this is strong evidence for the presence of EVs. As mentioned earlier and supported by other papers, EV-RNA constitutes <1% of the extracellular RNA found in circulating biofluids so it is important that the method used does not co-isolate other RNA-binding proteins. Our western blotting using a proprietary kit appeared to non-specifically express protein that was no different than unprocessed plasma (FIGURE 57B) and the concern is that many studies are reliant on rapid “exosome purification kits” that are not suitable for use.

One paper published in 2018 (212) appeared to support these findings, having compared SEC-EV with the exoEasy kit from Qiagen (Qiagen, Germany). This kit was compared by common characterisation methods using 15 patient samples and found to elute a purer (lower protein to EV particle ratio) EV preparation compared to the kit isolates. However comparable amounts of RNA were detected (median of 3.8ng) and this was not further studied for RNA composition. Since our method was optimised, there have been more

reports comparing SEC as a method for EV isolation (330, 331) and further improvements in the method are likely.

5.3.4 Lipoprotein as a confounding factor in the isolation of plasma EVs.

Our original protocol used in bile determined that fractions 7-12 would be suitable to isolate EVs without extracellular protein contamination. However, one of the significant limitations of the method has been the reported presence of lipoproteins which co-isolate with vesicles. Lipoprotein particles as well as extracellular protein does also contain bound RNA which could be a confounding factor and the source of our candidate miRNAs (332). In our study using plasma, we identified the presence of unilayered structures on TEM, a bimodal distribution on NTA and western blotting evidence using APOA-1 seemed to confirm the presence of lipoprotein as a co-contaminate in early fractions. We found that there was a significant difference in APOA-1 concentrations in pooled fractions 7-10 compared with fractions 11-12 and that this would reduce but not eliminate lipoprotein as a confounding co-isolate. To evaluate these fractions for the presence of miR-200 family, we ran additional verification experiments and found that they are significantly downregulated in lipoprotein compared with EV fractions (FIGURE 66).

5.3.5 Plasma EV-RNA compared with analysis of Plasma cfRNA markers.

Differential analysis of RNAs identified from our plasma cfRNA analysis did not show any statistically significant candidates whilst the presence of upregulated EV miRNAs was shown in a smaller cohort of samples (36 circulating plasma samples vs 20 plasma EV samples). This indicates that differential expression within the plasma cfRNA compartment may be preferentially occurring in the EV-RNA and this would therefore be a more suitable place to identify candidates for liquid biopsy.

Other studies have found upregulation of miR-200 (specifically miR-200a, miR-200b and miR-141) in the blood cfRNA of patients with both HCC (333) and PDAC (334) which can be contrasted with strong evidence that that miR-200 is downregulated in from PDAC UICC Stage II macro dissected tissue samples. It is possible that the difference detected is solely due to EV-RNA as cfRNA encompasses this. By identifying the specific compartment, smaller studies will be able to determine changes in cancer related expression.

5.3.5 EV parameters in plasma EVs

Interestingly, EV parameters as measured by NTA appeared to show a difference between PDAC and Benign populations. The median EV concentration in the Benign population was 2.3 x 10^{10} particles/mL whilst the mean EV concentration in the PDAC population was measured at 7.8 x 10^{10} particles/mL with PDAC samples containing more EVs ($p=0.0012$). Median particle size was also larger in PDAC samples with Benign samples having an average particle size of 149.4nm whilst PDAC had an average size of 163.0nm ($p<0.0001$). As a potential marker for PDAC, EV concentration demonstrated an AUC of 0.7742 (95% CI 0.6504 - 0.8980, $p=0.0002$) whilst median particle size demonstrated an AUC of 0.8349 (95% CI 0.7307 - 0.9391, $p<0.0001$). Together they were able to predict PDAC with an AUC of 0.8484 (95% CI 0.7496 - 0.9472, $p<0.0001$). Finally, if combined with the miR-200 family, we were able to predict PDAC with an outstanding AUC 0.9118 (95% CI 0.8387 - 0.9849, $p<0.0001$).

Other EV studies have shown increases in EV protein concentration associated with disease activity (335, 336) and this would be consistent with an increase in cellular turnover in active malignancy. EV markers such as CD63 have also been associated with cancer and it may be hypothesised that many reported overexpressed markers may be a surrogate for increased EV number. We attempted to normalise our validation results to miRNAs found stably expressed, but no study as yet has found a consistent RNA marker to evaluate this. Possible future research may be able to measure EV-RNA per vesicle, but until then we can only estimate numbers and normalise indirectly.

5.3.6 tRFs in plasma EVs as a novel diagnostic marker

RNA derived from tRNA fragments have been poorly evaluated in plasma EVs until this point. Zhu *et al.* (337) profiled SK-Hep1 liver cancer cells and showed that 5% of exosomal small RNAs were generated from tRNA, which was higher than the 0.2–2% of total small RNAs found in plasma exosomes from liver cancer patients in the same study but consistent with our findings (mean 3.4% in plasma EVs). We also agreed with their findings which was of overexpression of 5'-tRFs, and the presence of four tRNA isotypes. However whilst they found tRNA-Glu-CTC-5, tRNA-Val-TAC-3, tRNA-Gly-TCC-5, tRNA-Val-AAC-5, we found our samples overexpressed with tRNA-Glu-CTC-5, tRNA-Val-AAC-5, tRNA-Gly-GCC-5, tRNA-Val-CAC-5, matching in 50%. Kumar *et al.* (338) is the only other published study (n=4) which investigated the profiles of mRNA, piRNA, and tRNAs in the exosomes of blood serum of healthy individuals, patients with intraductal papillary

mucinous neoplasms (IPMN), and PDAC patients identifying differentially expressed candidates but with no further validation. It is therefore a promising novel area of investigation which is likely to be of future interest.

5.3.7 Analysis of pre- and post-operative miR-200 family from EV-RNA shows potential for prognostication.

Pre- and post-operative blood samples were obtained from patients (n=10) undergoing surgery for PDAC. Comparison with controls which either underwent a laparotomy and trial dissection but no resection of the cancer (palliative) or those undergoing for benign disease showed a significant decrease only in miR-200b-3p. Further investigation for this would require a much larger population to detect a difference with an average decrease in expression of -0.549 log fold change in the PDAC population and an average increase in expression of 0.421 log fold change in the control population. This has been done previously by other studies and has shown changes within 24 hours of surgery in plasma EVs sampled (275). Further analysis of EV-miR-200 in a larger cohort as a measure of potential early recurrence would be warranted.

Chapter Six – General Discussion and Conclusions

6.1 Discussion

6.1.1 Bile EVs show upregulated expression of let-7a and miR-21 in PDAC.

We have detected let-7a and miR-21 as upregulated in PDAC bile EV samples ($p<0.0001$; See Appendix Table 1 for Summary of Findings). Let-7a (lethal-7a) is a member of the let-7 family of which there are 12 members located on eight chromosomes. There are at least 3 paralogs of let-7a (let-7a-1, let-7a-2, let-7a-3) which occur in clusters let-7a-1/let-7f-1/let-7d on chromosome 9q22.3, cluster let-7a-2/miR-100/miR-125b-1 on chromosome 11q24.1 and cluster let-7a-3/let-7b on chromosome 22q13.31 (339). These are formed by different precursors which at an EV level are indistinguishable. In C. elegans, this miRNA is part of a core set of genes that function sequentially to control developmental larvae-to-adult transition (340).

A systematic review undertaken in 2012 evaluated tumour specimens in 20 different cancers and identified decreased let-7 associated with poor outcome in Gastric, lung, HCC, ovarian cancers (and increased miR-21 associated with poor outcome in patients with Astrocytoma, colon, lung, ovarian) (341). This only evaluated oligonucleotide arrays and RT-qPCR studies evaluating a median number of 328 miRNAs from 43 studies. Let-7a is believed to be downregulated in PDAC cells (342) although upregulated during the progression of PanIN with peak production of let-7a (alongside miR-21, miR-221, miR-222, miR-200, and miR-205) in PanIN 2/3 lesions (164). It is also downregulated in CCA tissues and cells (343).

As a blood marker, cell-free let-7a levels from plasma were compared in a cohort of 208 patients and shown to be downregulated when PDAC was compared with a control group that included patients with CP (344). This was also found to be the case in UC-derived exosomes obtained by Lai *et al.* (275) although use of a random single normaliser (miR-425-5p) could be questioned. Our findings in plasma EVs were that let-7a-5p was in fact found highly expressed and stable across all samples which made it potentially an endogenous miRNA in plasma EVs.

Let-7a is found to be involved in many recognised cancer pathways, and is repressed by MYC, an oncogenic transcription factor that leads to repression of CDKN1A and CDKN2B

cell cycle-inhibitory genes (345). The KRAS 3' UTR contains multiple MREs for let-7 (157) which was found to be reciprocally expressed in lung squamous cell carcinomas. It might be hypothesised that this could be an important mechanism in the squamous subtype of PDAC. The LIN28 gene was also first identified in C. elegans and has similar functions in development. It generates two proteins LIN28A and LIN28B which have been found to act on let-7 as well as containing MREs for let-7 leading to a feedback loop (346). These oncogenes work in different ways to repress let-7 (347) and have been shown to be significantly upregulated in human PDAC tissues and correlated with poor prognosis/cancer metastasis (348). Interestingly, circulating tumour cells (CTCs) in PDAC have recently been identified as containing LIN28B which was correlated with improved patient survival in a cohort of 56 patients (349). With approximately 3-20 CTCs per mL(350) of blood and ~10^{10} EVs present, it may be more feasible that let-7 expressed in circulating EVs has more of a role to play physiologically on cancer metastasis and progression (351). To-date no previous study has identified let-7a in bile-derived EVs and the physiological role of this EV-miRNA in bile remains to be determined.

MiR-21 on the other hand has been the most consistently described in PDAC tissue (352). Whilst let-7a is downregulated in pancreatic tissues, expression of miR-21 has been associated with early stages of PDAC from early precursor PanIN lesions (164) to invasive IPMNs (165) and additionally in stromal cancer-associated fibroblasts (CAFs) in resected PDAC tumours (166). Frampton *et al.* published an association between raised miR-21 (as well as miR-23a and miR-27a) and enhanced cell growth/proliferation *in vitro* leading to poor differentiation and OS in a patient cohort (353). This is supported by other studies including Moriyama et al. who investigated miR-21 overexpression in both cancer tissues and 14 PDAC cell lines (354). It is therefore considered an oncogene and has been shown to work by mediating matrix metalloproteinase (MMP)-2 and MMP-9 mRNA expression to promote cell invasion. The effect of miR-21 also extends to PDAC survival outcomes, being associated with PFS and OS in advanced PDAC patients treated with gemcitabine (355). It is furthermore found upregulated in cell-free plasma from multiple studies (356-358) where it has been described as negatively impacting OS and response to adjuvant chemotherapy. However, it is also found raised in CP (359, 360), metabolic disorders such as type 2 diabetes (361) and other solid tumours including osteosarcoma, breast, and colorectal cancers (362-365).

We were able to determine that miR-21 was significantly upregulated in bile-derived EVs and it had an outstanding individual AUC (0.9560) for the detection of PDAC. It was also

identified in our plasma EVs but we did not perform RT-qPCR validation to confirm its presence. Its role as a signalling molecule in EVs is intriguing with several possible mechanisms already established. CAFs (derived from patient resection tissue) with increased miR-21 expression have been found to interact to increase epithelial cell proliferation and survival and that this was through exosome signalling, although the cargo of EV-miRNA was not investigated (167). Circulating EV miR-21 has also been found as an early biomarker for type 1 diabetic development in beta cells and murine models which raises the interesting possibility that this may be a biomarker for patients with recent-onset diabetes, a recognised sign of PDAC. Again, our discovery in bile derived EVs may constitute an early marker of pancreatic dysfunction which can occur in both CP and PDAC.

6.1.2 Bile and Plasma EVs show upregulated expression of miR-200 family in PDAC and CCA

We have detected for the first time the presence of miR-200a and miR-429 ($p<0.0001$), two related miRNAs both from the miR-200 family and co-located on chromosome 1, in bile EVs from patients with PDAC and CCA. It was also clear that the miR-200 family was important in plasma EVs and we have shown that all 5 members are upregulated in plasma EVs. Interestingly, our preliminary work with surgical bile EVs found miR-141-3p (member of the miR-200 family) to be significantly downregulated in the bile of patients with cancer but was not found to be differentially expressed in our MIRABILE bile EV samples despite being well expressed (data not shown).

6.1.2.1 The evidence for EV-miR-200 in relation to its role in tissue.

The miR-200 family consists of nine mature miRNAs in two clusters located on chromosome 1p36.33 (miR-200a, miR-200b and miR-429) and chromosome 12p13.31 (miR-200c, and miR-141) (366).

Table 42: Selected summary of evidence for the miR-200 family in PDAC/CCA.

AUC – Area under the curve, PDAC – Pancreatic ductal adenocarcinoma, CCA – Cholangiocarcinoma, CP – Chronic pancreatitis, EpCam - epithelial cellular adhesion molecule, FFPE – Formalin-fixed paraffin embedded, miR – microRNA, NP – Normal pancreas, OS – Overall Survival.

Author	Ref.	Year	Cancer	Sample Size	Methodology	miRNAs	Notes
Tissue studies							
Szafranska *et al.*	(169)	2007	PDAC	6 PDAC cell lines 10 PDAC, 7 CP, 7 NP, 33 normal tissues as control.	Macrodissected tissue, mirVana miRNA bioarrays and RT-qPCR	**miR-141** **miR-200a** **miR-200b** **miR-200c upregulated in pancreas but downregulated in PDAC**	Compared with chronic pancreatitis to evaluate role of desmoplasia
Zhang *et al.*	(324)	2009	PDAC	3 PDAC cell lines 22 PDAC tumours. Adjacent normal tissue as control	Macrodissected tissue, RT-qPCR using QuantiMiR system	**miR-200b upregulated**	
Kent *et al.*	(322)	2009	PDAC	21 PDAC cell lines. HPDE and HPNE as control.	TRIzol, Northern blot and RT-qPCR	**miR-200b/200a/429 cluster upregulated**	Also compared 5 different patient derived xenografts and 4 normal bulk pancreas tissues showing miR-200a and miR-200b as upregulated in vivo.
Yu *et al.*	(367)	2010	PDAC	15 PDAC cell lines and 99 PDAC tumours. Adjacent normal tissue as control	Macrodissected FFPE tissue and RT-qPCR	**miR-200c upregulated**	High levels of miR-200c expression inhibit cancer invasion and stimulate cancer cell proliferation. High miR-200c expression was associated with a longer OS (p = 0.03).
Paik *et al.*	(368)	2015	PDAC	84 tissue samples (79 PDAC tumours). HPDE as control.	Macrodissected FFPE tissue and Rt-qPCR	**miR-200a/ 200b/141 upregulated** **miR-200c downregulated** compared to HPDE cells	Mainly patients who were Stage II and for surgical resection. Overall survival was longer in the patient group with low miR-200c.
Liu *et al.*	(369)	2016	PDAC	1 cell line and 75 PDAC tumours. Adjacent normal tissue as control	Macrodissected tissue and RT-qPCR	**miR-200c downregulated**	5-year overall survival of low miR-200c expression group was significantly shorter (RR: 2.314, 95% CI: 1.731-6.378, P = 0.008).
Dhayat *et al.*	(334)	2018	PDAC	96 PDAC tumours. 9 IPMN, 21 CP, 29 NP.	Macrodissected FFPE tissue, Qiazol and RT-qPCR	**miR-141 (p < 0.001)** **miR-200a (p = 0.005)** **miR-200b (p < 0.001)** **miR-200c (p < 0.001)** **miR-429 (p = 0.017)** **down-regulated in PDAC**	80% of patients were non metastatic. Comparisons with normal pancreas tissue.
Diaz-Riascos *et al.*	(370)	2019	PDAC	31 PDAC tumours. Adjacent normal tissue as control	Macro and laser capture micro dissected ductal cells, TRIzol and RT-qPCR	**miR-200a (45%, p = 0.03)** **miR-429 (61%, p = 0.001)** **upregulated**	Functional studies in PANC-1 (PDAC) spheroids and orthotopic xenografts demonstrated miR-429, and miR-141, decreased PDAC tumor growth and metastatic spread.
Plasma studies							

Shen *et al.*	(284)	2019	CCA	CCA (n=36) Controls (n=12)	Serum EV	**miR-141-3p, miR-200a-3p**, **miR-200b-3p**, **miR-200c-3p**, and **miR-205-5p upregulated**	miR-141-3p, miR-200a-3p, miR-200b-3p, miR-200c-3p together showed higher AUC (all > 0.80) than CA19-9 (0.78) in the diagnosis of CCA. Positive correlation with tumour characteristics and CA 19-9.
Reese *et al.*	(329)	2020	PDAC	PDAC (n=56) CP (n=11) Controls (n=22)	Serum EV	**miR-200b upregulated** (AUC 0.79; p = 0.0001) PDAC vs CP **miR-200c upregulated** (AUC of 0.67; p = 0.0239) PDAC vs Controls	Also compared miR-200b from EpCAM positive exosomes but was less accurate. High expression of miR-200b in EpCAM-positive serum exosomes were associated with shorter OS of patients.
Nakamura *et al.*	(214)	2022	PDAC	Discovery: PDAC (n=44) Controls (n=57), Training/validation: PDAC (n=124), Controls (n=67)	Plasma and serum	13 markers (5 cf- and 8 exo-miRNAs) which include **miR200a-3p** **miR200b-3p** **miR200c-3p** **miR429** **miR141–3p** **upregulated**	A combined signature (13 markers) was significantly superior and yielded an AUC value of 0.93 in the validation cohort (exo-miRNA alone was 0.89). Validation was particularly good in identification of patients with stage I and II cancers.

There is good evidence that the miR-200 family is well-expressed in pancreatic tissue with early studies from 2007, (See Table 42) identifying the miR-200 family (hsa-miR-141, -148a, -200a, -200b, -200c) as upregulated in macro dissected pancreatic tissue compared with normal tissue in a small cohort of samples (n=25) (169). In 2009, Zhang *et al.* further characterised 95 miRNAs by RT-qPCR and found only that miR-200b (not other miR-200 members) was significantly upregulated in 22 PDAC tumours (324). Whilst Szafranska *et al.* used tissue samples of patients with chronic pancreatitis and normal pancreatic tissues as controls, Zhang *et al.* used adjacent macro dissected normal tissue from the same patients, which may have been more suitable for identifying tissue specific expression patterns, and therefore resulting in different results. A later study by Diaz-Riasco *et al.* investigated both macro and micro-dissected PDAC tissues (n=31) and identified that both miR-200a and miR-429 were upregulated in 45% and 61% of tumours respectively when compared to adjacent normal tissue (370). However, this conflicted with Dhayat *et al.* in 2018 who performed a similar study focussing on Stage II disease in 96 PDAC tumours compared with normal controls as well as patients with IPMN and chronic pancreatitis and this was reported that all 5 miR-200 members were in face downregulated in cancer (334). It appears therefore that whilst miR-200 is overexpressed in pancreatic PDAC tissue when macro dissected and compared with healthy controls,

these differences are less apparent (and may even be reversed) when more specific microdissection together with comparison to adjacent normal tissue takes place.

Interestingly, the downregulation of miR-200 (miR-200c in particular) was supported by two other studies from Korea and China respectively. Paik *et al.* investigated 84 FFPE tissue samples from surgical resection (79 PDAC but also an assortment of other cancers e.g. adenosquamous, anaplastic and undifferentiated) and found that miR-200a/200b/141 was upregulated whilst miR-200c was downregulated in comparison with the expression in HPDE cells (368). Liu *et al.* found low overall miR-200c mean expression in 75 PDAC tissues compared with adjacent normal pancreas. When miR-200c expression was dichotomised in patients, low expression was observed to be closely correlated with advanced clinical stage, higher incidence of lymph node metastasis and liver metastasis (369). This was contrary to the findings from Korea, who found instead that overall survival was longer in the patient group with low miR-200c expression.

Conversely, Yu *et al.* examined a large number (n=99) of PDAC tumours for miR-200c expression by RT-qPCR and finding upregulation in PDAC. They noted the positive correlation between miR-200c levels and E-cadherin mRNA levels (i.e. higher miR-200c levels are associated with a more epithelial phenotype) and that it inhibited cancer cell invasion leading to a tumour suppressive phenotype with a positive effect on overall survival (similar to the other Chinese study) (367). All these studies have a particular limitation in trying to evaluate expression in macro dissected historical samples fixed in formalin and embedded in paraffin (FFPE). They are likely to contain degraded RNA but this should be overcome through the use of adjacent tissue as an appropriate control. they pancreatic tissue, it may be that these findings are reflective of the underlying sample population. It is interesting to note that these three studies finding varying expression (decreased in two and increased in one) were from a cohort of patients eligible for surgery and therefore with earlier (Stage II) disease.

This heterogeneity of results is also mirrored by the cell line work. A study of 21 PDAC cell lines was undertaken by Kent *et al.* in the same year who used a custom microarray for the expression of 321 human miRNAs. They found that miR-200a was the most highly expressed miRNA in PDAC cell lines, with miR-200b and miR-429 (which are co-located with miR-200a in a cluster), also upregulated in the majority of cell lines examined. Yu et

al. however found that in their 15 cell lines, miR-200c specifically was downregulated in the majority of cells (despite having found them upregulated in tissue earlier).

6.1.2.2 The role of miR-200 as a regulator in EMT plasticity.

The miR-200 family has been shown to be an important modulator of the epithelial-to-mesenchymal transition (EMT), inhibiting it by acting on ZEB1 and ZEB2 (325). This function was demonstrated by overexpressing members of the miR-200 family using lentiviral plasmids in PANC-1 (PDAC) cells grown as spheroids. According to them, only overexpression of miR-429 affected growth and furthermore led to a decrease in protein levels of ZEB1, N-cadherin, and Vimentin expression, whilst E-cadherin expression was upregulated. This was also recapitulated in orthotopic xenografts of these cells in nude mice (n=13), leading to a reduction in PDAC tumour growth and metastatic spread with overexpression of miR-429 (and to a lesser extent miR-141). There is also strong evidence that ZEB1 and miR-200 family members repress expression of each other in a reciprocal feedback loop which is dysregulated in "RT2" and "KPC" mice, a preclinical model used for neuroendocrine tumours and PDAC respectively (371).

However, evidence that all members of the miR-200 family act via ZEB1 and ZEB2 is not definitive. Although both contain 3´UTRs regulated by miR-200a, miR-200b (372) and miR-200c (373), it is likely that their function is more complex and may require cooperative function with other miRNAs. For example, Kent *et al.* undertook follow up studies in a HPNE-Kras^{G12D} cell line (immortalised normal pancreas with a KRAS mutation common to PDAC) co-expressed with both miR-200a and miR-200b (322). They identified that enforced expression of the miR-200 family in (at least some) PDAC cell lines is positively correlated with E-cadherin expression and negatively correlated with ZEB1 expression, but that the effect on ZEB2 was less clear.

Activation of the EMT pathway through signals such as TGFß and FGF lead to stimulation of a group of well-recognised transcription factors (i.e. SNAI1, SLUG, TWIST, ZEB1/2) and allows for cancer cells to metastasise. This is through dedifferentiation, loss of apicobasal polarity, cell-cell adhesion and increasing motility, acquiring instead a mesenchymal phenotype (374). Initially it was thought that this mesenchymal phenotype with loss of cell-cell adherence through the repression of E-cadherin (Cdh1) expression was necessary to metastasis but in fact, reversal of this process, i.e. Mesenchymal to Epithelial transition (MET) via expression of miR-200 family members either

endogenously or by retroviral transduction can also paradoxically increase tumour invasiveness (albeit shown in breast cancer cells by Dykxhoorn *et al. (375)*). Cancers are unlikely to be present as a binary process, either epithelial or mesenchymal, but may instead exist in a continuum of partial EMT (pEMT) states (376). Even in PDAC cell lines, there is a range of phenotypes with more mesenchymal cell lines such as Mia-PaCa-2 showing lower miR-200 expression compared with epithelial-like BxPC-3 (with consequent higher miR-200 expression) (334).

This "EMT-MET plasticity" has been found to be a key feature in PDAC metastatic organotropism (i.e. whether they metastasise to liver or lung) with cancer cells that remain in a fixed mesenchymal state forming undifferentiated tumours but fewer liver metastases in murine models (377). It is therefore possible that miR-200 tissue expression can lead to both an epithelial phenotype and paradoxically increased metastatic disease in models.

The role of EMT in PDAC is not only limited to metastatic potential but also to its chemoresistance, a relative feature of both PDAC and CCA. A study using KPC mice with genetic knockouts of EMT transcription factors TWIST1 and SNAI1 did not reduce metastatic disease but did confer additional resistance to drugs such as GEM (378). Thus, although miR-200 has a known role in ZEB1/ZEB2 and EMT, its dysregulation in the primary tumour may not be the main mechanism by which it negatively impacts overall survival. As far back as 2009, we were aware that inhibition of miR-200 with antisense oligonucleotides in both CCA (379) or transfection of miR-200 to induce re-expression in PDAC (380) led to changes in GEM sensitivity *in vitro*. It is therefore likely that cancer related EV-miR-200 expression should act to induce a phenotypic change which could be detectable in either a murine or *in vitro* PDAC model.

Finally, it is worth mentioning that there are actions reported in the literature of miR-200 actions away from the ZEB1 axis. Bracken *et al.* showed strong evidence using the HITS-CLIP procedure to indicate sites of interaction between Ago and mRNA that both miR-200a and miR-200b coordinated effects on breast cancer cells leading to changes in cell shape and motility via Rho-ROCK signalling, invadopodia formation, MMP activity, and focal adhesions (381). Additionally, miR-200 (in particular miR-200c) has been shown to lead to beta cell apoptosis through a ZEB1 independent pathway and possibly through Ypel2 a potential oncogene, leading to type 2 diabetes. Increased miR-200c may therefore be one mechanism that leads to impaired glucose metabolism and new onset diabetes in patients with PDAC (382).

6.1.2.3 The presence of miR-200 in plasma EVs and a comparison between EV and cfRNA.

The miR-200 family is found to be increased in a number of solid cancers (breast, prostate, oesophageal, gastric, ovarian, and metastatic colorectal cancers) and is associated with poor prognosis when present in the plasma (326). Whilst no previous studies have identified EV-miR-200 in the bile, three studies have shown evidence for miR-200 family as present in EVs in blood and plasma in patients with PDAC/CCA and are discussed here.

As mentioned earlier in Chapter Five, Reese *et al.* (329), Nakamura et al (214) and Shen *et al.* (284) have examined blood based EVs from 429 patients altogether (PDAC n=224, CCA n=36) and identified members of the miR-200 family as overexpressed in plasma EVs. Whilst Reese et al. found overexpression of miR-200b and miR-200c in PDAC (and no difference for miR-200a), we have confirmed that all members of the miR-200 family are overexpressed in plasma EVs isolated by SEC. Although our study did not get the chance to follow patients up, two studies were able to correlate survival rates with miR-200c correlated with shorter OS in PDAC ($p = 0.038$) (329) and reduced tumour free survival in CCA ($p=0.005$) (284).

Interestingly, Dhayat *et al.* (the same group as Reese *et al.*) had examined the miR-200 family in serum cfRNA samples alongside tissue and although the miR-200 family was downregulated in PDAC compared to healthy controls, he did not find any relationship in their serum cfRNA miR-200 expression (334). This correlates with our findings which evaluated both plasma cfRNA and plasma EV-RNA, finding that the miR-200 family is differentially expressed in EVs but not statistically in the cfRNA. The first evidence for selective packaging of RNA into EVs comes from the landmark study by Valadi *et al.* who showed albeit indirectly, that EV-RNA was differentially expressed from their donor mast cells and were functional in recipient cells (200). Studies have consistently demonstrated this upregulation of particular miRNAs into EVs (383) but the evidence for active transport is not fully elucidated.

The main mechanism for protein sorting to exosomes is broadly described as a ubiquitin-dependent manner into MVBs via what is collectively known as the ESCRT (endosomal sorting complex required for transport) pathway (384). Initially, it was thought that this was the main pathway for miRNAs too, since common EV associated miRNAs (specifically

miR-16 and let-7a) are co-located in endosomal-MVB cellular fractions (385). However direct knockdown of an ESCRT-associated proteins (ALIX) appears to not alter extracellular miRNA levels in HEK293 cells (386). Kosaka et al. instead proposed that the main pathway for exosomal release of endogenous miR-16 and exogenous miR-146a was through the ceramide-dependent secretory machinery (CERT), which is modulated by knock-down of nSMase2 (neutral sphingomyelinase 2). These two processes (ESCRT and CERT) have interestingly since been reported to regulate secretion of EVs differently in polarised epithelial cells, with ceramide dependent transport on the basolateral aspect, and ESCRT controlling the apical release of EVs (387). Whether or not this is a process that occurs in the biliary epithelium would be of great interest to investigating the functionality of EVs in PDAC/CCA.

We have confirmed however that a large proportion of extracellular RNA is not vesicle-bound, and this cfRNA is co-purified with AGO2, a protein which is an effector in the RISC (192). By separating our vesicular fraction from extracellular protein by SEC we hoped to clearly distinguish these two compartments and therefore identify a source of biomarkers for the detection of PDAC/CCA. Whilst cfRNA would be a much more straightforward test, heterogeneity in results and present lack of validated markers from this avenue are notable (388). Our analysis shows that investigation of the EV compartment gives rise to a much more statistically significant differential expression than is present in the extracellular plasma alone.

If we hypothesise that there is active secretion of EV-miR-200 into the extracellular space then we have not yet identified why this might be a feature of PDAC or CCA. Incidentally, the miR-141/miR-200c cluster on chromosome 12 has been shown to be more important in regulating dedifferentiation and invasion than the miR-200a/miR-200b/miR-429 cluster (371). In our experiments however, we show stronger evidence that the miR-200a/miR-200b/miR-429 cluster is differentially expressed in our plasma and bile EVs (although both are expressed).

One possibility is that EV-miRNAs such as miR-200 are one of the mechanisms by which pre-metastatic niche formation is regulated alongside MIF and other cargo (389) to cause MET of circulating tumour cells contributing to distant metastatic formation (390). This seed and soil hypothesis was first noted by Stephen Paget in the 19th century, with only 2% of cells injected in mouse models lead to macroscopic tumours in experimental murine models (391). A landmark study showed that EVs from PDAC cells can "educate" murine models to increase metastatic burden (244) mediated by integrins, such as ITGα6 for lung-

tropic exosomes, ITGβ5 for liver-tropism and ITGβ3 in brain-tropism. One of the main differentially expressed markers in CCA EVs are integrins and this may be one of the determining features for cancer invasiveness (285). However it is yet to be determined whether or not EV-miRNA occurs at a functional effective concentration in the blood plasma of patients to cause this effect at distant sites.

6.2 Limitations

Our clinical data suggests that despite attempts to match age and sex in our control population, significant differences between our cancer and non-cancer cohorts including degree of biliary obstruction (serum bilirubin levels) and age, remained. The choice of control population to minimise differences is a potential drawback to any biomarker study and a feature of virtually every other study investigating EV-miRNAs. By recruiting patients agnostically at the time of endoscopy, it was hoped that we would be sampling a population that were as naïve as possible and would be close to the general population in terms of demographics. Furthermore, none of these patients received preoperative cancer treatment, and all blood specimens were obtained before initiation of any treatment.

Nodal status varied from 50-67% depending on the cohort of study and this was information obtained from radiological staging at the time of MDT. Lymph node staging is an important prognostic factor in PDAC and correlates with patient outcomes (392). Size threshold at CT for suspicion of lymph node involvement is 1cm but this has very poor sensitivity (14% in one study (393)) and it is therefore likely that the figures presented under call the true nodal involvement (in our study recorded as up to 67%). Radiological staging of nodal disease has been shown at best to have a sensitivity of 58% when counting nodes larger than 8mm (394). Reducing the number of samples to those patients who proceeded to surgery, particularly in PDAC and CCA would have made the project unfeasible (10-13% of patients were T2).

Our choice to perform small RNA Sequencing to discover potential candidates and the subsequent analysis will be likely to have introduced some element of bias, and no single method is optimal. This leads to a necessity for RT-qPCR validation which was performed in each of our studies (395). In our analysis we performed filtering to remove any candidates that had zero expression in >50% of samples provided limiting us to well-expressed miRNAs and tRFs that could be compared. Indeed, the use of any filter must be carefully taken into consideration, as this would eliminate potential candidates leading to a Type I error, but can also improve power (396). Alternative filtering could be undertaken to check whether there were any candidates that were both well-expressed only in cancer, and this is something that could be investigated in future.

Cell-free RNA as a source of markers was shown to have some drawbacks clinically. Heterogeneity of small RNA species was identified with proportions of miRNAs varying from 0 to 85% in bile samples. Bile-derived EV-miRNAs also exhibited variable

composition which led to 2 samples being excluded on the basis of low RNA input and outlier results when differentially analysed. Plasma cfRNA showed more consistent miRNA expression (mean miRNA composition 61% and range 37-90%) which indicated that plasma might be more suitable for miRNA analysis, but we were only able to identify two significant candidates for further evaluation (miR-144-5p and miR-29-3p). Over the course of the project, the method for RNA isolation and RT-qPCR was altered, which may have reduced the study's ability to make cross-comparisons between the cfRNA and EV-RNA. With improvements in sequencing technology and speed, it is hoped that small input will not be a barrier to future biomarker analysis.

The invasiveness of ERCP remains a barrier to diagnosis using ERCP bile aspiration and it is possible that future clinical practice will move further away from endoscopic drainage in order to reduce complications and improve survival in potentially resectable disease. Furthermore, difficulties with bile sampling volume means that the procedure is likely to yield a low clinical utility if the input is too low or too technically challenging. We demonstrated that PTC samples are also a potential source of EV-miRNAs with no differences detected.

Our study was also limited to one centre for the recruitment of patients and therefore would benefit from independent validation in a multicentre prospective study. The correlation of our results with other independent studies is reassuring and it is likely that bioinformatics analysis will be able to draw further conclusions from the increasing amounts of RNA-Seq data. By making our data publicly available, it will hopefully be a repository for future hypothesis driven discoveries.

6.3 Future Directions

With the results of this thesis having confirmed the role of EV and cfRNA as containing miRNAs with diagnostic potential in PDAC and CCA, there are several aspects which would be worth investigating.

6.3.1 Outstanding questions arising from this project.

Has the study has met its objective of identifying a clinically-meaningful diagnostic ncRNA signature for PDAC/CCA? Our aims were to identify and validate differentially expressed small non-coding RNAs within bile and plasma which would be able to distinguish between patients with PDAC and CCA. We have therefore evaluated both the bile and plasma extracellular RNA milieu and identified certain significant markers, particularly in EVs, that are upregulated in cancer.

Our method of isolation by SEC, was shown to be a quick method which demonstrated verifiable EVs and can be further shown to contain miRNAs (and tRFs) which may have a role in tumourigenesis. Their presence in relation to these EVs remains unclear, and further studies may warrant evaluation in conjunction with enzymatic treatment of these EVs to see whether they are truly present as cargo, present on the "surfaceome" (397) or simply co-localised. More sophisticated solutions may be utilised such as flow cytometry or super-resolution microscopy combining fluorophores with computational reconstruction of multiple images over time (398).

However as the focus was on a clinically meaningful diagnostic signature, the direction of our project will be primarily to verify plasma EV miR-200 in a separate centre and with a different sample population of patients with early-stage pancreatic cancer. Our bile EV miR-21/let-7a/miR-200a/miR-429 signature could be further evaluated in patients presenting with inflammatory benign conditions such as PSC or pancreatitis, as there were few patients from this cohort in our study. This is a clinically challenging area as the prevalence of PSC is low (6.1 per 100,000) and potentially underdiagnosed (399), meaning that population wide studies would be needed Definitive follow up for patients recruited was also not available, and our data was based on final overall diagnosis rather than intention-to-treat, which makes it difficult to know how these tests will perform in a real-world scenario.

6.3.2 Identifying the precursor gene cluster responsible for let-7a upregulation.

Let-7a exists as 3 paralogs, and it may be that only one cluster 9q22.3, 11q24.1 or 22q13.31 is responsible for this upregulation in bile. Cross-validation with tissue for the pri-miRNA responsible in samples that are upregulated for let-7a could focus attention on a particular cluster which is responsible for the differential expression and therefore could indicate associated regulatory pathways. It may also be possible to analyse this expression in existing gene data sets. It is unclear why let-7a should be stable in plasma and upregulated in our bile EVs. It would be interesting to examine whether cholangiocytes release EV-miRNAs with differential expression from their apical or basolateral surfaces. It may be that let-7a is preferentially secreted into the bile in cancer, or that this is a by-product of a "leaky" epithelium during tumourigenesis.

6.3.3 Determine any functional role for miR-200, miR-21 and let-7a in bile and plasma.

Whilst studies have been done *in vitro* to demonstrate miR-200 over-expression in PDAC and CCA, with more time, it would be helpful to evaluate SEC-EV preparations on PDAC cells to determine if these have a phenotypic effect at circulating concentrations on cell growth, proliferation and survival. EV preparations with high expression of EV-miR-200/let-7a/miR-21 could be compared with low expressed EVs and applied to PDAC cells. We could also evaluate their role on tumour-associated macrophages. A more sophisticated study might involve murine models utilising the principle of EV education to inspect the role of EV-miRNAs. The first genetically engineered mouse model (GEMM) of PDAC is termed "KC" and involves a targeted expression of an endogenous (knocked-in) $KRAS^{G12D}$ oncogene in mice pancreata, through interbreeding animals that express Cre recombinase from the pancreatic-specific promoters, PDX-1 or P48, using the well-known Cre-LoxP process (400). This generates mutant mice with ductal lesions identical to human PanINs with complete penetrance. Subsequent progression through addition of mutations such as the Li-Fraumeni human ortholog, $TP53^{R175H}$ are present in the "KPC" murine model which leads to a dramatically shortened survival (100% mortality by 12 months) (401). EV education by retroorbital injection in KC and KPC mice (402) with confirmed EV-miR-200/let-7a/miR-21 cargo followed by tumour and metastatic burden analysis could be a way to evaluate the role of EVs in organotropism and/or chemoresistance. We hypothesise that these miRNAs are likely to be pro-metastatic and lead to upregulation of MET markers such as E-cadherin.

6.3.4 EV miR-200, miR-21 and let-7a within early precursor lesions.

Our study was limited by the natural history of PDAC/CCA which led to a predominance of advanced disease in our clinical cohort. Earlier models of PanIN progression and EV-miRNA expression from these precursor lesions are warranted. A clinical study could investigate prospectively patients with early lesions such as IPMNs and PanINs who are undergoing EUS-FNA for diagnosis. This would be a much lengthier study due to the lower presentation rate of these diseases.

It is thought that miR-21 is upregulated in early invasive IPMN and at PanIN 2/3 but this has not been demonstrated *in vitro*. The role of organoids, particularly PanIN organoids are a potential novel source of information. Cell-conditioned media from these might shed light on the source of EV-miR-200.

Patient derived or GEMM derived organoids can also generate ductule models and could be a source of EV-miRNA which would recapitulate the miRNA expression of PDAC or CCA (279). Tumours are believed to be made up of clonal subtypes which are complicated to sequence and lead to heterogeneity. The role of single cell RNA-Seq in early organoid models would be the next step to understanding the life cycle of progressive disease but is currently not possible for miRNA analysis.

6.3.5 Relationship of EV-miRNAs to molecular subtype and survival.

Our study was also limited by the lack of patient tissue typing. With new knowledge on molecular subtypes and trials underway to sequence PDAC and CCA tumours, it would be helpful to evaluate let-7a/miR-21/miR-200 expression in this context. Survival data was not available for all patients, and this also limited the outcomes measured.

Pre- and post-operative plasma samples appeared to indicate a role for miR-200 prognostication. We would need to undertake a larger study to evaluate if decreased miR-200 expression led to a better prognosis in the same way as CA 19-9. Our small feasibility study demonstrated that using patients who have had a laparotomy, but no surgical resection would be a suitable control in this study. We hypothesise that these markers will be downregulated in patients and could be an early sign of tumour recurrence.

6.3.6 Novel tRFs may be a future promising source of biomarkers.

This was the first study to have analysed for differential expression the presence of tRFs within EVs. In our bile samples we found that tRFs comprised 10.416% of overall reads compared with an average of 8.525% for miRNAs in our discovery sequencing. Validating the presence of tRFs in our EV samples from bile and plasma would be a novel direction and potentially lead to a new area of differential expression in cancer. Our four tRF analyses (bile cfRNA, MIRABILE EV, surgical bile EV and plasma EV) all identified a significant 5' tRF bias, with very few 3' tRFs identified. This has been shown in other studies, including by Nolte-'t Hoen et al. in dendritic cell derived EVs (201), Tosar et al. in breast cancer derived EVs (403) and Zhu et al. in EVs isolated from SK-Hep1 liver cancer cells (337). One possible explanation is that these are preferentially bound to protective protein complexes, which has been found to occur to 5'tRNA halves in blood (404). As tRNA sequencing is in its infancy, there are also ongoing methods being developed to prevent any bias caused by technical factors such as ligation of adapters during library preparation (405).

In summary, our results showed that differential expression of small RNAs (miRNAs and tRFs) can be both identified through RNA sequencing of bile and plasma derived EVs isolated using SEC and validated as suitable for the diagnosis of PDAC and CCA.

Our final bile EV models generated an outstanding AUC value of 0.8249 in the detection of malignancy, rising to 0.9150 when combined with CA 19-9 but with a reduction in specificity. This four-marker panel (let-7a/miR-21/miR-200a/miR-429) had a validated AUC of 0.961 in the detection of PDAC only with a sensitivity of 85.19% and specificity of 91.89%.

Our final plasma EV model generated a validated AUC of 0.8226 in the detection of PDAC, rising to 0.9617 when combined with CA 19-9 giving a sensitivity of 89.29% and a specificity of 100%. Individually miR-200a-3p demonstrated an AUC of 0.7828, miR-200b-3p demonstrated an AUC of 0.7022, miR-200c-3p demonstrated an AUC of 0.7280, miR-141-3p demonstrated an AUC of 0.7645 and miR-429 demonstrated a lower AUC of 0.6232.

We were also able to detect a significant increase in EV concentration in the plasma of patients with PDAC. Individually, EV concentration demonstrated an AUC of 0.7742

whilst median particle size demonstrated an AUC of 0.8349. If combined with the miR-200 family, we were able to predict PDAC with an outstanding AUC 0.9118.

Our benign controls also included patients with chronic pancreatitis and was undertaken amongst a population which has a similar age-sex composition. Although this did not appear to corroborate with any previous studies investigating bile EV-miRNAs, the results here show novel biomarkers that may aid the detection of PDAC and biliary tract cancers. The presence of well-established oncomiRs in combination with miR-200 family members indicates a wealth of functional significance with roles in EMT and chemoresistance building on the results of other studies already published.

Bibliography

1. Kumar V, Abbas AK, Fausto N, Aster JC. Robbins and Cotran pathologic basis of disease, professional edition e-book: Elsevier health sciences; 2014.
2. Luo G, Fan Z, Gong Y, Jin K, Yang C, Cheng H, et al. Characteristics and Outcomes of Pancreatic Cancer by Histological Subtypes. Pancreas. 2019;48(6):817-22.
3. Ordonez NG. Pancreatic acinar cell carcinoma. Adv Anat Pathol. 2001;8(3):144-59.
4. UK CR. Pancreatic Cancer Statistics [Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/pancreatic-cancer.
5. Ferlay J, Steliarova-Foucher E, Lortet-Tieulent J, Rosso S, Coebergh JW, Comber H, et al. Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012. Eur J Cancer. 2013;49(6):1374-403.
6. Hamada T, Yuan C, Yurgelun MB, Perez K, Khalaf N, Morales-Oyarvide V, et al. Family history of cancer, Ashkenazi Jewish ancestry, and pancreatic cancer risk. Br J Cancer. 2019;120(8):848-54.
7. Wolpin BM, Chan AT, Hartge P, Chanock SJ, Kraft P, Hunter DJ, et al. ABO blood group and the risk of pancreatic cancer. J Natl Cancer Inst. 2009;101(6):424-31.
8. Yadav D, Lowenfels AB. The epidemiology of pancreatitis and pancreatic cancer. Gastroenterology. 2013;144(6):1252-61.
9. Parkin DM, Boyd L, Walker LC. 16. The fraction of cancer attributable to lifestyle and environmental factors in the UK in 2010. Br J Cancer. 2011;105 Suppl 2(Suppl 2):S77-81.
10. Calle EE, Rodriguez C, Walker-Thurmond K, Thun MJ. Overweight, obesity, and mortality from cancer in a prospectively studied cohort of U.S. adults. N Engl J Med. 2003;348(17):1625-38.
11. Aune D, Greenwood DC, Chan DS, Vieira R, Vieira AR, Navarro Rosenblatt DA, et al. Body mass index, abdominal fatness and pancreatic cancer risk: a systematic review and non-linear dose-response meta-analysis of prospective studies. Ann Oncol. 2012;23(4):843-52.
12. Jiao L, Mitrou PN, Reedy J, Graubard BI, Hollenbeck AR, Schatzkin A, et al. A combined healthy lifestyle score and risk of pancreatic cancer in a large cohort study. Arch Intern Med. 2009;169(8):764-70.
13. Park SK, Oh CM, Kim MH, Ha E, Choi YS, Ryoo JH. Metabolic syndrome, metabolic components, and their relation to the risk of pancreatic cancer. Cancer. 2020;126(9):1979-86.
14. Batabyal P, Vander Hoorn S, Christophi C, Nikfarjam M. Association of diabetes mellitus and pancreatic adenocarcinoma: a meta-analysis of 88 studies. Ann Surg Oncol. 2014;21(7):2453-62.
15. Wang YT, Gou YW, Jin WW, Xiao M, Fang HY. Association between alcohol intake and the risk of pancreatic cancer: a dose-response meta-analysis of cohort studies. BMC Cancer. 2016;16(1):212.
16. Midha S, Chawla S, Garg PK. Modifiable and non-modifiable risk factors for pancreatic cancer: A review. Cancer Lett. 2016;381(1):269-77.
17. Raimondi S, Lowenfels AB, Morselli-Labate AM, Maisonneuve P, Pezzilli R. Pancreatic cancer in chronic pancreatitis; aetiology, incidence, and early detection. Best Pract Res Clin Gastroenterol. 2010;24(3):349-58.
18. Maisonneuve P, FitzSimmons SC, Neglia JP, Campbell PW, 3rd, Lowenfels AB. Cancer risk in nontransplanted and transplanted cystic fibrosis patients: a 10-year study. J Natl Cancer Inst. 2003;95(5):381-7.

19. Enjuto DT, Herrera N, C JC, Ramos Bonilla A, Llorente-Lazaro R, Gonzalez Guerreiro J, et al. Hereditary Pancreatitis Related to SPINK-1 Mutation. Is There an Increased Risk of Developing Pancreatic Cancer? J Gastrointest Cancer. 2023;54(1):268-9.
20. Weiss FU. Pancreatic cancer risk in hereditary pancreatitis. Front Physiol. 2014;5:70.
21. Koorstra JB, Feldmann G, Habbe N, Maitra A. Morphogenesis of pancreatic cancer: role of pancreatic intraepithelial neoplasia (PanINs). Langenbecks Arch Surg. 2008;393(4):561-70.
22. Klein WM, Hruban RH, Klein-Szanto AJ, Wilentz RE. Direct correlation between proliferative activity and dysplasia in pancreatic intraepithelial neoplasia (PanIN): additional evidence for a recently proposed model of progression. Modern pathology. 2002;15(4):441-7.
23. Hackeng WM, Hruban RH, Offerhaus GJ, Brosens LA. Surgical and molecular pathology of pancreatic neoplasms. Diagn Pathol. 2016;11(1):47.
24. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science. 2008;321(5897):1801-6.
25. Hruban RH, Goggins M, Parsons J, Kern SE. Progression model for pancreatic cancer. Clin Cancer Res. 2000;6(8):2969-72.
26. Almoguera C, Shibata D, Forrester K, Martin J, Arnheim N, Perucho M. Most human carcinomas of the exocrine pancreas contain mutant c-K-ras genes. Cell. 1988;53(4):549-54.
27. Guerra C, Barbacid M. Genetically engineered mouse models of pancreatic adenocarcinoma. Mol Oncol. 2013;7(2):232-47.
28. Bournet B, Muscari F, Buscail C, Assenat E, Barthet M, Hammel P, et al. KRAS G12D Mutation Subtype Is A Prognostic Factor for Advanced Pancreatic Adenocarcinoma. Clin Transl Gastroenterol. 2016;7(3):e157.
29. Waters AM, Der CJ. KRAS: The Critical Driver and Therapeutic Target for Pancreatic Cancer. Cold Spring Harb Perspect Med. 2018;8(9).
30. McWilliams RR, Wieben ED, Rabe KG, Pedersen KS, Wu Y, Sicotte H, et al. Prevalence of CDKN2A mutations in pancreatic cancer patients: implications for genetic counseling. Eur J Hum Genet. 2011;19(4):472-8.
31. Wilentz RE, Iacobuzio-Donahue CA, Argani P, McCarthy DM, Parsons JL, Yeo CJ, et al. Loss of expression of Dpc4 in pancreatic intraepithelial neoplasia: Evidence that DPC4 inactivation occurs late in neoplastic progression. Cancer Res. 2000;60(7):2002-6.
32. Maltzman W, Czyzyk L. UV irradiation stimulates levels of p53 cellular tumor antigen in nontransformed mouse cells. Mol Cell Biol. 1984;4(9):1689-94.
33. Sonoyama T, Sakai A, Mita Y, Yasuda Y, Kawamoto H, Yagi T, et al. TP53 codon 72 polymorphism is associated with pancreatic cancer risk in males, smokers and drinkers. Molecular medicine reports. 2011;4(3):489-95.
34. Chang C-J, Chao C-H, Xia W, Yang J-Y, Xiong Y, Li C-W, et al. p53 regulates epithelial–mesenchymal transition and stem cell properties through modulating miRNAs. Nat Cell Biol. 2011;13(3):317-23.
35. Jahedi H, Fahud AL, Lim CL. Role of p53 family isoforms in enhancing aggressiveness and chemoresistance in pancreatic cancer. World Academy of Sciences Journal. 2019;1(5):236-46.
36. Zhang J, Tian X-J, Xing J. Signal transduction pathways of EMT induced by TGF-β, SHH, and WNT and their crosstalks. Journal of clinical medicine. 2016;5(4):41.
37. Aloni-Grinstein R, Shetzer Y, Kaufman T, Rotter V. p53: the barrier to cancer stem cell formation. FEBS Lett. 2014;588(16):2580-9.

38. Li C, Heidt DG, Dalerba P, Burant CF, Zhang L, Adsay V, et al. Identification of pancreatic cancer stem cells. Cancer Res. 2007;67(3):1030-7.
39. Patil K, Khan FB, Akhtar S, Ahmad A, Uddin S. The plasticity of pancreatic cancer stem cells: implications in therapeutic resistance. Cancer Metastasis Rev. 2021;40(3):691-720.
40. McGrath NA, Fu J, Gu SZ, Xie C. Targeting cancer stem cells in cholangiocarcinoma (Review). Int J Oncol. 2020;57(2):397-408.
41. Zhong Z, Harmston N, Wood KC, Madan B, Virshup DM. A p300/GATA6 axis determines differentiation and Wnt dependency in pancreatic cancer models. J Clin Invest. 2022;132(12).
42. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature. 2010;467(7319):1114-7.
43. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med. 2011;17(4):500-3.
44. Aung KL, Fischer SE, Denroche RE, Jang GH, Dodd A, Creighton S, et al. Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS Trial. Clin Cancer Res. 2018;24(6):1344-54.
45. Yao W, Maitra A, Ying H. Recent insights into the biology of pancreatic cancer. EBioMedicine. 2020;53:102655.
46. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 2015;47(10):1168-78.
47. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016;531(7592):47-52.
48. Hoadley KA, Yau C, Wolf DM, Cherniack AD, Tamborero D, Ng S, et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. Cell. 2014;158(4):929-44.
49. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature. 2015;518(7540):495-501.
50. Janky R, Binda MM, Allemeersch J, Van den Broeck A, Govaere O, Swinnen JV, et al. Prognostic relevance of molecular subtypes and master regulators in pancreatic ductal adenocarcinoma. BMC Cancer. 2016;16(1):632.
51. Puleo F, Nicolle R, Blum Y, Cros J, Marisa L, Demetter P, et al. Stratification of Pancreatic Ductal Adenocarcinomas Based on Tumor and Microenvironment Features. Gastroenterology. 2018;155(6):1999-2013 e3.
52. Collisson EA, Bailey P, Chang DK, Biankin AV. Molecular subtypes of pancreatic cancer. Nat Rev Gastroenterol Hepatol. 2019;16(4):207-20.
53. Dougan SK. The Pancreatic Cancer Microenvironment. Cancer J. 2017;23(6):321-5.
54. Robin F, Angenard G, Cano L, Courtin-Tanguy L, Gaignard E, Khene ZE, et al. Molecular profiling of stroma highlights stratifin as a novel biomarker of poor prognosis in pancreatic ductal adenocarcinoma. Br J Cancer. 2020;123(1):72-80.
55. Erkan M, Reiser-Erkan C, Michalski CW, Kong B, Esposito I, Friess H, et al. The impact of the activated stroma on pancreatic ductal adenocarcinoma biology and therapy resistance. Curr Mol Med. 2012;12(3):288-303.
56. Ozdemir BC, Pentcheva-Hoang T, Carstens JL, Zheng X, Wu CC, Simpson TR, et al. Depletion of carcinoma-associated fibroblasts and fibrosis induces

immunosuppression and accelerates pancreas cancer with reduced survival. Cancer Cell. 2014;25(6):719-34.
57. Riquelme E, Zhang Y, Zhang L, Montiel M, Zoltan M, Dong W, et al. Tumor Microbiome Diversity and Composition Influence Pancreatic Cancer Outcomes. Cell. 2019;178(4):795-806 e12.
58. Cardinale V, Carpino G, Reid L, Gaudio E, Alvaro D. Multiple cells of origin in cholangiocarcinoma underlie biological, epidemiological and clinical heterogeneity. World J Gastrointest Oncol. 2012;4(5):94-102.
59. Nakeeb A, Pitt HA, Sohn TA, Coleman J, Abrams RA, Piantadosi S, et al. Cholangiocarcinoma. A spectrum of intrahepatic, perihilar, and distal tumors. Ann Surg. 1996;224(4):463-73; discussion 73-5.
60. England PH. National Cancer Intelligence Network: Rare and less common cancers - Incidence and Mortality in England, 2010 to 2013. 2015.
61. Khan SA, Emadossadaty S, Ladep NG, Thomas HC, Elliott P, Taylor-Robinson SD, et al. Rising trends in cholangiocarcinoma: is the ICD classification system misleading us? Journal of hepatology. 2012;56(4):848-54.
62. Chapman MH, Thorburn D, Hirschfield GM, Webster GGJ, Rushbrook SM, Alexander G, et al. British Society of Gastroenterology and UK-PSC guidelines for the diagnosis and management of primary sclerosing cholangitis. Gut. 2019;68(8):1356-78.
63. Alsaleh M, Leftley Z, Barbera TA, Sithithaworn P, Khuntikeo N, Loilome W, et al. Cholangiocarcinoma: a guide for the nonspecialist. Int J Gen Med. 2019;12:13-23.
64. Sastry AV, Abbadessa B, Wayne MG, Steele JG, Cooperman AM. What is the incidence of biliary carcinoma in choledochal cysts, when do they develop, and how should it affect management? World journal of surgery. 2015;39(2):487-92.
65. Churi CR, Shroff R, Wang Y, Rashid A, Kang HC, Weatherly J, et al. Mutation profiling in cholangiocarcinoma: prognostic and therapeutic implications. PLoS One. 2014;9(12):e115383.
66. Andersen JB, Thorgeirsson SS. Genetic profiling of intrahepatic cholangiocarcinoma. Curr Opin Gastroenterol. 2012;28(3):266-72.
67. Kipp BR, Voss JS, Kerr SE, Barr Fritcher EG, Graham RP, Zhang L, et al. Isocitrate dehydrogenase 1 and 2 mutations in cholangiocarcinoma. Hum Pathol. 2012;43(10):1552-8.
68. Nakanuma Y. A novel approach to biliary tract pathology based on similarities to pancreatic counterparts: is the biliary tract an incomplete pancreas? Pathology international. 2010;60(6):419-29.
69. Zen Y, Fujii T, Itatsu K, Nakamura K, Minato H, Kasashima S, et al. Biliary papillary tumors share pathological features with intraductal papillary mucinous neoplasm of the pancreas. Hepatology. 2006;44(5):1333-43.
70. Lee H, Wang K, Johnson A, Jones DM, Ali SM, Elvin JA, et al. Comprehensive genomic profiling of extrahepatic cholangiocarcinoma reveals a long tail of therapeutic targets. J Clin Pathol. 2016;69(5):403-8.
71. Galdy S, Lamarca A, McNamara MG, Hubner RA, Cella CA, Fazio N, et al. HER2/HER3 pathway in biliary tract malignancies; systematic review and meta-analysis: a potential therapeutic target? Cancer Metastasis Rev. 2017;36(1):141-57.
72. Banales JM, Marin JJG, Lamarca A, Rodrigues PM, Khan SA, Roberts LR, et al. Cholangiocarcinoma 2020: the next horizon in mechanisms and management. Nat Rev Gastroenterol Hepatol. 2020;17(9):557-88.
73. Liao W, Clift AK, Patone M, Coupland C, Gonzalez-Izquierdo A, Pereira SP, et al. Identifying symptoms associated with diagnosis of pancreatic exocrine and neuroendocrine neoplasms: a nested case-control study of the UK primary care population. Br J Gen Pract. 2021;71(712):e836-e45.

74. Kalser MH, Barkin J, MacIntyre JM. Pancreatic cancer. Assessment of prognosis by clinical presentation. Cancer. 1985;56(2):397-402.
75. Hippisley-Cox J, Coupland C. Identifying patients with suspected pancreatic cancer in primary care: derivation and validation of an algorithm. Br J Gen Pract. 2012;62(594):e38-45.
76. Walter FM, Mills K, Mendonca SC, Abel GA, Basu B, Carroll N, et al. Symptoms and patient factors associated with diagnostic intervals for pancreatic cancer (SYMPTOM pancreatic study): a prospective cohort study. Lancet Gastroenterol Hepatol. 2016;1(4):298-306.
77. Ducreux M, Cuhna AS, Caramella C, Hollebecque A, Burtin P, Goere D, et al. Cancer of the pancreas: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2015;26 Suppl 5:v56-68.
78. Elbanna KY, Jang HJ, Kim TK. Imaging diagnosis and staging of pancreatic ductal adenocarcinoma: a comprehensive review. Insights Imaging. 2020;11(1):58.
79. Best LM, Rawji V, Pereira SP, Davidson BR, Gurusamy KS. Imaging modalities for characterising focal pancreatic lesions. Cochrane Database Syst Rev. 2017;4(4):CD010213.
80. Burnett AS, Calvert TJ, Chokshi RJ. Sensitivity of endoscopic retrograde cholangiopancreatography standard cytology: 10-y review of the literature. J Surg Res. 2013;184(1):304-11.
81. Navaneethan U, Njei B, Lourdusamy V, Konjeti R, Vargo JJ, Parsi MA. Comparative effectiveness of biliary brush cytology and intraductal biopsy for detection of malignant biliary strictures: a systematic review and meta-analysis. Gastrointest Endosc. 2015;81(1):168-76.
82. De Moura DTH, Moura EGH, Bernardo WM, De Moura ETH, Baraca FI, Kondo A, et al. Endoscopic retrograde cholangiopancreatography versus endoscopic ultrasound for tissue diagnosis of malignant biliary stricture: Systematic review and meta-analysis. Endosc Ultrasound. 2018;7(1):10-9.
83. Sadeghi A, Mohamadnejad M, Islami F, Keshtkar A, Biglari M, Malekzadeh R, et al. Diagnostic yield of EUS-guided FNA for malignant biliary stricture: a systematic review and meta-analysis. Gastrointest Endosc. 2016;83(2):290-8 e1.
84. Hewitt MJ, McPhail MJ, Possamai L, Dhar A, Vlavianos P, Monahan KJ. EUS-guided FNA for diagnosis of solid pancreatic neoplasms: a meta-analysis. Gastrointest Endosc. 2012;75(2):319-31.
85. Heimbach JK, Sanchez W, Rosen CB, Gores GJ. Trans-peritoneal fine needle aspiration biopsy of hilar cholangiocarcinoma is associated with disease dissemination. HPB (Oxford). 2011;13(5):356-60.
86. Yarandi SS, Runge T, Wang L, Liu Z, Jiang Y, Chawla S, et al. Increased Incidence of Benign Pancreatic Pathology following Pancreaticoduodenectomy for Presumed Malignancy over 10 Years despite Increased Use of Endoscopic Ultrasound. Diagn Ther Endosc. 2014;2014:701535.
87. Pancreatic S. Guidelines for the management of patients with pancreatic cancer periampullary and ampullary carcinomas. Gut. 2005;54(Suppl 5):v1.
88. Koprowski H, Herlyn M, Steplewski Z, Sears HF. Specific antigen in serum of patients with colon carcinoma. Science. 1981;212(4490):53-5.
89. Patel AH, Harnois DM, Klee GG, LaRusso NF, Gores GJ. The utility of CA 19-9 in the diagnoses of cholangiocarcinoma in patients without primary sclerosing cholangitis. Am J Gastroenterol. 2000;95(1):204-7.
90. Goonetilleke KS, Siriwardena AK. Systematic review of carbohydrate antigen (CA 19-9) as a biochemical marker in the diagnosis of pancreatic cancer. Eur J Surg Oncol. 2007;33(3):266-70.

91. Humphris JL, Chang DK, Johns AL, Scarlett CJ, Pajic M, Jones MD, et al. The prognostic and predictive value of serum CA19.9 in pancreatic cancer. Ann Oncol. 2012;23(7):1713-22.
92. Rosen A, Linder S, Harmenberg U, Pegert S. Serum Levels of CA 19-9 and CA 50 in Relation to Lewis Blood Cell Status in Patients with Malignant and Benign Pancreatic Disease. Pancreas. 1993;8(2):160-5.
93. Brody JR, Witkiewicz AK, Yeo CJ. The past, present, and future of biomarkers: a need for molecular beacons for the clinical management of pancreatic cancer. Adv Surg. 2011;45(1):301-21.
94. Mann DV, Edwards R, Ho S, Lau WY, Glazer G. Elevated tumour marker CA19-9: clinical interpretation and influence of obstructive jaundice. Eur J Surg Oncol. 2000;26(5):474-9.
95. Ermiah E, Eddfair M, Abdulrahman O, Elfagieh M, Jebriel A, Al-Sharif M, et al. Prognostic value of serum CEA and CA19-9 levels in pancreatic ductal adenocarcinoma. Mol Clin Oncol. 2022;17(2):126.
96. Meng Q, Shi S, Liang C, Xiang J, Liang D, Zhang B, et al. Diagnostic Accuracy of a CA125-Based Biomarker Panel in Patients with Pancreatic Cancer: A Systematic Review and Meta-Analysis. J Cancer. 2017;8(17):3615-22.
97. Meng Q, Shi S, Liang C, Liang D, Xu W, Ji S, et al. Diagnostic and prognostic value of carcinoembryonic antigen in pancreatic cancer: a systematic review and meta-analysis. Onco Targets Ther. 2017;10:4591-8.
98. Arkadopoulos N, Kyriazi MA, Papanikolaou IS, Vasiliou P, Theodoraki K, Lappas C, et al. Preoperative biliary drainage of severely jaundiced patients increases morbidity of pancreaticoduodenectomy: results of a case-control study. World J Surg. 2014;38(11):2967-72.
99. French JJ, Mansfield SD, Jaques K, Jaques BC, Manas DM, Charnley RM. Fast-track management of patients undergoing proximal pancreatic resection. Ann R Coll Surg Engl. 2009;91(3):201-4.
100. Pande R, Hodson J, Marudanayagam R, Chatzizacharias N, Dasari B, Muiesan P, et al. Survival Advantage of Upfront Surgery for Pancreatic Head Cancer Without Preoperative Biliary Drainage. Front Oncol. 2020;10:526514.
101. Malleo G, Maggino L, Capelli P, Gulino F, Segattini S, Scarpa A, et al. Reappraisal of Nodal Staging and Study of Lymph Node Station Involvement in Pancreaticoduodenectomy with the Standard International Study Group of Pancreatic Surgery Definition of Lymphadenectomy for Cancer. J Am Coll Surg. 2015;221(2):367-79 e4.
102. Vernuccio F, Messina C, Merz V, Cannella R, Midiri M. Resectable and Borderline Resectable Pancreatic Ductal Adenocarcinoma: Role of the Radiologist and Oncologist in the Era of Precision Medicine. Diagnostics (Basel). 2021;11(11).
103. Diener MK, Knaebel HP, Heukaufer C, Antes G, Buchler MW, Seiler CM. A systematic review and meta-analysis of pylorus-preserving versus classical pancreaticoduodenectomy for surgical treatment of periampullary and pancreatic carcinoma. Ann Surg. 2007;245(2):187-200.
104. Klaiber U, Mihaljevic A, Hackert T. Radical pancreatic cancer surgery-with arterial resection. Transl Gastroenterol Hepatol. 2019;4:8.
105. Miyazaki M, Yoshitomi H, Takano S, Shimizu H, Kato A, Yoshidome H, et al. Combined hepatic arterial resection in pancreatic resections for locally advanced pancreatic cancer. Langenbecks Arch Surg. 2017;402(3):447-56.
106. Schnelldorfer T, Ware AL, Sarr MG, Smyrk TC, Zhang L, Qin R, et al. Long-term survival after pancreatoduodenectomy for pancreatic adenocarcinoma: is cure possible? Ann Surg. 2008;247(3):456-62.

107. Ghafouri-Fard S, Abak A, Tondro Anamag F, Shoorei H, Fattahi F, Javadinia SA, et al. 5-Fluorouracil: A Narrative Review on the Role of Regulatory Mechanisms in Driving Resistance to This Chemotherapeutic Agent. Front Oncol. 2021;11:658636.
108. Neoptolemos JP, Stocken DD, Friess H, Bassi C, Dunn JA, Hickey H, et al. A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer. N Engl J Med. 2004;350(12):1200-10.
109. Burris HA, 3rd, Moore MJ, Andersen J, Green MR, Rothenberg ML, Modiano MR, et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: a randomized trial. J Clin Oncol. 1997;15(6):2403-13.
110. Neoptolemos JP, Palmer DH, Ghaneh P, Psarelli EE, Valle JW, Halloran CM, et al. Comparison of adjuvant gemcitabine and capecitabine with gemcitabine monotherapy in patients with resected pancreatic cancer (ESPAC-4): a multicentre, open-label, randomised, phase 3 trial. Lancet. 2017;389(10073):1011-24.
111. Von Hoff DD, Ervin T, Arena FP, Chiorean EG, Infante J, Moore M, et al. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. N Engl J Med. 2013;369(18):1691-703.
112. Frese KK, Neesse A, Cook N, Bapiro TE, Lolkema MP, Jodrell DI, et al. nab-Paclitaxel potentiates gemcitabine activity by reducing cytidine deaminase levels in a mouse model of pancreatic cancer. Cancer Discov. 2012;2(3):260-9.
113. Conroy T, Hammel P, Hebbar M, Ben Abdelghani M, Wei AC, Raoul JL, et al. FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer. N Engl J Med. 2018;379(25):2395-406.
114. Katayama ES, Hue JJ, Bajor DL, Ocuin LM, Ammori JB, Hardacre JM, et al. A comprehensive analysis of clinical trials in pancreatic cancer: what is coming down the pike? Oncotarget. 2020;11(38):3489.
115. Lythgoe MP, Krell J, Mahmoud S, Mills EC, Vasudevan A, Savage P. Development and economic trends in anticancer drugs licensed in the UK from 2015 to 2019. Drug Discov Today. 2021;26(2):301-7.
116. Lythgoe MP, Desai A, Gyawali B, Savage P, Krell J, Warner JL, et al. Cancer Therapy Approval Timings, Review Speed, and Publication of Pivotal Registration Trials in the US and Europe, 2010-2019. JAMA Netw Open. 2022;5(6):e2216183.
117. Zhao L, Singh V, Ricca A, Lee P. Survival Benefit of Pembrolizumab for Patients With Pancreatic Adenocarcinoma: A Case Series. J Med Cases. 2022;13(5):240-3.
118. Fietkau R, Ghadimi M, Grützmann R, Wittel UA, Jacobasch L, Uhl W, et al. Randomized phase III trial of induction chemotherapy followed by chemoradiotherapy or chemotherapy alone for nonresectable locally advanced pancreatic cancer: First results of the CONKO-007 trial. American Society of Clinical Oncology; 2022.
119. Hammel P, Huguet F, van Laethem JL, Goldstein D, Glimelius B, Artru P, et al. Effect of Chemoradiotherapy vs Chemotherapy on Survival in Patients With Locally Advanced Pancreatic Cancer Controlled After 4 Months of Gemcitabine With or Without Erlotinib: The LAP07 Randomized Clinical Trial. JAMA. 2016;315(17):1844-53.
120. Strauss VY, Shaw R, Virdee PS, Hurt CN, Ward E, Tranter B, et al. Study protocol: a multi-centre randomised study of induction chemotherapy followed by capecitabine +/- nelfinavir with high- or standard-dose radiotherapy for locally advanced pancreatic cancer (SCALOP-2). BMC Cancer. 2019;19(1):121.
121. Primrose JN, Fox RP, Palmer DH, Malik HZ, Prasad R, Mirza D, et al. Capecitabine compared with observation in resected biliary tract cancer (BILCAP): a randomised, controlled, multicentre, phase 3 study. Lancet Oncol. 2019;20(5):663-73.
122. Lamarca A, Palmer DH, Wasan HS, Ross PJ, Ma YT, Arora A, et al. Second-line FOLFOX chemotherapy versus active symptom control for advanced biliary tract cancer

(ABC-06): a phase 3, open-label, randomised, controlled trial. Lancet Oncol. 2021;22(5):690-701.
123. Oh DY, Lee KH, Lee DW, Yoon J, Kim TY, Bang JH, et al. Gemcitabine and cisplatin plus durvalumab with or without tremelimumab in chemotherapy-naive patients with advanced biliary tract cancer: an open-label, single-centre, phase 2 study. Lancet Gastroenterol Hepatol. 2022;7(6):522-32.
124. Akhoundova Sanoyan D, McNamara MG, Lamarca A, Valle JW. Adjuvant chemotherapy in biliary tract cancer: state of the art and future perspectives. Curr Opin Oncol. 2020;32(4):364-9.
125. McMasters KM, Tuttle TM, Leach SD, Rich T, Cleary KR, Evans DB, et al. Neoadjuvant chemoradiation for extrahepatic cholangiocarcinoma. Am J Surg. 1997;174(6):605-8; discussion 8-9.
126. Yadav S, Xie H, Bin-Riaz I, Sharma P, Durani U, Goyal G, et al. Neoadjuvant vs. adjuvant chemotherapy for cholangiocarcinoma: A propensity score matched analysis. Eur J Surg Oncol. 2019;45(8):1432-8.
127. Lee RC, Feinbaum RL, Ambros V. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell. 1993;75(5):843-54.
128. Hertel J, Bartschat S, Wintsche A, Otto C, Students of the Bioinformatics Computer L, Stadler PF. Evolution of the let-7 microRNA family. RNA Biol. 2012;9(3):231-41.
129. Kozomara A, Birgaoanu M, Griffiths-Jones S. miRBase: from microRNA sequences to function. Nucleic Acids Res. 2019;47(D1):D155-D62.
130. Ambros V, Bartel B, Bartel DP, Burge CB, Carrington JC, Chen X, et al. A uniform system for microRNA annotation. RNA. 2003;9(3):277-9.
131. Kabekkodu SP, Shukla V, Varghese VK, J DS, Chakrabarty S, Satyamoorthy K. Clustered miRNAs and their role in biological functions and diseases. Biol Rev Camb Philos Soc. 2018;93(4):1955-86.
132. Lee Y, Kim M, Han J, Yeom KH, Lee S, Baek SH, et al. MicroRNA genes are transcribed by RNA polymerase II. EMBO J. 2004;23(20):4051-60.
133. Nguyen TA, Jo MH, Choi YG, Park J, Kwon SC, Hohng S, et al. Functional Anatomy of the Human Microprocessor. Cell. 2015;161(6):1374-87.
134. Auyeung VC, Ulitsky I, McGeary SE, Bartel DP. Beyond secondary structure: primary-sequence determinants license pri-miRNA hairpins for processing. Cell. 2013;152(4):844-58.
135. Tang X, Zhang Y, Tucker L, Ramratnam B. Phosphorylation of the RNase III enzyme Drosha at Serine300 or Serine302 is required for its nuclear localization. Nucleic Acids Res. 2010;38(19):6610-9.
136. Ha M, Kim VN. Regulation of microRNA biogenesis. Nat Rev Mol Cell Biol. 2014;15(8):509-24.
137. Ruby JG, Jan CH, Bartel DP. Intronic microRNA precursors that bypass Drosha processing. Nature. 2007;448(7149):83-6.
138. Wen J, Ladewig E, Shenker S, Mohammed J, Lai EC. Analysis of nearly one thousand mammalian mirtrons reveals novel features of dicer substrates. PLoS computational biology. 2015;11(9):e1004441.
139. Lund E, Guttinger S, Calado A, Dahlberg JE, Kutay U. Nuclear export of microRNA precursors. Science. 2004;303(5654):95-8.
140. Cai X, Hagedorn CH, Cullen BR. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. RNA. 2004;10(12):1957-66.
141. Pratt AJ, MacRae IJ. The RNA-induced silencing complex: a versatile gene-silencing machine. J Biol Chem. 2009;284(27):17897-901.
142. Bernstein E, Caudy AA, Hammond SM, Hannon GJ. Role for a bidentate ribonuclease in the initiation step of RNA interference. Nature. 2001;409(6818):363-6.

143. MacRae IJ, Ma E, Zhou M, Robinson CV, Doudna JA. In vitro reconstitution of the human RISC-loading complex. Proc Natl Acad Sci U S A. 2008;105(2):512-7.
144. Yang JS, Maurin T, Robine N, Rasmussen KD, Jeffrey KL, Chandwani R, et al. Conserved vertebrate mir-451 provides a platform for Dicer-independent, Ago2-mediated microRNA biogenesis. Proc Natl Acad Sci U S A. 2010;107(34):15163-8.
145. Jo MH, Shin S, Jung SR, Kim E, Song JJ, Hohng S. Human Argonaute 2 Has Diverse Reaction Pathways on Target RNAs. Mol Cell. 2015;59(1):117-24.
146. Roberts TC. The MicroRNA Biology of the Mammalian Nucleus. Mol Ther Nucleic Acids. 2014;3(8):e188.
147. Fromm B, Billipp T, Peck LE, Johansen M, Tarver JE, King BL, et al. A Uniform System for the Annotation of Vertebrate microRNA Genes and the Evolution of the Human microRNAome. Annu Rev Genet. 2015;49:213-42.
148. Tomasello L, Distefano R, Nigita G, Croce CM. The MicroRNA Family Gets Wider: The IsomiRs Classification and Role. Front Cell Dev Biol. 2021;9:668648.
149. Wyman SK, Knouf EC, Parkin RK, Fritz BR, Lin DW, Dennis LM, et al. Post-transcriptional generation of miRNA variants by multiple nucleotidyl transferases contributes to miRNA transcriptome complexity. Genome Res. 2011;21(9):1450-61.
150. Saunders MA, Liang H, Li W-H. Human polymorphism at microRNAs and microRNA target sites. Proceedings of the National Academy of Sciences. 2007;104(9):3300-5.
151. Marceca GP, Distefano R, Tomasello L, Lagana A, Russo F, Calore F, et al. MiREDiBase, a manually curated database of validated and putative editing events in microRNAs. Sci Data. 2021;8(1):199.
152. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature. 2010;466(7308):835-40.
153. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proc Natl Acad Sci U S A. 2002;99(24):15524-9.
154. Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M, Shimizu M, et al. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proc Natl Acad Sci U S A. 2005;102(39):13944-9.
155. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, et al. A MicroRNA signature associated with prognosis and progression in chronic lymphocytic leukemia. N Engl J Med. 2005;353(17):1793-801.
156. Calin GA, Sevignani C, Dumitru CD, Hyslop T, Noch E, Yendamuri S, et al. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proc Natl Acad Sci U S A. 2004;101(9):2999-3004.
157. Johnson SM, Grosshans H, Shingara J, Byrom M, Jarvis R, Cheng A, et al. RAS is regulated by the let-7 microRNA family. Cell. 2005;120(5):635-47.
158. Hammond SM. RNAi, microRNAs, and human disease. Cancer Chemother Pharmacol. 2006;58 Suppl 1(1):s63-8.
159. Peng Y, Croce CM. The role of MicroRNAs in human cancer. Signal Transduct Target Ther. 2016;1(1):15004.
160. Longnecker DS. Anatomy and Histology of the Pancreas (version 1.0). Pancreapedia: The Exocrine Pancreas Knowledge Base. 2014.
161. Lee EJ, Gusev Y, Jiang J, Nuovo GJ, Lerner MR, Frankel WL, et al. Expression profiling identifies microRNA signature in pancreatic cancer. Int J Cancer. 2007;120(5):1046-54.
162. Bloomston M, Frankel WL, Petrocca F, Volinia S, Alder H, Hagan JP, et al. MicroRNA expression patterns to differentiate pancreatic adenocarcinoma from normal pancreas and chronic pancreatitis. JAMA. 2007;297(17):1901-8.

163. Calatayud D, Dehlendorff C, Boisen MK, Hasselby JP, Schultz NA, Werner J, et al. Tissue MicroRNA profiles as diagnostic and prognostic biomarkers in patients with resectable pancreatic ductal adenocarcinoma and periampullary cancers. Biomark Res. 2017;5(1):8.
164. Du Rieu MC, Torrisani J, Selves J, Al Saati T, Souque A, Dufresne M, et al. MicroRNA-21 is induced early in pancreatic ductal adenocarcinoma precursor lesions. Clinical chemistry. 2010;56(4):603-12.
165. Caponi S, Funel N, Frampton AE, Mosca F, Santarpia L, Van der Velde AG, et al. The good, the bad and the ugly: a tale of miR-101, miR-21 and miR-155 in pancreatic intraductal papillary mucinous neoplasms. Ann Oncol. 2013;24(3):734-41.
166. Kadera BE, Li L, Toste PA, Wu N, Adams C, Dawson DW, et al. MicroRNA-21 in pancreatic ductal adenocarcinoma tumor-associated fibroblasts promotes metastasis. PLoS One. 2013;8(8):e71978.
167. Richards KE, Zeleniak AE, Fishel ML, Wu J, Littlepage LE, Hill R. Cancer-associated fibroblast exosomes regulate survival and proliferation of pancreatic cancer cells. Oncogene. 2017;36(13):1770-8.
168. Müller S, Raulefs S, Bruns P, Afonso-Grunz F, Plötner A, Thermann R, et al. Next-generation sequencing reveals novel differentially regulated mRNAs, lncRNAs, miRNAs, sdRNAs and a piRNA in pancreatic cancer. Molecular cancer. 2015;14:1-18.
169. Szafranska AE, Davison TS, John J, Cannon T, Sipos B, Maghnouj A, et al. MicroRNA expression alterations are linked to tumorigenesis and non-neoplastic processes in pancreatic ductal adenocarcinoma. Oncogene. 2007;26(30):4442-52.
170. Bauer AS, Keller A, Costello E, Greenhalf W, Bier M, Borries A, et al. Diagnosis of pancreatic ductal adenocarcinoma and chronic pancreatitis by measurement of microRNA abundance in blood and tissue. PLoS One. 2012;7(4):e34151.
171. Schultz NA, Werner J, Willenbrock H, Roslind A, Giese N, Horn T, et al. MicroRNA expression profiles associated with pancreatic adenocarcinoma and ampullary adenocarcinoma. Mod Pathol. 2012;25(12):1609-22.
172. Han J, DePinho RA, Maitra A. Single-cell RNA sequencing in pancreatic cancer. Nat Rev Gastroenterol Hepatol. 2021;18(7):451-2.
173. Lee JJ, Bernard V, Semaan A, Monberg ME, Huang J, Stephens BM, et al. Elucidation of Tumor-Stromal Heterogeneity and the Ligand-Receptor Interactome by Single-Cell Transcriptomics in Real-world Pancreatic Cancer BiopsiesscRNA-seq of Pancreatic Cancer Real-world Biopsies. Clinical Cancer Research. 2021;27(21):5912-21.
174. Aung KL, Fischer SE, Denroche RE, Jang G-H, Dodd A, Creighton S, et al. Genomics-Driven Precision Medicine for Advanced Pancreatic Cancer: Early Results from the COMPASS TrialReal-Time Genomic Profiling of Advanced Pancreatic Cancer. Clinical Cancer Research. 2018;24(6):1344-54.
175. Alwine JC, Kemp DJ, Stark GR. Method for detection of specific RNAs in agarose gels by transfer to diazobenzyloxymethyl-paper and hybridization with DNA probes. Proc Natl Acad Sci U S A. 1977;74(12):5350-4.
176. Zhu H, Zhang H, Xu Y, Lassakova S, Korabecna M, Neuzil P. PCR past, present and future. Biotechniques. 2020;69(4):317-25.
177. Jensen SG, Lamy P, Rasmussen MH, Ostenfeld MS, Dyrskjot L, Orntoft TF, et al. Evaluation of two commercial global miRNA expression profiling platforms for detection of less abundant miRNAs. BMC Genomics. 2011;12(1):435.
178. Ballantyne KN, van Oorschot RA, Mitchell RJ. Locked nucleic acids in PCR primers increase sensitivity and performance. Genomics. 2008;91(3):301-5.
179. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463-7.

180. Giraldez MD, Spengler RM, Etheridge A, Godoy PM, Barczak AJ, Srinivasan S, et al. Comprehensive multi-center assessment of small RNA-seq methods for quantitative miRNA profiling. Nature biotechnology. 2018;36(8):746-57.
181. Tan EM, Schur PH, Carr RI, Kunkel HG. Deoxybonucleic acid (DNA) and antibodies to DNA in the serum of patients with systemic lupus erythematosus. J Clin Invest. 1966;45(11):1732-40.
182. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res. 1977;37(3):646-50.
183. Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, Pulford K, et al. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. Br J Haematol. 2008;141(5):672-5.
184. Machida T, Tomofuji T, Maruyama T, Yoneda T, Ekuni D, Azuma T, et al. miR-1246 and miR-4644 in salivary exosome as potential biomarkers for pancreatobiliary tract cancer. Oncol Rep. 2016;36(4):2375-81.
185. Bryzgunova OE, Laktionov PP. Extracellular Nucleic Acids in Urine: Sources, Structure, Diagnostic Potential. Acta Naturae. 2015;7(3):48-54.
186. Song Q, Zhang Y, Liu H, Du Y. Potential of Using Cell-Free DNA and miRNA in Breast Milk to Screen Early Breast Cancer. Biomed Res Int. 2020;2020:8126176.
187. Boissiere A, Gala A, Ferrieres-Hoa A, Mullet T, Baillet S, Petiton A, et al. Cell-free and intracellular nucleic acids: new non-invasive biomarkers to explore male infertility. Basic Clin Androl. 2017;27(1):7.
188. Weber JA, Baxter DH, Zhang S, Huang DY, Huang KH, Lee MJ, et al. The microRNA spectrum in 12 body fluids. Clin Chem. 2010;56(11):1733-41.
189. Turchinovich A, Weiz L, Burwinkel B. Extracellular miRNAs: the mystery of their origin and function. Trends Biochem Sci. 2012;37(11):460-5.
190. Godoy PM, Bhakta NR, Barczak AJ, Cakmak H, Fisher S, MacKenzie TC, et al. Large Differences in Small RNA Composition Between Human Biofluids. Cell Rep. 2018;25(5):1346-58.
191. Turchinovich A, Weiz L, Langheinz A, Burwinkel B. Characterization of extracellular circulating microRNA. Nucleic Acids Res. 2011;39(16):7223-33.
192. Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proc Natl Acad Sci U S A. 2011;108(12):5003-8.
193. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res. 2008;18(10):997-1006.
194. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, et al. Circulating microRNAs as stable blood-based markers for cancer detection. Proc Natl Acad Sci U S A. 2008;105(30):10513-8.
195. Sohail AM, Khawar MB, Afzal A, Hassan A, Shahzaman S, Ali A. Multifaceted roles of extracellular RNAs in different diseases. Mil Med Res. 2022;9(1):43.
196. Larson MH, Pan W, Kim HJ, Mauntz RE, Stuart SM, Pimentel M, et al. A comprehensive characterization of the cell-free transcriptome reveals tissue- and subtype-specific biomarkers for cancer detection. Nat Commun. 2021;12(1):2357.
197. Chen S, Jin Y, Wang S, Xing S, Wu Y, Tao Y, et al. Cancer type classification using plasma cell-free RNAs derived from human and microbes. Elife. 2022;11:e75181.
198. Zaporozhchenko IA, Ponomaryova AA, Rykova EY, Laktionov PP. The potential of circulating cell-free RNA as a cancer biomarker: challenges and opportunities. Expert Rev Mol Diagn. 2018;18(2):133-45.
199. Raposo G, Nijman HW, Stoorvogel W, Liejendekker R, Harding CV, Melief CJ, et al. B lymphocytes secrete antigen-presenting vesicles. J Exp Med. 1996;183(3):1161-72.

200. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nat Cell Biol. 2007;9(6):654-9.
201. Nolte-'t Hoen EN, Buermans HP, Waasdorp M, Stoorvogel W, Wauben MH, t Hoen PA. Deep sequencing of RNA from immune cell-derived vesicles uncovers the selective incorporation of small non-coding RNA biotypes with potential regulatory functions. Nucleic Acids Res. 2012;40(18):9272-85.
202. Lotvall J, Hill AF, Hochberg F, Buzas EI, Di Vizio D, Gardiner C, et al. Minimal experimental requirements for definition of extracellular vesicles and their functions: a position statement from the International Society for Extracellular Vesicles. J Extracell Vesicles. 2014;3:26913.
203. Thery C, Witwer KW, Aikawa E, Alcaraz MJ, Anderson JD, Andriantsitohaina R, et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): a position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. J Extracell Vesicles. 2018;7(1):1535750.
204. Thery C, Boussac M, Veron P, Ricciardi-Castagnoli P, Raposo G, Garin J, et al. Proteomic analysis of dendritic cell-derived exosomes: a secreted subcellular compartment distinct from apoptotic vesicles. J Immunol. 2001;166(12):7309-18.
205. Thery C, Ostrowski M, Segura E. Membrane vesicles as conveyors of immune responses. Nat Rev Immunol. 2009;9(8):581-93.
206. Witwer KW, Théry C. Extracellular vesicles or exosomes? On primacy, precision, and popularity influencing a choice of nomenclature. Taylor & Francis; 2019. p. 1648167.
207. Tosar JP, Cayota A, Witwer K. Exomeres and Supermeres: monolithic or diverse? J Extracell Biol. 2022;1(6):e45.
208. Consortium E-T, Van Deun J, Mestdagh P, Agostinis P, Akay Ö, Anand S, et al. EV-TRACK: transparent reporting and centralizing knowledge in extracellular vesicle research. Nature Methods. 2017;14:228.
209. Gardiner C, Di Vizio D, Sahoo S, Thery C, Witwer KW, Wauben M, et al. Techniques used for the isolation and characterization of extracellular vesicles: results of a worldwide survey. J Extracell Vesicles. 2016;5(1):32945.
210. Gyorgy B, Modos K, Pallinger E, Paloczi K, Pasztoi M, Misjak P, et al. Detection and isolation of cell-derived microparticles are compromised by protein complexes resulting from shared biophysical parameters. Blood. 2011;117(4):e39-48.
211. Brennan K, Martin K, FitzGerald SP, O'Sullivan J, Wu Y, Blanco A, et al. A comparison of methods for the isolation and separation of extracellular vesicles from protein and lipid particles in human serum. Sci Rep. 2020;10(1):1039.
212. Stranska R, Gysbrechts L, Wouters J, Vermeersch P, Bloch K, Dierickx D, et al. Comparison of membrane affinity-based method with size-exclusion chromatography for isolation of exosome-like vesicles from human plasma. J Transl Med. 2018;16(1):1.
213. Nishiwada S, Cui Y, Sho M, Jun E, Akahori T, Nakamura K, et al. Transcriptomic Profiling Identifies an Exosomal microRNA Signature for Predicting Recurrence Following Surgery in Patients With Pancreatic Ductal Adenocarcinoma. Ann Surg. 2022;276(6):e876-e85.
214. Nakamura K, Zhu Z, Roy S, Jun E, Han H, Munoz RM, et al. An Exosome-based Transcriptomic Signature for Noninvasive, Early Detection of Patients With Pancreatic Ductal Adenocarcinoma: A Multicenter Cohort Study. Gastroenterology. 2022;163(5):1252-66 e2.
215. Liu DSK, Upton FM, Rees E, Limb C, Jiao LR, Krell J, et al. Size-Exclusion Chromatography as a Technique for the Investigation of Novel Extracellular Vesicles in Cancer. Cancers (Basel). 2020;12(11).

216. Boing AN, van der Pol E, Grootemaat AE, Coumans FA, Sturk A, Nieuwland R. Single-step isolation of extracellular vesicles by size-exclusion chromatography. J Extracell Vesicles. 2014;3.
217. Nordin JZ, Lee Y, Vader P, Mager I, Johansson HJ, Heusermann W, et al. Ultrafiltration with size-exclusion liquid chromatography for high yield isolation of extracellular vesicles preserving intact biophysical and functional properties. Nanomedicine. 2015;11(4):879-83.
218. de Menezes-Neto A, Saez MJ, Lozano-Ramos I, Segui-Barber J, Martin-Jaular L, Ullate JM, et al. Size-exclusion chromatography as a stand-alone methodology identifies novel markers in mass spectrometry analyses of plasma-derived vesicles from healthy individuals. J Extracell Vesicles. 2015;4:27378.
219. Thery C, Amigorena S, Raposo G, Clayton A. Isolation and characterization of exosomes from cell culture supernatants and biological fluids. Curr Protoc Cell Biol. 2006;Chapter 3(1):Unit 3 22.
220. Zhang H, Lyden D. Asymmetric-flow field-flow fractionation technology for exomere and small extracellular vesicle separation and characterization. Nat Protoc. 2019;14(4):1027-53.
221. Ko J, Bhagwat N, Black T, Yee SS, Na YJ, Fisher S, et al. miRNA Profiling of Magnetic Nanopore-Isolated Extracellular Vesicles for the Diagnosis of Pancreatic Cancer. Cancer Res. 2018;78(13):3688-97.
222. Yang Z, LaRiviere MJ, Ko J, Till JE, Christensen T, Yee SS, et al. A Multianalyte Panel Consisting of Extracellular Vesicle miRNAs and mRNAs, cfDNA, and CA19-9 Shows Utility for Diagnosis and Staging of Pancreatic Ductal Adenocarcinoma. Clin Cancer Res. 2020;26(13):3248-58.
223. Severino V, Dumonceau JM, Delhaye M, Moll S, Annessi-Ramseyer I, Robin X, et al. Extracellular Vesicles in Bile as Markers of Malignant Biliary Stenoses. Gastroenterology. 2017;153(2):495-504 e8.
224. Dragovic RA, Gardiner C, Brooks AS, Tannetta DS, Ferguson DJ, Hole P, et al. Sizing and phenotyping of cellular vesicles using Nanoparticle Tracking Analysis. Nanomedicine. 2011;7(6):780-8.
225. Bachurski D, Schuldner M, Nguyen PH, Malz A, Reiners KS, Grenzi PC, et al. Extracellular vesicle measurements with nanoparticle tracking analysis - An accuracy and repeatability comparison between NanoSight NS300 and ZetaView. J Extracell Vesicles. 2019;8(1):1596016.
226. Groot Kormelink T, Arkesteijn GJ, Nauwelaers FA, van den Engh G, Nolte-'t Hoen EN, Wauben MH. Prerequisites for the analysis and sorting of extracellular vesicle subpopulations by high-resolution flow cytometry. Cytometry A. 2016;89(2):135-47.
227. Maas SL, Broekman ML, Vrij Jd. Tunable resistive pulse sensing for the characterization of extracellular vesicles. Exosomes and microvesicles: Springer; 2017. p. 21-33.
228. Parisse P, Rago I, Ulloa Severino L, Perissinotto F, Ambrosetti E, Paoletti P, et al. Atomic force microscopy analysis of extracellular vesicles. Eur Biophys J. 2017;46(8):813-20.
229. Wu Y, Deng W, Klinke DJ, 2nd. Exosomes: improved methods to characterize their morphology, RNA content, and surface protein biomarkers. Analyst. 2015;140(19):6631-42.
230. Chuo ST, Chien JC, Lai CP. Imaging extracellular vesicles: current and emerging methods. J Biomed Sci. 2018;25(1):91.
231. Hoog JL, Lotvall J. Diversity of extracellular vesicles in human ejaculates revealed by cryo-electron microscopy. J Extracell Vesicles. 2015;4:28680.

232. Chen C, Zong S, Wang Z, Lu J, Zhu D, Zhang Y, et al. Imaging and Intracellular Tracking of Cancer-Derived Exosomes Using Single-Molecule Localization-Based Super-Resolution Microscope. ACS Appl Mater Interfaces. 2016;8(39):25825-33.
233. Karimi N, Cvjetkovic A, Jang SC, Crescitelli R, Hosseinpour Feizi MA, Nieuwland R, et al. Detailed analysis of the plasma extracellular vesicle proteome after separation from lipoproteins. Cell Mol Life Sci. 2018;75(15):2873-86.
234. Larios J, Mercier V, Roux A, Gruenberg J. ALIX- and ESCRT-III-dependent sorting of tetraspanins to exosomes. J Cell Biol. 2020;219(3).
235. Davidson SM, Takov K, Yellon DM. Exosomes and Cardiovascular Protection. Cardiovasc Drugs Ther. 2017;31(1):77-86.
236. Ludwig N, Razzo BM, Yerneni SS, Whiteside TL. Optimization of cell culture conditions for exosome isolation using mini-size exclusion chromatography (mini-SEC). Exp Cell Res. 2019;378(2):149-57.
237. Guerreiro EM, Vestad B, Steffensen LA, Aass HCD, Saeed M, Ovstebo R, et al. Efficient extracellular vesicle isolation by combining cell media modifications, ultrafiltration, and size-exclusion chromatography. PLoS One. 2018;13(9):e0204276.
238. He C, Fischer S, Meyer G, Muller I, Jungst D. Two-dimensional electrophoretic analysis of vesicular and micellar proteins of gallbladder bile. J Chromatogr A. 1997;776(1):109-15.
239. Enderle D, Spiel A, Coticchia CM, Berghoff E, Mueller R, Schlumpberger M, et al. Characterization of RNA from Exosomes and Other Extracellular Vesicles Isolated by a Novel Spin Column-Based Method. PLoS One. 2015;10(8):e0136133.
240. Muller L, Hong CS, Stolz DB, Watkins SC, Whiteside TL. Isolation of biologically-active exosomes from human plasma. J Immunol Methods. 2014;411:55-65.
241. Sokolova V, Ludwig AK, Hornung S, Rotan O, Horn PA, Epple M, et al. Characterisation of exosomes derived from human cells by nanoparticle tracking analysis and scanning electron microscopy. Colloids Surf B Biointerfaces. 2011;87(1):146-50.
242. Jayachandran M, Miller VM, Heit JA, Owen WG. Methodology for isolation, identification and characterization of microvesicles in peripheral blood. J Immunol Methods. 2012;375(1-2):207-14.
243. Chevillet JR, Kang Q, Ruf IK, Briggs HA, Vojtech LN, Hughes SM, et al. Quantitative and stoichiometric analysis of the microRNA content of exosomes. Proc Natl Acad Sci U S A. 2014;111(41):14888-93.
244. Hoshino A, Costa-Silva B, Shen TL, Rodrigues G, Hashimoto A, Tesic Mark M, et al. Tumour exosome integrins determine organotropic metastasis. Nature. 2015;527(7578):329-35.
245. Costa-Silva B, Aiello NM, Ocean AJ, Singh S, Zhang H, Thakur BK, et al. Pancreatic cancer exosomes initiate pre-metastatic niche formation in the liver. Nat Cell Biol. 2015;17(6):816-26.
246. Peinado H, Zhang H, Matei IR, Costa-Silva B, Hoshino A, Rodrigues G, et al. Pre-metastatic niches: organ-specific homes for metastases. Nat Rev Cancer. 2017;17(5):302-17.
247. Boyer JL. Bile formation and secretion. Compr Physiol. 2013;3(3):1035-78.
248. Hofmann AF. The continuing importance of bile acids in liver and intestinal disease. Arch Intern Med. 1999;159(22):2647-58.
249. Amigo L, Covarrubias C, Nervi F. Separation and quantitation of cholesterol carriers in native bile by ultracentrifugation. Hepatology. 1990;12(3 Pt 2):130S-3S.
250. Somjen GJ, Gilat T. Contribution of vesicular and micellar carriers to cholesterol transport in human bile. J Lipid Res. 1985;26(6):699-704.
251. Sahlin S, Thyberg P, Ahlberg J, Angelin B, Einarsson K. Distribution of cholesterol between vesicles and micelles in human gallbladder bile: Influence of

treatment with chenodeoxycholic acid and ursodeoxycholic acid. Hepatology. 1991;13(1):104-10.
252. Kaplun A, Talmon Y, Konikoff FM, Rubin M, Eitan A, Tadmor M, et al. Direct visualization of lipid aggregates in native human bile by light- and cryo-transmission electron-microscopy. FEBS Lett. 1994;340(1-2):78-82.
253. Masyuk AI, Huang BQ, Ward CJ, Gradilone SA, Banales JM, Masyuk TV, et al. Biliary exosomes influence cholangiocyte regulatory mechanisms and proliferation through interaction with primary cilia. Am J Physiol Gastrointest Liver Physiol. 2010;299(4):G990-9.
254. Li L, Masica D, Ishida M, Tomuleasa C, Umegaki S, Kalloo AN, et al. Human bile contains microRNA-laden extracellular vesicles that can be used for cholangiocarcinoma diagnosis. Hepatology. 2014;60(3):896-907.
255. Plebani M, Banfi G, Bernardini S, Bondanini F, Conti L, Dorizzi R, et al. Serum or plasma? An old question looking for new answers. Clin Chem Lab Med. 2020;58(2):178-87.
256. Wang K, Yuan Y, Cho JH, McClarty S, Baxter D, Galas DJ. Comparing the MicroRNA spectrum between serum and plasma. PLoS One. 2012;7(7):e41561.
257. Tsui NB, Ng EK, Lo YM. Stability of endogenous and added RNA in blood specimens, serum, and plasma. Clin Chem. 2002;48(10):1647-53.
258. Zhou J, Liu Z, Yang S, Li X. Identification of microRNAs as biomarkers for cholangiocarcinoma detection: A diagnostic meta-analysis. Clin Res Hepatol Gastroenterol. 2017;41(2):156-62.
259. Letelier P, Riquelme I, Hernandez AH, Guzman N, Farias JG, Roa JC. Circulating MicroRNAs as Biomarkers in Biliary Tract Cancers. Int J Mol Sci. 2016;17(5):791.
260. Correa-Gallego C, Maddalo D, Doussot A, Kemeny N, Kingham TP, Allen PJ, et al. Circulating Plasma Levels of MicroRNA-21 and MicroRNA-221 Are Potential Diagnostic Markers for Primary Intrahepatic Cholangiocarcinoma. PLoS One. 2016;11(9):e0163699.
261. Li G, Pu Y. MicroRNA signatures in total peripheral blood of gallbladder cancer patients. Tumour Biol. 2015;36(9):6985-90.
262. Zhou X, Lu Z, Wang T, Huang Z, Zhu W, Miao Y. Plasma miRNAs in diagnosis and prognosis of pancreatic cancer: A miRNA expression analysis. Gene. 2018;673:181-93.
263. Duell EJ, Lujan-Barroso L, Sala N, Deitz McElyea S, Overvad K, Tjonneland A, et al. Plasma microRNAs as biomarkers of pancreatic cancer risk in a prospective cohort study. Int J Cancer. 2017;141(5):905-15.
264. Schultz NA, Dehlendorff C, Jensen BV, Bjerregaard JK, Nielsen KR, Bojesen SE, et al. MicroRNA biomarkers in whole blood for detection of pancreatic cancer. JAMA. 2014;311(4):392-404.
265. Couch Y, Buzas EI, Di Vizio D, Gho YS, Harrison P, Hill AF, et al. A brief history of nearly EV-erything - The rise and rise of extracellular vesicles. J Extracell Vesicles. 2021;10(14):e12144.
266. Toyofuku M, Nomura N, Eberl L. Types and origins of bacterial membrane vesicles. Nat Rev Microbiol. 2019;17(1):13-24.
267. Casadevall A, Nosanchuk JD, Williamson P, Rodrigues ML. Vesicular transport across the fungal cell wall. Trends Microbiol. 2009;17(4):158-62.
268. Wolf P. The nature and significance of platelet products in human plasma. Br J Haematol. 1967;13(3):269-88.
269. Crawford N. The presence of contractile proteins in platelet microparticles isolated from human and animal platelet-free plasma. Br J Haematol. 1971;21(1):53-69.

270. Prince AM, Adams WR. Virus-like particles in human plasma and serum: role of platelet lysosomes. J Natl Cancer Inst. 1966;37(2):153-66.
271. Harding C, Heuser J, Stahl P. Receptor-mediated endocytosis of transferrin and recycling of the transferrin receptor in rat reticulocytes. J Cell Biol. 1983;97(2):329-39.
272. Johnstone RM, Adam M, Hammond JR, Orr L, Turbide C. Vesicle formation during reticulocyte maturation. Association of plasma membrane activities with released vesicles (exosomes). J Biol Chem. 1987;262(19):9412-20.
273. Pietrowska M, Wlosowicz A, Gawin M, Widlak P. MS-Based Proteomic Analysis of Serum and Plasma: Problem of High Abundant Components and Lights and Shadows of Albumin Removal. Adv Exp Med Biol. 2019;1073:57-76.
274. Que R, Ding G, Chen J, Cao L. Analysis of serum exosomal microRNAs and clinicopathologic features of patients with pancreatic adenocarcinoma. World J Surg Oncol. 2013;11:219.
275. Lai X, Wang M, McElyea SD, Sherman S, House M, Korc M. A microRNA signature in circulating exosomes is superior to exosomal glypican-1 levels for diagnosing pancreatic cancer. Cancer Lett. 2017;393:86-93.
276. Goto T, Fujiya M, Konishi H, Sasajima J, Fujibayashi S, Hayashi A, et al. An elevated expression of serum exosomal microRNA-191, - 21, -451a of pancreatic neoplasm is considered to be efficient diagnostic marker. BMC Cancer. 2018;18(1):116.
277. Kawamura S, Iinuma H, Wada K, Takahashi K, Minezaki S, Kainuma M, et al. Exosome-encapsulated microRNA-4525, microRNA-451a and microRNA-21 in portal vein blood is a high-sensitive liquid biomarker for the selection of high-risk pancreatic ductal adenocarcinoma patients. J Hepatobiliary Pancreat Sci. 2019;26(2):63-72.
278. Kim MW, Koh H, Kim JY, Lee S, Lee H, Kim Y, et al. Tumor-Specific miRNA Signatures in Combination with CA19-9 for Liquid Biopsy-Based Detection of PDAC. Int J Mol Sci. 2021;22(24).
279. Zeold A, Sandor GO, Kiss A, Soos AA, Tolgyes T, Bursics A, et al. Shared extracellular vesicle miRNA profiles of matched ductal pancreatic adenocarcinoma organoids and blood plasma samples show the power of organoid technology. Cell Mol Life Sci. 2021;78(6):3005-20.
280. Melo SA, Luecke LB, Kahlert C, Fernandez AF, Gammon ST, Kaye J, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. Nature. 2015;523(7559):177-82.
281. Lucien F, Lac V, Billadeau DD, Borgida A, Gallinger S, Leong HS. Glypican-1 and glycoprotein 2 bearing extracellular vesicles do not discern pancreatic cancer from benign pancreatic diseases. Oncotarget. 2019;10(10):1045.
282. Frampton AE, Prado MM, Lopez-Jimenez E, Fajardo-Puerta AB, Jawad ZAR, Lawton P, et al. Glypican-1 is enriched in circulating-exosomes in pancreatic cancer and correlates with tumor burden. Oncotarget. 2018;9(27):19006-13.
283. Melo S, Kahlert C, LeBleu V, Kalluri R. Multiple antibodies identify Glypican-1 on serum exosomes from patients with pancreatic cancer. bioRxiv. 2017:145706.
284. Shen L, Chen GP, Xia QF, Shao SJ, Fang HX. Exosomal miR-200 family as serum biomarkers for early detection and prognostic prediction of cholangiocarcinoma. International Journal of Clinical and Experimental Pathology. 2019;12(10):3870-+.
285. Arbelaiz A, Azkargorta M, Krawczyk M, Santos-Laso A, Lapitz A, Perugorria MJ, et al. Serum extracellular vesicles contain protein biomarkers for primary sclerosing cholangitis and cholangiocarcinoma. Hepatology. 2017;66(4):1125-43.
286. Schimmel P. The emerging complexity of the tRNA world: mammalian tRNAs beyond protein synthesis. Nat Rev Mol Cell Biol. 2018;19(1):45-58.
287. Liu DSK, Yang QZC, Asim M, Krell J, Frampton AE. The Clinical Significance of Transfer RNAs Present in Extracellular Vesicles. Int J Mol Sci. 2022;23(7):3692.

288. Novoa EM, Pavon-Eternod M, Pan T, Ribas de Pouplana L. A role for tRNA modifications in genome structure and codon usage. Cell. 2012;149(1):202-13.
289. Gebetsberger J, Polacek N. Slicing tRNAs to boost functional ncRNA diversity. RNA Biol. 2013;10(12):1798-806.
290. Huang X, Yuan T, Tschannen M, Sun Z, Jacob H, Du M, et al. Characterization of human plasma-derived exosomal RNAs by deep sequencing. BMC Genomics. 2013;14:319.
291. Turchinovich A, Drapkina O, Tonevitsky A. Transcriptome of Extracellular Vesicles: State-of-the-Art. Front Immunol. 2019;10:202.
292. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet journal. 2011;17(1):10-2.
293. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550.
294. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res. 2004;64(15):5245-50.
295. Poruk KE, Gay DZ, Brown K, Mulvihill JD, Boucher KM, Scaife CL, et al. The clinical utility of CA 19-9 in pancreatic adenocarcinoma: diagnostic and prognostic updates. Curr Mol Med. 2013;13(3):340-51.
296. Rosch T, Meining A, Fruhmorgen S, Zillinger C, Schusdziarra V, Hellerhoff K, et al. A prospective comparison of the diagnostic accuracy of ERCP, MRCP, CT, and EUS in biliary strictures. Gastrointest Endosc. 2002;55(7):870-6.
297. Thomas MJ, Pellegrini CA, Way LW. Usefulness of diagnostic tests for biliary obstruction. Am J Surg. 1982;144(1):102-8.
298. UK CR. About Bile Duct Cancer [updated 11 Jan 2018. Available from: https://www.cancerresearchuk.org/about-cancer/bile-duct-cancer/about.
299. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological). 2018;57(1):289-300.
300. Scheepbouwer C, Aparicio-Puerta E, Gomez-Martin C, Verschueren H, van Eijndhoven M, Wedekind LE, et al. ALL-tRNAseq enables robust tRNA profiling in tissue samples. Genes Dev. 2023;37(5-6):243-57.
301. Cruzat V, Macedo Rogero M, Noel Keane K, Curi R, Newsholme P. Glutamine: Metabolism and Immune Function, Supplementation and Clinical Translation. Nutrients. 2018;10(11).
302. Johnson MR, Barnes S, Kwakye JB, Diasio RB. Purification and characterization of bile acid-CoA:amino acid N-acyltransferase from human liver. J Biol Chem. 1991;266(16):10227-33.
303. Tibshirani R. Regression Shrinkage and Selection Via the Lasso. Journal of the Royal Statistical Society: Series B (Methodological). 2018;58(1):267-88.
304. Kern F, Fehlmann T, Solomon J, Schwed L, Grammes N, Backes C, et al. miEAA 2.0: integrating multi-species microRNA enrichment analysis and workflow management systems. Nucleic Acids Res. 2020;48(W1):W521-W8.
305. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.
306. Gordon KJ, Kirkbride KC, How T, Blobe GC. Bone morphogenetic proteins induce pancreatic cancer cell invasiveness through a Smad1-dependent mechanism that involves matrix metalloproteinase-2. Carcinogenesis. 2009;30(2):238-48.

307. Goulley J, Dahl U, Baeza N, Mishina Y, Edlund H. BMP4-BMPR1A signaling in ß cells is required for and augments glucose-stimulated insulin secretion. Cell metabolism. 2007;5(3):207-19.
308. Fabbri M, Paone A, Calore F, Galli R, Gaudio E, Santhanam R, et al. MicroRNAs bind to Toll-like receptors to induce prometastatic inflammatory response. Proc Natl Acad Sci U S A. 2012;109(31):E2110-6.
309. Lanki M, Seppanen H, Mustonen H, Hagstrom J, Haglund C. Toll-like receptor 1 predicts favorable prognosis in pancreatic cancer. PLoS One. 2019;14(7):e0219245.
310. Shigehara K, Yokomuro S, Ishibashi O, Mizuguchi Y, Arima Y, Kawahigashi Y, et al. Real-time PCR-based analysis of the human bile microRNAome identifies miR-9 as a potential diagnostic biomarker for biliary tract cancer. PLoS One. 2011;6(8):e23584.
311. Cote GA, Gore AJ, McElyea SD, Heathers LE, Xu H, Sherman S, et al. A pilot study to develop a diagnostic test for pancreatic ductal adenocarcinoma based on differential expression of select miRNA in plasma and bile. Am J Gastroenterol. 2014;109(12):1942-52.
312. Tsen A, Barbara M, Rosenkranz L. Dilemma of elevated CA 19-9 in biliary pathology. Pancreatology. 2018;18(8):862-7.
313. Coenen-Stass AML, Magen I, Brooks T, Ben-Dov IZ, Greensmith L, Hornstein E, et al. Evaluation of methodologies for microRNA biomarker detection by next generation sequencing. RNA Biol. 2018;15(8):1133-45.
314. Holmes AD, Chan PP, Chen Q, Ivanov P, Drouard L, Polacek N, et al. A Standardized Ontology for Naming tRNA-derived RNAs Based on Molecular Origin. bioRxiv. 2022:2022.05.06.490965.
315. Gamez-Valero A, Monguio-Tortajada M, Carreras-Planella L, Franquesa M, Beyer K, Borras FE. Size-Exclusion Chromatography-based isolation minimally alters Extracellular Vesicles' characteristics compared to precipitating agents. Sci Rep. 2016;6(1):33641.
316. Iwamura T, Katsuki T, Ide K. Establishment and characterization of a human pancreatic cancer cell line (SUIT-2) producing carcinoembryonic antigen and carbohydrate antigen 19-9. Jpn J Cancer Res. 1987;78(1):54-62.
317. Webber J, Clayton A. How pure are your vesicles? J Extracell Vesicles. 2013;2.
318. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 2002;3(7):RESEARCH0034.
319. Ge X, Wang Y, Nie J, Li Q, Tang L, Deng X, et al. The diagnostic/prognostic potential and molecular functions of long non-coding RNAs in the exosomes derived from the bile of human cholangiocarcinoma. Oncotarget. 2017;8(41):69995-70005.
320. Xu Y, Kang P, Leng K, Yao Y, Liao G, Han Y, et al. Circ_ASPH promotes cholangiocarcinoma growth and metastasis through the miR-581/ATP-binding cassette transporter G1 signaling pathway. Cancer Commun (Lond). 2020;40(10):545-50.
321. Konoshenko MY, Lekchnov EA, Vlassov AV, Laktionov PP. Isolation of Extracellular Vesicles: General Methodologies and Latest Trends. Biomed Res Int. 2018;2018:8545347.
322. Kent OA, Mullendore M, Wentzel EA, Lopez-Romero P, Tan AC, Alvarez H, et al. A resource for analysis of microRNA expression and function in pancreatic ductal adenocarcinoma cells. Cancer Biol Ther. 2009;8(21):2013-24.
323. Chu CH, Chou W, Wang F, Yeh CN, Chen TC, Yeh TS. Expression profile of microRNA-200 family in cholangiocarcinoma arising from choledochal cyst. J Gastroenterol Hepatol. 2016;31(5):1052-9.
324. Zhang Y, Li M, Wang H, Fisher WE, Lin PH, Yao Q, et al. Profiling of 95 microRNAs in pancreatic cancer cell lines and surgical specimens by real-time PCR analysis. World J Surg. 2009;33(4):698-709.

325. Korpal M, Lee ES, Hu G, Kang Y. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. J Biol Chem. 2008;283(22):14910-4.
326. Lee JS, Ahn YH, Won HS, Sun S, Kim YH, Ko YH. Prognostic Role of the MicroRNA-200 Family in Various Carcinomas: A Systematic Review and Meta-Analysis. Biomed Res Int. 2017;2017:1928021.
327. Que RS, Lin C, Ding GP, Wu ZR, Cao LP. Increasing the immune activity of exosomes: the effect of miRNA-depleted exosome proteins on activating dendritic cell/cytokine-induced killer cells against pancreatic cancer. J Zhejiang Univ Sci B. 2016;17(5):352-60.
328. Xu YF, Hannafon BN, Zhao YD, Postier RG, Ding WQ. Plasma exosome miR-196a and miR-1246 are potential indicators of localized pancreatic cancer. Oncotarget. 2017;8(44):77028-40.
329. Reese M, Flammang I, Yang Z, Dhayat SA. Potential of Exosomal microRNA-200b as Liquid Biopsy Marker in Pancreatic Ductal Adenocarcinoma. Cancers (Basel). 2020;12(1).
330. Guo J, Wu C, Lin X, Zhou J, Zhang J, Zheng W, et al. Establishment of a simplified dichotomic size-exclusion chromatography for isolating extracellular vesicles toward clinical applications. J Extracell Vesicles. 2021;10(11):e12145.
331. Takov K, Yellon DM, Davidson SM. Comparison of small extracellular vesicles isolated from plasma by ultracentrifugation or size-exclusion chromatography: yield, purity and functional potential. J Extracell Vesicles. 2019;8(1):1560809.
332. Vickers KC, Palmisano BT, Shoucri BM, Shamburek RD, Remaley AT. MicroRNAs are transported in plasma and delivered to recipient cells by high-density lipoproteins. Nat Cell Biol. 2011;13(4):423-33.
333. Dhayat SA, Husing A, Senninger N, Schmidt HH, Haier J, Wolters H, et al. Circulating microRNA-200 Family as Diagnostic Marker in Hepatocellular Carcinoma. PLoS One. 2015;10(10):e0140066.
334. Dhayat SA, Traeger MM, Rehkaemper J, Stroese AJ, Steinestel K, Wardelmann E, et al. Clinical Impact of Epithelial-to-Mesenchymal Transition Regulating MicroRNAs in Pancreatic Ductal Adenocarcinoma. Cancers (Basel). 2018;10(9):328.
335. Ludwig S, Floros T, Theodoraki MN, Hong CS, Jackson EK, Lang S, et al. Suppression of Lymphocyte Functions by Plasma Exosomes Correlates with Disease Activity in Patients with Head and Neck Cancer. Clin Cancer Res. 2017;23(16):4843-54.
336. Theodoraki MN, Yerneni SS, Hoffmann TK, Gooding WE, Whiteside TL. Clinical Significance of PD-L1(+) Exosomes in Plasma of Head and Neck Cancer Patients. Clin Cancer Res. 2018;24(4):896-905.
337. Zhu L, Li J, Gong Y, Wu Q, Tan S, Sun D, et al. Exosomal tRNA-derived small RNA as a promising biomarker for cancer diagnosis. Mol Cancer. 2019;18(1):74.
338. Kumar SR, Kimchi ET, Manjunath Y, Gajagowni S, Stuckel AJ, Kaifi JT. RNA cargos in extracellular vesicles derived from blood serum in pancreas associated conditions. Sci Rep. 2020;10(1):2800.
339. Yazarlou F, Kadkhoda S, Ghafouri-Fard S. Emerging role of let-7 family in the pathogenesis of hematological malignancies. Biomed Pharmacother. 2021;144:112334.
340. Roush S, Slack FJ. The let-7 family of microRNAs. Trends Cell Biol. 2008;18(10):505-16.
341. Nair VS, Maeda LS, Ioannidis JP. Clinical outcome prediction by microRNAs in human cancer: a systematic review. J Natl Cancer Inst. 2012;104(7):528-40.
342. Torrisani J, Bournet B, du Rieu MC, Bouisson M, Souque A, Escourrou J, et al. let-7 MicroRNA transfer in pancreatic cancer-derived cells inhibits in vitro cell proliferation but fails to alter tumor progression. Hum Gene Ther. 2009;20(8):831-44.

343. Pan X, Wang G, Wang B. MicroRNA-1182 and let-7a exert synergistic inhibition on invasion, migration and autophagy of cholangiocarcinoma cells through down-regulation of NUAK1. Cancer Cell Int. 2021;21(1):161.
344. Nweke EE, Brand M. Downregulation of the let-7 family of microRNAs may promote insulin receptor/insulin-like growth factor signalling pathways in pancreatic ductal adenocarcinoma. Oncol Lett. 2020;20(3):2613-20.
345. Chang TC, Yu D, Lee YS, Wentzel EA, Arking DE, West KM, et al. Widespread microRNA repression by Myc contributes to tumorigenesis. Nat Genet. 2008;40(1):43-50.
346. Guo Y, Chen Y, Ito H, Watanabe A, Ge X, Kodama T, et al. Identification and characterization of lin-28 homolog B (LIN28B) in human hepatocellular carcinoma. Gene. 2006;384:51-61.
347. Piskounova E, Polytarchou C, Thornton JE, LaPierre RJ, Pothoulakis C, Hagan JP, et al. Lin28A and Lin28B inhibit let-7 microRNA biogenesis by distinct mechanisms. Cell. 2011;147(5):1066-79.
348. Wang Y, Li J, Guo S, Ouyang Y, Yin L, Liu S, et al. Lin28B facilitates the progression and metastasis of pancreatic ductal adenocarcinoma. Oncotarget. 2017;8(36):60414-28.
349. Franses JW, Philipp J, Missios P, Bhan I, Liu A, Yashaswini C, et al. Pancreatic circulating tumor cell profiling identifies LIN28B as a metastasis driver and drug target. Nat Commun. 2020;11(1):3303.
350. Banko P, Lee SY, Nagygyorgy V, Zrinyi M, Chae CH, Cho DH, et al. Technologies for circulating tumor cell separation from whole blood. J Hematol Oncol. 2019;12(1):48.
351. Wang T, Wang G, Hao D, Liu X, Wang D, Ning N, et al. Aberrant regulation of the LIN28A/LIN28B and let-7 loop in human malignant tumors and its effects on the hallmarks of cancer. Mol Cancer. 2015;14:125.
352. Frampton AE, Krell J, Jamieson NB, Gall TM, Giovannetti E, Funel N, et al. microRNAs with prognostic significance in pancreatic ductal adenocarcinoma: A meta-analysis. Eur J Cancer. 2015;51(11):1389-404.
353. Frampton AE, Castellano L, Colombo T, Giovannetti E, Krell J, Jacob J, et al. MicroRNAs cooperatively inhibit a network of tumor suppressor genes to promote pancreatic tumor growth and progression. Gastroenterology. 2014;146(1):268-77 e18.
354. Moriyama T, Ohuchida K, Mizumoto K, Yu J, Sato N, Nabae T, et al. MicroRNA-21 modulates biological functions of pancreatic cancer cells including their proliferation, invasion, and chemoresistance. Molecular cancer therapeutics. 2009;8(5):1067-74.
355. Giovannetti E, Funel N, Peters GJ, Del Chiaro M, Erozenci LA, Vasile E, et al. MicroRNA-21 in pancreatic cancer: correlation with clinical outcome and pharmacologic aspects underlying its role in the modulation of gemcitabine activity. Cancer Res. 2010;70(11):4528-38.
356. Ali S, Almhanna K, Chen W, Philip PA, Sarkar FH. Differentially expressed miRNAs in the plasma may provide a molecular signature for aggressive pancreatic cancer. Am J Transl Res. 2010;3(1):28-47.
357. Abue M, Yokoyama M, Shibuya R, Tamai K, Yamaguchi K, Sato I, et al. Circulating miR-483-3p and miR-21 is highly expressed in plasma of pancreatic cancer. Int J Oncol. 2015;46(2):539-47.
358. Karasek P, Gablo N, Hlavsa J, Kiss I, Vychytilova-Faltejskova P, Hermanova M, et al. 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):321-7.
359. Hu L, Han D, Yu D, Ao D, Yang Z. Circulating Blood miR-155 and miR-21 Promote the Development of Acute Pancreatitis and Can Be Used to Assess the Risk Stratification of Pancreatitis. J Healthc Eng. 2021;2021:2064162.

360. Li H, Yang X, Cao B, Guan J. Increased plasma clusterin and miR-21 in acute pancreatitis. Br J Biomed Sci. 2021;78(4):229-32.
361. Yazdanpanah Z, Kazemipour N, Kalantar SM, Vahidi Mehrjardi MY. Plasma miR-21 as a potential predictor in prediabetic individuals with a positive family history of type 2 diabetes mellitus. Physiol Rep. 2022;10(2):e15163.
362. Motamedi M, Hashemzadeh Chaleshtori M, Ghasemi S, Mokarian F. Plasma Level Of miR-21 And miR-451 In Primary And Recurrent Breast Cancer Patients. Breast Cancer (Dove Med Press). 2019;11:293-301.
363. Nakka M, Allen-Rhoades W, Li Y, Kelly AJ, Shen J, Taylor AM, et al. Biomarker significance of plasma and tumor miR-21, miR-221, and miR-106a in osteosarcoma. Oncotarget. 2017;8(57):96738-52.
364. Chen H, Liu H, Zou H, Chen R, Dou Y, Sheng S, et al. Evaluation of Plasma miR-21 and miR-152 as Diagnostic Biomarkers for Common Types of Human Cancers. J Cancer. 2016;7(5):490-9.
365. Kanaan Z, Rai SN, Eichenberger MR, Roberts H, Keskey B, Pan J, et al. Plasma miR-21: a potential diagnostic marker of colorectal cancer. Ann Surg. 2012;256(3):544-51.
366. Humphries B, Yang C. The microRNA-200 family: small molecules with novel roles in cancer development, progression and therapy. Oncotarget. 2015;6(9):6472-98.
367. Yu J, Ohuchida K, Mizumoto K, Sato N, Kayashima T, Fujita H, et al. MicroRNA, hsa-miR-200c, is an independent prognostic factor in pancreatic cancer and its upregulation inhibits pancreatic cancer invasion but increases cell proliferation. Mol Cancer. 2010;9(1):169.
368. Paik WH, Song BJ, Kim HW, Kim HR, Hwang J-H. MicroRNA-200c as a prognostic biomarker for pancreatic cancer. The Korean Journal of Gastroenterology. 2015;66(4):215-20.
369. Liu J-P, Shang J, Kang Y, Mao C-S, Ning H-B, Ma C. Expression of microRNA-200c in human pancreatic ductal adenocarcinoma and its prognostic significance. Int J Clin Exp Pathol. 2016;9(9):9455-60.
370. Diaz-Riascos ZV, Ginesta MM, Fabregat J, Serrano T, Busquets J, Buscail L, et al. Expression and Role of MicroRNAs from the miR-200 Family in the Tumor Formation and Metastatic Propensity of Pancreatic Cancer. Mol Ther Nucleic Acids. 2019;17:491-503.
371. Title AC, Hong SJ, Pires ND, Hasenohrl L, Godbersen S, Stokar-Regenscheit N, et al. Genetic dissection of the miR-200-Zeb1 axis reveals its importance in tumor differentiation and invasion. Nat Commun. 2018;9(1):4671.
372. Gregory PA, Bert AG, Paterson EL, Barry SC, Tsykin A, Farshid G, et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nat Cell Biol. 2008;10(5):593-601.
373. Burk U, Schubert J, Wellner U, Schmalhofer O, Vincan E, Spaderna S, et al. A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO reports. 2008;9(6):582-9.
374. Brabletz T, Kalluri R, Nieto MA, Weinberg RA. EMT in cancer. Nat Rev Cancer. 2018;18(2):128-34.
375. Dykxhoorn DM, Wu Y, Xie H, Yu F, Lal A, Petrocca F, et al. miR-200 enhances mouse breast cancer cell colonization to form distant metastases. PLoS One. 2009;4(9):e7181.
376. Lo HC, Zhang XH. EMT in Metastasis: Finding the Right Balance. Dev Cell. 2018;45(6):663-5.
377. Reichert M, Bakir B, Moreira L, Pitarresi JR, Feldmann K, Simon L, et al. Regulation of Epithelial Plasticity Determines Metastatic Organotropism in Pancreatic Cancer. Dev Cell. 2018;45(6):696-711 e8.

378. Zheng X, Carstens JL, Kim J, Scheible M, Kaye J, Sugimoto H, et al. Epithelial-to-mesenchymal transition is dispensable for metastasis but induces chemoresistance in pancreatic cancer. Nature. 2015;527(7579):525-30.
379. Meng F, Henson R, Lang M, Wehbe H, Maheshwari S, Mendell JT, et al. Involvement of human micro-RNA in growth and response to chemotherapy in human cholangiocarcinoma cell lines. Gastroenterology. 2006;130(7):2113-29.
380. Li Y, VandenBoom TG, 2nd, Kong D, Wang Z, Ali S, Philip PA, et al. Up-regulation of miR-200 and let-7 by natural agents leads to the reversal of epithelial-to-mesenchymal transition in gemcitabine-resistant pancreatic cancer cells. Cancer Res. 2009;69(16):6704-12.
381. Bracken CP, Li X, Wright JA, Lawrence DM, Pillman KA, Salmanidis M, et al. Genome-wide identification of miR-200 targets reveals a regulatory network controlling cell invasion. EMBO J. 2014;33(18):2040-56.
382. Korc M. Pathogenesis of Pancreatic Cancer-Related Diabetes Mellitus: Quo Vadis? Pancreas. 2019;48(5):594-7.
383. Ohshima K, Inoue K, Fujiwara A, Hatakeyama K, Kanto K, Watanabe Y, et al. Let-7 microRNA family is selectively secreted into the extracellular environment via exosomes in a metastatic gastric cancer cell line. PloS one. 2010;5(10):e13247.
384. Babst M. MVB vesicle formation: ESCRT-dependent, ESCRT-independent and everything in between. Current opinion in cell biology. 2011;23(4):452-7.
385. Gibbings DJ, Ciaudo C, Erhardt M, Voinnet O. Multivesicular bodies associate with components of miRNA effector complexes and modulate miRNA activity. Nat Cell Biol. 2009;11(9):1143-9.
386. Kosaka N, Iguchi H, Yoshioka Y, Takeshita F, Matsuki Y, Ochiya T. Secretory mechanisms and intercellular transfer of microRNAs in living cells. J Biol Chem. 2010;285(23):17442-52.
387. Matsui T, Osaki F, Hiragi S, Sakamaki Y, Fukuda M. ALIX and ceramide differentially control polarized small extracellular vesicle release from epithelial cells. EMBO Rep. 2021;22(5):e51475.
388. Kirschner MB, van Zandwijk N, Reid G. Cell-free microRNAs: potential biomarkers in need of standardized reporting. Front Genet. 2013;4:56.
389. Dong Q, Liu X, Cheng K, Sheng J, Kong J, Liu T. Pre-metastatic Niche Formation in Different Organs Induced by Tumor Extracellular Vesicles. Front Cell Dev Biol. 2021;9:733627.
390. Ocana OH, Corcoles R, Fabra A, Moreno-Bueno G, Acloque H, Vega S, et al. Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. Cancer Cell. 2012;22(6):709-24.
391. Luzzi KJ, MacDonald IC, Schmidt EE, Kerkvliet N, Morris VL, Chambers AF, et al. Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. Am J Pathol. 1998;153(3):865-73.
392. Kayahara M, Nagakawa T, Ohta T, Kitagawa H, Ueno K, Tajima H, et al. Analysis of paraaortic lymph node involvement in pancreatic carcinoma: a significant indication for surgery? Cancer. 1999;85(3):583-90.
393. Roche CJ, Hughes ML, Garvey CJ, Campbell F, White DA, Jones L, et al. CT and pathologic assessment of prospective nodal staging in patients with ductal adenocarcinoma of the head of the pancreas. AJR Am J Roentgenol. 2003;180(2):475-80.
394. Loch FN, Asbach P, Haas M, Seeliger H, Beyer K, Schineis C, et al. Accuracy of various criteria for lymph node staging in ductal adenocarcinoma of the pancreatic head by computed tomography and magnetic resonance imaging. World Journal of Surgical Oncology. 2020;18:1-10.

395. Benesova L, Halkova T, Bunganic B, Belsanova B, Zavoral M, Traboulsi E, et al. Comparison of Native Aspirates and Cytological Smears Obtained by EUS-Guided Biopsies for Effective DNA/RNA Marker Testing in Pancreatic Cancer. Pathol Oncol Res. 2020;26(1):379-85.
396. Bourgon R, Gentleman R, Huber W. Independent filtering increases detection power for high-throughput experiments. Proc Natl Acad Sci U S A. 2010;107(21):9546-51.
397. Xu R, Greening DW, Chen M, Rai A, Ji H, Takahashi N, et al. Surfaceome of Exosomes Secreted from the Colorectal Cancer Cell Line SW480: Peripheral and Integral Membrane Proteins Analyzed by Proteolysis and TX114. Proteomics. 2019;19(8):e1700453.
398. Betzig E, Patterson GH, Sougrat R, Lindwasser OW, Olenych S, Bonifacino JS, et al. Imaging intracellular fluorescent proteins at nanometer resolution. science. 2006;313(5793):1642-5.
399. Liang H, Manne S, Shick J, Lissoos T, Dolin P. Incidence, prevalence, and natural history of primary sclerosing cholangitis in the United Kingdom. Medicine (Baltimore). 2017;96(24):e7116.
400. Hingorani SR, Petricoin EF, Maitra A, Rajapakse V, King C, Jacobetz MA, et al. Preinvasive and invasive ductal pancreatic cancer and its early detection in the mouse. Cancer Cell. 2003;4(6):437-50.
401. Hingorani SR, Wang L, Multani AS, Combs C, Deramaudt TB, Hruban RH, et al. Trp53R172H and KrasG12D cooperate to promote chromosomal instability and widely metastatic pancreatic ductal adenocarcinoma in mice. Cancer Cell. 2005;7(5):469-83.
402. Ariston Gabriel AN, Jiao Q, Yvette U, Yang X, Al-Ameri SA, Du L, et al. Differences between KC and KPC pancreatic ductal adenocarcinoma mice models, in terms of their modeling biology and their clinical relevance. Pancreatology. 2020;20(1):79-88.
403. Tosar JP, Gámbaro F, Sanguinetti J, Bonilla B, Witwer KW, Cayota A. Assessment of small RNA sorting into different extracellular fractions revealed by high-throughput sequencing of breast cell lines. Nucleic acids research. 2015;43(11):5601-16.
404. Dhahbi JM, Spindler SR, Atamna H, Yamakawa A, Boffelli D, Mote P, et al. 5' tRNA halves are present as abundant complexes in serum, concentrated in blood cells, and modulated by aging and calorie restriction. BMC Genomics. 2013;14:298.
405. Xu H, Yao J, Wu DC, Lambowitz AM. Improved TGIRT-seq methods for comprehensive transcriptome profiling with decreased adapter dimer formation and bias correction. Sci Rep. 2019;9(1):7953.
406. Bernard V, Kim DU, San Lucas FA, Castillo J, Allenson K, Mulu FC, et al. Circulating Nucleic Acids Are Associated With Outcomes of Patients With Pancreatic Cancer. Gastroenterology. 2019;156(1):108-18 e4.
407. Yang S, Che SP, Kurywchak P, Tavormina JL, Gansmo LB, Correa de Sampaio P, et al. Detection of mutant KRAS and TP53 DNA in circulating exosomes from healthy individuals and patients with pancreatic cancer. Cancer Biol Ther. 2017;18(3):158-65.
408. Allenson K, Castillo J, San Lucas FA, Scelo G, Kim DU, Bernard V, et al. High prevalence of mutant KRAS in circulating exosome-derived DNA from early-stage pancreatic cancer patients. Ann Oncol. 2017;28(4):741-7.
409. Kahlert C, Melo SA, Protopopov A, Tang J, Seth S, Koch M, et al. Identification of double-stranded genomic DNA spanning all chromosomes with mutated KRAS and p53 DNA in the serum exosomes of patients with pancreatic cancer. J Biol Chem. 2014;289(7):3869-75.
410. Feng Z, Li K, Qin K, Liang J, Shi M, Ma Y, et al. The LINC00623/NAT10 signaling axis promotes pancreatic cancer progression by remodeling ac4C modification of mRNA. J Hematol Oncol. 2022;15(1):112.

411. Li Z, Jiang P, Li J, Peng M, Zhao X, Zhang X, et al. Tumor-derived exosomal lnc-Sox2ot promotes EMT and stemness by acting as a ceRNA in pancreatic ductal adenocarcinoma. Oncogene. 2018;37(28):3822-38.
412. Kitagawa T, Taniuchi K, Tsuboi M, Sakaguchi M, Kohsaki T, Okabayashi T, et al. Circulating pancreatic cancer exosomal RNAs for detection of pancreatic cancer. Mol Oncol. 2019;13(2):212-27.
413. Ko J, Bhagwat N, Yee SS, Ortiz N, Sahmoud A, Black T, et al. Combining Machine Learning and Nanofluidic Technology To Diagnose Pancreatic Cancer Using Exosomes. ACS Nano. 2017;11(11):11182-93.
414. Shao H, Zhang Y, Yan J, Ban X, Fan X, Chang X, et al. Upregulated MicroRNA-483-3p is an Early Event in Pancreatic Ductal Adenocarcinoma (PDAC) and as a Powerful Liquid Biopsy Biomarker in PDAC. Onco Targets Ther. 2021;14:2163-75.
415. Wang C, Wang J, Cui W, Liu Y, Zhou H, Wang Y, et al. Serum Exosomal miRNA-1226 as Potential Biomarker of Pancreatic Ductal Adenocarcinoma. Onco Targets Ther. 2021;14:1441-51.
416. Xu YF, Xu X, Bhandari K, Gin A, Rao CV, Morris KT, et al. Isolation of extracellular vesicles in the context of pancreatic adenocarcinomas: Addition of one stringent filtration step improves recovery of specific microRNAs. PLoS One. 2021;16(11):e0259563.
417. Sato Y, Suzuki R, Takagi T, Sugimoto M, Ohira H. Circulating extracellular vesicle-encapsulated microRNA as screening biomarkers for intraductal papillary mucinous neoplasm. Oncol Lett. 2020;20(6):315.
418. Vicentini C, Calore F, Nigita G, Fadda P, Simbolo M, Sperandio N, et al. Exosomal miRNA signatures of pancreatic lesions. BMC Gastroenterol. 2020;20(1):137.
419. Guo S, Qin H, Liu K, Wang H, Bai S, Liu S, et al. Blood small extracellular vesicles derived miRNAs to differentiate pancreatic ductal adenocarcinoma from chronic pancreatitis. Clin Transl Med. 2021;11(9):e520.
420. Zou X, Wei J, Huang Z, Zhou X, Lu Z, Zhu W, et al. Identification of a six-miRNA panel in serum benefiting pancreatic cancer diagnosis. Cancer Med. 2019;8(6):2810-22.
421. Li Q, Geng S, Yuan H, Li Y, Zhang S, Pu L, et al. Circular RNA expression profiles in extracellular vesicles from the plasma of patients with pancreatic ductal adenocarcinoma. FEBS Open Bio. 2019;9(12):2052-62.
422. Wang W, Lou W, Ding B, Yang B, Lu H, Kong Q, et al. A novel mRNA-miRNA-lncRNA competing endogenous RNA triple sub-network associated with prognosis of pancreatic cancer. Aging (Albany NY). 2019;11(9):2610-27.
423. Takahasi K, Iinuma H, Wada K, Minezaki S, Kawamura S, Kainuma M, et al. Usefulness of exosome-encapsulated microRNA-451a as a minimally invasive biomarker for prediction of recurrence and prognosis in pancreatic ductal adenocarcinoma. J Hepatobiliary Pancreat Sci. 2018;25(2):155-61.
424. Li Z, Yanfang W, Li J, Jiang P, Peng T, Chen K, et al. Tumor-released exosomal circular RNA PDE8A promotes invasive growth via the miR-338/MACC1/MET pathway in pancreatic cancer. Cancer Lett. 2018;432:237-50.
425. Li Z, Tao Y, Wang X, Jiang P, Li J, Peng M, et al. Tumor-Secreted Exosomal miR-222 Promotes Tumor Progression via Regulating P27 Expression and Re-Localization in Pancreatic Cancer. Cell Physiol Biochem. 2018;51(2):610-29.
426. Li J, Li Z, Jiang P, Peng M, Zhang X, Chen K, et al. Circular RNA IARS (circ-IARS) secreted by pancreatic cancer cells and located within exosomes regulates endothelial monolayer permeability to promote tumor metastasis. J Exp Clin Cancer Res. 2018;37(1):177.
427. Bartsch DK, Gercke N, Strauch K, Wieboldt R, Matthai E, Wagner V, et al. The Combination of MiRNA-196b, LCN2, and TIMP1 is a Potential Set of Circulating

Biomarkers for Screening Individuals at Risk for Familial Pancreatic Cancer. J Clin Med. 2018;7(10).
428. Mikamori M, Yamada D, Eguchi H, Hasegawa S, Kishimoto T, Tomimaru Y, et al. MicroRNA-155 Controls Exosome Synthesis and Promotes Gemcitabine Resistance in Pancreatic Ductal Adenocarcinoma. Sci Rep. 2017;7:42339.
429. Chen D, Wu X, Xia M, Wu F, Ding J, Jiao Y, et al. Upregulated exosomic miR-23b-3p plays regulatory roles in the progression of pancreatic cancer. Oncol Rep. 2017;38(4):2182-8.
430. Madhavan B, Yue S, Galli U, Rana S, Gross W, Muller M, et al. Combined evaluation of a panel of protein and miRNA serum-exosome biomarkers for pancreatic cancer diagnosis increases sensitivity and specificity. Int J Cancer. 2015;136(11):2616-27.

Appendix

Table 1: Summary of results from overall project.

In order to clarify the positive findings from this project, each arm/study of the overall project is shown according to whether there have been validated potential biomarkers (miRNAs in bile cfRNA, bile EV-RNA and plasma EV-RNA), studies with identified biomarkers but not yet validated (bile cfRNA tRFs, bile EV markers from surgical samples, plasma cfRNA miRNAs) and finally those with no statistically significant differential expression.

cfRNA- cell free RNA, tRFs – Transfer RNA fragments.

Summary Of Overall Experimental Findings From Project

Study Name	RNA subtype	n	Discovery	n	Validation Candidates	Normalisers	Results
Validated Studies							
Bile cfRNA	miRNA	78	*Malignant vs. Benign* 88 candidates (38 upregulated)	87	miR-182-5p miR-199a-5 miR-196a-5p miR-335-5p miR-340-5p miR-424-5p	miR-26a let-7b	*Malignant vs. Benign* **miR-182-5p upregulated** (p=0.0179) **miR-340-5p downregulated** (p=0.0002) miR-196a-5p upregulated (p=0.0671) miR-199a-5p, miR-424-5p, miR-335-5p not validated.
Bile EV RNA (MIRABILE)	miRNA	30	*Malignant vs. Benign* 30 candidates (4 upregulated) *PDAC vs. Benign* 42 candidates (5 upregulated) *CCA vs. Benign* 21 candidates (4 upregulated)	81	*PDAC vs. Benign* let-7a-5p miR-21-5p miR-200a-3p miR-200b-3p miR-200c-3p miR-429 miR-320a-3p miR-320b miR-320c miR-320d *CCA vs. Benign* miR-181a-2-3p miR-182-5p	miR-125a-5p miR-194-5p UniSp6	*PDAC vs. Benign* **let-7a-5p, miR-21-5p, miR-200a-3p and miR-429 upregulated** (p<0.0001) **miR-320a upregulated** (p=0.424) *CCA vs. Benign* **miR-200a-3p upregulated** (p=0.0057) **miR-429 upregulated** (p=0.0468) **miR-320a upregulated** (p= 0.0491) *Malignant vs. Benign* **miR-200a-3p/miR-429/miR-320a** AUC of 0.9150 miR-181a-2-3p, miR-182-5p not validated.
Plasma EV RNA	miRNA	20	*PDAC vs. Benign* 15 candidates (12 upregulated)	61	miR-141-3p miR-200a-3p miR-200b-3p miR-200c-3p miR-429 Additional miRNAs detected: miR-21-5p miR-92b-3p miR-148a-3p miR-150-5p miR-192-5p miR-194-5p miR-196a-5p miR-342-3p miR-375-3p miR-423-5p	miR-23a miR-26a UniSp6	*PDAC vs. Benign* **miR-200a-3p and miR-141-3p upregulated** (p<0.0001) **miR-200b-3p upregulated** (p=0.0007) **miR-200c-3p upregulated** (p=0.0011) **miR-429 upregulated** (p=0.0112) Overall AUC 0.8226 (5-miR)
Non-validated Studies (RNA-Seq performed)							
Bile cfRNA	tRF	78	*Malignant vs. Benign* 264 candidates (156 upregulated)	n/a	See Table 14		
Plasma cfRNA	miRNA	41	*PDAC vs. Benign* 2 candidates (1 upregulated)	n/a	miR-144-5p miR-29-3p		
Bile EV RNA (Surgical)	miRNA	18	*Malignant vs. Benign* 8 candidates (5 upregulated) *PDAC vs. Benign* 2 candidates (1 upregulated) *CCA vs. Benign* 6 candidates (4 upregulated)	n/a	miR-4516 miR-141-3p (downregulated)		
Bile EV RNA (Surgical)	tRF	18	*Malignant vs. Benign* 92 candidates (15 upregulated) *PDAC vs. Benign* 69 candidates (13 upregulated) *CCA vs. Benign* 163 candidates (54 upregulated)	n/a	See Table 24		
Studies with negative findings							
Bile EV RNA (MIRABILE)	tRF	30	No significant candidates found.	n/a			
Plasma EV RNA	tRF	20	No significant candidates found.	n/a			

Table 2: Summary of Plasma EV studies in PDAC.

EV DNA studies

First Author	Year	Sample Size	Biofluid	Plasma Biomarker	Significant comments
Bernard *et al.* (406)	2019	PDAC (n=194)	Plasma	**Mutant KRAS allelic fraction (MAF)**	Analysis of plasma samples showed that a KRAS MAF peak of ≥1% in any on-treatment serial exoDNA sample was significantly associated with eventual disease progression, as determined by RECIST 1.1 (P < .0001). However, 12%–25% with preneoplastic pancreatic cysts or nonneoplastic pancreatic diseases had detectable circulating mutants. KRAS mutations are present in up to 80% of preneoplastic pancreatic cysts (including low-grade mucinous cysts).
Yang *et al.* (407)	2017	PDAC (n=48), IPMN (n=7), CP (n=9), Controls (n=114)	Serum	**$KRAS^{G12D}$** and **$TP53^{R273H}$**	In PDAC patients, digital PCR analyses of exosomal DNA identified $KRAS^{G12D}$ mutation in 39.6% of cases, and $TP53^{R273H}$ mutation in 4.2% of cases.
Allenson *et al.* (408)	2017	Discovery: PDAC all stages (n=68), PDAC localized (n=20), Controls (n=54) Validation: PDAC early stage (n=39), Controls (n=82)	Plasma	**Mutant KRAS allelic fraction (MAF)**	In the validation cohort, mutant KRAS exoDNA was detected in 43.6% of early-stage PDAC patients and 20% of healthy controls. Higher KRAS MAF was also associated with reduced Disease Free Survival (DFS) in patients.
Kahlert *et al.* (409)	2014	PDAC (n=2), Controls (n=2)	Serum	Mutations in **KRAS** and **p53**	Study provides evidence that exosomes can carry large fragments (~10 kb) of double-stranded genomic DNA.

EV long non-coding RNA studies

First Author	Year	No. of Patients	Biofluid	Plasma Biomarker	Significant comments
Feng *et al.* (410)	2022	Discovery: PDAC (n=5) Control (n=5) Validation: PDAC (n=73) Controls (n = 42)	Serum	**LINC00623**	AUC1 = 0.874. Additionally, LINC00623 had a higher AUC value for distinguishing PDAC patients with normal CA19-9 levels.
Li *et al.*(411)	2018	PDAC (n=20) Control (n=20)	Plasma	**lncRNA-Sox2ot** upregulated	Expression of Sox2ot was significantly associated with TNM stage (P= 0.014) and was also related to lymphatic and vascular invasion. Sox2ot competitively binds to the miR-200 family to regulate the expression of Sox2, thus promoting invasion and metastasis of PDAC.

EV mRNA studies

Yang *et al.* (222)	2020	Discovery: Controls (n=12) PDAC (n=17) Training: Controls (n=27) PDAC (n=20) Test: Controls (n=79) PDAC (n=57)	Plasma	Combined EV mRNA and miRNA. **CA 19-9, EV CK18, EV CD63, EV miR-409 and circulating DNA**	Best EV mRNA marker was CK18 (Accuracy = 66%, 95% CI 58–73%) and best performing EV miRNA marker was miR-409 (Accuracy=59%, 95% CI 55–63%).
Kitagawa *et al.*(412)	2018	PDAC (n=27) Controls (n=13)	Serum	**2 mRNAs (WASF2, ARF6) and 2 snoRNAs (SNORA74A, SNORA25)**	The AUCs of WASF2, ARF6, SNORA74A, and SNORA25 in serum from patients in the early stages of PDAC (stages 0, I, and IIA) were > 0.90, compared with an AUC of 0.93 for serum CA19-9
Ko *et al.* (413)	2017	Training: PDAC (n=5) Controls (n=5). Test: PDAC (n=12) Controls (n=12)		**ARG1, CD63, CK18, Erbb3, GAPDH, HGF3A, KRAS, ODC1**	EpCAM isolation compared with pan-exosome and found to be better for discrimination.
EV non-coding RNA studies					
Nishiwada *et al.* (213)	2022	Discovery (n = 25) (training cohort: n = 82, validation cohort: n = 57) clinical cohort (n = 46)	Plasma	6-exo-miRNA risk-prediction model - **miR-130b-5p, miR-133a-3p, miR-195-5p, miR-432-5p, miR-1229-3p, and miR-1273f**	Upregulated in early recurrence, predicting this with AUC value of 0.81 (95% CI = 0.70–0.89), and a corresponding specificity of 0.85 and a sensitivity of 0.72. Validation AUC was 0.78 (95% CI = 0.65–0.88)
Nakamura *et al.* (214)	2022	Discovery: PDAC (n=44) Controls (n=57), Training/validation: PDAC (n=124), Controls (n=67)	Plasma and serum	13 markers (5 cf- and 8 exo-miRNAs) **miR200c-3p, miR148a-3p, miR216a-5p, miR145-5p, miR200b-3p, miR143-3p, miR34a-5p, miR429, miR141–3p, miR1260b, miR145–3p, miR216b-5p, miR200a-3p, miR1260a, miR217-5p**	A combined signature was significantly superior and yielded an AUC value of 0.93 in the validation cohort (exo-miRNA alone was 0.89). Validation was particularly good in identification of patients with stage I and II cancers (AUC 0.93; sensitivity, 80%; specificity, 91%; PPV, 93%; NPV, 76%; and accuracy 84%). Additional diagnostic potential in CA 19-9 subgroup.
Zeöld *et al.*(279)	2021	PDAC (n=9) CP (n=8) Controls (n=8)	Plasma	**miR-21** and **miR-195**	Not statistically significant after adjustment by Benjamini–Hochberg FDR. Evaluated expression of organoid EVs.

Shao *et al.* (414)	2021	PDAC (n=63) Controls (n=22)	Serum	**miR-483-3p**	Exosomal miR-483-3p had an AUC of 0.69 (95% CI, 0.58–0.81) and was not correlated with serum miR-483-3p levels which had an AUC of 0.81 (95% CI, 0.71–0.91).
Wang *et al.* (415)	2021	PDAC (n=5) Controls (n=8)	Serum	**miR-203b-5p, miR-342-5p, miR-337-5p, miR-149-5p, miR-877-5p, miR-203a-3p** upregulated	
Xu *et al.* (416)	2021	PDAC (n=7) Controls (n=7)	Plasma	Increased **miR-196a** expression in PDAC	miR-1246 levels are not elevated in patients with early stage PDAC
Kim *et al.*(278)	2021	PDAC (n=20) Controls (n=20)	Plasma	**miR-10b, miR-21, miR-155, miR-429,** and **miR-1290** (AUC>0.65)	AUC values for miR-10b, miR-21, miR-155, miR-429, and miR-1290 were 0.693 (95% CI 0.529–0.827), 0.771 (95% CI, 0.614–0.888), 0.843 (95% CI, 0.695–0.937), 0.867 (95% CI, 0.529–0.827), and 0.786 (95% CI, 0.630–0.898) respectively.
Sato *et al.* (417)	2020	Discovery: IPMN (n=2), IPMN-derived PDAC (n=2) Controls (n=4) Validation: IPMN (n=38) Controls (n=21)	Serum	**miR-6132, hsa-miR-22-3p** and **hsa-miR-4539**	IPMNs vs control, only EV-miR-4539 showed a significant difference (P=0.004) with AUC of 0.72 for EV-miR-4539 (95% CI: 0.59–0.85).
Vicentini *et al.* (418)	2020	PDAC (n=58), NET (n=42), IPMN (n=20), Ampullary (n=20), CP (n=15)	Plasma	**miR-372-3p; miR-140-3p; miR-644a** upregulated in PDAC.	Selected miRNAs (miR-106-5p; miR-4454; miR-7975; miR-320e; miR-19b-3p; miR-16-5p and miR-17-5p) were validated by RT-qPCR after NanoString technique for miRNAs profiling and three selected miRNAs (miR-4454; miR-106a-5p and miR-17-5p) were then validated also by in situ hybridization.
Kumar *et al.* (338)	2020	PDAC (n=2), Controls (n=2)	Serum	None specified	
Reese *et al.* (329)	2020	PDAC (n=56) CP (n=11) Controls (n=22)	Serum	**MiR-200b** (AUC 0.79; $p = 0.0001$) PDAC vs CP **MiR-200c** (AUC of 0.67; $p = 0.0239$) PDAC vs Controls	Also compared miR-200b from EpCAM positive exosomes but was less accurate. Combining all four miR fractions with CA.19-9 entailed an outstanding diagnostic accuracy of 97% ($p < 0.0001$). High expression of miR-200b in EpCAM-positive serum exosomes ($p = 0.032$, median OS 9 months (high) vs. 18 months (low)) and miR-200c in total serum exosomes ($p = 0.038$, median OS 11 months (high) vs. 18 months (low)) were associated with shorter OS of patients.

Guo *et al.* (419)	2020	Training cohort: PDAC (n = 30), CP (n=18) Test cohort: PDAC (n = 27), CP (n=15)	Plasma	**EV miR-95-3p/EV miR-26b-5p** ratio	In the independent test cohort the AUC for the miR-95-3p/miR-26b-5p was 0.875 with a sensitivity to distinguish PDAC from CP patients of 81.5% and specificity of 93.3%. Additionally, miR-335-5p/miR-340-5p ratio was identified to associate with PDAC metastasis and poor prognosis.
Kawamura *et al.* (277)	2019	PDAC (n=55) Controls (n=20)	Portal vein plasma	**miR-451a, miR-21, miR-4525**	Exosomal miR-4525, miR-451a and miR-21 levels in portal vein blood from PDAC patients were significantly higher than those found in the peripheral blood.
Zou *et al.* (420)	2019	Discovery: PDAC (n=20), Controls (n=10) Training: PDAC (n=36), Controls (n=36) Testing phase: PDAC (n=93) Controls (n=71) Ext Validation: PDAC (n=30), Controls (n=30)	Serum	**miR-192-5p, miR-19a-3p,** and **miR-19b-3p in exosomes**	Predominately a circulating marker analysis which identified let-7b-5p, miR-192-5p, miR-19a-3p, miR-19b-3p, miR-223-3p, and miR-25-3p with AUC of 0.910 (95% CI: 0.872-0.948; sensitivity = 95.3%, specificity = 76.7%).
Li *et al.* (421)	2019	PDAC (n=8), Controls (n=8)	Plasma	**hsa_circ_0002130, hsa_circ_0000896, hsa_circ_0101692, hsa_circ_0005882, hsa_circ_0001250, hsa_circ_0037096, hsa_circ_0000128, hsa_circ_0000606, hsa_circ_0103896, hsa_circ_0006662, hsa_circ_0092763, hsa_circ_0094190**	10 out of 13 circRNAs ($P < 0.05$) validated by RT-qPCR.
Zhou *et al.* (262)	2018	Training: PDAC/Control (n=40, n=40) Testing: PDAC/Control (n=112, n=116) Exosomal: PDAC/Control (n=31, n=37)	Plasma	**miR-122-5p** and **miR-193b-3p** were up-regulated**, miR-221-3p** was down-regulated	The AUCs for exosomal miR-122-5p (0.722; 95% CI: 0.591–0.853), exosomal miR-193b-3p (0.651; 95% CI: 0.51–0.792) and the signature of the two exosomal miRNAs combined (0.849; 95% CI: 0.756–0.942).
Wang *et al.* (422)	2018	PDAC (n=50) Controls (n=12)	Serum	**miR-301a-3p**	Shown to predict TNM classification.

Takahasi *et al.*(423)	2018	PDAC (n=56) Control (n=3)	Plasma	**miR-451a**	Divided into high and low expression. Positively associated with tumour size, stage, negatively associated with disease free (P=0.004) and overall survival (P=0.001).
Li *et al.*(424)	2018	PDAC (n=93)	Plasma	**Circ-PDE8A**	Up-regulation was significantly associated with lymphatic invasion (*P=0.014), T factor (P=0.049) and TNM stage (P =0.005).*
Li *et al.* (425)	2018	PDAC (n=73)	Plasma	**miR-222**	High in PDAC patients and significantly correlated to tumour size and TNM stage, and was an independent risk factor for survival.
Li *et al.* (426)	2018	PDAC (n=40)	Plasma	**Circ-IARS**	Circ-IARS associated with tumour vessel invasion, liver metastasis, and TNM stage. It was also shown to competitively bind miR-122, inhibit its expression and release inhibition of downstream target gene RhoA activity, increase the expression of F-actin, and promote cell contraction.
Goto *et al.*(276)	2018	PDAC (n=32), IPMN (n=29), controls (n=22)	Serum	**miR-191, miR-21** and **miR-451a**	Significantly up-regulated in patients with PDAC and IPMN compared to controls (*P<0.05). The AUC, diagnostic accuracy and specificity of the 3 exosomal miRs were superior to circulating miRs. However, CA19–9 was still superior for the diagnosis of advanced PDAC.*
Bartsch *et al.*(427)	2018	PDAC (n=96) Controls (n=20)	Serum	The panel **miR-196b/LCN2/TIMP1**	Could distinguish high-grade lesions and stage I PDAC from controls with absolute specificity and sensitivity.
Xu *et al.*(328)	2017	PDAC (n=15) Controls (n=15)	Plasma	**miR-196a, miR-1246**	These were enriched in localized PDAC. Immunoaffinity isolation using GPC-1 antibodies for plasma exosome miRNA analysis did not improve results.
Mikamori *et al.*(428)	2017	PDAC (n=23)	Plasma	**miR-155**	Up-regulation correlated with reduced DFS, but not OS and could be used as a clinical marker in gemcitabine resistance.
Chen *et al.* (429)	2017	PC (n=16) CP (n=18) Controls (n=20)	Serum	**miR-23b-3p**	Was verified to be the only up-regulated miRNA in both PDAC and CP groups, as compared to normal controls.
Lai *et al.* (275)	2017	PDAC (n=29) CP (n=11)	Plasma	**miR-10b, miR-20a, miR-21, miR-30c, miR-106b, and miR-181a**	Appeared to have AUCs approaching 1.0. Post-resection, the elevated levels of exosomal miR-10b, -20a, -21, -30c, -106b decreased to normal values.

Madhavan *et al.* (430)	2015	PDAC (n=131) CP (n=25) Benign disease (n=34) Control (n=30)	Serum	**miR-1246, miR-4644, miR-3976,** and **miR-4306**	For miRNA markers the sensitivity was 0.81 (95% CI: 0.71–0.89) and specificity was 0.94 (95% CI: 0.84–0.99).
Kanwar *et al.*	2014	PDAC (n=5) Controls (n=5)	Serum	**hsa-miR-130a, hsa-miR-29b, hsa-miR-30b, hsa-miR-518d, hsa-miR-551b** and **hsa-miR-646**	Utilised an "ExoChip", a microfluidic device fabricated in polydimethylsiloxane (PDMS) and functionalized with antibodies against CD63.
Que *et al.* (327)	2013	PDAC (n=22) Benign (n=6) Ampullary (n=7) CP (n=6) Controls (n=8)	Serum	**miR-17-5p** and **miR-21**	AUC of miR-17-5p and miR-21 were 0.887 (0.796 to 0.978) and 0.897 (0.803 to 0.991), respectively.

cancers

IMPACT FACTOR 6.575

Indexed in: PubMed

Review

Size-Exclusion Chromatography as a Technique for the Investigation of Novel Extracellular Vesicles in Cancer

Daniel S. K. Liu, Flora M. Upton, Eleanor Rees, Christopher Limb, Long R. Jiao, Jonathan Krell and Adam E. Frampton

Special Issue

Molecular Biomarkers and Precision Medicine for Pancreatic, Liver and Biliary Tract Cancers

Edited by

Dr. Adam E. Frampton

MDPI

https://doi.org/10.3390/cancers12113156

cancers

MDPI

Review

Size-Exclusion Chromatography as a Technique for the Investigation of Novel Extracellular Vesicles in Cancer

Daniel S. K. Liu [1,2], Flora M. Upton [1], Eleanor Rees [1], Christopher Limb [3], Long R. Jiao [2], Jonathan Krell [1] and Adam E. Frampton [1,3,4,*]

1 Department of Surgery & Cancer, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London W12 0HS, UK; daniel.liu08@imperial.ac.uk (D.S.K.L.); f.upton19@imperial.ac.uk (F.M.U.); e.rees@imperial.ac.uk (E.R.); j.krell@imperial.ac.uk (J.K.)
2 Hepato-Pancreato-Biliary (HPB) Surgical Unit, Department of Surgery & Cancer, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London W12 0HS, UK; l.jiao@imperial.ac.uk
3 Hepato-Pancreato-Biliary (HPB) Surgical Unit, Royal Surrey County Hospital NHS Foundation Trust, Guildford, Surrey GU2 7XX, UK; c.limb@surrey.ac.uk
4 Department of Clinical and Experimental Medicine, Faculty of Health and Medical Sciences, The Leggett Building, University of Surrey, Guildford, Surrey GU2 7WG, UK
* Correspondence: a.frampton@imperial.ac.uk; Tel.: +44-207-594-2125

Received: 10 September 2020; Accepted: 15 October 2020; Published: 27 October 2020



Simple Summary: Extracellular vesicles (EVs) are small particles that are released by cancer cells, and they may hold vital information for researchers looking for early markers for diagnosis. Size-exclusion chromatography (SEC) is a classical technique that has become increasingly popular and can be used for rapid isolation and investigation of both their cargo and functionality. This systematic review highlights its main technical aspects, the type of materials involved and by covering the findings of the identified papers hopes to demonstrate the utility of this method in cancer research to date.

Abstract: Cancer cells release extracellular vesicles, which are a rich target for biomarker discovery and provide a promising mechanism for liquid biopsy. Size-exclusion chromatography (SEC) is an increasingly popular technique, which has been rediscovered for the purposes of extracellular vesicle (EV) isolation and purification from diverse biofluids. A systematic review was undertaken to identify all papers that described size exclusion as their primary EV isolation method in cancer research. In all, 37 papers were identified and discussed, which showcases the breadth of applications in which EVs can be utilised, from proteomics, to RNA, and through to functionality. A range of different methods are highlighted, with Sepharose-based techniques predominating. EVs isolated using SEC are able to identify cancer cells, highlight active pathways in tumourigenesis, clinically distinguish cohorts, and remain functionally active for further experiments.

Keywords: size-exclusion chromatography; SEC; exosomes; extracellular vesicles; cancer; biomarkers

1. Introduction

Extracellular vesicles (EVs) are small particles defined by a lipid bilayer and an inability to replicate, and they lack organelles or a functional nucleus. Historically, they have been termed various names, such as "shedding membrane vesicles", "membrane fragments" [1], "microvesicles" [2], or more recently "exosomes". We now make a broad distinction between subsets of EVs by their biogenesis, size, or density, and this is dependent on the experimental isolation method used.

Whilst larger structures (200–1000 nm) tend to be plasma membrane-derived apoptotic bodies and microvesicles [3], smaller structures (30–150 nm), which are of endocytic origin (released into the extracellular environment from fusion of multivesicular bodies (MVB) with the plasma membrane), are termed exosomes [4]. However, there remains a considerable overlap in what is isolated for experimental characterisation, and studies investigating vesicular biogenesis show that both large and small vesicles can demonstrate direct budding from the plasma membrane [5]. Exosomes are the dominant area of research, having been shown to act as a distinct intercellular signalling mechanism, shuttling messenger RNAs (mRNAs) and microRNAs (miRNAs) with the potential to regulate functional expression in other cells [6]. When there is no clear demonstration of endosomal origin, the International Society for Extracellular Vesicles (ISEV) [7,8] suggests use of the term EVs to improve clarity.

The release of all types of EVs by tumour cells is of particular interest to researchers investigating potential biomarkers, as these may be a source of several targets in the growing field of "liquid biopsy" for cancers that are difficult to access or diagnose. This idea involves the investigation of circulating tumour cells, free nucleic acids, and proteins, which are detectable in blood and biofluids, including from samples as diverse as breast milk [9], saliva [10], ascites [11], and bile [12].

1.1. Isolation Methods in Brief

There are several methods for the isolation and enrichment of EVs from both cell-conditioned media (CCM) and biofluids. Historically, differential ultracentrifugation (dUC) has been popular, as it uses commonly available laboratory facilities, with low costs and simple protocols to obtain pelleted EVs. In 2016, a survey showed that 81% of all experiments registered used dUC as their primary EV isolation method [13]. First, low speeds are used to remove cellular debris and subcellular organelles, then protocols using speeds of approximately 100,000 × g for 1–2 h are utilised to retrieve EV pellets. Large volumes of cell-conditioned medium can be processed, and the use of a washing step and repeat spin, as well as filtration, can further enhance the purity of the EVs and remove any contaminating soluble proteins.

This has given way to several other techniques, designed to exploit the different characteristics of EVs, including by size, density, and known surface markers. By 2017, dUC accounted for 45% of all experiments with a reported 190 unique isolation methods registered on EV-TRACK, an online tool designed to improve standardisation across the field [14]. The choice of method is fundamentally based on a trade-off between EV yield and purity from contaminants such as albumin that may be highly represented in samples [15]. Differential ultracentrifugation separates based on a sedimentation coefficient, which is in turn determined by the mass of the particle and the distance travelled. Therefore, small changes in centrifugation protocols are likely to lead to a wide variation in pelleted material [16].

Exosomes have been found to float at specific densities ranging from 1.05 to 1.25 g/mL on continuous sucrose gradients [17], and this has been used as a method for separation of a highly pure EV fraction. This results in a relatively low yield and remains labour intensive as well as requiring eventual separation from the gradient buffer made of iodixanol or sucrose.

As well as separation by density and/or size, immunoaffinity using biotin-labelled antibody capture beads [18] can be undertaken as well as polyethylene glycol (PEG) precipitation as commonly used in many commercial kits (ExoQuick by System Biosciences, Palo Alto, CA, USA, is one example). Whilst the former is highly selective and, therefore, has a low overall yield, the latter is both rapid and technically easy but co-isolates non-vesicular contaminants to a greater degree [19,20].

Size-exclusion chromatography (SEC) was shown in 2016 to account for 5.6% of experiments but has become more popular as interest in EV isolation methods with diagnostic potential has accelerated in the past four years. In this regard, the advantage of a single-step technique able to isolate from small sample sizes may be of great interest.

1.2. Size-Exclusion Chromatography, an Old but New Technique

Size-exclusion chromatography (SEC) was initially designed as a technique to fractionate amino acids and protein fragments by molecular weight and first utilised starch in a process termed "partition chromatography" [21].

The process of chromatography involves the use of a stationary phase (in this case a gel; see Figure 1) held in a column, which allows a liquid mobile phase to pass through. This contains the analyte in aqueous buffer solution, which leaves the column at a rate proportional to its size, or more accurately the hydrodynamic volume [19]. With complex mixtures made up of different-sized particles, larger molecules are excluded from the gel and are recovered quickly, whilst smaller particles are impeded and elute much later.

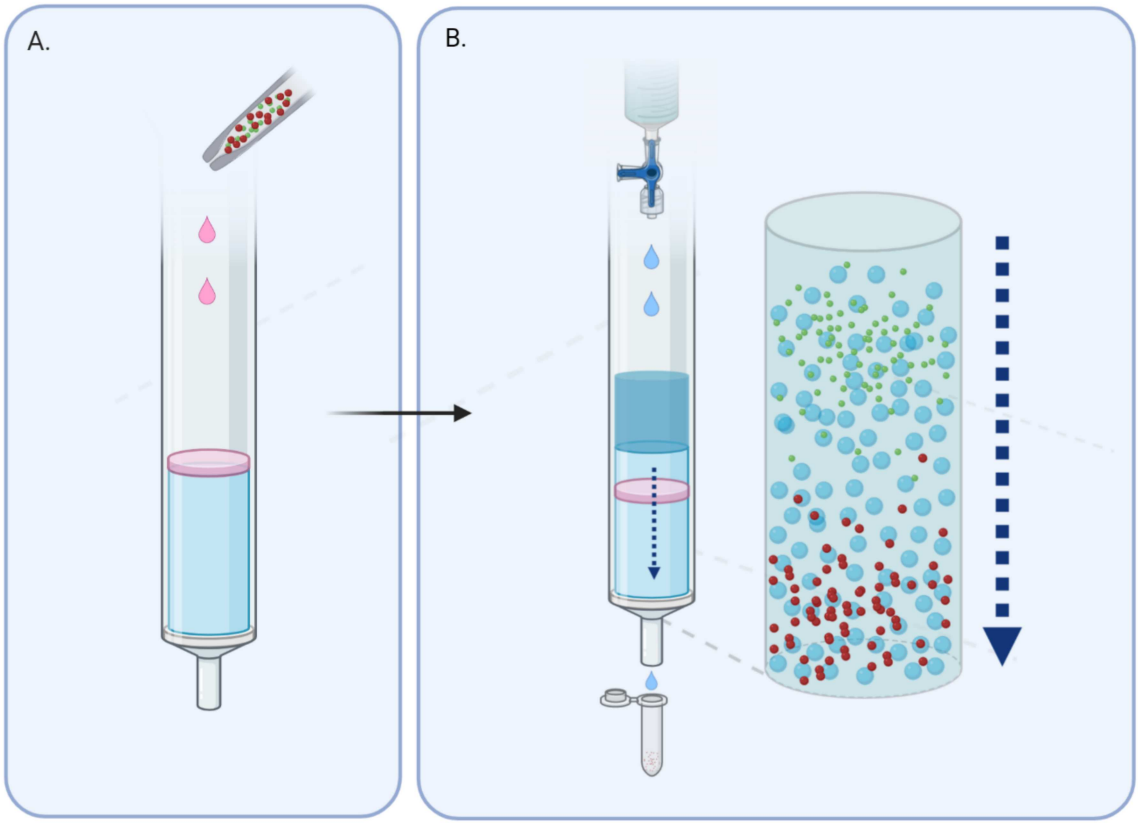


Figure 1. Larger particles take a more direct path through the gel and elute first. (**A**) Sample is layered onto column prepared with gel matrix. (**B**) With continuous addition of wash buffer, separation of the sample according to size occurs along the column.

The initial experiments by researchers working at Queen Charlotte's Hospital in London used potato and maize starch and were noted to be inconsistent in size and behaviour leading to different elution volumes in each column [22]. This was improved by the subsequent development of dextran and crosslinked agarose, which is a polysaccharide polymer material extracted from seaweed. Today, gel matrices are known by tradenames and include dextran polymers (Sephadex (Cytiva, Marlborough, MA, USA)), agarose (Sepharose (Cytiva, Marlborough, MA, USA)), or polyacrylamide (Sephacryl (Cytiva, Marlborough, MA, USA) or BioGel P (Bio-Rad, Hercules, CA, USA)) manufactured as beads and optimised for experimental consistency.

The transit time of an analyte must be experimentally determined and is roughly based on the size of the pores within the gel matrix. Additionally, different gels have an optimal molecular weight range, which will allow good linear separation, and it is recommended that this is usually calibrated with a standard curve based on known weights. Other factors that will affect the process are the

dimensions of the column, usage of column frits and consistency of the process of forming the gel within the column, which can also be minimised by use of industrially standardised columns on the market. Larger bed volumes can lead to an improved resolution of larger particles, but this must be balanced by a longer time taken to elute each fraction [23].

Literature on SEC as a method for EV isolation has become more prevalent, with this current review identifying more papers year-on-year. The benefits of SEC are that it is rapid, taking 10 to 20 min per sample; relatively inexpensive; and able to efficiently remove over 95% of blood protein, which makes it extremely clinically applicable when compared with dUC or other techniques [24].

The difficulties inherent in the process include preparatory steps to remove large contaminants, the possible co-elution of similar sized lipoproteins, and any concentration steps that may be required either before or after. Lipoproteins such as very-low-density lipoprotein (VLDL) have a similar size to EVs and are a recognised co-precipitant from blood plasma or serum samples, which should be identified by the presence of apolipoproteins and a lipid monolayer [25]. This is an important factor to consider, as lipoproteins are carriers of potentially functional miRNAs and, so, could complicate potential biomarker discovery [26]. Combining SEC with other methods, such as immunoprecipitation is possible and has been demonstrated in several papers within this review. Further work may also be needed to determine the interaction between EVs and gels commonly in use. EVs have a large surface area to volume ratio for interaction, and this is a potential area of future research [27].

Uses of SEC within the EV field also include separation of vesicles from free dye such as removing PKH or CFSE (CarboxyFluorescein diacetate Succinimidyl Ester) from treated EVs prior to subsequent analysis [28], as well as separating EVs from iodixanol used in gradient density methods [29]. This use of orthogonal approaches, i.e., both size and density purification in EV isolation and purification, is increasing as protocols are shared and familiarity with the different methods improves.

2. Methodology of Review

2.1. Objectives

The overall aim of the review was to explore the use of size-exclusion chromatography (SEC) (also known as gel filtration) systematically in the context of EV isolation and its use in the cancer biomarker discovery field.

2.2. Search Methods

Databases searched included MEDLINE (https://www.nlm.nih.gov/bsd/medline.html), EMBASE (https://www.embase.com/), and Journals@Ovid using Ovid SP (https://www.ovid.com/; accessed last on 20 April 2020; see Appendix A for search terms). Additional papers were identified by iteratively searching through reference lists and previous work by the same authors. Finally, a search of EV-TRACK for all experiments using "size-exclusion chromatography" as a separation protocol was undertaken, and the relevant papers were included.

2.3. Selection Criteria

The focus of this review is on original studies that used SEC, or gel filtration, as a primary method of EV isolation in either a cancer cell line or clinical sample from a cancer cohort. After excluding duplicates, abstracts were reviewed and we excluded any descriptive protocol articles, conference abstracts, or non-original research. Forty full-text articles were reviewed, and those found to use SEC as an adjunct method (e.g., removal of dye from pre-isolated EVs) were also excluded (see Figure S1).

3. Results

3.1. Comparison of Isolation Methods

We looked at 37 studies, of which there were 39 different isolation methods described (shown in Table 1).

Table 1. Separation methods in brief.

Method	Frequency
Sepharose (either 2B or 4B)	13
MiniSEC (using Sepharose 2B)	10
qEV column (proprietary filtration matrix)	10
Sephacryl (S-400 or 500 HR)	2
Microfluidic Platforms	2
EVSecond (modified styrene-divinylbenzene matrix)	1
Bio-Gel A (Agarose)	1

MiniSEC, Method utilised by the Whiteside Lab (Pittsburgh, PA, USA); qEV, Proprietary columns trademarked by Izon Science (Christchurch, New Zealand); HR, High Resolution.

Sephacryl and Sepharose were two of the most common polymers identified in the papers examined. Sepharose appeared to be the most popular, with 2B and 4B referring to the percentage of crosslinked agarose (2–4%). MiniSEC, the method that was coined and used solely by the Whiteside lab group, uses small-size ion exchange columns and has resulted in a significant number of publications.

Proprietary methods include the qEV column introduced in 2015 by Izon Science (Christchurch, New Zealand), and the EVSecond (GL Science, Tokyo, Japan). One of the theoretical benefits of industrially manufactured columns is reduced operator dependence in the making of Sepharose columns, leading to better uniformity. The qEV columns come with ISO 13485:2016 compliance, which requires that the certified organisation demonstrates a quality management system to maintain certain standards and systems during the manufacturing and delivery process. Although this does not guarantee that EVs are isolated, it is important that elution patterns are characterised for all types of columns prior to use.

3.2. EV Characterisation, Concordance with MISEV Guidelines

Minimal Information for Studies of Extracellular Vesicles (MISEV2018) [8] was an updated version of guidelines developed in 2014 [7] to sensitise researchers to the experimental and reporting requirements in the EV field. By standardising nomenclature and experimental methods, it is hoped that this will improve the reliability and reproducibility of studies.

EV-TRACK is an open-source knowledgebase which was endorsed by MISEV2018 and uses a key feature, termed an EV-METRIC (taken from http://www.evtrack.org/about.php) to highlight nine specific areas that can be used to support critical appraisal of an EV study (Table 2) [30].

Table 2. The EV-METRIC.

Experimental Parameters Related to EV Separation	
Density Gradient	Type and concentrations of the gradient performed, use as validation of EVs in experiments.
EV Density	EV density should be reported if undertaken.
Ultracentrifugation Specifics	When ultracentrifugation is used, there should be a description of the *g*-force, duration and rotor type of these steps.

Table 2. *Cont.*

Protein Analysis	
EV-Enriched Proteins	3 or more EV-enriched proteins.
Non EV-Enriched Protein	Assessment of one or more EV-enriched protein.
Antibody Specifics	Antibodies used for immunoblotting should be reported to include clone/reference number and dilutions.
Lysate Preparation	Lysis of EVs should involve reporting of buffer composition.
Particle Analysis	
Qualitative and Quantitative Analysis	Implementation of both qualitative (e.g., electron microscopy) and quantitative methods (e.g., nanoparticle tracking analysis, high-resolution flow cytometry). Quantitative methods should include reporting of particle concentration.
Electron Microscopy Images	Inclusion of both wide-field and a close-up electron microscopy image.

EV, extracellular vesicle. Bold is used to mark the different table headings.

Tables S1 and S2 show the isolation methods identified and the characterisation used, which is a vital part of calculating the EV-METRIC. Whilst centrifugation specifics have been noted, density experiments were only undertaken to assess SEC-isolated EVs in four studies [31–33]. We have therefore focused our assessment of the quality of studies using the EV-METRIC, specifically the analysis of proteins and particle characteristics (i.e., size, distribution, concentration, morphology), in the hope that this will stimulate discussion of the differences between the studies.

Protein analysis whether by Western blot, flow cytometry or mass spectroscopy for EV-enriched proteins is a feature of nearly every study reviewed (84% of studies), and Western blots for the main tetraspanins, such as CD63, CD9, and CD81 have become a surrogate for the demonstration of EVs, alongside electron microscopy (EM) and nanoparticle tracking analysis (NTA). Whilst these have been established for some time [34], it is worth bearing in mind that they can be enriched in vesicles that are formed in both an endosomal and non-endosomal fashion (i.e., not just exosomal) [35]. Alix (assessed in 11% of studies), and tumour susceptibility gene 101 (TSG101; noted in 41% of studies), which both relate to initial intraluminal vesicle (ILV) formation in the endosomal sorting complex required for transport (ESCRT) machinery, are used as common markers of EVs. There is also significant heterogeneity when it comes to clinical samples, and it is not expected that all markers are constitutively expressed in EVs. Furthermore, only three studies investigated non-EV markers, which may include lipoproteins and serum-derived materials, or proteins derived from subcellular structures not present in EVs. Examples of this might be proteins derived from the endoplasmic reticulum (such as grp94 or calnexin) or mitochondrial proteins such as cytochrome C.

3.3. EV-Associated Cargo

EVs have been found to contain a diverse array of cargo, which includes lipids, soluble and transmembrane proteins, RNAs, and DNA. Additionally, some studies have focussed on the function of EVs and their effect on recipient cells. In order to analyse the findings, we have divided the studies by the general -omics approach they have used (see Table 3). This split 37 studies into those looking at functionality of EVs; proteomics; feasibility studies; and a small number investigating the RNA content of isolated EVs.

Table 3. Omics approaches of studies reviewed.

Study Type	Number
Proteomics	13
Biological Function	9
Feasibility of Method	11
RNA	4

3.3.1. EV Proteomics

EVs are associated with pro-apoptotic protein markers and stimulate an immune response that may promote the tumour microenvironment

The study of protein cargo from EVs has been a key area of research, with studies investigating the contents of "shedding membrane vesicles" obtained by gel filtration dating back to 2003 [36]. Initial studies looked into Fas ligand (FasL), and its immunosuppressive effects on T-cells, an area of significant interest in the 1990s, as EVs had been shown to demonstrate antigen-presenting ability to T-cells [37]. It had been known for some time that cancer cells produced a large number of EVs, and that these were detectable in patients with cancer [38]. However, it was unclear whether they were the cause or effect of increases in synthetic, metabolic, or catabolic pathways [39].

Taylor et al. [36] is the earliest study covered by our review and showed that the Fas/FasL pathway plays a dual role in vivo, mediating proinflammatory effects, as well as immune cell apoptosis. The 42 kDa cell-membrane-associated protein was detected in these vesicles and led to the apoptosis of activated normal T lymphocytes ex vivo. These data were supported in both oral and ovarian cancer cohorts [2].

Since those early investigations, EVs have been found to play a much wider role to include cell signalling and metabolism. This includes studies such as Hong et al. [40] and Kawakami et al. [31], which focussed on the presence of proteins such as TGF-β1 (Transforming growth factor beta 1) and GGT (Gamma-glutamyltransferase). TGF-β1 can be carried by exosomes and be found to be raised in patients with acute myeloid leukemia (AML) but normalised after long-term complete remission. This was better correlated with clinical disease status than broad measurements of total exosomal protein levels. However, they were unable to determine a subset of EVs, which were responsible for TGF-β1 carriage.

GGT, the activity of which is upregulated in colon, ovary, and liver cancer; astrocytic glioma; soft tissue sarcoma; melanoma; and leukaemia, was drawn from a selection of highly upregulated proteins identified from an androgen-resistant cell line and found to be a serum biomarker raised in exosomes that could distinguish patients with prostate cancer from those with benign prostatic hypertrophy (BPH). However, they used multiple methodologies for their isolation, alternating between dUC and SEC, as well as density gradient and immunocapture. Although EVs are obtainable by all these methods, in multiple comparative studies, these appear to isolate EV populations with different size distributions and protein content, which suggests that the use of a consistent isolation methodology is best practice [41,42].

EV research has begun to talk less of shedding (i.e., a passive excretory mechanism for EVs), and instead has focussed on EVs as perhaps being actively released as part of cancer progression. One study that looked into this was reported by Djusberg et al. [43] and showed that overexpression of YIPF6 (a protein co-expressed with androgen receptors) in 22R v1 prostate cancer cells resulted in enhanced secretion of EVs.

The report by Broggi et al. [44] used K14-VEGFR-3-Ig transgenic mice lacking dermal lymphatics, to show that lymphatics were a major route of EV trafficking from the interstitial space into the lymph fluid, lymph nodes (LNs), and systemic circulation, and that this could be a source of tumour-derived EVs.

However, this study suffered from methodological flaws, using dUC to investigate miRNAs, and switching to an in vivo study using SEC. It is possible that although their clinical study detected metastasis-related miRNAs specifically expressed in the EVs of metastatic melanoma patients, they were unable to detect a statistical difference between cohorts since they investigated EVs with lower purity. Another criticism is the use of ultracentrifugation supernatant as a "non-EV" comparator. Indeed, it cannot be assumed that supernatant generated after UC is a suitable non-EV sample for comparison, as it has been shown that single spins of 1–2 h may only remove 40–50% of EVs [45].

EV Protein Markers Are Differentially Expressed by Cancer Cells

SEC can be used to effectively isolate EVs from conditioned cell culture medium, and this has led to the development of several specific EV-associated markers that could be implicated in tumourigenesis.

Theodoraki et al. [46–48] have written three papers looking at patients with head and neck cancer (HNSCC) using a process termed "immune affinity capture" to magnetically label specific subsets of EVs isolated by SEC with antibodies. This technique enables downstream analysis of EVs that express specific protein markers and, therefore, may be attributed to T-cells or otherwise. T-cell specific EVs were isolated based on CD3+ expression and found to constitute between 38% and 54% of total exosomes isolated from plasma samples, a higher proportion than that found in a healthy cohort. Conversely, CD3− EVs (potentially tumour-derived) were found to contain high levels of CD44v 3 protein and to be associated with unfavourable clinicopathological parameters.

They also focussed on the PD-1/PD-L1 (Programmed death-ligand 1) immunosuppression pathway, with high PD-L1 expression in tumours being associated with poor prognosis. They were able to show expression of PD-L1 on EVs was associated with more disease activity and UICC (Union for International Cancer Control) staging, as well as activity on normal T-cells in vitro.

Interestingly, they measured both surface PD-L1 cargo in exosomes and PD-L1 levels in patient plasma but did not find a correlation. This may indicate that exosomal PD-L1 expression is an independent biomarker distinct from soluble protein levels, with proven activity on peripheral cells. Unfortunately, tissue correlation was never investigated. It would have also been especially interesting to see whether this activity was modulated by PD-1 inhibitors, such as pembrolizumab, although these experiments have subsequently been shown with EVs isolated via UC in melanoma cells [49], and appear to be a promising area for future research [50].

Stress-induced pathways, such as p53 activation (non-mutant), lead to increased EV secretion via *TSAP6* and *CHMP4C* gene transcription [51,52]. Differential EV expression in cell injury caused by ionising radiation (IR) may therefore be a good model to investigate EVs as biomarkers during cancer treatment. Abramowicz et al. [53] showed that in a p53-wildtype Human Papillomavirus (HPV)- associated head and neck cancer cell line, EVs released by irradiated cells had increased levels of proteins, such as CD63 and CD81, whereas other surface protein levels remained unchanged (i.e., CD9 and TSG101). Gene ontology analyses showed associations between upregulated proteins and DNA repair, regulation of reactive oxygen species metabolism, and de novo protein folding. Unusually, histones and chromatin-associated factors were also identified, which may be related to the increased loading of exosomal DNA during cellular senescence [54].

Freitas et al. [55] demonstrated that altering the glycosylation profile of gastric adenocarcinoma cells, MKN45, also leads to a change in EV glycosylation. Comparing well-known isolation methods, they found that only 16.3% of proteins were common to all EV isolation methods, with SEC and Optiprep Density Gradients (ODG) resulting in EVs with the highest concentration of EV markers. Interestingly, the EVs derived were particularly enriched for specific glycoproteins carrying the Sialyl-Tn antigen (STn), compared with cell lysate. Synthesis of STn is associated with poor prognosis in gastric cancer and is a widely expressed epithelial cancer marker [56].

Furthermore, SEC can also be used to distinguish the proteome of EVs from cell lines representing HPV+ and HPV− head and neck cancer. Ludwig et al. [57] showed differential protein expression, particularly with upregulation of CD47 in HPV+ HNSCC cells, which inhibited exosome uptake/phagocytosis by human monocytes, and may play a role in the more robust, virus-primed anti-tumour immunity present in HPV+ malignancies [58].

EV Expression Highlights Differences in Clinical Outcome and Surgical Resectability

The recent paper by Theodoraki et al. [48] demonstrated how SEC can be a highly sensitive method to accurately determine EV protein expression. In this study, they developed a custom antibody microarray to further characterise the EV surface antigen expression in blood plasma from a cohort of HNSCC patients in a longitudinal manner, before and after treatment with cetuximab, radiotherapy,

and ipilimumab. They then followed these patients up for 24 months to assess EVs as a predictive marker for disease recurrence.

This microarray consisted of classical exosome surface markers such as CD9, CD63, CD81, and tumour-enriched exosome "TEX" markers, which included Epidermal Growth Factor Receptor 1 (EGFR1), Melanoma-associated antigen 3 (MAGEA3), Epithelial cell adhesion molecule (EpCAM), and Chondroitin sulfate proteoglycan 4 (CSPG4). TEX markers were not detected in their small cohort of healthy controls (n = 5), whilst in the cancer patients, the TEX/Total exosome ratio, as well as total exosomal protein levels, were significantly different in patients who had recurrence after treatment. In assessing subsets of EVs, they returned to the immunoaffinity capture method to look at CD3, PD-L1, and CTLA4 (cytotoxic T-lymphocyte-associated protein 4) expression on T-cell derived EVs, and highlighted that high levels of CD3-/PD-L1+ and CD3-/CTLA4+ exosomes at baseline might predict patients who would benefit from immunotherapy.

EVs from SEC can also correlate with a patient's tumour grade and the extent of surgical removal, in patients presenting with brain glioma [59]. Liquid chromatography mass spectroscopy (LC-MS) identified 12 plasma EV-associated proteins that were significantly different between patients with glioblastoma and low-grade glioma. The authors were especially interested in Syndecan-1, a membrane protein, which is now known to interact with the well-known EV markers Syntenin-ALIX, which facilitate membrane budding [60]. Syndecan-1 had an exceptionally high AUC (Area Under the Curve on the receiver operating characteristic) value of 0.81 and a sensitivity and specificity of 71% and 91%, respectively, for distinguishing these high- and low-grade tumours.

However, in both these studies, they did not compare EV concentration between the cohorts, which could account for the increase in protein expression between disease groups. Normalisation by either EV number, or an established EV marker, would improve the detection of differentially expressed EV-associated proteins.

Nevertheless, it appears that there are EV-associated proteins increased in cancer, and these are readily detectable in small cohort clinical trials using SEC isolation. The majority of proteins involved in transformation to cancer are usually downregulated [61], so focussing on a subset that cause EV release may be promising. With the dynamic range of blood plasma protein concentrations varying by >10 orders of magnitude [62], these studies show overall that SEC is an efficient technique for the removal of a high abundance protein, thus allowing for greater sensitivity and detection of specific proteins secreted by cancer cells.

3.3.2. EV RNA

Clinical Studies Show That EVs from SEC Are Enriched with Functionally Active Small RNAs

A small number of studies (n = 4; see Table S2) have investigated RNA expression within EVs after using SEC as an isolation method. Two were clinical studies, whilst the other two were in vitro studies [63,64].

The first study by Rabinowits et al. [65], (published two years after the paper by Valadi et al. highlighting the role of EVs in transferring functional mRNAs and miRNAs [6]) used EpCAM immunoaffinity capture to purify for EVs present in the plasma of patients with lung cancer. EpCAM has been shown to be ubiquitously expressed on epithelial tumours [66] and on EVs from these cancers [67,68]. A microarray containing probes for 467 human mature miRNAs was used to rapidly identify potential miRNA candidates. As has been demonstrated in other papers [69], EV concentration was higher in the cancer cohort, with a higher RNA concentration. Interestingly total exosome and total miRNA levels were also significantly higher in their cancer cohort, but the paper did not report on differentially expressed miRNA candidates.

Van Eijndhoven et al. [33] investigated EVs obtained from patients with classical Hodgkin's lymphoma (cHL), comparing these patients with healthy controls and patients with Systemic Lupus Erythematosus. This study had several advantages. Firstly, the authors used RNA sequencing to

compare EVs from the clinical cohorts, and secondly, utilising SEC they were able to compare EV fractions with RNA extracted from the protein/High-Density Lipoprotein (HDL) fraction generated from the same samples. This protein/HDL fraction contained a significant number of miRNA reads but at a relatively low complexity (<120 unique species), compared with the comprehensive miRNA repertoires of cHL patients obtained from EVs, with miR-486-5p and miR-92a-3p being highly abundant. They were then able to identify a panel of five cHL-related miRNAs (i.e., miR-127-3p, miR-155-5p, miR-21-5p, let-7a-5p). Next, using RT-qPCR validation of miR-127-3p and miR-155-5p, they were able to show that miRNA fold differences were more robust in EVs, compared with those seen in total plasma. Additionally, in a small cohort (n = 7) that was followed longitudinally, levels of miR-127-3p, miR-155-5p, and miR-21-5p in EVs decreased markedly after treatment, whilst levels of these miRNAs in total plasma did not.

The major limitation of this study was the lack of assessment for EV-associated protein markers, which should be included if we want to conclude that this is truly an EV study as per ISEV guidelines. Instead, PKH67-labelled (a type of lipophilic dye) flow cytometry was used to show evidence of EVs. This is generally not recommended as non-EV species, such as lipoproteins, are not distinguished separately [70,71].

EVs from Cell Lines Can also Be a Useful Source of RNA-Based Biomarkers

Cell lines are a useful model for EV studies, because a large number of cells can be cultured for the purposes of EV isolation. This raises an issue with larger volumes of cell culture supernatant, which can be overcome using a concentration step, usually with centrifugation, in order to then load SEC columns with a manageable volume. Furthermore, if EV-depleted media is used [72], and there is careful consideration of culture supplements [73], then the expression of EV related molecules can be directly linked to the cell type under investigation.

Sakha et al. [63] used SEC to investigate a previously established highly metastatic human oral squamous cell carcinoma (SCC) cell line (HOC313-LM) and found it to be superior to dUC, or the proprietary kit by Thermo Fisher (Waltham, MA, USA). Subsequently, they were able to compare this cell line with its less metastatic parental cells (HOC313-P). Microarray analysis identified a greater level of miRNA expression in the exosomal cargo, and 7 oncogenic miRNA candidates (miR-17, miR-30a-3p, miR-30a-5p, miR-92a, miR-181a, miR-342-3p, and miR-1246) were differentially expressed in both cellular and exosomal compartments. Next, they showed that miR-1246 did not affect cell growth but instead enhanced migration and invasion in HOC313-P, TSU, and HeLa cells, possibly through reducing levels of the tumour suppressive protein DENND2D. However, it should be commented that they used small nucleolar RNA U6B as an endogenous control for normalising their experiments, which can be variably expressed in cancer [74], and levels in EVs may not be stable between healthy and cancer cohorts [75]. Current guidelines suggest the use of multiple controls for normalisation, with RNAs demonstrated as suitable with preliminary studies, or alternatively using a mean expression value normalisation strategy [76].

Peacock et al. [64] looked at oropharyngeal SCC and HPV and its association with EV miRNA expression. EV RNA was sequenced, and validated by RT-qPCR, with highly expressed miRNAs further investigated for their potential functions using miRTarBase. RNA sequencing identified a fairly low number of RNAs (2–15%) aligned to miRNA sequences, with the rest found to be mRNA, rRNA, tRNA, snoRNA, snRNA, and lincRNA. Fourteen miRNAs were enriched in HPV+ cell-derived EVs, whereas 19 miRNAs were enriched in HPV− EVs. As HPV+ oral cancer has a better prognosis than HPV− disease but is rising in incidence, these candidates would benefit from further functional investigation as this may allow the development of potential theranostics [77].

3.3.3. EV Functionality

Functional Studies Demonstrate Tumour-Derived EV Mechanisms Involved in Immunomodulation

SEC-isolated EVs have been used to demonstrate unusual properties of EVs. Indeed, ex vivo studies by the Whiteside Lab have shown that tumour-derived EVs were immunosuppressive, and induced apoptosis of activated primary CD8+ cells, whilst T-cell- or dendritic-cell-derived EVs were immunostimulatory and promoted ex vivo proliferation of resting T-cells [78].

The purity of samples prepared in this way is important, as incubation experiments may give confounding results if the protein of interest is co-isolated as a contaminant. Frozen plasma does lead to electron-microscopy evidence of vesicular damage, and string-like aggregates, which are not cost-effectively removed by exogenous enzymatic treatments [32].

Tumour-derived EVs have been strongly implicated in the maintenance of the tumour microenvironment and can be shown to act in several ways to suppress T-cells. One study investigated immunosuppression via adenosine production, which is found at high concentrations in cancer tissues and known to be a crucial mediator in the alteration of immune cell functions in cancer [79]. The study by Schuler et al. [80] was able to demonstrate that EVs can deliver enzymatically active membrane-tethered CD73 to CD4+, CD39+, and regulatory T cells (Treg) lacking this enzyme. This co-operative interaction is a novel mechanism that could explain immunosuppression in the tumour microenvironment.

Hong et al. [69] showed that AML exosomes co-incubated with human natural killer (NK) cells inhibit expression of NKG2D, a major recognition receptor involved in tumour cell recognition and HNSCC exosomes' suppressed activation and proliferation of activated T lymphocytes by downregulating expression levels of CD69 in co-culture. This was further expanded by Ludwig et al. [81], who used densitometry analysis of Western blots for an array of apoptosis-associated protein markers to show that the cargos of exosomes from patients with active cancer were enriched in proteins such as COX-2 (cyclooxygenase-2 also known as Prostaglandin-endoperoxide synthase 2), TGFβ-LAP (TGF-β Latency Associated Peptide), PD-1, CTLA-4, and TRAIL (Tumour necrosis Factor-related apoptosis-inducing ligand). They highlighted from this a potential strategy in the development of plasma exosomes, as future cancer biomarkers could be based on functional evaluations of total exosome fractions ex vivo (i.e., using flow-based apoptosis assays after co-incubation of ex vivo EVs with normal immune cell subsets, such as CD8+ Jurkat cells).

The emerging evidence for the role of EVs in the tumour microenvironment includes effects on angiogenesis in endothelial cells. Ludwig et al. [82] used cell-line-derived EVs to stimulate human umbilical vein endothelial cells, leading to an increased formation of vascular structures both in vitro and in vivo. Longitudinal analysis of plasma EVs showed distinct alterations in their ability to alter epithelial-to-mesenchymal transition (EMT) in recipient cells, a position that was supported by the work of Theodoraki et al. who investigated time dependent changes in plasma EV functionality after photodynamic therapy in patients with HNSCC [83]. By comparing functional activity before and after treatment, these papers suggest that this was directly related to the tumour phenotype. However, it remains to be seen whether this functional activity arises from EMT-related miRNAs or proteins.

Other mechanisms of immunomodulation proposed by EVs include the vesicular protein Arginase-1 (ARG1), which regulates the availability of L-arginine, and has been strongly associated as an immunomodulator in various settings [84,85]. This was investigated by Czystowska et al. [11] using SEC and immunoaffinity to enrich for EPCAM positive EVs in patients with ovarian cancer malignant ascites, using these EVs to suppress the proliferation of CD4+ and CD8+ T-cells, compared with EVs obtained from ovarian cyst fluid.

Sjöqvist et al. [86] took an unusual approach to investigating EV function, by isolating EVs from two types of healthy cells (keratinocytes and fibroblasts) and compared their anti-proliferative effects on cancer cells, in particular human SCC. This suppression was steroid resistant and was, therefore, identified as a rationale for cell sheet therapy (i.e., autologous transplantation of epithelial cells from the oral mucosa or otherwise) in oesophageal stricture prevention.

4. Conclusions

In summary, SEC can isolate EVs with a purity (as measured by levels of protein co-isolated) that is superior to dUC [32] as well as precipitation methods [20]. It can be used both for cell-conditioned medium and for clinical samples, thereby allowing the investigation of protein and RNA expression, as well as functional activity. Characterisation of EVs by this method, according to MISEV guidelines [8], demonstrates that they contain EV-related surface protein markers and structural morphology, making them suitable as a source for further research.

The high purity of SEC-isolated EVs leads to a clearer "signal" than is found for many other isolation methods, particularly with clinical samples. This allows a greater detection of potential cancer biomarkers particularly in view of the increased total EV protein and RNA levels that have been consistently identified by these studies in the plasma/sera of cancer patients. Reducing highly abundant miRNAs detected in total plasma is likely to generate benefits in EV research, allowing for better biomarker identification [33]. Protein complexes may involve RNA-binding proteins, such as the Argonaute family of proteins (AGO), and SEC allows protein-enriched fractions of the same sample to be simultaneously investigated [87]. Methods that lead to high protein co-isolation may also affect the linearity of high-sensitivity assays such as flow cytometry [88].

Of particular interest is the growing use of SEC for the isolation of biologically active EVs, that remain active even after freeze/thaw cycles [89]. This may be due to the lower operator-dependence, and reduced damage due to shearing forces caused by centrifugation at high speeds [19]. Additionally, the use of precipitation methods may have negative aspects on co-incubation studies, making some kits unsuitable as an isolation method if active EVs are required [90].

It is important to note that SEC has some limitations highlighted by these studies, which include the requirement for meticulous preparation of samples prior to loading [69], a maximal loading volume of 1–1.5 mL (dependent on column sizing), requirements for a concentration step if protein concentrations are low, and standardisation of column preparation. Contamination by plasma-associated proteins and lipoproteins is well documented (see Figure 2) and will have potentially impacted all studies that did not measure these in their SEC fractions [91].

Despite the small number of studies, 39% of the Minimal Information for Studies of Extracellular Vesicles (MISEV) criteria for characterisation was incomplete, which decreased to 34% when including only papers from 2014 onwards. RNA based studies were particularly poor at adhering to MISEV, with at least two features missing from all the studies. Without each of the components, it is impossible to guarantee that any RNA biomarker can be confidently associated with EVs and not the contaminants. This is consistent with the previous EV-TRACK findings, which noted that 30% (n = 5/17) of studies did not fully characterise EVs used for downstream RNA analysis. No experiments implemented other recommendations, such as using enzymatic digestion (lipase, proteinase, DNase, or RNase) to compare samples for non-EV-associated carriage.

The choice of EV isolation method has a significant impact on the EV yield [42], and this remains true both within SEC-based studies and compared with other methods. As this technique gains interest, it is likely that more comparative studies will also shed light on EV alterations that may occur during centrifugation compared with the much gentler gravity-dependent process of SEC. This will also require minimum reporting of parameters (as is the case with dUC, i.e., spin speed, time, rotor used/k-factor, and machine) [16], and it is hoped that future experiments using SEC will describe the type of material used for separation, dimensions of the elution column including the void volume, and the fractions ultimately classed as EVs to improve standardisation.

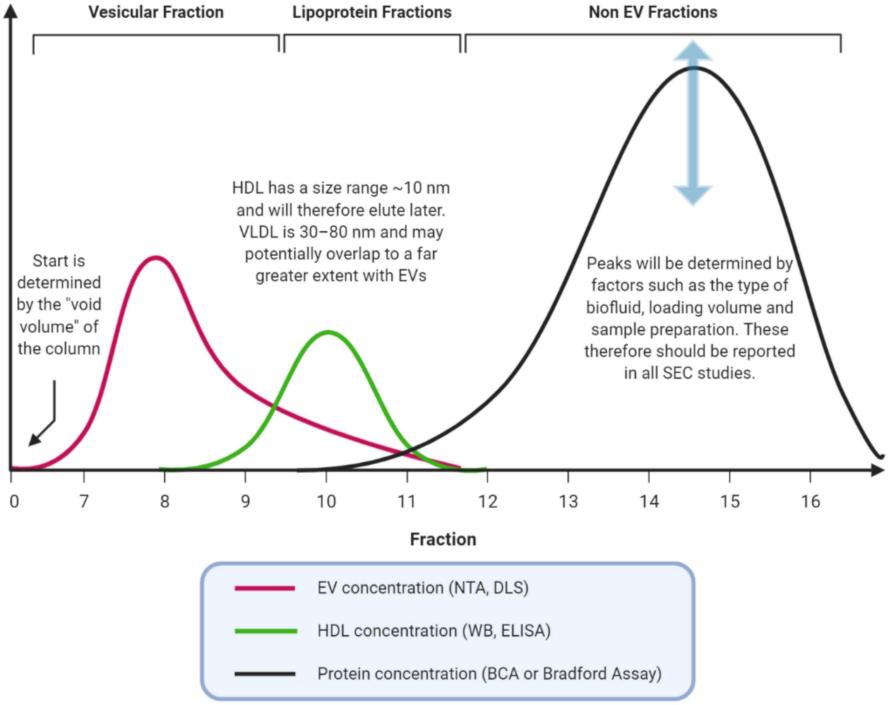


Figure 2. Schematic graph showing make up of consecutive fractions. EVs elute first but there is often an overlap between similarly sized particles e.g., small EVs and lipoproteins. Assessing these potential contaminants should be a mandatory feature of size-exclusion chromatography (SEC) EV studies. NTA, Nanoparticle Tracking Analysis; DLS, Dynamic Light Scattering; HDL, High-Density Lipoproteins; WB, Western Blot; ELISA, Enzyme-Linked ImmunoSorbent Assay; BCA, BicinChoninic Acid.

There continues to be innovative methods discovered for isolation, and some of these papers have been included in this review [92–95]. These demonstrate the feasibility for the isolation of EVs based on size, particularly in potential "lab-on-chip" applications. However, at present, the published results show that complex biofluids are problematic, leading to multiple preparatory steps to reduce sample complexity and various complications such as clogging and poor flow rates.

Asymmetric field flow fractionation is one such technology, which has been recently well described [96] and can be defined also as a size based EV isolation method. In this method, the solvent flows along a narrow channel (several hundreds of μm) together with a perpendicular cross flow that is created by a semipermeable membrane. Particles diffuse along a semipermeable membrane to be eluted according to size, and this is combined with real-time measurement of size by dynamic light scattering. It requires both significant time (1–2 h per sample) and specialised equipment but has the major advantage of enabling greater flexibility in separation parameters.

It is hoped that in the near future, these minor problems with SEC will be overcome, as the EVs isolated appear to be of high purity and a potential resource for the discovery of clinically useful cancer biomarkers.

Supplementary Materials: The following are available online at http://www.mdpi.com/2072-6694/12/11/3156/s1, Figure S1: PRISMA flow chart, Table S1: Size-exclusion methodology (chronological), Table S2: Papers reviewed showing EV characterization.

Author Contributions: Writing—original draft preparation, D.S.K.L. and F.M.U.; writing—review and editing, D.S.K.L., E.R., C.L. and F.M.U.; supervision, L.R.J., J.K., and A.E.F. Figures—Created with Biorender.com. All authors have read and agreed to the published version of the manuscript.

Funding: Royal College of Surgeons, England; J.P. Moulton Charitable Foundation Trust, U.K.; and the S.A.L. Charitable Fund, U.K.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

1. exosom*.ab:ti.
2. EVs.ab,ti.
3. (extracellular adj vesicl*).ab,ti.
4. microsom*.ab,ti.
5. vesicular.ab,ti.
6. microvesic*.af.
7. 1 or 2 or 3 or 4 or 5 or 6
8. (size adj exclusion).af.
9. gel filtration.af.
10. 8 or 9
11. cancer*.ab,ti.
12. neoplas*.ab,ti.
13. carcinoma*.ab,ti.
14. adenocarcinoma*.ab,ti.
15. 11 or 12 or 13 or 14
16. 7 and 10 and 15

References

1. Taylor, U.D.; Chou, I.-N.; Black, P.H. Isolation of plasma membrane fragments from cultured murine melanoma cells. *Biochem. Biophys. Res. Commun.* **1983**, *113*, 470–476. [CrossRef]
2. Kim, J.W.; Wieckowski, E.; Taylor, D.D.; E Reichert, T.; Watkins, S.; Whiteside, T.L. Fas ligand-positive membranous vesicles isolated from sera of patients with oral cancer induce apoptosis of activated T lymphocytes. *Clin. Cancer Res.* **2005**, *11*, 1010–1020. [PubMed]
3. Théry, C.; Boussac, M.; Véron, P.; Ricciardi-Castagnoli, P.; Raposo, G.; Garin, J.; Amigorena, S. Proteomic Analysis of Dendritic Cell-Derived Exosomes: A Secreted Subcellular Compartment Distinct from Apoptotic Vesicles. *J. Immunol.* **2001**, *166*, 7309–7318. [CrossRef] [PubMed]
4. Théry, C.; Ostrowski, M.; Segura, E. Membrane vesicles as conveyors of immune responses. *Nat. Rev. Immunol.* **2009**, *9*, 581–593. [CrossRef] [PubMed]
5. Booth, A.M.; Fang, Y.; Fallon, J.K.; Yang, J.-M.; Hildreth, J.E.; Gould, S.J. Exosomes and HIV Gag bud from endosome-like domains of the T cell plasma membrane. *J. Cell Biol.* **2006**, *172*, 923–935. [CrossRef] [PubMed]
6. Valadi, H.; Ekström, K.; Bossios, A.; Sjöstrand, M.; Lee, J.J.; Tvall, J.O.L.O. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* **2007**, *9*, 654–659. [CrossRef]
7. Lötvall, J.; Hill, A.; Hochberg, F.; Buzás, E.I.; Di Vizio, L.; Gardiner, C.; Gho, Y.S.; Kurochkin, I.V.; Mathivanan, S.; Quesenberry, P.; et al. *Minimal Experimental Requirements for Definition of Extracellular Vesicles and their Functions: A Position Statement from the International Society for Extracellular Vesicles*; Taylor & Francis: Oxfordshire, UK, 2014. [CrossRef]
8. Théry, C.; Witwer, K.W.; Aikawa, E.; Alcaraz, M.J.; Anderson, J.D.; Andriantsitohaina, R.; Antoniou, A.; Arab, T.; Archer, F.; Atkin-Smith, G.K.; et al. Minimal information for studies of extracellular vesicles 2018 (MISEV2018): A position statement of the International Society for Extracellular Vesicles and update of the MISEV2014 guidelines. *J. Extracell. Vesicles* **2018**, *7*, 1535750. [CrossRef]
9. Vaswani, K.; Mitchell, M.D.; Holland, O.J.; Koh, Y.Q.; Hill, R.J.; Harb, T.; Davies, P.S.W.; Peiris, H. A Method for the Isolation of Exosomes from Human and Bovine Milk. *J. Nutr. Metab.* **2019**, *2019*, 1–6. [CrossRef]
10. Lässer, C.; Alikhani, V.S.; Ekström, K.; Eldh, M.; Paredes, P.T.; Bossios, A.; Sjöstrand, M.; Gabrielsson, S.; Lötvall, J.; Valadi, H. Human saliva, plasma and breast milk exosomes contain RNA: Uptake by macrophages. *J. Transl. Med.* **2011**, *9*, 9. [CrossRef]
11. Czystowska-Kuzmicz, M.; Sosnowska, A.; Nowis, D.; Ramji, K.; Szajnik, M.; Chlebowska-Tuz, J.; Wolinska, E.; Gaj, P.; Grazul, M.; Pilch, Z.; et al. Small extracellular vesicles containing arginase-1 suppress T-cell responses and promote tumor growth in ovarian carcinoma. *Nat. Commun.* **2019**, *10*, 3000. [CrossRef]

12. Masyuk, A.I.; Huang, B.Q.; Ward, C.J.; Gradilone, S.A.; Banales, J.M.; Masyuk, T.V.; Radtke, B.; Splinter, P.L.; LaRusso, N.F. Biliary exosomes influence cholangiocyte regulatory mechanisms and proliferation through interaction with primary cilia. *Am. J. Physiol. Liver Physiol.* **2010**, *299*, G990–G999. [CrossRef]
13. Gardiner, C.; Di Vizio, D.; Sahoo, S.; Théry, C.; Witwer, K.W.; Wauben, M.; Hill, A.F. Techniques used for the isolation and characterization of extracellular vesicles: Results of a worldwide survey. *J. Extracell. Vesicles* **2016**, *5*, 32945. [CrossRef] [PubMed]
14. Van Deun, J.; Mestdagh, P.; Agostinis, P.; Akay, Ö.; Anand, S.; Anckaert, J.; Martinez, Z.A.; Baetens, T.; Beghein, E.; EV-TRACK Consortium; et al. EV-TRACK: Transparent reporting and centralizing knowledge in extracellular vesicle research. *Nat. Methods* **2017**, *14*, 228–232. [CrossRef]
15. Buschmann, D.; Kirchner, B.; Hermann, S.; Märte, M.; Wurmser, C.; Brandes, F.; Kotschote, S.; Bonin, M.; Steinlein, O.K.; Pfaffl, M.W.; et al. Evaluation of serum extracellular vesicle isolation methods for profiling miRNAs by next-generation sequencing. *J. Extracell. Vesicles* **2018**, *7*, 1481321. [CrossRef] [PubMed]
16. Cvjetkovic, A.; Lötvall, J.; Lässer, C. The influence of rotor type and centrifugation time on the yield and purity of extracellular vesicles. *J. Extracell. Vesicles* **2014**, *3*. [CrossRef]
17. Théry, C.; Amigorena, S.; Raposo, G.; Clayton, A. Isolation and Characterization of Exosomes from Cell Culture Supernatants and Biological Fluids. *Curr. Protoc. Cell Biol.* **2006**, *30*, 3.22.1. [CrossRef] [PubMed]
18. Clayton, A.; Court, J.; Navabi, H.; Adams, M.; Mason, M.D.; A Hobot, J.; Newman, G.R.; Jasani, B. Analysis of antigen presenting cell derived exosomes, based on immuno-magnetic isolation and flow cytometry. *J. Immunol. Methods* **2001**, *247*, 163–174. [CrossRef]
19. Taylor, D.D.; Shah, S. Methods of isolating extracellular vesicles impact down-stream analyses of their cargoes. *Methods* **2015**, *87*, 3–10. [CrossRef] [PubMed]
20. Gámez-Valero, A.; Monguió-Tortajada, M.; Carreras-Planella, L.; Franquesa, M.; Beyer, K.; Borràs, F.E. Size-Exclusion Chromatography-based isolation minimally alters Extracellular Vesicles' characteristics compared to precipitating agents. *Sci. Rep.* **2016**, *6*, srep33641. [CrossRef]
21. Lathe, G.H.; Ruthven, C.R. The separation of substances and estimation of their relative molecular sizes by the use of colums of starch in water. *Biochem. J.* **1956**, *62*, 665–674. [CrossRef]
22. Eisenstein, M. Adventures in the matrix. *Nat. Chem. Biol.* **2006**, *3*, 410. [CrossRef]
23. Lane, R.E.; Korbie, D.; Trau, M.; Hill, M.M. Optimizing Size Exclusion Chromatography for Extracellular Vesicle Enrichment and Proteomic Analysis from Clinically Relevant Samples. *Proteom.* **2019**, *19*, e1800156. [CrossRef] [PubMed]
24. Coumans, F.A.W.; Brisson, A.R.; Buzás, E.I.; Dignat-George, F.; Drees, E.E.E.; El Andaloussi, S.; Emanueli, C.; Gasecka, A.; Hendrix, A.; Hill, A.F.; et al. Methodological Guidelines to Study Extracellular Vesicles. *Circ. Res.* **2017**, *120*, 1632–1648. [CrossRef] [PubMed]
25. Böing, A.N.; Van Der Pol, E.; Grootemaat, A.E.; Coumans, F.A.W.; Sturk, A.; Nieuwland, R. Single-step isolation of extracellular vesicles by size-exclusion chromatography. *J. Extracell. Vesicles* **2014**, *3*, 23430. [CrossRef]
26. Michell, D.L.; Vickers, K.C. Lipoprotein carriers of microRNAs. *Biochim. et Biophys. Acta BBA Mol. Cell Biol. Lipids* **2016**, *1861*, 2069–2074. [CrossRef]
27. Buzás, E.I.; Tóth, E.Á.; Sódar, B.W.; Szabó-Taylor, K.É. Molecular interactions at the surface of extracellular vesicles. *Semin. Immunopathol.* **2018**, *40*, 453–464. [CrossRef]
28. Morales-Kastresana, A.; Telford, B.; Musich, T.A.; McKinnon, K.; Clayborne, C.; Braig, Z.; Rosner, A.; Demberg, T.; Watson, D.C.; Karpova, T.S.; et al. Labeling Extracellular Vesicles for Nanoscale Flow Cytometry. *Sci. Rep.* **2017**, *7*, 1–10. [CrossRef]
29. Dhondt, B.; Geeurickx, E.; Tulkens, J.; Van Deun, J.; Vergauwen, G.; Lippens, L.; Miinalainen, I.; Rappu, P.; Heino, J.; Ost, P.; et al. Unravelling the proteomic landscape of extracellular vesicles in prostate cancer by density-based fractionation of urine. *J. Extracell. Vesicles* **2020**, *9*, 1736935. [CrossRef]
30. Van Deun, J.; Hendrix, A.; On behalf of the EV-TRACK Consortium. Is your article EV-TRACKed? *J. Extracell. Vesicles* **2017**, *6*, 1379835. [CrossRef]
31. Kawakami, K.; Fujita, Y.; Matsuda, Y.; Arai, T.; Horie, K.; Kameyama, K.; Kato, T.; Masunaga, K.; Kasuya, Y.; Tanaka, M.; et al. Gamma-glutamyltransferase activity in exosomes as a potential marker for prostate cancer. *BMC Cancer* **2017**, *17*, 1–12. [CrossRef]
32. Muller, L.; Hong, C.-S.; Stolz, D.B.; Watkins, S.C.; Whiteside, T.L. Isolation of biologically-active exosomes from human plasma. *J. Immunol. Methods* **2014**, *411*, 55–65. [CrossRef] [PubMed]

33. Van Eijndhoven, M.A.J.; Zijlstra, J.M.; Groenewegen, N.J.; Drees, E.E.E.; Van Niele, S.; Baglio, S.R.; Koppers-Lalic, D.; Van Der Voorn, H.; Libregts, S.F.W.M.; Wauben, M.H.M.; et al. Plasma vesicle miRNAs for therapy response monitoring in Hodgkin lymphoma patients. *JCI Insight* **2016**, *1*, e89631. [CrossRef] [PubMed]
34. Théry, C.; Regnault, A.; Garin, J.; Wolfers, J.; Zitvogel, L.; Ricciardi-Castagnoli, P.; Raposo, G.; Amigorena, S. Molecular Characterization of Dendritic Cell-Derived Exosomes. *J. Cell Biol.* **1999**, *147*, 599–610. [CrossRef] [PubMed]
35. Andreu, Z.; Ýañez, M. Tetraspanins in Extracellular Vesicle Formation and Function. *Front. Immunol.* **2014**, *5*, 442. [CrossRef]
36. Taylor, D.D.; Gerçel-Taylor, C.; Lyons, K.S.; Stanson, J.; Whiteside, T.L. T-cell apoptosis and suppression of T-cell receptor/CD3-zeta by Fas ligand-containing membrane vesicles shed from ovarian tumors. *Clin. Cancer Res.* **2003**, *9*, 5113–5119.
37. Raposo, G.; Nijman, H.W.; Stoorvogel, W.; Liejendekker, R.; Harding, C.V.; Melief, C.J.; Geuze, H.J. B lymphocytes secrete antigen-presenting vesicles. *J. Exp. Med.* **1996**, *183*, 1161–1172. [CrossRef]
38. Baran, J.; Baj-Krzyworzeka, M.; Weglarczyk, K.; Szatanek, R.; Zembala, M.; Barbasz, J.; Czupryna, A.; Szczepanik, A.; Zembala, M. Circulating tumour-derived microvesicles in plasma of gastric cancer patients. *Cancer Immunol. Immunother.* **2009**, *59*, 841–850. [CrossRef]
39. Taylor, D.D.; Black, P.H. Shedding of plasma membrane fragments. Neoplastic and developmental importance. *Dev. Boil.* **1986**, *3*, 33–57.
40. Hong, C.-S.; Muller, L.; Whiteside, T.L.; Boyiadzis, M. Plasma Exosomes as Markers of Therapeutic Response in Patients with Acute Myeloid Leukemia. *Front. Immunol.* **2014**, *5*, 160. [CrossRef]
41. Lobb, R.J.; Becker, M.; Wen, S.W.; Wong, C.S.F.; Wiegmans, A.P.; Leimgruber, A.; Möller, A. Optimized exosome isolation protocol for cell culture supernatant and human plasma. *J. Extracell. Vesicles* **2015**, *4*, 27031. [CrossRef]
42. Brennan, K.; Martin, K.; Fitzgerald, S.P.; O'Sullivan, J.; Wu, Y.; Blanco, A.; Richardson, C.; Mc Gee, M.M. A comparison of methods for the isolation and separation of extracellular vesicles from protein and lipid particles in human serum. *Sci. Rep.* **2020**, *10*, 1–13. [CrossRef] [PubMed]
43. Djusberg, E.; Jernberg, E.; Thysell, E.; Golovleva, I.; Lundberg, P.; Crnalic, S.; Widmark, A.; Bergh, A.; Brattsand, M.; Wikström, P. High levels of the AR-V7 Splice Variant and Co-Amplification of the Golgi Protein CodingYIPF6inARAmplified Prostate Cancer Bone Metastases. *Prostate* **2017**, *77*, 625–638. [CrossRef] [PubMed]
44. Broggi, M.A.; Maillat, L.; Clement, C.C.; Bordry, N.; Corthésy, P.; Auger, A.; Matter, M.; Hamelin, R.; Potin, L.; Demurtas, D.; et al. Tumor-associated factors are enriched in lymphatic exudate compared to plasma in metastatic melanoma patients. *J. Exp. Med.* **2019**, *216*, 1091–1107. [CrossRef]
45. Shelke, G.V.; Lässer, C.; Gho, Y.S.; Lötvall, J. Importance of exosome depletion protocols to eliminate functional and RNA-containing extracellular vesicles from fetal bovine serum. *J. Extracell. Vesicles* **2014**, *3*, 24783. [CrossRef]
46. Theodoraki, M.; Hoffmann, T.K.; Whiteside, T.L. Separation of plasma-derived exosomes into CD3(+) and CD3(-) fractions allows for association of immune cell and tumour cell markers with disease activity in HNSCC patients. *Clin. Exp. Immunol.* **2018**, *192*, 271–283. [CrossRef] [PubMed]
47. Theodoraki, M.-N.; Yerneni, S.S.; Hoffmann, T.K.; Gooding, W.E.; Whiteside, T.L. Clinical Significance of PD-L1+Exosomes in Plasma of Head and Neck Cancer Patients. *Clin. Cancer Res.* **2017**, *24*, 896–905. [CrossRef] [PubMed]
48. Theodoraki, M.-N.; Yerneni, S.; Gooding, W.E.; Ohr, J.; Clump, D.A.; Bauman, J.E.; Ferris, R.L.; Whiteside, T.L. Circulating exosomes measure responses to therapy in head and neck cancer patients treated with cetuximab, ipilimumab, and IMRT. *OncoImmunology* **2019**, *8*, e1593805. [CrossRef]
49. Chen, G.; Huang, A.C.; Zhang, W.; Zhang, G.; Wu, M.; Xu, W.; Yu, Z.; Yang, J.; Wang, B.; Sun, H.; et al. Exosomal PD-L1 contributes to immunosuppression and is associated with anti-PD-1 response. *Nat. Cell Biol.* **2018**, *560*, 382–386. [CrossRef]
50. Daassi, D.; Mahoney, K.M.; Freeman, G.J. The importance of exosomal PDL1 in tumour immune evasion. *Nat. Rev. Immunol.* **2020**, *20*, 209–215. [CrossRef]
51. Yu, X.; Harris, S.L.; Levine, A.J. The Regulation of Exosome Secretion: A Novel Function of the p53 Protein. *Cancer Res.* **2006**, *66*, 4795–4801. [CrossRef]

52. Yu, X.; Riley, T.; Levine, A.J. The regulation of the endosomal compartment byp53the tumor suppressor gene. *FEBS J.* **2009**, *276*, 2201–2212. [CrossRef] [PubMed]
53. Abramowicz, A.; Wojakowska, A.; Marczak, L.; Lysek-Gladysinska, M.; Smolarz, M.; Story, M.D.; Polanska, J.; Widlak, P.; Pietrowska, M. Ionizing radiation affects the composition of the proteome of extracellular vesicles released by head-and-neck cancer cells in vitro. *J. Radiat. Res.* **2019**, *60*, 289–297. [CrossRef] [PubMed]
54. Takahashi, A.; Okada, R.; Nagao, K.; Kawamata, Y.; Hanyu, A.; Yoshimoto, S.; Takasugi, M.; Watanabe, S.; Kanemaki, M.T.; Obuse, C.; et al. Exosomes maintain cellular homeostasis by excreting harmful DNA from cells. *Nat. Commun.* **2017**, *8*, 15287. [CrossRef]
55. Freitas, D.; Balmaña, M.; Poças, J.; Campos, D.; Osório, H.; Konstantinidi, A.; Vakhrushev, S.Y.; Magalhães, A.; Reis, C.A. Different isolation approaches lead to diverse glycosylated extracellular vesicle populations. *J. Extracell. Vesicles* **2019**, *8*, 1621131. [CrossRef]
56. Julien, S.; Videira, P.A.; Delannoy, P. Sialyl-Tn in Cancer: (How) Did We Miss the Target? *Biomolecules* **2012**, *2*, 435–466. [CrossRef] [PubMed]
57. Ludwig, S.; Marczak, Ł.; Sharma, P.; Abramowicz, A.; Gawin, M.; Widlak, P.; Whiteside, T.L.; Pietrowska, M. Proteomes of exosomes from HPV(+) or HPV(-) head and neck cancer cells: Differential enrichment in immunoregulatory proteins. *OncoImmunology* **2019**, *8*, 1593808. [CrossRef]
58. Krupar, R.; Robold, K.; Gaag, D.; Spanier, G.; Kreutz, M.; Renner, K.; Hellerbrand, C.; Hofstaedter, F.; Bosserhoff, A.K. Immunologic and metabolic characteristics of HPV-negative and HPV-positive head and neck squamous cell carcinomas are strikingly different. *Virchows Archiv* **2014**, *465*, 299–312. [CrossRef] [PubMed]
59. Chandran, V.I.; Welinder, C.; Månsson, A.-S.; Offer, S.; Freyhult, E.; Pernemalm, M.; Lund, S.M.; Pedersen, S.; Lehtiö, J.; Marko-Varga, G.; et al. Ultrasensitive Immunoprofiling of Plasma Extracellular Vesicles Identifies Syndecan-1 as a Potential Tool for Minimally Invasive Diagnosis of Glioma. *Clin. Cancer Res.* **2019**, *25*, 3115–3127. [CrossRef]
60. Baietti, M.F.; Zhang, Z.; Mortier, E.; Melchior, A.; DeGeest, G.; Geeraerts, A.; Ivarsson, Y.; Depoortere, F.; Coomans, C.; Vermeiren, E.; et al. Syndecan–syntenin–ALIX regulates the biogenesis of exosomes. *Nat. Cell Biol.* **2012**, *14*, 677–685. [CrossRef]
61. Danielsson, F.; Skogs, M.; Huss, M.; Rexhepaj, E.; O'Hurley, G.; Klevebring, D.; Pontén, F.; Gad, A.K.B.; Uhlén, M.; Lundberg, E. Majority of differentially expressed genes are down-regulated during malignant transformation in a four-stage model. *Proc. Natl. Acad. Sci.* **2013**, *110*, 6853–6858. [CrossRef]
62. Anderson, N.L.; Polanski, M.; Pieper, R.; Gatlin, T.; Tirumalai, R.S.; Conrads, T.P.; Veenstra, T.D.; Adkins, J.N.; Pounds, J.G.; Fagan, R.; et al. The Human Plasma Proteome. *Mol. Cell. Proteom.* **2004**, *3*, 311–326. [CrossRef]
63. Sakha, S.; Muramatsu, T.; Ueda, K.; Inazawa, J. Exosomal microRNA miR-1246 induces cell motility and invasion through the regulation of DENND2D in oral squamous cell carcinoma. *Sci. Rep.* **2016**, *6*, 38750. [CrossRef] [PubMed]
64. Peacock, B.; Rigby, A.; Bradford, J.R.; Pink, R.C.; Hunter, K.D.; Lambert, D.W.; Hunt, S. Extracellular vesicle micro RNA cargo is correlated with HPV status in oropharyngeal carcinoma. *J. Oral Pathol. Med.* **2018**, *47*, 954–963. [CrossRef] [PubMed]
65. Rabinowits, G.; Gerçel-Taylor, C.; Day, J.M.; Taylor, D.D.; Kloecker, G.H. Exosomal MicroRNA: A Diagnostic Marker for Lung Cancer. *Clin. Lung Cancer* **2009**, *10*, 42–46. [CrossRef] [PubMed]
66. Went, P.; Vasei, M.; Bubendorf, L.; Terracciano, L.; Tornillo, L.; Riede, U.; Kononen, J.; Simon, R.; Sauter, G.; A Baeuerle, P. Frequent high-level expression of the immunotherapeutic target Ep-CAM in colon, stomach, prostate and lung cancers. *Br. J. Cancer* **2006**, *94*, 128–135. [CrossRef]
67. Taylor, D.D.; Gercel-Taylor, C. MicroRNA signatures of tumor-derived exosomes as diagnostic biomarkers of ovarian cancer. *Gynecol. Oncol.* **2008**, *110*, 13–21. [CrossRef] [PubMed]
68. Tauro, B.J.; Greening, D.W.; Mathias, R.A.; Mathivanan, S.; Ji, H.; Simpson, R.J. Two Distinct Populations of Exosomes Are Released from LIM1863 Colon Carcinoma Cell-derived Organoids. *Mol. Cell. Proteom.* **2012**, *12*, 587–598. [CrossRef]
69. Hong, C.-S.; Funk, S.; Muller, L.; Boyiadzis, M.; Whiteside, T.L. Isolation of biologically active and morphologically intact exosomes from plasma of patients with cancer. *J. Extracell. Vesicles* **2016**, *5*, 29289. [CrossRef]
70. Simonsen, J.B. Pitfalls associated with lipophilic fluorophore staining of extracellular vesicles for uptake studies. *J. Extracell. Vesicles* **2019**, *8*, 1582237. [CrossRef]

71. Takov, K.; Yellon, D.M.; Davidson, S.M. Confounding factors in vesicle uptake studies using fluorescent lipophilic membrane dyes. *J. Extracell. Vesicles* **2017**, *6*, 1388731. [CrossRef]
72. Lehrich, B.M.; Liang, Y.; Khosravi, P.; Federoff, H.J.; Fiandaca, M.S. Fetal Bovine Serum-Derived Extracellular Vesicles Persist within Vesicle-Depleted Culture Media. *Int. J. Mol. Sci.* **2018**, *19*, 3538. [CrossRef] [PubMed]
73. Auber, M.; Fröhlich, D.; Drechsel, O.; Karaulanov, E.; Krämer-Albers, E.-M. Serum-free media supplements carry miRNAs that co-purify with extracellular vesicles. *J. Extracell. Vesicles* **2019**, *8*, 1656042. [CrossRef] [PubMed]
74. Appaiah, H.N.; Goswami, C.P.; A Mina, L.; Badve, S.; Sledge, J.G.W.; Liu, Y.; Nakshatri, H. Persistent upregulation of U6:SNORD44 small RNA ratio in the serum of breast cancer patients. *Breast Cancer Res.* **2011**, *13*, R86. [CrossRef] [PubMed]
75. Xiang, M.; Zeng, Y.; Yang, R.; Xu, H.; Chen, Z.; Zhong, J.; Xie, H.; Xu, Y.; Zeng, X. U6 is not a suitable endogenous control for the quantification of circulating microRNAs. *Biochem. Biophys. Res. Commun.* **2014**, *454*, 210–214. [CrossRef] [PubMed]
76. Mestdagh, P.; Van Vlierberghe, P.; De Weer, A.; Muth, D.; Westermann, F.; Speleman, F.; Vandesompele, J. A novel and universal method for microRNA RT-qPCR data normalization. *Genome Biol.* **2009**, *10*, R64. [CrossRef] [PubMed]
77. Pan, C.; Issaeva, N.; Yarbrough, W.G. HPV-driven oropharyngeal cancer: Current knowledge of molecular biology and mechanisms of carcinogenesis. *Cancers Head Neck* **2018**, *3*, 1–11. [CrossRef] [PubMed]
78. Wieckowski, E.U.; Visus, C.; Szajnik, M.; Szczepanski, M.J.; Storkus, W.J.; Whiteside, T.L. Tumor-Derived Microvesicles Promote Regulatory T Cell Expansion and Induce Apoptosis in Tumor-Reactive Activated CD8+ T Lymphocytes. *J. Immunol.* **2009**, *183*, 3720–3730. [CrossRef]
79. Antonioli, L.; Blandizzi, C.; Pacher, P.; Haskó, G. Immunity, inflammation and cancer: A leading role for adenosine. *Nat. Rev. Cancer* **2013**, *13*, 842–857. [CrossRef]
80. Schuler, P.J.; Saze, Z.; Hong, C.-S.; Muller, L.; Gillespie, D.G.; Cheng, D.; Harasymczuk, M.; Mandapathil, M.; Lang, S.; Jackson, E.K.; et al. Human CD4+CD39+regulatory T cells produce adenosine upon co-expression of surface CD73 or contact with CD73+exosomes or CD73+cells. *Clin. Exp. Immunol.* **2014**, *177*, 531–543. [CrossRef]
81. Ludwig, S.; Floros, T.; Theodoraki, M.-N.; Hong, C.-S.; Jackson, E.K.; Lang, S.; Whiteside, T.L. Suppression of Lymphocyte Functions by Plasma Exosomes Correlates with Disease Activity in Patients with Head and Neck Cancer. *Clin. Cancer Res.* **2017**, *23*, 4843–4854. [CrossRef]
82. Ludwig, N.; Yerneni, S.S.; Razzo, B.M.; Whiteside, T.L. Exosomes from HNSCC Promote Angiogenesis through Reprogramming of Endothelial Cells. *Mol. Cancer Res.* **2018**, *16*, 1798–1808. [CrossRef] [PubMed]
83. Theodoraki, M.; Yerneni, S.; Lorenz, K.; Laban, S.; Schuler, P.; Brunner, C.; Hoffmann, T.; Theodorakis, I.; Whiteside, T. PO-243 Exosomes from plasma of HNSCC patients treated with photodynamic therapy are biomarkers for epithelial-mesenchymal transition. *Poster Presentation* **2018**, *3*, A322. [CrossRef]
84. Royo, F.; Moreno, L.; Mleczko, J.; Palomo, L.; Gonzalez, E.; Cabrera, D.; Cogolludo, A.; Vizcaino, F.P.; Van-Liempd, S.; Falcón-Pérez, J.M. Hepatocyte-secreted extracellular vesicles modify blood metabolome and endothelial function by an arginase-dependent mechanism. *Sci. Rep.* **2017**, *7*, srep42798. [CrossRef] [PubMed]
85. Zhang, H.; Liu, J.; Qu, D.; Wang, L.; Wong, C.M.; Lau, C.-W.; Huang, Y.; Wang, Y.F.; Huang, H.; Xia, Y.; et al. Serum exosomes mediate delivery of arginase 1 as a novel mechanism for endothelial dysfunction in diabetes. *Proc. Natl. Acad. Sci.* **2018**, *115*, E6927–E6936. [CrossRef] [PubMed]
86. Sjöqvist, S.; Kasai, Y.; Shimura, D.; Ishikawa, T.; Ali, N.; Iwata, T.; Kanai, N. Oral keratinocyte-derived exosomes regulate proliferation of fibroblasts and epithelial cells. *Biochem. Biophys. Res. Commun.* **2019**, *514*, 706–712. [CrossRef] [PubMed]
87. Arroyo, J.D.; Chevillet, J.R.; Kroh, E.M.; Ruf, I.K.; Pritchard, C.C.; Gibson, D.F.; Mitchell, P.S.; Bennett, C.F.; Pogosova-Agadjanyan, E.L.; Stirewalt, D.L.; et al. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc. Natl. Acad. Sci.* **2011**, *108*, 5003–5008. [CrossRef]
88. Suárez, H.; Gámez-Valero, A.; Reyes, R.; López-Martín, S.; Rodríguez, M.J.; Carrascosa, J.L.; Cabañas, C.; Borràs, F.E.; Yáñez-Mo, M. A bead-assisted flow cytometry method for the semi-quantitative analysis of Extracellular Vesicles. *Sci. Rep.* **2017**, *7*, 11271. [CrossRef]
89. Jeyaram, A.; Jay, S.M. Preservation and Storage Stability of Extracellular Vesicles for Therapeutic Applications. *AAPS J.* **2017**, *20*, 1–7. [CrossRef]

90. Patel, G.K.; Khan, M.A.; Zubair, H.; Srivastava, S.K.; Khushman, M.; Singh, S.; Singh, A.P. Comparative analysis of exosome isolation methods using culture supernatant for optimum yield, purity and downstream applications. *Sci. Rep.* **2019**, *9*, 1–10. [CrossRef]
91. Welton, J.L.; Webber, J.P.; Botos, L.-A.; Jones, M.; Clayton, A. Ready-made chromatography columns for extracellular vesicle isolation from plasma. *J. Extracell. Vesicles* **2015**, *4*, 27269. [CrossRef]
92. Dong, X.; Chi, J.; Zheng, L.; Ma, B.; Li, Z.; Wang, S.; Zhao, C.; Liu, H. Efficient isolation and sensitive quantification of extracellular vesicles based on an integrated ExoID-Chip using photonic crystals. *Lab a Chip* **2019**, *19*, 2897–2904. [CrossRef]
93. Smith, J.T.; Wunsch, B.H.; Dogra, N.; Ahsen, M.E.; Lee, K.; Yadav, K.K.; Weil, R.; Pereira, M.A.; Patel, J.V.; Duch, E.A.; et al. Integrated nanoscale deterministic lateral displacement arrays for separation of extracellular vesicles from clinically-relevant volumes of biological samples. *Lab a Chip* **2018**, *18*, 3913–3925. [CrossRef] [PubMed]
94. Wunsch, B.H.; Smith, J.T.; Gifford, S.M.; Wang, C.; Brink, M.; Bruce, R.L.; Austin, R.H.; Stolovitzky, G.; Astier, B.H.W.J.T.S.S.M.G.C.W.M.B.R.L.B.G.S.Y. Nanoscale lateral displacement arrays for the separation of exosomes and colloids down to 20 nm. *Nat. Nanotechnol.* **2016**, *11*, 936–940. [CrossRef] [PubMed]
95. Davies, R.T.; Kim, J.; Jang, S.C.; Choi, E.-J.; Gho, Y.S.; Park, J. Microfluidic filtration system to isolate extracellular vesicles from blood. *Lab a Chip* **2012**, *12*, 5202–5210. [CrossRef] [PubMed]
96. Zhang, H.; Lyden, D. Asymmetric-flow field-flow fractionation technology for exomere and small extracellular vesicle separation and characterization. *Nat. Protoc.* **2019**, *14*, 1027–1053. [CrossRef] [PubMed]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

International Journal of
Molecular Sciences

IMPACT FACTOR 6.208

Indexed in: PubMed

Review

The Clinical Significance of Transfer RNAs Present in Extracellular Vesicles

Daniel S. K. Liu, Qi Zhi Clayton Yang, Mohammad Asim, Jonathan Krell and Adam E. Frampton

Special Issue
The Role of Exosomes in Health and Disease

Edited by
Dr. Vincenzo Lionetti

MDPI

https://doi.org/10.3390/ijms23073692

International Journal of
Molecular Sciences

MDPI

Review

The Clinical Significance of Transfer RNAs Present in Extracellular Vesicles

Daniel S. K. Liu [1,†], Qi Zhi Clayton Yang [1,†], Mohammad Asim [2], Jonathan Krell [1] and Adam E. Frampton [1,2,3,*]

[1] Division of Cancer, Department of Surgery & Cancer, Imperial College London, Hammersmith Hospital Campus, Du Cane Road, London W12 0HS, UK; daniel.liu08@imperial.ac.uk (D.S.K.L.); qi.yang17@imperial.ac.uk (Q.Z.C.Y.); j.krell@imperial.ac.uk (J.K.)
[2] Department of Clinical and Experimental Medicine, Faculty of Health and Medical Sciences, University of Surrey, Guildford, Surrey GU2 7WG, UK; m.asim@surrey.ac.uk
[3] HPB Surgical Unit, Royal Surrey County Hospital NHS Foundation Trust, Guildford, Surrey GU2 7XX, UK
* Correspondence: adam.frampton@surrey.ac.uk or a.frampton@imperial.ac.uk
† These authors contributed equally to this work.

Abstract: Extracellular vesicles (EVs) are important for intercellular signalling in multi-cellular organisms. However, the role of mature transfer RNAs (tRNAs) and tRNA fragments in EVs has yet to be characterised. This systematic review aimed to identify up-to-date literature on tRNAs present within human EVs and explores their potential clinical significance in health and disease. A comprehensive and systematic literature search was performed, and the study was conducted in accordance with PRISMA guidelines. Electronic databases MEDLINE and EMBASE were searched up until 1 January 2022. From 685 papers, 60 studies were identified for analysis. The majority of papers reviewed focussed on the role of EV tRNAs in cancers (31.7%), with numerous other conditions represented. Blood and cell lines were the most common EV sources, representing 85.9% of protocols used. EV isolation methods included most known methods, precipitation being the most common (49.3%). The proportion of EV tRNAs was highly variable, ranging between 0.04% to >95% depending on tissue source. EV tRNAs are present in a multitude of sources and show promise as disease markers in breast cancer, gastrointestinal cancers, and other diseases. EV tRNA research is an emerging field, with increasing numbers of papers highlighting novel methodologies for tRNA and tRNA fragment discovery.

Keywords: extracellular vesicle; exosome; transfer RNA; tRNA fragment; tRNA half; cancer

check for updates

Citation: Liu, D.S.K.; Yang, Q.Z.C.; Asim, M.; Krell, J.; Frampton, A.E. The Clinical Significance of Transfer RNAs Present in Extracellular Vesicles. *Int. J. Mol. Sci.* **2022**, *23*, 3692. https://doi.org/10.3390/ijms23073692

Academic Editor: Vincenzo Lionetti

Received: 28 February 2022
Accepted: 23 March 2022
Published: 28 March 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Extracellular vesicles (EVs) are small cargo-containing structures with a lipid bilayer but do not have the cellular machinery required to replicate. They have been shown to play a role in cell-to-cell communication, as they can be found to transport biological material including proteins, lipids, ribonucleic acid (RNA), and deoxyribonucleic acid (DNA) between cells, leading to cellular changes within multi-cellular organisms [1]. It is a process that has been conserved through evolution, found both in prokaryotes and eukaryotes [2].

EVs are an umbrella term for a heterogenous array of secreted membrane vesicles and can be further distinguished by cell source, biogenesis pathways, and size ranges, thus contributing to diverse nomenclature [2]. Two broad categories that are often used distinguish microvesicles (MVs) and exosomes on the basis of cellular biogenesis [2]. Microvesicles are generated by the outward budding of the phospholipid bilayer generating vesicles made directly from the plasma membrane. These vesicles tend to be larger with MVs having a size range of up to 1000 nm or larger in some cases [2,3]. Exosomes, on the other hand, involve intracellular mechanisms forming multivesicular endosomes (MVEs) which contain intraluminal vesicles (ILVs). ILVs are formed from the budding of MVEs

which then fuse with the plasma membrane and get released extracellularly as exosomes. Exosomes are usually smaller in size (<150 nm) closely reflecting that of ILVs [2,3] and therefore also express specific markers that relate to their endosomal origin [2].

EVs have been found to be present in virtually all types of biofluids, including cerebrospinal fluid (CSF) [4], urine [5], blood [6], bile [7], and breast milk [8] as well as being released by other organisms and even plants [9]. Numerous studies focus on their carrier material which contains lipids, protein, and RNA in an endeavour to link them to the pathophysiology of a wide range of diseases [1]. With the current advances in RNA sequencing technologies and bioinformatic strategies, this has drawn attention to the wide range of small RNA (sRNA) content in EVs.

Whilst much of the literature has focused on studying micro RNAs (miRNAs) [10], short 16–22 nucleotide (nt) single-stranded molecules with a role in canonical transcription and translational pathways, less is known about the role of transfer RNAs (tRNA) despite being the most abundant RNA in the human genome [11], and the first non-coding RNA to have been discovered [12]. The canonical cytoplasmic tRNA molecule is 76 nt long and well-conserved, with a cloverleaf secondary structure and a tertiary L-shaped structure that separates the anticodon triplet for recognition of template mRNA from the amino acid attachment site. Whilst tRNAs are present in all species for protein translation, tRNA genetic expression is species-specific and with over 270 different tRNA sequences present among approximately 450 tRNA human genes, highly complex in eukaryotes [13]. Furthermore, the genetic expression of tRNAs is subject to modifications such as base-specific methylation, altering function, and fragmentation which can generate whole new subspecies of RNA molecules. This has led to, over the past decade [14,15], a wide expansion in tRNA biology with regulatory functions affecting gene expression, protein synthesis, and the stress response as significant downstream functions [11].

The fragmentation products of tRNAs are a potential source of new functionality away from conventional protein synthesis. tRNA-halves and tRNA-derived fragments (tRFs) are two such fragmentation products of tRNAs and have distinct biogenesis pathways (Figure 1). tRNA-halves are thought to be formed by nucleases such as angiogenin as well as Dicer and RNase Z which are upregulated under varying cellular conditions including stress [11,14,16]. The simplest fragmentation is a cleavage of the anti-codon loop forming either 5′ or 3′ tRNA-halves with lengths corresponding to half the full-length of mature tRNAs [14]. tRFs on the other hand, includes 5′ and 3′-tRFs with shorter nucleotide sequences and are named based on where the cleavage occurred, and are thought to be produced from the cleavage occurring in or around the D-loop or T-loop structure of mature tRNAs, respectively [16]. However, tRFs can also arise from 3′ nuclease cleavage of 5′ tRNA-halves and as such, the enzymes responsible for the production of smaller tRFs remain not well understood [16]. There are also tRFs that do not correspond to the 5′ or 3′ end and are called internal tRFs. Downstream effects of tRFs are thought to be related to RNA machinery, associating with argonaute proteins to exhibit miRNA-like regulatory and silencing activity on mRNA to facilitate post-transcriptional repression [16]. This represents a new spectrum of molecular pathways that converge to modify gene expression and thus influence cells in health and disease.

In this review, we aimed to comprehensively search the literature for studies that have identified tRNA and tRNA-derived cargo present in human EVs from studies performed either in vivo or in vitro. This has enabled us to clearly define the current state of EV tRNA research, defining the EVs present in these studies and highlighting functional and clinical significance where noted.

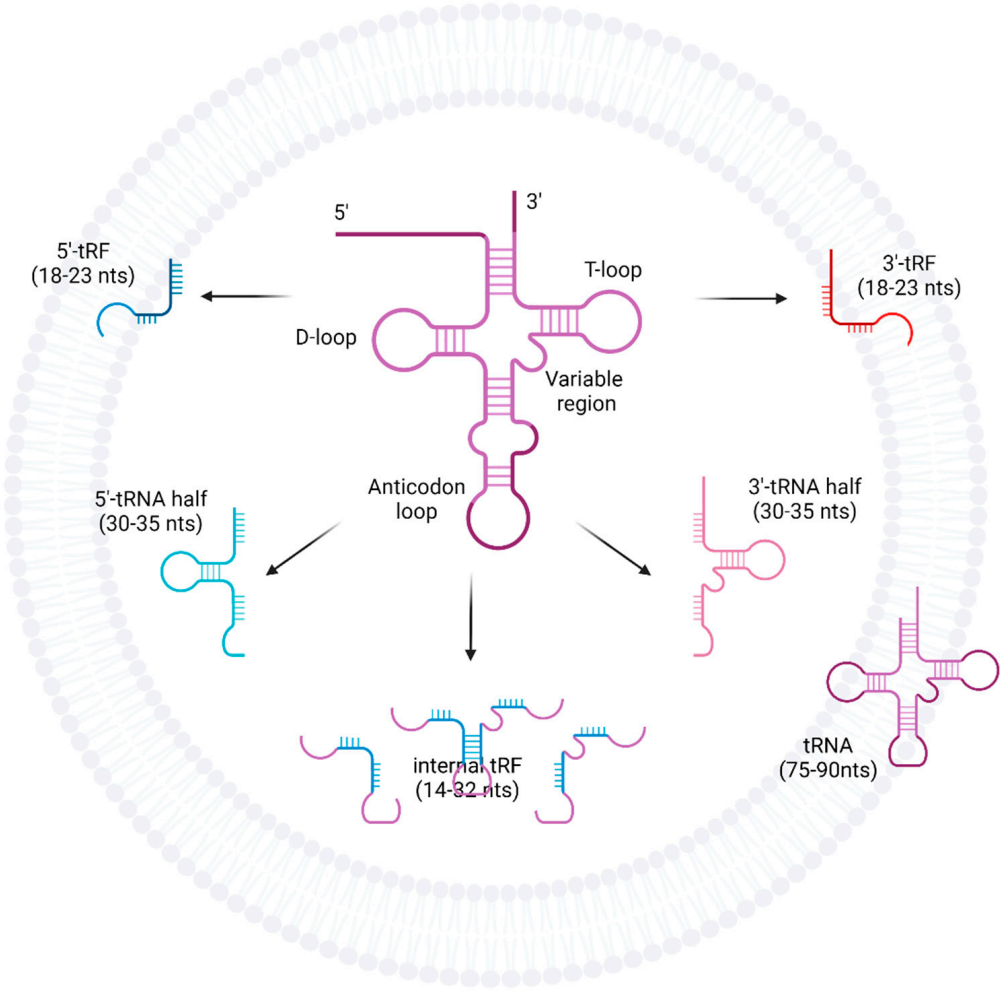


Figure 1. Fragmentation of tRNAs into tRFs, tRNA-halves, and internal tRFs. tRNAs are cleaved at the anti-codon loop by enzymes such as angiogenin, Dicer, or RNase Z into either 5′ or 3′-tRNA-halves. tRFs can arise from mature tRNA, pre-tRNA as well as tRNA-halves, and are formed when cleavage occurs at either the D-loop or T-loop. However, the enzymes responsible for tRF production are less well understood.

2. Results

A total of 60 studies were identified as relevant for this review. Forty-four studies (73.3%) focussed on pathological diseases, whilst the remaining 16 studies (26.7%) did not investigate a particular disease but instead focussed investigations on optimising methodologies such as choice of EV source, isolation method, characterisation of tRNA cargo, as well as on biofluids from healthy individuals.

Of the 44 disease-focussed studies, 43.2% of them were primarily investigating cancers (Figure 2A). The remaining studies are mainly comprised of papers focussing on respiratory, renal, bone, sepsis, autoimmune, reproductive tract, and infectious diseases. Notably, gastrointestinal cancers were the most common cancer studied (31.6%), followed by breast cancer, pan-cancer, and HPV-induced cancers representing 15.8% each (Figure 2A).

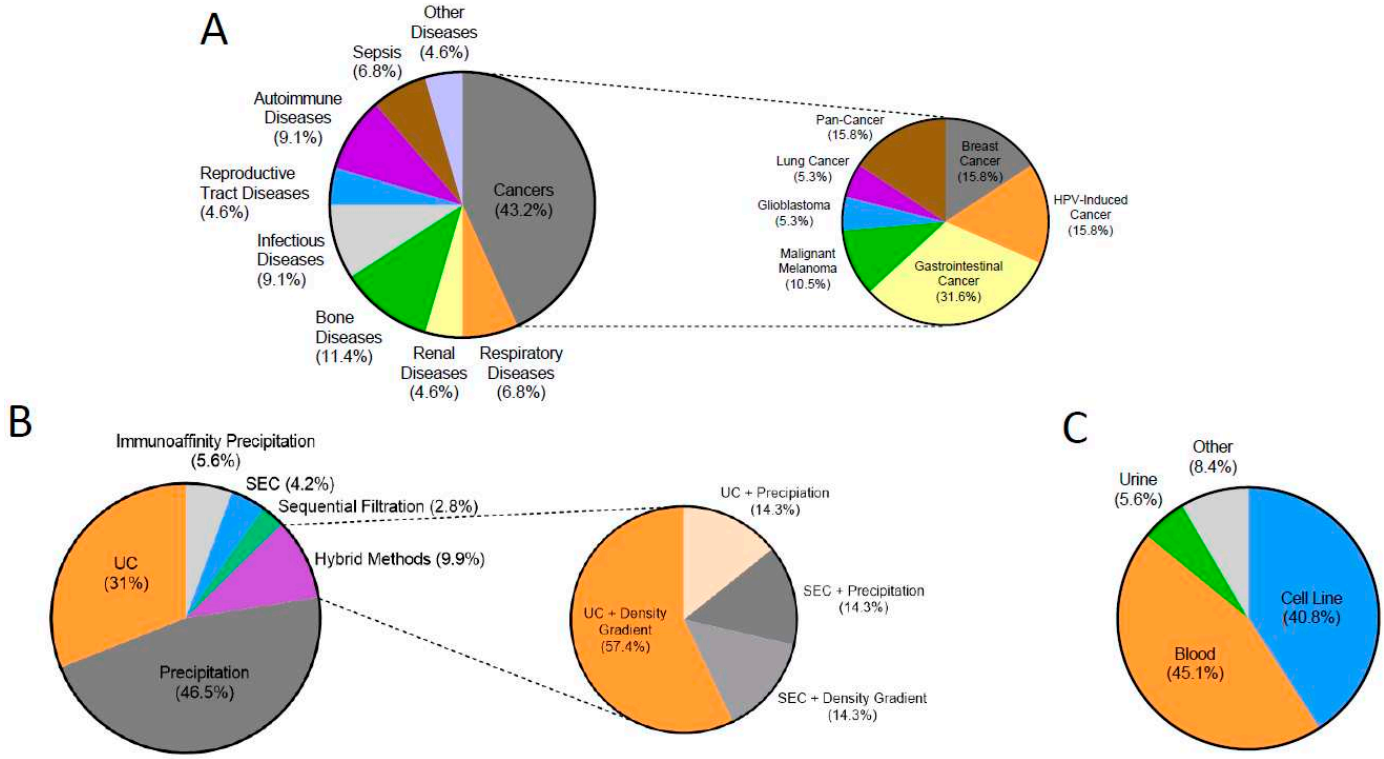


Figure 2. Types of diseases included in the review and the EV methodologies used. (**A**) Pie chart showing the primary disease focus of the studies (n = 44) included in the review. Studies which investigated EVs in healthy biofluids or methodology optimisation were not included in the pie chart; (**B**) Pie chart showing the different EV isolation methodologies used in the protocols investigated (n = 71); (**C**) Pie chart showing the different types of biofluids used for EV isolation in the protocols investigated (n = 71).

2.1. Overview of EV Isolation Methodologies and Characterisation

A total of 71 different EV isolation protocols were identified from 60 studies (Figure 2B). The main methodologies used were ultracentrifugation (UC), precipitation, affinity-based purification, size exclusion chromatography (SEC), density gradient, and sequential filtration, with the majority using either UC or precipitation in isolation (77.5%). In order to simplify these protocols for comparison, we excluded all centrifugation steps lower than 16,500 g and did not include large pore filtration as a distinct step. The extracted protocols are shown in Supplementary Table S1.

Precipitation was the most popular EV isolation method identified, with 46.5% of protocols using it as the sole EV isolation method (Figure 2B). Six different precipitation methods were used with ExoQuick (System Biosciences, Palo Alto, CA, USA) being the most common kit used. Other kits included Total Exosome Isolation Reagent (Invitrogen, Waltham, MA, USA), PureExo Exosome Isolation Kit (101Bio, Mountain View, CA, USA), miRCURY Exosome Isolation Kit (QIAGEN, Hilden, Germany), and Ribo Exosome Isolation Reagent Kit (Ribobio, Guangzhou, China) and Plasma/Serum Exosome Purification Kit (Norgen Biotek Corporation, Thorold, ON, Canada), Cell Culture Media Exosome Purification Mini Kit (Norgen Biotek Corporation), Umibio Exosome Isolation Kit (Umibio, Shanghai, China) and Exo2D RNA solution (Exosomeplus, Seoul, Korea).

The second most common methodology to isolate EVs was UC (31%) as a sole method, with affinity purification methods and SEC, third and fourth, respectively (Figure 2B). Protocols for UC varied and there was minimal reporting of rotor sizes and centrifuges used which has an effect on pellet aggregation and EV isolation. Most (90.5%) used a final spin speed of >100,000 g and of the 22 UC protocols (used as the sole method), 36.4% of protocols involved two consecutive UC spins. In addition, seven protocols used a hybrid of

methods with the most common (four of them) being the combination of density gradients and ultracentrifugation (Figure 2B).

Minimal information guidelines for the reporting of EV studies were published in 2014 and updated in 2018 which highlighted characterisation criteria needed to demonstrate the presence of EVs in a study. We therefore reviewed the studies for Western blots showing EV markers, transmission electron microscopy pictures, and evidence of nanoparticle tracking analysis (or equivalent quantitative analysis). In total, 35% of the studies identified employed these three EV characterisation techniques.

2.2. Overview of EV Sources

The vast proportion of studies identified used either cell lines or blood as their source of EVs, with the majority still obtained from cell-conditioned media. A total of 40.8% of protocols used cell lines, and 45.1% of protocols used blood (Figure 2C). Other sources were surprisingly diverse, coming from semen, saliva, ovarian follicular fluid, bile, sweat, and urine.

2.3. EV tRNAs Present in Non-Pathological Samples

2.3.1. Optimisation of EV Experimental Methodology

Several studies are now discussed which had the broad remit of optimising various stages of the pipeline required for EV tRNA research. This includes determining the parameters of EV source material, tRNA isolation, and RNA sequencing methodologies from primarily healthy samples.

Srinivasan et al. [17] investigated comprehensively the RNA profiles of different biofluids in a large study that involved 20 healthy volunteer donors and 3 patients with cholangiocarcinoma (CCA). They undertook RNA sequencing on five different biofluids (bile, cell culture, plasma, serum, and urine) and applied ten different extracellular RNA isolation methods with the compatibility of each protocol matched to the type of biofluid investigated. Additionally, this study was unique in being performed across six different laboratories to determine inter-laboratory variability. tRNA expression profiles were grouped according to the type of fragment (5′-half, 3′-half, 5′-tRF, 3′-tRF, or internal, i-tRFs) and they noticed significant differences between samples, with bile and urine samples almost completely comprised of 5′-halves and 5′-tRFs, whilst cell-line-derived EV tRNA sequences were predominantly made up of 5′-tRFs, compared to their intracellular profile which contained mainly 3′-tRFs and i-tRFs. According to amino acid distribution, each biofluid type had broadly similar profiles, with 30–40% being represented by both tRNA-Gly and tRNA-Glu and 2–5% by tRNA-His and tRNA-Val. By comparing isolation methods, they identified that across EV samples, these were less variable in profile compared with the analysis undertaken for miRNA and mRNAs.

This variability in EV isolation methods was also examined by Lasser et al. [18], who looked at just one type of EV released by mast cells using UC and subsequent sucrose density gradient to isolate high density (HD) and low density (LD) EVs. The density gradient is a highly selective, less commonly used technique for EV isolation. They found that in the HD fraction, mature miRNA (23%) was mainly found, whereas the LD fraction showed a predominance of tRNAs (28%). Moreover, 96% of the tRFs found in the LD fraction were also mainly from the 5′ terminal of tRNAs. This is significant as UC inherently takes into account particle and vesicle density, which means that different UC protocols or even rotors may contribute to an alteration in the size and density of vesicles isolated, leading to variability in RNA profiles. Further analysis of centrifugation techniques was shown by Tosar et al. [19] who analysed breast epithelial EVs (from MCF-7 and MCF-10A cells) dividing them into a 16,000 g pellet (p16) and 100,000 g pellet (p100), and comparing these to the supernatant (S100). The authors found that the S100 portion contained an 18-fold increase in tRNA fragments compared to the intracellular fraction, with the p16 and p100 fractions having a 3-fold and 2-fold increase, respectively. These extracellular tRNA-derived sequences showed a narrow size distribution, with a mode at 30–31 nt and

again a preference for 5′ sequences. In contrast, miRNAs decreased in relative abundance between the intracellular and extracellular (vesicular and non-vesicular) compartments. What was interesting was that they noted that extracellular tRNAs may be present and bound to ribonucleoprotein-associated complexes which complicates evaluation of the EV compartment. Additionally, they identified that 5′ halves from tRNA-Glu-CUC were detectable in foetal bovine serum, a common supplement to cell culture media and so care must be taken in excluding this from EVs isolated from cell lines.

Sork et al. [20] investigated five immortalized cell lines (HEK293T (human embryonic kidney), RD4 (human skeletal muscle cells), Neuro2a (neuroblastoma cells), C17.2 (immortalised mouse neural progenitor cells), and C2C12 (immortalised mouse myoblasts)) to gain a better comprehensive understanding of the differences and overlap in RNA species and abundance in EVs, using transcriptomic and proteomic approaches. They found that tRNAs were one of the most predominant RNAs, after ribosomal RNA (rRNA) and sRNA, found in the EVs across all five cell lines. Specifically, tRNA-Gly-GCC, tRNA-His-GTG, and tRNA-Glu-CTC were the most prevalent fragments. The authors were, however, not able to further subclassify these tRNA fragments and were also not able to attribute any to specific biological functions.

Small RNA library preparation was investigated by Huang et al. [21] in 2013, as this can be a source of RNA sequencing variability. They compared the applicability of three RNA sequencing methods within the plasma EVs of healthy volunteers obtained using ExoQuick. Three preparation kits designed to obtain small non-coding RNAs (30–40 nts) were investigated (NEBNext multiplex small RNA library preparation kit, NEXTflex small RNA sequencing kit, and TruSeq small RNA sample preparation kit) and appeared to show good technical reproducibility. However, each kit had preferential enrichment of low abundant RNAs. Importantly, this inherent bias in RNA library preparation would be present for the tRNA analysis performed in their study as tRNAs accounted for just 1.24% of total mapped reads. It is highly recommended that all studies investigating EV RNAs should therefore include validation by qRT-PCR.

Since then, a couple of new techniques for analysis of EV tRNAs have arisen, with a new pipeline—One-Step Total Transcriptome Protocol proposed by Amorim et al. [22] to address the issue of analysing both short and long RNAs found in EVs in a single next-generation sequencing (NGS) library. This modified protocol used RNA fragmentation by RNase-III treatment to minimise adaptor-dimer formation and changes in the ligation and PCR step. The authors found that within the plasma EVs of five healthy participants included in the study, short non-coding RNA represented the second most abundant RNA population after rRNA reads were filtered. Of note, tRNAs were the major species, representing 57.29% of mapped reads. This pipeline reduced bias from amplification due to heterogenous samples and therefore has the potential for applications in liquid biopsies, as it allowed for accurate total transcriptome profiling of diverse long and short RNAs. Shurtleff et al. [23], used a novel thermostable group II intron reverse transcriptase (TGIRT) sequencing method to investigate the RNA profiles of EVs secreted from human embryonic kidney 293T cells in order to identify the role of YBX1 RNA-binding protein in sorting small non-coding RNAs into EVs which they had previously identified as being necessary for the sorting of miR-223 in the same cells [24]. This method allowed RNA species with post-transcriptionally modified secondary structures to be better reverse transcribed for library preparation. EVs were isolated using a highly selective hybrid method, using UC and sucrose gradient, and appeared to contain mainly full-length mature tRNA transcripts representing 50% of the total RNA reads. Surprisingly, EV miRNAs constituted <0.5% of total reads using this methodology. Additional work to check for 3′ phosphates showed that tRNA fragments were not abundant in these EVs. Their paper also made note of evidence for an exosome-specific tRNA modification involving a tRNA modification at the D-loop (see Figure 1). Finally, to characterise the function of YBX1 in EVs, they conducted sequencing on wild-type cells and YBX1-null cells, finding a 69% decrease in tRNAs in

YBX1-null exosomes and a corresponding 2.3-fold increase in tRNAs in YBX1-null cells, suggesting that YBX1 is a key protein in the sorting of tRNAs into EVs in HEK293T cells.

Therefore, recent focus has now been on how selectively tRNAs are sorted into EVs. A study from 2020 by Gambaro et al. [25] used YBX1-null MCF-7 cells and appeared to contradict the findings of Shurtleff et al. [23] with regard to the importance of YBX1. Synthetic oligonucleotide 5′-tiRNA-Gly was used to show that tRNA-halves could be secreted from the parent cell into EVs in a concentration-dependent manner without the need of specific sequence motifs. 5′-tiRNA-Gly are capable of forming RNase-resistant homodimers and this was previously demonstrated to be important to the process. They were able to also demonstrate cell-to-cell transfection of synthetic 5′-tiRNA-Gly transported in EVs which led to almost two hundred gene alterations seen by the microarray panel. This provides the exciting groundwork for EV-mediated therapeutics to incorporate synthetic oligonucleotides. Kaur et al. [26] investigated EV surface markers as another potential mechanism of RNA sorting to EVs through differences in RNA populations present in EVs with CD63, CD47, and/or MHC1 surface markers and secreted by Jurkat T cells. The authors found that MHC1+ EVs compared to CD63+ and CD47+ EVs, showed a distinct RNA profile with an increased abundance of tRNAs expressed. tRNA-Glu-CTC 1-5 and tRNA-Glu-TTC 3-1 (reported as TRE-CTC and TRE-TRR) were the full-length tRNAs that were specifically present in higher levels in MHC1+ EVs. The majority of EVs captured in the study, however, did not have any of the aforementioned surface proteins, which, therefore, limited the analysis from clearly identifying the role of CD63, CD47, and MHC1 in EV tRNA loading. Nozaki et al. [27] postulated that the nuclear protein Poly ADP-ribose polymerase (PARP) 1 may also have a role in intercellular signalling using EVs as well as its known roles in tumour proliferation, DNA repair, and inflammation. They investigated EVs produced from embryonic stem cells (ES) and found EVs from PARP1 knockout cells had higher levels of tRNAs than EVs from wild-type PARP1 cells (49.1% and 39.5% of the total sRNAs, respectively). No further functional analysis was demonstrated.

2.3.2. EV tRNAs Are Present in Diverse Biofluids

Ferrero et al. [28] analysed the RNA profiles across four different biofluids (blood plasma, urine, cervical scrapes, and stool) from healthy individuals to determine if the RNA profile remained consistent across different samples. However, blood plasma was the only biofluid where EVs were isolated, whilst the other biofluids were subject to RNA isolation methods directly. Here, tRNAs were found to make up a high proportion of the circulating RNA species present in urine (45.1%) compared with other specimens (1–3.3%) including plasma EVs. A study by Miranda et al. [29], published in 2014 also analysed the RNA repertoire present within the microvesicles of urine of an individual healthy male isolated by UC. They found that a large proportion of reads was ascribed to ribosomal RNA (87%), which in some studies is normally filtered, and although a wide variety of species were present including tRNA, snRNAs, small cytoplasmic RNAs, satellite repeats, and others, this comprised less than 6–6.1% of the total number of aligned reads in the sample.

We identified three studies that analysed EV tRNAs in unique biofluids. Bart et al. [30] characterised for the first time the nucleic acid profile associated with EVs from sweat produced under rigorous exercise. EV-associated-tRNAs were measured individually and pooled, finding that tRNAs were found to be the most abundant RNA species found (accounting for over 70% of annotated reads) and corroborated by individual sweat EV analysis. Vojtech et al. [31] was the only study identified characterising the non-coding RNA profile of EVs from semen in order to investigate their potential immunomodulatory functions. Six samples were isolated using UC with a sucrose cushion and sequenced as two libraries based on size (15–40 nt and 40–100 nt) fractionated using a 10% acrylamide-urea gel. Full-length mature tRNAs accounted for 5.93% of the reads from the larger size library whilst tRNA fragments accounted for 16.02% of the small size library. Different tRNA fragment isoacceptor types were also found selectively enriched as full-length or fragments, with tRNA-Glu and tRNA-Asp showing the biggest difference between the two

in proportion. Only nine different 5′-tRNA fragments were tested, but 5′-tRNA-Gly and tRNA-Val predominated. Most tRNA fragments from both libraries were again noted as originating from the 5′-tRNA ends. These profiles appeared to be fairly consistent in the small sample that was tested. Additionally, a study by Cooke et al. [32] characterised the tRNA profile in the EVs isolated from the syncytiotrophoblast (STB) of the placenta. Six samples of placental tissue were collected during elective caesarean section and perfusate was obtained through a collecting system. STB EVs were compared to placental tissue and showed tRNAs representing over 95% of mappable reads, in contrast to just 50% of the reads at a cellular level. 5′-tRNA halves were found to be the predominant tRNA species found in STB EVs, which was validated by qRT-PCR.

One notable final study by Zhao et al. [33] investigated the differences and similarities in the serum exosomes between pooled samples from 4–5 humans, mice, and rats. They found tRNAs and miRNAs were the most abundant RNAs across the three species investigated.

2.4. EV tRNAs Implicated in Disease

2.4.1. Enrichment and Differential Expression of EV tRNAs in Cancer

One of the largest areas for EV tRNA research is in cancer diagnosis and pathogenesis with 19 studies identified in our review. EVs are one way by which malignant cells communicate and signal, thus underpinning several of the critical hallmarks of cancer first detailed by Hanahan et al. [34].

One of the earliest and broadest studies to be identified is Yuan et al. [6], who investigated the EV RNA cargo by sequencing the blood plasma of 50 healthy patients and 142 pan-cancer patients, including those with colon, prostate, and pancreatic cancer. This analysis found a very low abundance of tRNAs, representing just 2.1% of total mappable reads, and a high proportion of miRNA and piwi-interacting RNA (piRNA). No mention was made of differentially expressed tRNA markers. The authors did however note that different methods for vesicle isolation, RNA extraction, sequencing library preparation, or gel size selection all contributed to significant batch effects which could contribute to bias in such a large dataset. Similar EV tRNA proportions were noted by a smaller pan-cancer study in 2021 conducted by Jia et al. [35]. They investigated the difference in extracellular RNA in different plasma fractions. Plasma was obtained from 10 patients with a variety of cancers, namely lung cancer, tongue cancer, prostate cancer, and colon cancer. They used ExoQuick precipitation to generate a series of samples together with a final supernatant. The authors found that tRNAs showed a decrease in proportion, accounting for 4.3% of the RNA content in the "EV" precipitate. However, they did not characterise their EVs, which limits the conclusions we can draw from this. Vergauwen et al. [36] used a combination of SEC and density gradient to isolate EVs from blood plasma obtained from a broad range of patients with ovarian cancer, breast cancer, and HIV. This was in an attempt to define clinical context-dependent variations in the EV RNA profile in cancer patients. tRNAs were found to comprise 7.79% of non-ribosomal mapped reads compared with 7.18% for miRNAs. They also investigated the EV RNA profiles of pooled samples across different time points (several months after therapy or surgery) and found that the tRNA species carried in plasma EVs were indeed time-dependent.

The largest study identified in cancer included over 100 breast cancer patients and was published by Wang et al. [37] in 2020. This involved 176 early-stage breast cancer (BC) patients together with 140 female controls and is notable for its primary focus on the tRNA profile (tRFs and tRNA-halves) found from cell supernatants, plasma exosomes, and tissue samples. This study utilised several stages of sequencing followed by quantitative reverse transcription polymerase chain reaction (qRT-PCR) to narrow down potential candidates with differential expression. Six 5′-tRFs were eventually identified as downregulated in BC (tRF-Glu-CTC-003, tRF-Gly-CCC-007, tRF-Gly-CCC-008, tRF-Leu-CAA-003, tRF-Ser-TGA-001, and tRF-Ser-TGA-002) and this was verified in an external validation phase (samples from another hospital). Whilst this was focussed on extracellular plasma markers, a subset of 40 patients was used to examine the exosomal compartment, and two of these

markers (tRF-Ser-TGA-001 and tRF-Ser-TGA-002) exhibited significantly lower levels in exosomes. This was a less marked difference than circulating plasma expression which indicated that exosomes may not be the source of these particular tRFs. A supplementary paper [38] by the same authors reported on a similar cohort involving 128 breast cancer patients and 116 healthy controls, this time using BC cells compared with epithelial cells (MCF-10A) to validate upregulated tRF biomarkers identified from sequencing data. This study found three tRNA-derived small RNAs (tsRNAs; tRF-Arg-CCT-017, tRF-Gly-CCC-001, and tiRNA-Phe-GAA-003), which were all validated in a similar exosome subset of 40 patients as well as in circulating plasma samples. Target gene analysis for these tsRNAs showed them to be cancer-related involving TGF-β and Wnt-signalling-pathways as well as regulating genes AKT1, MRAS, and WNT4. The validation of these candidates by qRT-PCR was performed in both studies, strengthening the results found. A similar study by Koi et al. [39] investigated 32 BC patients and 20 controls and found a candidate tRF-Lys (TTT) significantly expressed in serum, but not quite significantly different in EVs ($p = 0.62$). There are a number of reasons for the difference in findings, including the method of EV isolation as well as the threshold for candidate markers. Additionally, there were differences in the proportion of breast cancer subtypes and stages of the included patients in each respective study.

Three studies from China regarding gastric cancer were identified initially. Ren et al. [40] sequenced EVs obtained from gastric cancer cell lines and normal gastric mucosal epithelial cell lines. Their library preparation protocols used were only targeted to reliably capture 20–40 nt sequences; thus, whilst they were able to accurately capture tRFs, the presence of full-length mature tRNAs was not evaluated. This study revealed low read counts of tRFs (<10%) across the five cell lines examined. Tang et al. [41] investigated RNA identified by sequencing from a cohort of 36 early-stage gastric cancer patients and healthy volunteers. They also reported a low tRNA abundance with only 0.04% of total mapped reads, but they were sequencing at low depth, obtaining average mappable reads of about 6.5 million per patient. A larger study on gastrointestinal cancers was conducted recently by Ge et al. [7]. The study comprised 82 patients, investigating the difference in exosomal miRNA between patients with malignant and benign biliary obstruction. The malignant group included pancreatic cancer, cholangiocarcinoma, and ampullary carcinoma patients. Interestingly, although the study focussed on exosomal miRNA, it was actually only the second most mapped sRNA at 24%. The proportion of exosomal sRNA mapped to tRNA was in fact the highest, at 27%.

One final paper we investigated in this area was subsequently retracted by the publishers due to similarities to another article regarding tRNAs [42,43].

Various other studies investigating cancers in the gastrointestinal system were identified. Zhu et al. [44] profiled SK-Hep1 liver cancer cells and showed that 5% of exosomal small RNAs were generated from tRNA, which was higher than the 0.2–2% of total small RNAs found in plasma exosomes from an unreported number of liver cancer patients. In both cases and supporting other identified studies, tRNA-5 (5′-tRF) was also the major type of tRNA with four tRNAs significantly higher in the plasma exosomes of liver cancer patients (tRNA-Val-TAC-3, tRNA-Gly-TCC-5, tRNA-Val-AAC-5, and tRNA-Glu-CTC-5). Kumar et al. [45], published a small study which investigated the profiles of mRNA, piRNA, and tRNAs in the exosomes of blood serum of healthy individuals, intraductal papillary mucinous neoplasms (IPMN), and pancreatic ductal adenocarcinoma (PDAC). Using two donors from each sample, they found five tRNAs differentially expressed between IPMN vs. healthy, PDAC vs. healthy, and PDAC vs. IPMN patients, but these were not validated. Human oesophageal cancer was investigated in a study by Liao et al. [46], reporting the class of sRNAs exosomal tRNA reads represented 9.74% and 32.33% of unique reads and total reads, respectively, although individual tRNAs as potential biomarkers were not mentioned. It was also observed that the number of unique reads and total reads for exosomal miRNAs, 0.63% and 1.31%, respectively, was lower than that of exosomal tRNAs.

The increased abundance of specific molecules of tRNAs compared to miRNAs was also noted by Wei et al. [47] in glioblastoma stem cells, highlighting that the tRF Glu-CTC was potentially present in EVs at a 1:1 ratio, whilst the most abundant miRNA molecule (miR-21) was present at a lower 1:10 ratio. Contrastingly, lower levels of tRNAs were found compared to miRNA in the study by Li et al. [48], where they investigated the RNA profile of the plasma EVs from 28 patients with lung adenocarcinomas (ADC), 13 lung squamous cell cancer (SQCC), and 13 healthy controls (CTRL). RNA sequencing analysis revealed SQCC, ADC, and CTRL had EV tRNA fragment proportions of 2.95%, 1.13%, and 1.13%, respectively. These results were validated by qRT-PCR using a larger cohort of patients, including some of the patients from the sequencing stage as well as an additional 16 ADC patients, 18 SQCC patients, and 14 CTRL patients. No further investigation into the different tRNA biotypes present within EVs was conducted.

Three studies relating to HPV-induced cancers were also identified. Tong et al. [49] investigated the RNA profile of EVs in head and neck squamous cell carcinoma (HNSCC) and cervical squamous cell carcinoma (CSCC) cell lines. Overall, exosomes and microvesicles were highly enriched with tRNAs when compared to their source cells (31% and 34% vs. 3%), with nearly 90% of the tRNAs made up of 5′-tRNA-halves and 5′-tRNA fragments. These extracellular sources had similar tRNA compositions but differed with respect to their source cells, with broad classifications of tRNA species indicating Glu-CTC being the most abundant species in exosomes and microvesicles (32.6% and 29.9%, respectively) and Gly-GCC in source cells (27.6%). No differences were found, however, across cancer type or HPV status. A large study by Xi et al. [50] in 2021 further investigated the level of expression of extracellular-vesicle-associated tRFs from the blood plasma samples of 30 HNSCC (termed hypopharyngeal in the paper) patients with no metastasis, 30 patients with lung metastasis, and 30 healthy controls. Differential expression analysis and qPCR validation revealed a particular tRF (tRF-1:30-Lys-CTT-1-M2) to be significantly expressed in the EVs of patients with cancer, and this was additionally associated with tumour stage and grade, as well as lung metastatic risk. In addition, Cho et al. [51] investigated the exosomal RNAs from the plasma of 30 cervical cancer patients who were undergoing chemoradiotherapy and compared them to healthy individuals. They found 474 differentially expressed exosomal tRNAs. However, when they used mRNAs to correlate with the relative changes in non-coding RNA, they found that only 9 of the 474 tRNAs were actually functionally relevant.

EV tRNAs have been identified in different subsets of EVs from malignant melanoma cell lines in two papers by Lunavat et al. [52,53]. Their 2015 paper separated different subsets of EVs namely apoptotic bodies and microvesicles from exosomes using differential centrifugation and analysed their content. They found that exosomes potentially contain relatively more RNAs than apoptotic bodies and microvesicles, but that tRNAs were relatively depleted in exosomes. They specifically found that mitochondrial tRNA (mt-tRNA) was enriched in cells but not in exosomes, but it is unclear if they distinguished this from cytoplasmic tRNA species. In their later study [53], the effects of treatment on malignant melanoma cells harbouring $BRAF^{V600}$ mutations were explored. The authors investigated the effect vemurafenib (BRAF inhibitor) had on the RNA cargo contained within the different subsets of EVs secreted by melanoma cell lines. They once again identified that different subsets of EVs had distinct tRNA enrichment profiles, but this time found that a larger proportion of the reads were mapped as tRNA in the extracellular vesicles compared with cells, with the majority of tRNAs found in microvesicles. Particularly enriched were glutamate and glycine tRNAs, whilst exosomes were enriched with valine tRNAs. Aspartic acid, proline, and alanine tRNAs were depleted in vesicles compared with cells. Interestingly, vemurafenib was not shown to alter the relative expression levels of these tRNAs in the different subsets of EVs. These studies [52,53] therefore noted that treatment did not alter proportions of tRNA found but was not able to shed further light on specific tRNA expression.

2.4.2. EV tRNAs in Respiratory Disease

Four studies were found in this field by Corsello et al. [54], Sundar et al. [55], and Singh et al. [56] investigating how different smoking methods had a differential impact on the EV tRNA content of cells and in blood samples. Corsello et al. [54] investigated the effect cigarette smoke condensate (CSC) had on the RNA profile of EVs derived from small airway epithelial (SAE) cells. The authors employed an Exospin method which incorporates both precipitation and SEC for EV isolation. Average tRNA proportions increased from 9.1% to 12.5% but this did not reach a statistical difference. They were the largest component of identified RNA species, with miRNAs in second place. Sundar et al. [55] also investigated if patients with chronic obstructive pulmonary disease (COPD), compared with smokers and non-smokers, had a different blood EV tRNA profile. Here, tRNAs were the second most common biotype and differential expression of five tRNAs was found between smokers and COPD, and between non-smokers and COPD. Singh et al. [56] was a further paper by the same research group using the same cohort of samples and went on to conduct a separate comparison based on whether they were non-smokers, cigarette smokers, waterpipe smokers, or e-cigarette smokers. An important point to note however is the lack of detail as to whether these tRNAs were fragments or halves.

2.4.3. EV tRNAs in Sepsis

Three studies were identified that investigated EV tRNAs in the blood samples of patients in intensive care (ICU) with sepsis and compared with healthy volunteers. The first was a comprehensive study by Buschmann et al. [57] which compared five different kits (four methods of EV isolation namely precipitation, size exclusion chromatography, membrane affinity, and ultracentrifugation) in combination with two RNA extraction methods to isolate EVs from blood plasma. Crucially, they found differences in RNA subtypes by different methods, with more tRNAs captured from EVs using the membrane affinity method (exoRNeasy Serum-Plasma Midi Kit by Qiagen). However, this was noted to be a relatively low specificity kit for EVs. One important conclusion from this paper was that isolation methods less specific for EVs tended to yield more RNAs, better libraries, and an increase in differential expression, compared with more specific methods which purify EVs rather than circulating cell-free material in general, resulting in less complex libraries. The proportion of tRNA yield could therefore be a surrogate measure for specificity (or lack thereof) of EV capture. Another study in this field was by Hermann et al. [58], who found that the majority of the expression profiles of miRNA and tRNAs matched between arterial and venous sera in 20 cardiac surgery patients. They assessed EVs by precipitation (termed crude) and additionally purified these EVs by SEC, thus suggesting that RNAs including tRNAs obtained from either a vein or artery should be suitable for the detection of EV biomarkers. Hermann et al. [59] was able to then conduct a large study including 30 patients with community-acquired pneumonia, 65 with sepsis, and 47 healthy individuals, using precipitation methodology (which had been noted to reliably separate sepsis from healthy controls) to investigate RNA biomarkers associated with EVs. In this paper, tRNAs were found to be less abundant than miRNA species across all three patient groups, with a large proportion of unmapped reads. Unfortunately, they did not further investigate potential tRNA markers in this well-conducted study.

2.4.4. EV tRNAs in the Urine of Patients with Renal Disease

Renal diseases represent an important avenue where EVs could act as diagnostic markers, and EVs isolated from the urine are being evaluated as a non-invasive metabolic snapshot to assess pathological renal cells. We highlight two small in vivo studies in this review investigating urinary EVs from patients with chronic kidney disease (CKD) and idiopathic membranous nephropathy (IMN).

Khurana et al. [5] compared the urinary exosomes obtained by UC from 15 patients with CKD compared with 10 healthy volunteers and identified two tRFs significantly decreased (5′-tRF^{Val} and 5′-tRF^{Leu}, displaying a size of about 32–36 nt, i.e., halves). Interest-

ingly, the authors used cell culture and Northern blotting to verify that these tRFs were present in a high abundance in the exosomes released by renal cells. In addition, mitochondrial tRNAs (mt-tRNA; tRNA-Cys, tRNA-Pro, and tRNA-Glu, sized about 60–68 nt) were also found to be significantly decreased in the urinary exosomes but this was not further investigated. Zhang et al. [60] was one of the few research groups to use an immunoaffinity precipitation method to isolate EVs based on affinity to phosphatidylserine (Wako Pure Chemical Industries, Osaka, Japan) and investigated the proportion of different small RNAs found from the urinary exosomes of idiopathic membranous nephropathy (IMN) in a small group of six patients and five controls. They found that small RNAs derived from tRNAs were the most prevalent RNA species found within the urinary EVs of IMN patients. The authors however did not perform a separate experiment to validate the source and origin of the RNA, nor did they identify specific candidates.

2.4.5. EV tRNAs in Metabolic Bone Disease

Degeneration of the bone and osteoblastogenesis are closely linked; thus, perturbations in these biological processes have important implications for bone health. Three EV studies were conducted in this field, where two studies conducted in vivo experiments for the investigation of bone degeneration whilst the other study conducted in vitro experiments on osteoblastogenesis, all reporting changes in the tRNA levels of exosomes as one of their key observations.

A large study of 80 patients conducted by Zhang et al. [43], using the ExoQuick kit, found potential diagnostic markers for osteoporosis by investigating the differences in exosomal tRNA fragments found in the blood plasma of osteoporosis patients compared to healthy individuals. tRF-25, tRF-38 and tRF-18 derived from plasma exosomes showed particular promise as their relative expression in osteoporosis patients, as validated and confirmed by qRT-PCR, was significantly higher than in healthy controls. Importantly, the tRF panel made up of these three tRFs underwent receiver operating characteristic (ROC) curve analysis and showed a remarkable average area under the curve (AUC) of 0.815, which is higher than other known miRNA biomarkers used for osteoporosis. Another degenerative bone disease where exosomal tRNAs showed significant changes in levels was identified in patients suffering from steroid-induced osteonecrosis of the femoral head (SONFH). Fang et al. [61] compared the expression profile of exosomal tsRNA in a small number (n = 10) of SONFH patients with healthy controls, and further sought to characterise the effect bone marrow-derived stem cell (BMSC) exosomes had on the pathogenesis of SONFH. The authors found three exosomal tsRNAs expressed at significantly lower levels in SONFH patients, validated by qRT-PCR. Importantly, their target genes were all implicated in the Wnt-signalling pathway and osteogenic differentiation. The authors further focussed on tsRNA-10277, as it had the largest fold change and found tsRNA-10277 containing BMSC exosomes had beneficial effects by decreasing adipogenesis in SONFH cells and promoting osteogenesis, directly contrasting the effect dexamethasone had on BMSC exosomes in SONFH patients. Confusingly, the authors have used an idiosyncratic nomenclature which does not correspond to any known tRF.

Mesenchymal stem cells (MSCs) are a common source of EVs for study due to their ability to be cultured easily as well as multi-lineage differentiation, giving rise to a variety of mesenchymal phenotypes such as osteoblasts (bone), adipocytes (fat), and chondrocytes (cartilage). A study by Baglio et al. [62] compared a number of primary BMSCs and adipose stem cells (ASCs) from elective plastic surgical samples to assess the EV RNA profiles released during culture. Sequencing analysis showed that whilst ASC exosomes showed a major RNA profile corresponding to tRNA-halves, BMSC exosomes showed two distinct RNA profiles. Further investigation into the genes involved in pluripotency revealed that the two RNA profiles observed in BMSC exosomes were each related to the different BMSC differentiation stages, with more differentiated BMSCs associated with a predominant tRNA-half profile (much like ASCs), and less differentiated BMSCs correlating with a profile of full-length tRNAs and 33-nt fragments. In addition, by comparing cells and their

corresponding EVs, they were able to highlight a weak correlation and therefore selective incorporation of RNA species into EVs. EV tRNAs accounted for 5–11% of RNA content in cells but represented >50% of total small RNAs in ASC EVs and 23–35% in BMSC EVs. The high abundance of tRNAs found in the Evs thus prompted further investigation into their subtypes, where 5′-tRNA halves were found to be the predominant species. EVs in BMSCs were also explored by Yan et al. [63] as part of a study into osteoblastogenesis, using UC to first isolate the EVs. Subsequent RNA sequencing was used to identify the miRNA and tRNA profiles from the Evs derived from ASCs and BMSCs, which were then compared with parental cells prior to differentiation. Similar results to the study by Baglio et al. were found, where although tRNA reads in the BMSC EVs decreased from 52% to 14%, from Day 0 to Day 7 of osteoblastic differentiation, this was not noted in ASC EVs. The striking fall in tRNA expression levels was found to be mainly contributed by a decrease in 5′-tRNA halves (Gly-CCC, Gly-GCC, and His-GTG). In addition, three tRNA fragments, which were all 5′-tRNA halves, constituted over half the tRNA reads in each cohort, indicating a degree of heterogeneity. Although this was a small study, it appeared to show that alterations in this pattern of expression were seen during cell-specific osteoblastic differentiation. Another study conducted by Kaur et al. [64] aimed to characterise the RNA profile of the EVs from adipose tissue mesenchymal stem cells (AT-MSCs) and human pluripotent stem cells. Interestingly, the isolation of EVs was performed via a combination of UC and miRCURY Exosome Isolation Kit (Exiqon now QIAGEN). The authors subsequently found that whilst most EV reads were attributed to rRNAs, most of the sRNA reads in hPSCs and AT-MSCs were still nevertheless attributed to tRNAs at 69% and 47%, respectively. The selectivity of AT-MSCs for tRNAs shown here is again in line with the findings of Baglio et al.

2.4.6. EV tRNAs in Infectious Disease

Human immunodeficiency virus (HIV) infection, tuberculosis, and malaria rank consistently among the top few infectious diseases which cause the greatest burden of disease worldwide. Interestingly, three of the four studies identified here were on these three diseases.

Chettimada et al. [65] investigated and compared the plasma EV RNA profile of 20 HIV-positive patients with 20 HIV-negative controls. Notably, the cohort of HIV-positive patients was on antiretroviral therapy and had viral loads <1000 HIV RNA copies/mL. They noted that the number of unique mapped reads corresponding to tRNAs were equal between HIV-positive patients and uninfected individuals but did not comment on the differential expression of particular tRNAs. Another infectious disease closely associated with HIV infection is tuberculosis. Pawar et al. [66] investigated the role of 5′ tRNA halves during infection of human monocyte-derived macrophages (HMDMs) using Mycobacterium bovis (BCG). EVs were isolated via ultracentrifugation and tRNAs were validated via qRT-PCR. The authors found that 5′-tRNA-His-GUG halves were in particular selectively packaged into EVs of HMDMs and present to a high degree, approximately 210-fold higher than the most abundant miRNA (miR-150) found. 5′-tRNA-His-GUG halves were found to stimulate Toll-like receptor 7 (TLR7) pathways significantly but were also highly specific for TLR7. In addition, 5′-tRNA-His-GUG halves and 5′-tRNA-Glu-CUC halves were found to be upregulated by TLR-activated NF-κB, where increased expression of ANG mRNA led to the increased cleavage of tRNAs. Furthermore, a feed-forward loop was suggested where the tRNA-halves could potentially activate TLR7 pathways, also leading to NF-κB activation and, thus, further tRNA-half production. Importantly, as NF-κB is linked to many inflammatory diseases and malignant conditions; this could, in turn, suggest the involvement of tRNA-halves found within EVs in such diseases. Malaria is another disease which is challenging to treat and requires long term management. Sisquella et al. [67] investigated the RNA profile from the EVs secreted by P. falciparum-infected red blood cells to determine whether EVs contained parasitic small RNA. They found a significant increase in RNA content after infection and distinct tRNA peaks when quantified by RNA Bioanalyser.

In addition, a study by He et al. [68] investigated the microbial–host relation in the oral cavity by exploring the tRNAs contained within EVs in the saliva and their effects on the growth of oral bacteria, particularly those that had high sequence similarity to tRNAs from a group of Gram-negative oral bacteria. The authors found that these 5′-tRNA halves (tsRNA-000794 and tsRNA-020498) in salivary exosomes inhibited growth and protein synthesis of Fusobacterium nucleatum, a key oral commensal and opportunistic pathogen. These EV tRNA fragments in turn were increased in salivary EVs in response to F. nucleatum. This study, therefore, suggests a possible mechanism evolved by the oral microbiota to suppress the growth of Gram-negative organisms to maintain oral health through microbial–host EV-mediated homeostasis.

2.4.7. EV tRNAs in Diseases of the Reproductive Tract

Two studies have been identified in this review where tRNAs in EVs play potential roles in the diseases of the reproductive tract. Zhang et al. [69] reported the use of exosomal tRFs to predict whether microdissection testicular sperm extraction (Micro-TESE) could be used to predict success as a sperm retrieval procedure for patients suffering from non-obstructive azoospermia (NOA). Differential expression of EV tRFs was analysed from the blood plasma of twelve patients with successful micro-TESE against eighteen patients where no sperm was retrieved. This initial screen for differentially expressed tRFs yielded three potential candidates. However, after qPCR validation, only two tRFs (tRF-Gly-GCC-002 and tRF-Glu-CTC-005) were found to reach statistical significance. ROC curve analysis was performed and revealed that tRF-Gly-GCC-002 and tRF-Glu-CTC-005 could predict NOA with successful sperm retrieval with an AUC of 0.921 and 0.954, respectively, indicating high predictive accuracy. The other study grouped into this category was an investigation into the role of human follicular fluid EVs in polycystic ovarian syndrome (PCOS) by Hu et al. [70]. Four PCOS patients and two healthy patients were included in the study. With a fold change >1, ten tRNAs were found upregulated in PCOS patients whilst another ten tRNAs were significantly downregulated. The authors, however, did not further investigate these findings as only miRNAs were the focus of the study.

2.4.8. EV tRNAs in Autoimmune Disease

Four studies reviewed EVs in autoimmune diseases including systemic lupus erythematosus (SLE), atopic dermatitis in the young, and dermatomyositis. Yang et al. [71] was a biomarker study performed on a very large sample of patients including 192 SLE patients and 109 normal controls to examine the difference in serum tsRNA profile between SLE patients with and without lupus nephritis and healthy controls. They identified tRF-His-GTG-1 as the most significantly upregulated in the serum, between SLE patients without lupus nephritis and control patients. They then conducted a separate experiment on an unspecified number of samples to show that this marker tRF-His-GTG-1 was also present in exosomes ($p < 0.0001$). Dou et al. [72] revealed the potential role of exosomal tRNAs as a way of inhibiting the pro-inflammatory process mediated by macrophages commonly seen in SLE. In order to investigate the effect MSC exosomes had on macrophages, the authors compared the tRNA expression profile of macrophages cultured with MSC exosomes with macrophages which were not. They found tsRNA-21109 as one of the prominent tRNAs differentially upregulated. They then investigated macrophage activation, using MSC exosomes which either contained or did not contain tsRNA-21109, and found MSC exosomes containing tsRNA-21109 inhibited the pro-inflammatory activation of macrophages and thus could represent a potential target for SLE treatment.

Meng et al. [73] for the first time attempted to elucidate the role of exosomal tRFs in paediatric patients with atopic dermatitis. Using plasma samples from three atopic dermatitis patients and three healthy controls, the authors revealed an impressive 135 exosomal tRFs which are differentially expressed. Of these, two select tRFs were chosen to be validated via qRT-PCR in 20 patients and 20 controls showing significant downregulation in patients compared to controls. These two tRFs then underwent ROC curve

analysis and, in particular, tRF-28-QSZ34KRQ590K showed a remarkable AUC of 0.900 for diagnosing atopic dermatitis. However, tRF-28-QSZ34KRQ590K did not show any significant correlation with the disease severity of atopic dermatitis. Another autoimmune condition was examined by Zhong et al. [74], where the exosomal RNA profile from the blood samples of 10 dermatomyositis patients, with half of the patients having simultaneous interstitial lung disease (ILD) involvement, was investigated. In all three patient groups, namely dermatomyositis patients with ILD and without ILD and control patients, exosomal tRNA represented one of the smallest proportions of sRNA, with 0.42%, 0.49% and 0.64%, respectively.

2.4.9. EV tRNAs in Other Diseases

Two studies finally highlight the diversity of EV tRNA in other diseases. Lee et al. [75] found that in patients suffering from Post-Traumatic Stress Disorder (PTSD), all patient samples contained EVs with significantly higher proportions ($p < 1 \times 10^{-4}$) of tRNAs compared to EV depleted plasma and plasma. Fitz et al. [76] investigated the potential of using plasma EV associated RNAs to differentiate between Alzheimer's disease (AD) patients and normal controls who have no dementia. tRNAs were found to contribute only 4.7% of the genes mapped. Small nucleolar RNAs on the other hand were found to have significant discriminatory power for identifying AD patients.

3. Discussion

In our review, we have focussed on the tRNAs, tRNA-halves, and tRFs associated with human EVs, in health and pathological disease states. We have described the various biofluids used for EV capture and characterised the different EV isolation methods used in protocols for the subsequent identification of EV tRNAs contained within. Importantly, we have shown that multiple pathologies and especially malignancies can be distinguished by differential EV tRNA expression and may thus be of clinical significance for both diagnosis and prognosis of the disease.

The abundance of tRNAs contained within EVs is highly variable and may depend on tissue and disease type, ranging from as low as 0.04% in the blood plasma EVs of gastric cancer patients [41] to as high as >95% from EVs obtained from healthy placental syncytiotrophoblasts [32]. However, tRNAs appear to be a consistent finding in EVs, occurring in sequencing samples with as high a prevalence (and sometimes higher) as miRNAs. Additionally, EV isolation methods will need to be tailored to the requirements of the experiment, with highly selective methods such as density gradient preferable when contamination with RBP is to be avoided. We found that 46.5% of the protocols used precipitation as their EV isolation method, and this has been recognised as having relatively low purity [77,78]. Moreover, the technological limitations of sequencing methods may further add to the complexity of interpreting the EV tRNAs found. Our review highlighted a few limitations of current sequencing technologies. In many studies, the library preparation is limited to specific size ranges [40] and low abundance species can lead to selective biases in enrichment. [21] Moreover, conventional sequencing methods may not effectively capture RNAs that are post-transcriptionally modified [23]. Finally, Tosar et al. [79] highlighted how piRNA transcripts differ from some tRNA-halves by only one nucleotide. Individual piRNA and as well as other small cytoplasmic RNA species are highly homologous to major tRNA fragments within commonly used databases; thus, mapping of reads may lead to an under- or over-representation of tRNA fragments. The limitations encountered with sequencing technologies could be addressed through novel pipelines such as that reported by Amorim et al. [22]. Ongoing collaborative work such as the Extracellular RNA Communication Consortium has made efforts to improve this and work towards strategies such as deconvolution to determine tRNA sources [80].

It is increasingly recognised that tRNA-derived fragments have a role in mammalian cells, with 5′-tRNA-halves having been associated with actions on the ribosome leading to stress granule formation [81] and translational inhibition. Smaller tRFs share a size

similarity to miRNAs and there is significant evidence in the wider (non-EV) literature that they can have actions in conjunction with Argonaute [82], the Piwi subfamily, and other RNA-binding proteins [82,83].

tRNAs and their fragmentation products have been found as an EV biomarker in circulating fluid and tissue for a number of malignancies including leukaemia, prostate, ovarian, pancreatic, and colorectal cancers [84,85]. In our study, breast cancer showed particular promise with several differentially expressed tRNAs (tRF-Arg-CCT-017, tRF-Gly-CCC-001, and tiRNA-Phe-GAA-003) associated with key regulatory pathways such as Wnt signalling, classically involved in cancer initiation and progression. [86] In fact, several other malignancies such as colorectal cancers [87], gastric cancers [88], and glioblastomas [89] also secrete cancer-stem-cell-associated EVs, which act on the β-catenin/Wnt-signalling pathways to increase stemness, in turn increasing their tumorigenic potential.

In addition, the majority of the studies in our review showed that the fragmentation products of tRNAs, namely 5′-tRFs and 5′ tRNA-halves, were the most abundant subspecies of tRNAs identified in EVs. A possible reason for the observation of 5′ products was shown to be potentially due to reverse transcription failure due to post-transcriptional modifications in otherwise full-length tRNAs by Shurtleff et al. who used an alternative thermostable sequencing system. What is also clear is that simple analysis of tRNA molecules by broad amino acid (isoacceptor types) may not be sophisticated enough to determine potentially clinically relevant changes in EV tRNA profiles. Further validation of sequencing by qRT-PCR is necessary to prove true differential expression and this means additionally incorporating the standardised use of tRNA molecular terminology.

We have also highlighted that several biological mechanisms have been clearly associated with changes in tRNA-halves levels. For instance, osteoblastogenesis revealed that BMSCs alter their cargo content in EVs from full-length tRNAs to tRNA-halves as their differentiation progresses [62]. We have noted the evidence that there are time-dependent changes in the EV tRNA profile of cancer patients and may perhaps reflect clinical changes such as disease burden [36]. In addition, tRFs have even been shown to possess potential diagnostic potential in a multitude of diseases.

tRNA-sorting mechanisms in EVs have also been the focus of various studies, with RNA-binding proteins [23], surface markers [26], and tRNA enzymatic stability as some of the mechanisms identified [25]. Potential therapeutic uses of EV tRNA include the loading of synthetic oligonucleotides into EVs via parent cells to alter downstream gene expression in donor cells [25]. This particular paper by Gambaro et al. [25] demonstrated key mechanisms for tRNA sorting, such as the concentration-dependent transfection and selection of stable synthetic oligonucleotides against enzyme degradation, through well-controlled transfection experiments.

Due to the heterogeneity of the studies included in our review, a meta-analysis could not be conducted; thus, a narrative synthesis of the literature was undertaken. Many studies may include tRNA data but not explicitly mention this and will have been excluded. We also restricted our search to articles involving human cell models or patient-derived biofluids which may exclude important studies using animal EV models that could have shed some light on tRNA markers across other species.

Further Work and Outlook

Our review shows that EV tRNAs are present in various healthy and diseased cells and tissues and confirms that tRNAs and their fragmentation products are one of the most abundant RNA species expressed in EVs and cells after miRNAs. Coupled with their presence in diverse biofluids, we believe this represents a clinical significance from an untapped source of potential markers. There is a vital need for future studies to determine the origin and loading of EV tRNAs, as well as the proteins or nucleases involved in their fragmentation and modification. In the future, we see combinations of tRNAs with other RNA species (such as miRNA) found from EVs will be more likely to yield better diagnostic and prognostic accuracy. Their differential expression presents an opportunity for further

clinical research where EV-associated tRFs are understood as a mechanism of cellular stress and signalling. It is in this context that future protocols must include size selection, as the question of whether EV tRNAs exist as fragmentation products or mature tRNAs is yet to be determined. Finally, collaborative work through online repositories of EV RNA sequencing data will ensure that future biomarkers remain available for data scientists and bioinformaticians. Some databases to manage and discover tRFs, such as trfdb (http://genome.bioch.virginia.edu/trfdb), tRFun (https://rna.sysu.edu.cn/tsRFun/), tRFexplorer (https://trfexplorer.cloud/), and MINTbase v2.0 (http://cm.jefferson.edu/MINTbase/) have arisen to promote this effort.

4. Materials and Methods

4.1. Objectives

Our primary aim was to perform a systematic review to identify all studies involving tRNA cargo present associated with human-derived EVs isolated in vivo or in vitro. Our objectives included defining the EV source, EV characteristics, and results of these studies in order to determine the most up-to-date knowledge on the functional and clinical significance of EV tRNA cargo.

4.2. Search Strategy

Electronic bibliographic databases MEDLINE (Medical Literature Analysis and Retrieval System Online) and EMBASE (Excerpta Medica Database) were searched according to a systematic pre-specified search strategy which included terms such as "extracellular vesicles", "microvesicles", "exosomes", "transfer ribonucleic acid", and "tRNA fragments" (see Appendix A). An iterative search involving bibliographies of the included articles was also manually searched to identify any additional relevant articles not found using the initial search strategy. The last search was conducted on 1 January 2022.

4.3. Screening and Selection Criteria

All articles identified from the search strategy were uploaded onto an online collaborative tool Covidence (Melbourne, Australia) and de-duplication was undertaken automatically. Title and abstract screening and full-text screening was conducted by (Q.Z.C.Y) and (D.L) independently and cross-referenced for accuracy. Any disagreements were reviewed and resolved after discussion. The screening process is shown in the PRISMA flowchart (Figure 3).

Papers that were included for data extraction included original research articles which investigated RNA sequencing data from EVs derived from human biofluids, human pathogens, or human cell lines, either healthy or pathological. Conference abstracts, poster presentations, review papers, and research articles that solely investigated EVs of non-human origin were excluded as well as bioinformatics papers that drew on prior datasets characterising or identifying tRNAs in EVs.

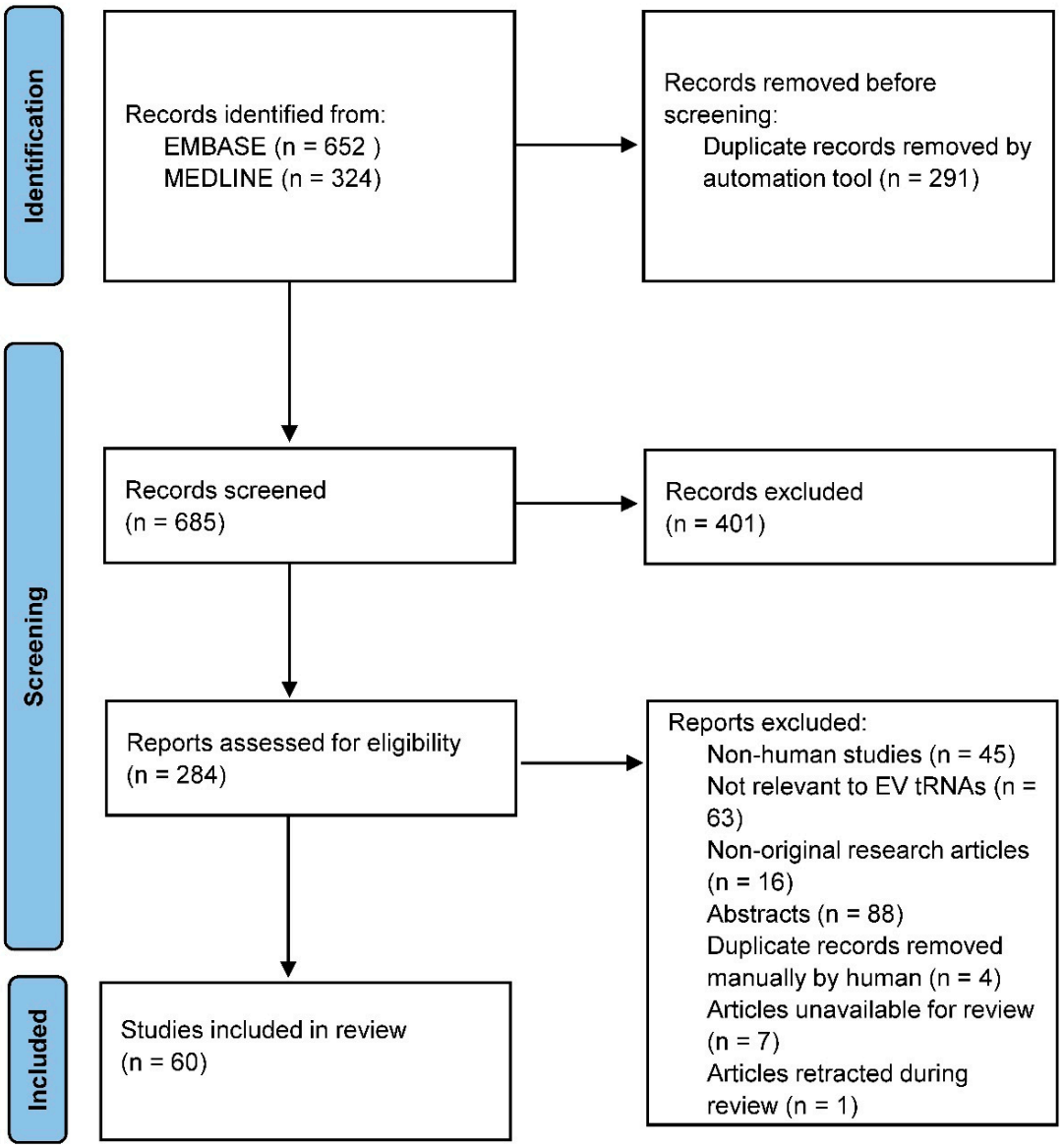


Figure 3. PRISMA flowchart showing the screening process of included articles.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms23073692/s1.

Author Contributions: Conceptualization, Q.Z.C.Y. and D.S.K.L.; formal analysis and systematic review, Q.Z.C.Y. and D.S.K.L.; writing—original draft preparation, Q.Z.C.Y. and D.S.K.L.; writing—review and editing, Q.Z.C.Y., D.S.K.L., M.A., J.K. and A.E.F.; figure creation, Q.Z.C.Y. and D.S.K.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: Figure 1 was created with BioRender.com.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A

MEDLINE (Last Search on 1 January 2022)

1. exp Extracellular Vesicles/
2. extracellular vesic*.ti,ab.
3. microvesic*.ti,ab.
4. exosom*.ti,ab.
5. microparticle*.ti,ab.
6. oncosom*.ti,ab.
7. cardiosom*.ti,ab.
8. prostatosom*.ti,ab.
9. tolerosom*.ti,ab.
10. vexosom*.ti,ab.
11. ectosom*.ti,ab.
12. vesicular.ti,ab.
13. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
14. exp RNA, Transfer/
15. tRNA*.ti,ab.
16. transfer RNA*.ti,ab.
17. transfer ribonucleic acid*.ti,ab.
18. tRF*.ti,ab.
19. tRNA derived fragment*.ti,ab.
20. 14 or 15 or 16 or 17 or 18 or 19
21. 13 and 20

EMBASE (Last Search on 1 January 2022)

1. exp exosome/
2. extracellular vesic*.ti,ab.
3. microvesic*.ti,ab.
4. exosom*.ti,ab.
5. microparticle*.ti,ab.
6. oncosom*.ti,ab.
7. cardiosom*.ti,ab.
8. prostatosom*.ti,ab.
9. tolerosom*.ti,ab.
10. vexosom*.ti,ab.
11. ectosom*.ti,ab.
12. vesicular.ti,ab.
13. 1 or 2 or 3 or 4 or 5 or 6 or 7 or 8 or 9 or 10 or 11 or 12
14. exp transfer RNA/
15. tRNA*.ti,ab.
16. transfer RNA*.ti,ab.
17. transfer ribonucleic acid*.ti,ab.
18. tRF*.ti,ab.
19. tRNA derived fragment*.ti,ab.
20. 14 or 15 or 16 or 17 or 18 or 19
21. 13 and 20

References

1. Andaloussi, S.E.L.; Mager, I.; Breakefield, X.O.; Wood, M.J. Extracellular vesicles: Biology and emerging therapeutic opportunities. *Nat. Rev. Drug Discov.* **2013**, *12*, 347–357. [CrossRef] [PubMed]
2. van Niel, G.; D'Angelo, G.; Raposo, G. Shedding light on the cell biology of extracellular vesicles. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 213–228. [CrossRef] [PubMed]
3. Moller, A.; Lobb, R.J. The evolving translational potential of small extracellular vesicles in cancer. *Nat. Rev. Cancer* **2020**, *20*, 697–709. [CrossRef]

4. Welton, J.L.; Loveless, S.; Stone, T.; von Ruhland, C.; Robertson, N.P.; Clayton, A. Cerebrospinal fluid extracellular vesicle enrichment for protein biomarker discovery in neurological disease; multiple sclerosis. *J. Extracell. Vesicles* **2017**, *6*, 1369805. [CrossRef] [PubMed]
5. Khurana, R.; Ranches, G.; Schafferer, S.; Lukasser, M.; Rudnicki, M.; Mayer, G.; Huttenhofer, A. Identification of urinary exosomal noncoding RNAs as novel biomarkers in chronic kidney disease. *RNA* **2017**, *23*, 142–152. [CrossRef] [PubMed]
6. Yuan, T.; Huang, X.; Woodcock, M.; Du, M.; Dittmar, R.; Wang, Y.; Tsai, S.; Kohli, M.; Boardman, L.; Patel, T.; et al. Plasma extracellular RNA profiles in healthy and cancer patients. *Sci. Rep.* **2016**, *6*, 19413. [CrossRef]
7. Ge, X.; Tang, L.; Wang, Y.; Wang, N.; Zhou, J.; Deng, X.; Zhong, Y.; Li, Q.; Wang, F.; Jiang, G.; et al. The diagnostic value of exosomal miRNAs in human bile of malignant biliary obstructions. *Dig. Liver Dis.* **2021**, *53*, 760–765. [CrossRef]
8. Admyre, C.; Johansson, S.M.; Qazi, K.R.; Filen, J.J.; Lahesmaa, R.; Norman, M.; Neve, E.P.; Scheynius, A.; Gabrielsson, S. Exosomes with immune modulatory features are present in human breast milk. *J. Immunol.* **2007**, *179*, 1969–1978. [CrossRef]
9. Raposo, G.; Stoorvogel, W. Extracellular vesicles: Exosomes, microvesicles, and friends. *J. Cell Biol.* **2013**, *200*, 373–383. [CrossRef]
10. Torres, A.G.; Marti, E. Toward an Understanding of Extracellular tRNA Biology. *Front. Mol. Biosci.* **2021**, *8*, 662620. [CrossRef]
11. Schimmel, P. The emerging complexity of the tRNA world: Mammalian tRNAs beyond protein synthesis. *Nat. Rev. Mol. Cell Biol.* **2018**, *19*, 45–58. [CrossRef] [PubMed]
12. Holley, R.W.; Apgar, J.; Everett, G.A.; Madison, J.T.; Marquisee, M.; Merrill, S.H.; Penswick, J.R.; Zamir, A. Structure of a Ribonucleic Acid. *Science* **1965**, *147*, 1462–1465. [CrossRef] [PubMed]
13. Novoa, E.M.; Pavon-Eternod, M.; Pan, T.; Ribas de Pouplana, L. A role for tRNA modifications in genome structure and codon usage. *Cell* **2012**, *149*, 202–213. [CrossRef] [PubMed]
14. Yamasaki, S.; Ivanov, P.; Hu, G.F.; Anderson, P. Angiogenin cleaves tRNA and promotes stress-induced translational repression. *J. Cell Biol.* **2009**, *185*, 35–42. [CrossRef]
15. Fujishima, K.; Sugahara, J.; Kikuta, K.; Hirano, R.; Sato, A.; Tomita, M.; Kanai, A. Tri-split tRNA is a transfer RNA made from 3 transcripts that provides insight into the evolution of fragmented tRNAs in archaea. *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 2683–2687. [CrossRef]
16. Gebetsberger, J.; Polacek, N. Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol.* **2013**, *10*, 1798–1806. [CrossRef]
17. Srinivasan, S.; Yeri, A.; Cheah, P.S.; Chung, A.; Danielson, K.; De Hoff, P.; Filant, J.; Laurent, C.D.; Laurent, L.D.; Magee, R.; et al. Small RNA Sequencing across Diverse Biofluids Identifies Optimal Methods for exRNA Isolation. *Cell* **2019**, *177*, 446. [CrossRef]
18. Lasser, C.; Shelke, G.V.; Yeri, A.; Kim, D.K.; Crescitelli, R.; Raimondo, S.; Sjostrand, M.; Gho, Y.S.; Van Keuren Jensen, K.; Lotvall, J. Two distinct extracellular RNA signatures released by a single cell type identified by microarray and next-generation sequencing. *RNA Biol.* **2017**, *14*, 58–72. [CrossRef]
19. Tosar, J.P.; Gambaro, F.; Sanguinetti, J.; Bonilla, B.; Witwer, K.W.; Cayota, A. Assessment of small RNA sorting into different extracellular fractions revealed by high-throughput sequencing of breast cell lines. *Nucleic Acids Res.* **2015**, *43*, 5601–5616. [CrossRef]
20. Sork, H.; Corso, G.; Krjutskov, K.; Johansson, H.J.; Nordin, J.Z.; Wiklander, O.P.B.; Lee, Y.X.F.; Westholm, J.O.; Lehtio, J.; Wood, M.J.A.; et al. Heterogeneity and interplay of the extracellular vesicle small RNA transcriptome and proteome. *Sci. Rep.* **2018**, *8*, 10813. [CrossRef]
21. Huang, X.; Yuan, T.; Tschannen, M.; Sun, Z.; Jacob, H.; Du, M.; Liang, M.; Dittmar, R.L.; Liu, Y.; Liang, M.; et al. Characterization of human plasma-derived exosomal RNAs by deep sequencing. *BMC Genom.* **2013**, *14*, 319. [CrossRef] [PubMed]
22. Amorim, M.G.; Valieris, R.; Drummond, R.D.; Pizzi, M.P.; Freitas, V.M.; Sinigaglia-Coimbra, R.; Calin, G.A.; Pasqualini, R.; Arap, W.; Silva, I.T.; et al. A total transcriptome profiling method for plasma-derived extracellular vesicles: Applications for liquid biopsies. *Sci. Rep.* **2017**, *7*, 14395. [CrossRef] [PubMed]
23. Shurtleff, M.J.; Yao, J.; Qin, Y.; Nottingham, R.M.; Temoche-Diaz, M.M.; Schekman, R.; Lambowitz, A.M. Broad role for YBX1 in defining the small noncoding RNA composition of exosomes. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E8987–E8995. [CrossRef] [PubMed]
24. Shurtleff, M.J.; Temoche-Diaz, M.M.; Karfilis, K.V.; Ri, S.; Schekman, R. Y-box protein 1 is required to sort microRNAs into exosomes in cells and in a cell-free reaction. *eLife* **2016**, *5*, e19276. [CrossRef]
25. Gambaro, F.; Calzi, M.L.; Fagundez, P.; Costa, B.; Greif, G.; Mallick, E.; Lyons, S.; Ivanov, P.; Witwer, K.; Cayota, A.; et al. Stable tRNA halves can be sorted into extracellular vesicles and delivered to recipient cells in a concentration-dependent manner. *RNA Biol.* **2020**, *17*, 1168–1182. [CrossRef]
26. Kaur, S.; Elkahloun, A.G.; Arakelyan, A.; Young, L.; Myers, T.G.; Otaizo-Carrasquero, F.; Wu, W.; Margolis, L.; Roberts, D.D. CD63, MHC class 1, and CD47 identify subsets of extracellular vesicles containing distinct populations of noncoding RNAs. *Sci. Rep.* **2018**, *8*, 2577. [CrossRef]
27. Nozaki, T.; Sasaki, Y.; Fukuda, I.; Isumi, M.; Nakamoto, K.; Onodera, T.; Masutani, M. Next-generation sequencing-based miRNA expression analysis in Parp1-deficient embryonic stem cell-derived exosomes. *Biochem. Biophys. Res. Commun.* **2018**, *499*, 410–415. [CrossRef]
28. Ferrero, G.; Cordero, F.; Tarallo, S.; Arigoni, M.; Riccardo, F.; Gallo, G.; Ronco, G.; Allasia, M.; Kulkarni, N.; Matullo, G.; et al. Small non-coding RNA profiling in human biofluids and surrogate tissues from healthy individuals: Description of the diverse and most represented species. *Oncotarget* **2018**, *9*, 3097–3111. [CrossRef]

29. Miranda, K.C.; Bond, D.T.; Levin, J.Z.; Adiconis, X.; Sivachenko, A.; Russ, C.; Brown, D.; Nusbaum, C.; Russo, L.M. Massively parallel sequencing of human urinary exosome/microvesicle RNA reveals a predominance of non-coding RNA. *PLoS ONE* **2014**, *9*, e96094. [CrossRef]
30. Bart, G.; Fischer, D.; Samoylenko, A.; Zhyvolozhnyi, A.; Stehantsev, P.; Miinalainen, I.; Kaakinen, M.; Nurmi, T.; Singh, P.; Kosamo, S.; et al. Characterization of nucleic acids from extracellular vesicle-enriched human sweat. *BMC Genom.* **2021**, *22*, 425. [CrossRef]
31. Vojtech, L.; Woo, S.; Hughes, S.; Levy, C.; Ballweber, L.; Sauteraud, R.P.; Strobl, J.; Westerberg, K.; Gottardo, R.; Tewari, M.; et al. Exosomes in human semen carry a distinctive repertoire of small non-coding RNAs with potential regulatory functions. *Nucleic Acids Res.* **2014**, *42*, 7290–7304. [CrossRef] [PubMed]
32. Cooke, W.R.; Cribbs, A.; Zhang, W.; Kandzija, N.; Motta-Mejia, C.; Dombi, E.; Ri, R.; Cerdeira, A.S.; Redman, C.; Vatish, M. Maternal circulating syncytiotrophoblast-derived extracellular vesicles contain biologically active 5′-tRNA halves. *Biochem. Biophys. Res. Commun.* **2019**, *518*, 107–113. [CrossRef] [PubMed]
33. Zhao, F.; Cheng, L.; Shao, Q.; Chen, Z.; Lv, X.; Li, J.; He, L.; Sun, Y.; Ji, Q.; Lu, P.; et al. Characterization of serum small extracellular vesicles and their small RNA contents across humans, rats, and mice. *Sci. Rep.* **2020**, *10*, 4197. [CrossRef] [PubMed]
34. Hanahan, D.; Weinberg, R.A. The hallmarks of cancer. *Cell* **2000**, *100*, 57–70. [CrossRef]
35. Jia, J.; Yang, S.; Huang, J.; Zheng, H.; He, Y.; Wang, L. Distinct Extracellular RNA Profiles in Different Plasma Components. *Front. Genet.* **2021**, *12*, 564780. [CrossRef]
36. Vergauwen, G.; Tulkens, J.; Pinheiro, C.; Avila Cobos, F.; Dedeyne, S.; De Scheerder, M.A.; Vandekerckhove, L.; Impens, F.; Miinalainen, I.; Braems, G.; et al. Robust sequential biophysical fractionation of blood plasma to study variations in the biomolecular landscape of systemically circulating extracellular vesicles across clinical conditions. *J. Extracell. Vesicles* **2021**, *10*, e12122. [CrossRef]
37. Wang, J.; Ma, G.; Li, M.; Han, X.; Xu, J.; Liang, M.; Mao, X.; Chen, X.; Xia, T.; Liu, X.; et al. Plasma tRNA Fragments Derived from 5′ Ends as Novel Diagnostic Biomarkers for Early-Stage Breast Cancer. *Mol. Therapy Nucleic Acids* **2020**, *21*, 954–964. [CrossRef]
38. Wang, J.; Ma, G.; Ge, H.; Han, X.; Mao, X.; Wang, X.; Veeramootoo, J.S.; Xia, T.; Liu, X.; Wang, S. Circulating tRNA-derived small RNAs (tsRNAs) signature for the diagnosis and prognosis of breast cancer. *NPJ Breast Cancer* **2021**, *7*, 4. [CrossRef]
39. Koi, Y.; Tsutani, Y.; Nishiyama, Y.; Ueda, D.; Ibuki, Y.; Sasada, S.; Akita, T.; Masumoto, N.; Kadoya, T.; Yamamoto, Y.; et al. Predicting the presence of breast cancer using circulating small RNAs, including those in the extracellular vesicles. *Cancer Sci.* **2020**, *111*, 2104–2115. [CrossRef]
40. Ren, J.; Zhou, Q.; Li, H.; Li, J.; Pang, L.; Su, L.; Gu, Q.; Zhu, Z.; Liu, B. Characterization of exosomal RNAs derived from human gastric cancer cells by deep sequencing. *Tumor Biol.* **2017**, *39*, 1–12. [CrossRef]
41. Tang, S.; Cheng, J.; Yao, Y.; Lou, C.; Wang, L.; Huang, X.; Zhang, Y. Combination of Four Serum Exosomal MiRNAs as Novel Diagnostic Biomarkers for Early-Stage Gastric Cancer. *Front. Genet.* **2020**, *11*, 237. [CrossRef] [PubMed]
42. Lin, C.; Zheng, L.; Huang, R.; Yang, G.; Chen, J.; Li, H. tRFs as Potential Exosome tRNA-Derived Fragment Biomarkers for Gastric Carcinoma. *Clin. Lab.* **2020**, *66*. [CrossRef] [PubMed]
43. Zhang, Y.; Cai, F.; Liu, J.; Chang, H.; Liu, L.; Yang, A.; Liu, X. Transfer RNA-derived fragments as potential exosome tRNA-derived fragment biomarkers for osteoporosis. *Int. J. Rheum. Dis.* **2018**, *21*, 1659–1669. [CrossRef] [PubMed]
44. Zhu, L.; Li, J.; Gong, Y.; Wu, Q.; Tan, S.; Sun, D.; Xu, X.; Zuo, Y.; Zhao, Y.; Wei, Y.Q.; et al. Exosomal tRNA-derived small RNA as a promising biomarker for cancer diagnosis. *Mol. Cancer* **2019**, *18*, 74. [CrossRef] [PubMed]
45. Kumar, S.R.; Kimchi, E.T.; Manjunath, Y.; Gajagowni, S.; Stuckel, A.J.; Kaifi, J.T. RNA cargos in extracellular vesicles derived from blood serum in pancreas associated conditions. *Sci. Rep.* **2020**, *10*, 2800. [CrossRef] [PubMed]
46. Liao, J.; Liu, R.; Yin, L.; Pu, Y. Expression profiling of exosomal miRNAs derived from human esophageal cancer cells by Solexa high-throughput sequencing. *Int. J. Mol. Sci.* **2014**, *15*, 15530–15551. [CrossRef]
47. Wei, Z.; Batagov, A.O.; Schinelli, S.; Wang, J.; Wang, Y.; El Fatimy, R.; Rabinovsky, R.; Balaj, L.; Chen, C.C.; Hochberg, F.; et al. Coding and noncoding landscape of extracellular RNA released by human glioma stem cells. *Nat. Commun.* **2017**, *8*, 1145. [CrossRef]
48. Li, C.; Qin, F.; Hu, F.; Xu, H.; Sun, G.; Han, G.; Wang, T.; Guo, M. Characterization and selective incorporation of small non-coding RNAs in non-small cell lung cancer extracellular vesicles. *Cell Biosci.* **2018**, *8*, 2. [CrossRef]
49. Tong, F.; Andress, A.; Tang, G.; Liu, P.; Wang, X. Comprehensive profiling of extracellular RNA in HPV-induced cancers using an improved pipeline for small RNA-seq analysis. *Sci. Rep.* **2020**, *10*, 19450. [CrossRef]
50. Xi, J.; Zeng, Z.; Li, X.; Zhang, X.; Xu, J. Expression and Diagnostic Value of tRNA-Derived Fragments Secreted by Extracellular Vesicles in Hypopharyngeal Carcinoma. *Onco Targets Ther.* **2021**, *14*, 4189–4199. [CrossRef]
51. Cho, O.; Kim, D.W.; Cheong, J.Y. Screening Plasma Exosomal RNAs as Diagnostic Markers for Cervical Cancer: An Analysis of Patients Who Underwent Primary Chemoradiotherapy. *Biomolecules* **2021**, *11*, 1691. [CrossRef] [PubMed]
52. Lunavat, T.R.; Cheng, L.; Kim, D.K.; Bhadury, J.; Jang, S.C.; Lasser, C.; Sharples, R.A.; Lopez, M.D.; Nilsson, J.; Gho, Y.S.; et al. Small RNA deep sequencing discriminates subsets of extracellular vesicles released by melanoma cells—Evidence of unique microRNA cargos. *RNA Biol.* **2015**, *12*, 810–823. [CrossRef] [PubMed]
53. Lunavat, T.R.; Cheng, L.; Einarsdottir, B.O.; Olofsson Bagge, R.; Veppil Muralidharan, S.; Sharples, R.A.; Lasser, C.; Gho, Y.S.; Hill, A.F.; Nilsson, J.A.; et al. BRAF(V600) inhibition alters the microRNA cargo in the vesicular secretome of malignant melanoma cells. *Proc. Natl. Acad. Sci. USA* **2017**, *114*, E5930–E5939. [CrossRef] [PubMed]

54. Corsello, T.; Kudlicki, A.S.; Garofalo, R.P.; Casola, A. Cigarette Smoke Condensate Exposure Changes RNA Content of Extracellular Vesicles Released from Small Airway Epithelial Cells. *Cells* **2019**, *8*, 1652. [CrossRef]
55. Sundar, I.K.; Li, D.; Rahman, I. Small RNA-sequence analysis of plasma-derived extracellular vesicle miRNAs in smokers and patients with chronic obstructive pulmonary disease as circulating biomarkers. *J. Extracell. Vesicles* **2019**, *8*, 1684816. [CrossRef]
56. Singh, K.P.; Maremanda, K.P.; Li, D.; Rahman, I. Exosomal microRNAs are novel circulating biomarkers in cigarette, waterpipe smokers, E-cigarette users and dual smokers. *BMC Med. Genom.* **2020**, *13*, 128. [CrossRef]
57. Buschmann, D.; Kirchner, B.; Hermann, S.; Marte, M.; Wurmser, C.; Brandes, F.; Kotschote, S.; Bonin, M.; Steinlein, O.K.; Pfaffl, M.W.; et al. Evaluation of serum extracellular vesicle isolation methods for profiling miRNAs by next-generation sequencing. *J. Extracell. Vesicles* **2018**, *7*, 1481321. [CrossRef]
58. Hermann, S.; Buschmann, D.; Kirchner, B.; Borrmann, M.; Brandes, F.; Kotschote, S.; Bonin, M.; Lindemann, A.; Reithmair, M.; Schelling, G.; et al. Transcriptomic profiling of cell-free and vesicular microRNAs from matched arterial and venous sera. *J. Extracell. Vesicles* **2019**, *8*, 1670935. [CrossRef]
59. Hermann, S.; Brandes, F.; Kirchner, B.; Buschmann, D.; Borrmann, M.; Klein, M.; Kotschote, S.; Bonin, M.; Reithmair, M.; Kaufmann, I.; et al. Diagnostic potential of circulating cell-free microRNAs for community-acquired pneumonia and pneumonia-related sepsis. *J. Cell. Mol. Med.* **2020**, *24*, 12054–12064. [CrossRef]
60. Zhang, J.; Zhu, Y.; Cai, R.; Jin, J.; He, Q. Differential Expression of Urinary Exosomal Small RNAs in Idiopathic Membranous Nephropathy. *BioMed Res. Int.* **2020**, *2020*, 3170927. [CrossRef]
61. Fang, S.; He, T.; Jiang, J.; Li, Y.; Chen, P. Osteogenic Effect of tsRNA-10277-Loaded Exosome Derived from Bone Mesenchymal Stem Cells on Steroid-Induced Osteonecrosis of the Femoral Head. *Drug Des. Dev. Ther.* **2020**, *14*, 4579–4591. [CrossRef] [PubMed]
62. Baglio, S.R.; Rooijers, K.; Koppers-Lalic, D.; Verweij, F.J.; Perez Lanzon, M.; Zini, N.; Naaijkens, B.; Perut, F.; Niessen, H.W.; Baldini, N.; et al. Human bone marrow- and adipose-mesenchymal stem cells secrete exosomes enriched in distinctive miRNA and tRNA species. *Stem Cell Res. Ther.* **2015**, *6*, 127. [CrossRef] [PubMed]
63. Yan, Y.; Chang, C.; Su, J.; Veno, M.T.; Kjems, J. Osteoblastogenesis Alters Small RNA Profiles in EVs Derived from Bone Marrow Stem Cells (BMSCs) and Adipose Stem Cells (ASCs). *Biomedicines* **2020**, *8*, 387. [CrossRef] [PubMed]
64. Kaur, S.; Abu-Shahba, A.G.; Paananen, R.O.; Hongisto, H.; Hiidenmaa, H.; Skottman, H.; Seppanen-Kaijansinkko, R.; Mannerstrom, B. Small non-coding RNA landscape of extracellular vesicles from human stem cells. *Sci. Rep.* **2018**, *8*, 15503. [CrossRef] [PubMed]
65. Chettimada, S.; Lorenz, D.R.; Misra, V.; Wolinsky, S.M.; Gabuzda, D. Small RNA sequencing of extracellular vesicles identifies circulating miRNAs related to inflammation and oxidative stress in HIV patients. *BMC Immunol.* **2020**, *21*, 57. [CrossRef] [PubMed]
66. Pawar, K.; Shigematsu, M.; Sharbati, S.; Kirino, Y. Infection-induced 5′-half molecules of tRNAHisGUG activate Toll-like receptor 7. *PLoS Biol.* **2020**, *18*, e3000982. [CrossRef]
67. Sisquella, X.; Ofir-Birin, Y.; Pimentel, M.A.; Cheng, L.; Abou Karam, P.; Sampaio, N.G.; Penington, J.S.; Connolly, D.; Giladi, T.; Scicluna, B.J.; et al. Malaria parasite DNA-harbouring vesicles activate cytosolic immune sensors. *Nat. Commun.* **2017**, *8*, 1985. [CrossRef]
68. He, X.; Li, F.; Bor, B.; Koyano, K.; Cen, L.; Xiao, X.; Shi, W.; Wong, D.T.W. Human tRNA-Derived Small RNAs Modulate Host-Oral Microbial Interactions. *J. Dent. Res.* **2018**, *97*, 1236–1243. [CrossRef]
69. Zhang, Q.; Liu, Z.; Han, X.; Li, Y.; Xia, T.; Zhu, Y.; Li, Z.; Wang, L.; Hao, L.; Hu, F.; et al. Circulatory exosomal tRF-Glu-CTC-005 and tRF-Gly-GCC-002 serve as predictive factors of successful microdissection testicular sperm extraction in patients with nonobstructive azoospermia. *Fertil. Steril.* **2021**, *117*, 512–521. [CrossRef]
70. Hu, J.; Tang, T.; Zeng, Z.; Wu, J.; Tan, X.; Yan, J. The expression of small RNAs in exosomes of follicular fluid altered in human polycystic ovarian syndrome. *PeerJ* **2020**, *8*, e8640. [CrossRef]
71. Yang, P.; Zhang, X.; Chen, S.; Tao, Y.; Ning, M.; Zhu, Y.; Liang, J.; Kong, W.; Shi, B.; Li, Z.; et al. A Novel Serum tsRNA for Diagnosis and Prediction of Nephritis in SLE. *Front. Immunol.* **2021**, *12*, 735105. [CrossRef] [PubMed]
72. Dou, R.; Zhang, X.; Xu, X.; Wang, P.; Yan, B. Mesenchymal stem cell exosomal tsRNA-21109 alleviate systemic lupus erythematosus by inhibiting macrophage M1 polarization. *Mol. Immunol.* **2021**, *139*, 106–114. [CrossRef] [PubMed]
73. Meng, L.; Jiang, L.; Chen, J.; Ren, H.; Gao, Z.; Wu, F.; Wen, Y.; Yang, L. Transfer RNA-derived fragment tRF-28-QSZ34KRQ590K in plasma exosomes may be a potential biomarker for atopic dermatitis in pediatric patients. *Exp. Ther. Med.* **2021**, *21*, 489. [CrossRef] [PubMed]
74. Zhong, D.; Wu, C.; Xu, D.; Bai, J.; Wang, Q.; Zeng, X. Plasma-Derived Exosomal hsa-miR-4488 and hsa-miR-1228-5p: Novel Biomarkers for Dermatomyositis-Associated Interstitial Lung Disease with Anti-Melanoma Differentiation-Associated Protein 5 Antibody-Positive Subset. *BioMed Res. Int.* **2021**, *2021*, 6676107. [CrossRef] [PubMed]
75. Lee, M.Y.; Baxter, D.; Scherler, K.; Kim, T.K.; Wu, X.; Abu-Amara, D.; Flory, J.; Yehuda, R.; Marmar, C.; Jett, M.; et al. Distinct profiles of cell-free microRNAs in plasma of veterans with post-traumatic stress disorder. *J. Clin. Med.* **2019**, *8*, 963. [CrossRef]
76. Fitz, N.F.; Wang, J.; Kamboh, M.I.; Koldamova, R.; Lefterov, I. Small nucleolar RNAs in plasma extracellular vesicles and their discriminatory power as diagnostic biomarkers of Alzheimer's disease. *Neurobiol. Dis.* **2021**, *159*, 105481. [CrossRef]
77. Yang, Y.; Wang, Y.; Wei, S.; Zhou, C.; Yu, J.; Wang, G.; Wang, W.; Zhao, L. Extracellular vesicles isolated by size-exclusion chromatography present suitability for RNomics analysis in plasma. *J. Transl. Med.* **2021**, *19*, 104. [CrossRef]

78. Brennan, K.; Martin, K.; FitzGerald, S.P.; O'Sullivan, J.; Wu, Y.; Blanco, A.; Richardson, C.; Mc Gee, M.M. A comparison of methods for the isolation and separation of extracellular vesicles from protein and lipid particles in human serum. *Sci. Rep.* **2020**, *10*, 1039. [CrossRef]
79. Tosar, J.P.; Rovira, C.; Cayota, A. Non-coding RNA fragments account for the majority of annotated piRNAs expressed in somatic non-gonadal tissues. *Commun. Biol.* **2018**, *1*, 2. [CrossRef]
80. Alexander, R.P.; Kitchen, R.R.; Tosar, J.P.; Roth, M.; Mestdagh, P.; Max, K.E.A.; Rozowsky, J.; Kaczor-Urbanowicz, K.E.; Chang, J.; Balaj, L.; et al. Open Problems in Extracellular RNA Data Analysis: Insights From an ERCC Online Workshop. *Front. Genet.* **2021**, *12*, 778416. [CrossRef]
81. A Emara, M.M.; Ivanov, P.; Hickman, T.; Dawra, N.; Tisdale, S.; Kedersha, N.; Hu, G.F.; Anderson, P. Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. *J. Biol. Chem.* **2010**, *285*, 10959–10968. [CrossRef] [PubMed]
82. Kuscu, C.; Kumar, P.; Kiran, M.; Su, Z.; Malik, A.; Dutta, A. tRNA fragments (tRFs) guide Ago to regulate gene expression post-transcriptionally in a Dicer-independent manner. *RNA* **2018**, *24*, 1093–1105. [CrossRef] [PubMed]
83. Magee, R.; Rigoutsos, I. On the expanding roles of tRNA fragments in modulating cell behavior. *Nucleic Acids Res* **2020**, *48*, 9433–9448. [CrossRef] [PubMed]
84. Yu, M.; Lu, B.; Zhang, J.; Ding, J.; Liu, P.; Lu, Y. tRNA-derived RNA fragments in cancer: Current status and future perspectives. *J. Hematol. Oncol.* **2020**, *13*, 121. [CrossRef]
85. Pekarsky, Y.; Balatti, V.; Palamarchuk, A.; Rizzotto, L.; Veneziano, D.; Nigita, G.; Rassenti, L.Z.; Pass, H.I.; Kipps, T.J.; Liu, C.G.; et al. Dysregulation of a family of short noncoding RNAs, tsRNAs, in human cancer. *Proc. Natl. Acad. Sci. USA* **2016**, *113*, 5071–5076. [CrossRef]
86. Zhan, T.; Rindtorff, N.; Boutros, M. Wnt signaling in cancer. *Oncogene* **2017**, *36*, 1461–1473. [CrossRef]
87. Cheng, W.C.; Liao, T.T.; Lin, C.C.; Yuan, L.E.; Lan, H.Y.; Lin, H.H.; Teng, H.W.; Chang, H.C.; Lin, C.H.; Yang, C.Y.; et al. RAB27B-activated secretion of stem-like tumor exosomes delivers the biomarker microRNA-146a-5p, which promotes tumorigenesis and associates with an immunosuppressive tumor microenvironment in colorectal cancer. *Int. J. Cancer* **2019**, *145*, 2209–2224. [CrossRef]
88. Mao, J.; Liang, Z.; Zhang, B.; Yang, H.; Li, X.; Fu, H.; Zhang, X.; Yan, Y.; Xu, W.; Qian, H. UBR2 Enriched in p53 Deficient Mouse Bone Marrow Mesenchymal Stem Cell-Exosome Promoted Gastric Cancer Progression via Wnt/beta-Catenin Pathway. *Stem Cells* **2017**, *35*, 2267–2279. [CrossRef]
89. Sun, Z.; Wang, L.; Zhou, Y.; Dong, L.; Ma, W.; Lv, L.; Zhang, J.; Wang, X. Glioblastoma Stem Cell-Derived Exosomes Enhance Stemness and Tumorigenicity of Glioma Cells by Transferring Notch1 Protein. *Cell. Mol. Neurobiol.* **2020**, *40*, 767–784. [CrossRef]

SPRINGER NATURE LICENSE
TERMS AND CONDITIONS

Jan 14, 2023

This Agreement between Mr. Daniel Liu ("You") and Springer Nature ("Springer Nature") consists of your license details and the terms and conditions provided by Springer Nature and Copyright Clearance Center.

License Number	5467580203306
License date	Jan 14, 2023
Licensed Content Publisher	Springer Nature
Licensed Content Publication	Nature Reviews Gastroenterology & Hepatology
Licensed Content Title	Molecular subtypes of pancreatic cancer
Licensed Content Author	Eric A. Collisson et al
Licensed Content Date	Feb 4, 2019
Type of Use	Thesis/Dissertation
Requestor type	academic/university or research institute
Format	print and electronic
Portion	figures/tables/illustrations
Number of figures/tables/illustrations	1
Will you be translating?	no

Circulation/distribution	1 - 29
Author of this Springer Nature content	no
Title	MicroRNAs In Extracellular Vesicles as Biomarkers for Pancreaticobiliary Cancers
Institution name	Imperial College London
Expected presentation date	Jan 2023
Portions	Figure 4 Phylotranscriptomic tree of pancreatic cancer.
Requestor Location	Mr. Daniel Liu 4th Floor IRDB Hammersmith Campus Imperial College London London, W12 0NN United Kingdom Attn: Mr. Daniel Liu
Total	0.00 GBP

Terms and Conditions

Springer Nature Customer Service Centre GmbH Terms and Conditions

The following terms and conditions ("Terms and Conditions") together with the terms specified in your [RightsLink] constitute the License ("License") between you as Licensee and Springer Nature Customer Service Centre GmbH as Licensor. By clicking 'accept' and completing the transaction for your use of the material ("Licensed Material"), you confirm your acceptance of and obligation to be bound by these Terms and Conditions.

1. Grant and Scope of License

1. 1. The Licensor grants you a personal, non-exclusive, non-transferable, non-sublicensable, revocable, world-wide License to reproduce, distribute, communicate to the public, make available, broadcast, electronically transmit or create derivative works using the Licensed Material for the purpose(s) specified in your RightsLink Licence Details only. Licenses are granted for the specific use requested in the order and for no other use, subject to these Terms and Conditions. You acknowledge and agree that the rights granted to you under this License do not include the right to

modify, edit, translate, include in collective works, or create derivative works of the Licensed Material in whole or in part unless expressly stated in your RightsLink Licence Details. You may use the Licensed Material only as permitted under this Agreement and will not reproduce, distribute, display, perform, or otherwise use or exploit any Licensed Material in any way, in whole or in part, except as expressly permitted by this License.

1. 2. You may only use the Licensed Content in the manner and to the extent permitted by these Terms and Conditions, by your RightsLink Licence Details and by any applicable laws.

1. 3. A separate license may be required for any additional use of the Licensed Material, e.g. where a license has been purchased for print use only, separate permission must be obtained for electronic re-use. Similarly, a License is only valid in the language selected and does not apply for editions in other languages unless additional translation rights have been granted separately in the License.

1. 4. Any content within the Licensed Material that is owned by third parties is expressly excluded from the License.

1. 5. Rights for additional reuses such as custom editions, computer/mobile applications, film or TV reuses and/or any other derivative rights requests require additional permission and may be subject to an additional fee. Please apply to journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

2. Reservation of Rights

Licensor reserves all rights not expressly granted to you under this License. You acknowledge and agree that nothing in this License limits or restricts Licensor's rights in or use of the Licensed Material in any way. Neither this License, nor any act, omission, or statement by Licensor or you, conveys any ownership right to you in any Licensed Material, or to any element or portion thereof. As between Licensor and you, Licensor owns and retains all right, title, and interest in and to the Licensed Material subject to the license granted in Section 1.1. Your permission to use the Licensed Material is expressly conditioned on you not impairing Licensor's or the applicable copyright owner's rights in the Licensed Material in any way.

3. Restrictions on use

3. 1. Minor editing privileges are allowed for adaptations for stylistic purposes or formatting purposes provided such alterations do not alter the original meaning or intention of the Licensed Material and the new figure(s) are still accurate and representative of the Licensed Material. Any other changes including but not limited to, cropping, adapting, and/or omitting material that affect the meaning, intention or moral rights of the author(s) are strictly prohibited.

3. 2. You must not use any Licensed Material as part of any design or trademark.

3. 3. Licensed Material may be used in Open Access Publications (OAP), but any such reuse must include a clear acknowledgment of this permission visible at the same time as the figures/tables/illustration or abstract and which must indicate that the Licensed Material is not part of the governing OA license but has been reproduced with permission. This may be indicated according to any standard referencing system but

must include at a minimum 'Book/Journal title, Author, Journal Name (if applicable), Volume (if applicable), Publisher, Year, reproduced with permission from SNCSC'.

4. STM Permission Guidelines

4. 1. An alternative scope of license may apply to signatories of the STM Permissions Guidelines ("STM PG") as amended from time to time and made available at https://www.stm-assoc.org/intellectual-property/permissions/permissions-guidelines/.

4. 2. For content reuse requests that qualify for permission under the STM PG, and which may be updated from time to time, the STM PG supersede the terms and conditions contained in this License.

4. 3. If a License has been granted under the STM PG, but the STM PG no longer apply at the time of publication, further permission must be sought from the Rightsholder. Contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

5. Duration of License

5. 1. Unless otherwise indicated on your License, a License is valid from the date of purchase ("License Date") until the end of the relevant period in the below table:

Reuse in a medical communications project	Reuse up to distribution or time period indicated in License
Reuse in a dissertation/thesis	Lifetime of thesis
Reuse in a journal/magazine	Lifetime of journal/magazine
Reuse in a book/textbook	Lifetime of edition
Reuse on a website	1 year unless otherwise specified in the License
Reuse in a presentation/slide kit/poster	Lifetime of presentation/slide kit/poster. Note: publication whether electronic or in print of presentation/slide kit/poster may require further permission.
Reuse in conference proceedings	Lifetime of conference proceedings
Reuse in an annual report	Lifetime of annual report
Reuse in training/CME materials	Reuse up to distribution or time period indicated in License
Reuse in newsmedia	Lifetime of newsmedia
Reuse in coursepack/classroom materials	Reuse up to distribution and/or time period indicated in license

6. Acknowledgement

6. 1. The Licensor's permission must be acknowledged next to the Licensed Material in print. In electronic form, this acknowledgement must be visible at the same time as the figures/tables/illustrations or abstract and must be hyperlinked to the journal/book's homepage.

6. 2. Acknowledgement may be provided according to any standard referencing system and at a minimum should include "Author, Article/Book Title, Journal name/Book imprint, volume, page number, year, Springer Nature".

7. Reuse in a dissertation or thesis

7. 1. Where 'reuse in a dissertation/thesis' has been selected, the following terms apply: Print rights of the Version of Record are provided for; electronic rights for use only on institutional repository as defined by the Sherpa guideline (www.sherpa.ac.uk/romeo/) and only up to what is required by the awarding institution.

7. 2. For theses published under an ISBN or ISSN, separate permission is required. Please contact journalpermissions@springernature.com or bookpermissions@springernature.com for these rights.

7. 3. Authors must properly cite the published manuscript in their thesis according to current citation standards and include the following acknowledgement: '*Reproduced with permission from Springer Nature*'.

8. License Fee

You must pay the fee set forth in the License Agreement (the "License Fees"). All amounts payable by you under this License are exclusive of any sales, use, withholding, value added or similar taxes, government fees or levies or other assessments. Collection and/or remittance of such taxes to the relevant tax authority shall be the responsibility of the party who has the legal obligation to do so.

9. Warranty

9. 1. The Licensor warrants that it has, to the best of its knowledge, the rights to license reuse of the Licensed Material. **You are solely responsible for ensuring that the material you wish to license is original to the Licensor and does not carry the copyright of another entity or third party (as credited in the published version).** If the credit line on any part of the Licensed Material indicates that it was reprinted or adapted with permission from another source, then you should seek additional permission from that source to reuse the material.

9. 2. EXCEPT FOR THE EXPRESS WARRANTY STATED HEREIN AND TO THE EXTENT PERMITTED BY APPLICABLE LAW, LICENSOR PROVIDES THE LICENSED MATERIAL "AS IS" AND MAKES NO OTHER REPRESENTATION OR WARRANTY. LICENSOR EXPRESSLY DISCLAIMS ANY LIABILITY FOR ANY CLAIM ARISING FROM OR OUT OF THE CONTENT, INCLUDING BUT NOT LIMITED TO ANY ERRORS, INACCURACIES, OMISSIONS, OR DEFECTS CONTAINED THEREIN, AND ANY IMPLIED OR EXPRESS WARRANTY AS TO MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE. IN NO EVENT SHALL LICENSOR BE LIABLE TO YOU OR ANY OTHER PARTY OR ANY OTHER PERSON OR FOR ANY SPECIAL, CONSEQUENTIAL, INCIDENTAL, INDIRECT, PUNITIVE, OR EXEMPLARY DAMAGES, HOWEVER CAUSED, ARISING OUT OF OR IN CONNECTION WITH THE DOWNLOADING, VIEWING OR USE OF THE LICENSED MATERIAL REGARDLESS OF THE FORM OF ACTION, WHETHER FOR BREACH OF CONTRACT, BREACH OF WARRANTY, TORT, NEGLIGENCE, INFRINGEMENT OR OTHERWISE (INCLUDING, WITHOUT LIMITATION,

DAMAGES BASED ON LOSS OF PROFITS, DATA, FILES, USE, BUSINESS OPPORTUNITY OR CLAIMS OF THIRD PARTIES), AND WHETHER OR NOT THE PARTY HAS BEEN ADVISED OF THE POSSIBILITY OF SUCH DAMAGES. THIS LIMITATION APPLIES NOTWITHSTANDING ANY FAILURE OF ESSENTIAL PURPOSE OF ANY LIMITED REMEDY PROVIDED HEREIN.

10. Termination and Cancellation

10. 1. The License and all rights granted hereunder will continue until the end of the applicable period shown in Clause 5.1 above. Thereafter, this license will be terminated and all rights granted hereunder will cease.

10. 2. Licensor reserves the right to terminate the License in the event that payment is not received in full or if you breach the terms of this License.

11. General

11. 1. The License and the rights and obligations of the parties hereto shall be construed, interpreted and determined in accordance with the laws of the Federal Republic of Germany without reference to the stipulations of the CISG (United Nations Convention on Contracts for the International Sale of Goods) or to Germany´s choice-of-law principle.

11. 2. The parties acknowledge and agree that any controversies and disputes arising out of this License shall be decided exclusively by the courts of or having jurisdiction for Heidelberg, Germany, as far as legally permissible.

11. 3. This License is solely for Licensor's and Licensee's benefit. It is not for the benefit of any other person or entity.

Questions? For questions on Copyright Clearance Center accounts or website issues please contact springernaturesupport@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777. For questions on Springer Nature licensing please visit https://www.springernature.com/gp/partners/rights-permissions-third-party-distribution

Other Conditions:

Version 1.4 - Dec 2022

Questions? customercare@copyright.com or +1-855-239-3415 (toll free in the US) or +1-978-646-2777.