DISSERTATION

submitted to the

Combined Faculties for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

M.Sc. Laura Kuhlmann

born in: Buzău, Romania

Oral-examination: 18.09.2014

Identification and validation of novel protein biomarkers in pancreatic ductal adenocarcinoma with the ability to distinguish molecular subtypes

Referees: Prof. Dr. Andreas Trumpp

Prof. Dr. Jonathan Sleeman

The work presented in this thesis was started in October 2010 and completed June 2014 under the supervision of Dr. Christoph Rösli in the research group “Biomarker discovery group”, Department of “Stem cells and cancer” at the German Cancer Research Center (DKFZ), Heidelberg, as well as at the Heidelberg Institute for Stem Cells and Experimental Medicine (HI-STEM) gGmbH, Heidelberg

Parts of this thesis have been published in:

Conference posters:

Postolache *et al.*

"Cell surface proteomics-driven identification of pancreatic ductal adenocarcinoma subtypes"

(**ASMS, 2012** – Vancouver, Canada)

Postolache *et al.*

"Proteomics identification of cell surface biomarkers discerning PDAC subtypes"

(**Proteomic Forum, 2013** – Berlin, Germany)

Postolache *et al.*

"In-Depth Proteomic Characterization of Pancreatic Ductal Adenocarcinoma Subtypes"

(**HUPO 12th Annual World Congress, 2013** – Yokhama, Japan)

Kuhlmann *et al.*

"Identification of novel protein biomarkers for subtype stratification and novel immunotherapies in pancreatic ductal adenocarcinoma"

(**Young Researchers in Life Sciences, 2014** – Paris, France)

Acknowledgements

First of all I would like to thank **Dr. Christoph Rösli**, for all his support and advices. He has been a great teacher! Thank you for guiding me in this new field and for having the patience to answer all my question – and there were really a lot of them in the beginning, when the mass spectrometer and all the proteomic software were so new to me. You have helped me grow and organize better, in the best Swiss way possible ☺. Thank you for always being available when I needed advice, or when the mass spec crashed on me. This has been a wonderful journey, and it would not have been possible without you!

I would also like to thank **Prof. Dr. Andreas Trumpp**, for giving me the opportunity to work in his exciting and stimulating lab. Thank you for leading those wonderful debates during our lab meetings and journal clubs and for promoting our lab-members with such passion. I really appreciate the friendly atmosphere you created and I will additionally cherish the memorable celebrations and parties we had with the lab ☺.

I would also like to thank my TAC members **Prof. Dr. Jonathan Sleemann** and **Dr. Martina Schnölzer**, for all the great advices during our advisory meetings. You helped me see my work from different perspectives and find solutions to improve it.

A big thank you goes to my lab colleagues – **Amelie** for sharing the experience of being the first PhD student of the lab with me, **Wiebke** for sweetening our lab, **Sabrina** for all the learning and funny moments we shared, **Alex** who we've dazzled with our huge data sets (and he dazzled us back with his wonderful software ☺), **Katharina** for all the advices during our collaboration and for making it possible to bring this project to the next level. I would also like to thank **Philipp** for creating a friendly work atmosphere during our collaborations and all his advices. Additionally, I would like to acknowledge all the great master students which chose our lab during their studies – and the most important one for me **Lisa Becker** ☺. Your dedication has brought this project one step closer to a clinical application. Thank you for being a wonderful student, it is amazing how much our collaboration has taught me. I wish you all the best during your PhD thesis, I somehow know you will do great ☺.

I am extremely grateful also to our wonderful collaborators – the lab of **Dr. Martin Sprick**. Thank you so much for sharing the PACO project with us! Many thanks to **Dr. Christian Eisen** for providing the PACO cells, the starting material for my work. Special thanks to **Corinna Klein** for guiding me through my first mouse experiments – I am your proud Padawan ☺. I would like to express additionally my gratitude for the two PACO ladies – **Elisa Noll** and **Dr. Elisa Espinet** – for the wonderful collaboration we shared. With your arrival, the project grew a new set of wings.

I would also like to thank the **Jasmin** and **Katja** for all their advices that helped me trouble shoot the staining experiments. Thank you also to and **Camille**, **Arnaud** and **Jacob** for familiarizing me with gene expression analysis. Your entire lab has been a source of great laughers and great science, I am so happy I could share that with you.

Thank you to all the **HI-STEM** and **A010** lab members, for creating such a wonderful and stimulating work environment. Your work is truly inspirational and I have learned so much from every single one of you!

I would also like to thank the graduate school and all my fellow PhD colleagues from different labs, with whom I participated in so many of the Helmholtz activities. I have made

so many great friends – to name just a few: **Katarina**, **Dharanidja**, **Nevyana**, **Beate**. Thank you for spicing up the PhD life and for sharing all those wonderful, but also those stressful moments.

A big thanks goes to **my family**. Words cannot express my gratitude for your support. I would like to thank my parents for believing in me and for the genes they gave me, that made me fell in love with life sciences, the same way they had ☺. I would also like to thank my grandparents, for contributing so much to my education and for supporting me in my ambitions. Thank you also to my husband Devin, for being there for me even for my stressful moments – I cannot express how much it means to me.

Abbreviations

2D	Two dimensional
2D PAGE	Two dimensional Polyacrylamide gel electrophoresis
5-FU	5-fluorouracil
A488	Alexa-488 dye
A546	Alexa-546 dye
A647	Alexa-647 dye
Ab	Antibody
ACN	Acetonitrile
ATX	ataxia telangiectasia
BRACA2	Breast cancer 2, early onset
BSA	Bovine seric albumin
CA125	Carbohydrate antigen 125
CA19-9	carbohydrate antigen 19-9/cancer antigen 19-9
CD31	cluster of differentiation 31/Platelet endothelial cell adhesion molecule
CDH17	Cadherin-17
CDKN2A	Cyclin-dependent kinase inhibitor 2A
cDNA	Complementary DNA (Deoxyribonucleic acid)
CHCA	α-Cyano-4-hydroxycinnamic acid
CID	Collision induced dissociation
CM	Conserved motif
Co-IP	Co-Immunorecipitation
CRD	Carbohydrate recognition domain
CSC	Cancer stem cell
Ct	Threshold cycle
DAPI	4´,6-Diamidino-2-phenylindole dihydrochloride
DKFZ	(translation) German Cancer Research Center
DMEM	Dulbecco's Modified Eagle's medium
DMSO	Dimethyl sulfoxide
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGF	Epidermal growth factor
ELISA	Enzyme-linked immunosorbent assay
EMT	Epithelial to mesenchymal transition
ER	Endoplasmic reticulum
ESI	Electro Spray Ionization
ETD	Electron transfer dissociation
FAK	Focal Adhesion Kinase
FASP	Filter aided sample preparation
FCS	Fetal Calf Serum
FDA	Food and Drug Administration (USA)
FGF	Firbrobalst growth factor
fM	Femtomolar
FRET	Fluorescence resonance energy transfer
g	gram
GAPDH	Glyceraldehyde 3-phosphate
GMCSF	Granulocyte-macrophage colony-stimulating factor
GNAS	Stimulatory G-protein alpha subunit
GO	Gene ontology
HE4	Human epididimus protein 4
HIF-1α	Hypoxia-inducible factor 1-alpha

HI-STEM	The Heidelberg Institute for Stem Cell Technology and Experimental Medicine gGmbH
HPDE	Human pancreatic ductal epithelial cells
HPNE	Human pancreatic nestin expressing cells
HRP	Horse radish peroxidase
HUVEC	Human umbilical vein endothelial cells
IF	Immunofluorescence
IFP	Interstitial fluid pressure
IGF	Insulin-like growth factor
IgG	Immunoglobulin G
IHC	Immunohistochemistry
IL1/IL6	Interleukin-1/Interleukin-6
IPMN	Intraductal papillary mucinous neoplasms
KRAS	V-Ki-ras2 Kirsten rat sarcoma viral oncogene homolo
L	Liter
LC	Liquid Chromatography
LCN2	Lipocalin-2
LGALS4	Galectin-4
M	Molar
mM	Milimolar
m/*z*	Mass to charge ratio
MALDI	Matrix-Assisted Laser Desorbtion/Ionization
MCN	Mucinous cystic neoplasms
miRNA	Micro RNA
Mist1	Basic helix-loop-helix family, member a15
mL	Mililiter
MLH1	MutL homolog 1, colon cancer, nonpolyposis type 2
MMP	Matrix metaloproteinase
MRM	Multiple reaction monitoring
mRNA	Messenger RNA (Ribonucleic acid)
MS	Mass spectrometry
MS^2	Tandem mass spectrometry
MUC1	Mucin-1
ng	Nanogram
NGAL	Lipocalin-2
NF-kB	Nuclear factor kappa B cells
nM	Nanomolar
NMR	Nuclear magnetic resonance
NSG	Nod SCID Gamma
OCT	Optimal Cutting Temperature compound
PACO	Pancreatic adenocarcinoma primary derived cell line (Dr. C. Eisen)
PAGE	Polyacrylamide gel electrophoresis
PanIN	Pancreatic intraepithelial neoplasia
PEB	PBS-EDTA + 1% bovine seric albumin buffer
PBS	Phosphate Buffer Saline
PCDH1	Protocadherin-1
PDAC	Pancreatic ductal adenocarcinoma
PDX1	Pancreatic and duodenal homeobox 1
pg	Picogram
PIGF	Placental Growth Factor
PKCθ	Protein kinase C θ
PLA	Proximity ligation assay
PRM	Parallel reaction monitoring
PTM	Post translational modification
Q1/Q2/Q3	$1^{st}/2^{nd}/3^{rd}$ quadrupole (quadrupole instruments)

QM	Quasimesenchymal
RapiGest	Sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxyl]-1-propanesulfonate
RPL13a	60S ribosomal protein L13a
RT-qPCR	Real time quantitative polymerase chain reaction
scFV	Single-chain variable fragment
SDS	Sodium dodecyl sulfate
SELDI	Surface-Enhanced Laser Desorption/Ionization
SEM	Standard error of mean
Shh	Sonic hedgehog
SILAC	Stable isotope labeling by amino acids in cell culture
SMAD4	Mothers against decapentaplegic homolog 4
SRM	Single reaction monitoring
STEAP1	Six-transmembrane epithelial antigen of prostate 1
sulfo-NHS	Sulfo-N-Hydroxysuccinimide
TDB	Trypsin digestion buffer
TFA	Trifluoroacetic acid
TGF-β1	Transforming growth factor-β1
TMA	Tissue micro array
TNFα	Tumor necrosis factor α
TOF	Time of flight
UPLC	Ultra performance liquid chromatography
VEGF	Vascular endothelial growth factor
WB	Western Blot
Wnt	Proto-oncogene protein Wnt
wt/v	Weight over volume
α-SMA	α smooth muscle actin
µg	Microgram
µm	Micrometer
µM	Micromolar

Table of contents

1. Abstracts 1

1.1 Summary 1

1.2 Zusammenfassung 3

2. Introduction 5

2.1 Pancreatic ductal adenocarcinoma 5

2.1.1 Molecular and genetic aspects of PDAC 6

2.1.2 Extracellular matrix characteristics in PDAC..... 7

2.1.3 Precursor lesions in PDAC..... 8

2.1.4 Molecular subtypes in PDAC..... 10

2.1.5 PDAC therapies – standard of care and novel approaches 11

2.2 Biomarker discovery and proteomics 13

2.2.1 Molecular biomarkers – definition and identification 13

2.2.2 Short introduction to proteomics and mass spectrometry 16

2.2.3 Clinical proteomics in cancer diagnostic and therapy – functional and methodological requirements 19

2.2.4 Using mass spectrometry to identify membrane associated protein biomarkers 21

2.2.5 Using mass spectrometry to identify secreted protein biomarkers 24

2.2.6 Validation methods and targeted proteomics 27

2.3 Current knowledge about novel PDAC biomarker candidates 31

2.3.1 Cadherin-17 31

2.3.2 Galectin-4 34

2.3.3 Protocadherin-1 36

2.3.4 Lipocalin-2 37

3. Aim of the study 41

4. Materials and methods 42

4.1 Materials 42

4.2 Methods 60

4.2.1 Cell culture methods 60

4.2.2 Generation of orthotopic pancreatic mouse tumors 61

4.2.3 Discovery of novel protein biomarkers *in vitro* 61

4.2.3.1 Discovery of novel cell surface biomarkers *in vitro* 61

4.2.3.2 Discovery of novel secreted biomarkers *in vitro* 62

4.2.4 Evaluation of *in vivo* vascular accessible protein markers 63

4.2.5 Mass spectrometric analysis 65

4.2.6 Validation of protein biomarkers 66

4.2.6.1 Antibody based validations 66

4.2.6.1.1 Western Blot 66

4.2.6.1.2 Immunocytofluorescence on methanol fixed cells 67
4.2.6.1.3 Immunocytofluorescence on formalin fixed paraffin embedded cells 67
4.2.6.1.4 Testing colocalization of putative exocrine-like biomerkers using confocal microscopy 68
4.2.6.1.5 Immunohistofluorescence 69
4.2.6.2 Antibody independent validations 70
4.2.6.2.1 MRM validations 70
4.2.6.2.2 Gene expression analysis 73
5. Results 75
5.1 Identification of novel protein biomarkers using mass spectrometry 77
5.1.1 Identification of novel cell surface protein biomarkers for PDAC using *in vitro* cultured primary, patient-matched cell lines 77
5.1.2 Identification of novel secreted protein biomarkers for PDAC *in vitro* …… 86
5.1.3 Overview of total protein list detected on the cell surface proteome and secretome of PDAC cell lines versus healthy controls ……..……..... 96
5.1.4 *In vivo* biotinylation of tumor xenograft mouse models for the identification of novel subtype specific biomarker candidates 97
5.1.4.1 Evaluation of perfusion success using immunofluorescence …… 98
5.1.4.2 Evaluation and quantification of perfusion success using ELISA 100
5.1.4.3 Identification of novel protein biomarkers *in vivo* using mass spectrometry…….. 101
5.2 Validation of detected pan-PDAC and exocrine-like PDAC protein biomarker candidates 107
5.2.1 Antibody based validations 107
5.2.1.1 *In vitro v*alidations…… 107
5.2.1.1.1 *In vitro* validations using immunocytofluorescence 107
5.2.1.1.2 Assessment of potential co-localization of exocrine-like protein biomarker candidates using confocal laser microscopy..... 112
5.2.1.1.3 Assessment of isoform distribution of pan-PDAC biomarker candidate PCDH1 using Western Blotting 115
5.2.1.2 *In vivo* immunohistofluorescence validations using mouse derived xenografts…..…… 117
5.2.1.3 *In vivo* validations on patient material..... 129
5.2.2 Antibody independent methods 133
5.2.2.1 *In vitro* validation and quantification of exocrine specific biomarkers using MRM 133
5.2.2.2 Evaluation of gene expression levels for the markers of interest using RT-qPCR 135
5.2.2.2.1 *In vitro* evaluation of gene expression levels for pan-PDAC and Exocrine like biomarker candidates using RT-qPCR 135

5.2.2.2.2 *In vivo* evaluation of gene expression levels for pan-PDAC and exocrine-like biomarker candidates using RT-qPCR on total tumor tissue and EpCAM+ cells 137

6. Discussions 139

6.1 Mass spectrometric based analysis of the cell surface proteome and secretome of cultured cells reveal novel putative PDAC markers, with the potential to distinguish between molecular subtypes 141

6.2 Validation of exocrine-like protein biomarker candidates 145

6.2.1 *In vitro* validations 145

6.2.2 *In vivo* validations 146

6.3 Validations of pan-PDAC protein biomarker candidates 149

6.3.1 *In vitro* validations 149

6.3.2 *In vivo* validations 150

7. Conclusions and outlook 152

8. Supplementary information 154

9. References 165

1. Abstracts

1.1 Summary

Pancreatic ductal adenocarcinoma (PDAC) is one of the deadliest forms of cancer, with a medium overall survival of only six months. The standard of care treatment, consisting of surgical resection followed by adjuvant chemotherapy, which incidentally can be offered to as little as 20% of the newly diagnosed patients, shows only marginal benefit for patients and no significant improvements to PDAC therapy have been made over the past two decades. PDAC is currently treated as a single, homogenous disease. Recently, three molecular subtypes termed classical, exocrine-like and quasimesenchymal (QM) have been defined, based on transcriptomic profiling of micro-dissected cancer cells from primary tumors and commercially available cell lines. The newly defined molecular subtypes were associated with differences in patient overall survival and are suggested to predict response to different chemotherapeutical agents. However, no clinical markers are available for the stratification of patients according to these molecular subtypes.

The aim of this study was the identification and validation of novel pan-PDAC as well as subtype-specific protein biomarker candidates, which can be employed in the development of novel clinical applications, such as diagnostics and/or targeted PDAC therapies. Using a panel of twelve patient derived primary pancreatic cancer cell lines, grown in a chemically defined serum free medium, we employed LC-MALDI-MS to identify novel cell surface and secreted PDAC protein biomarker candidates. Two commercially available healthy pancreatic cell lines served as controls. The cell surface proteome was analyzed following *in vitro* biotinylation and subsequent streptavidin pull down of the covalently biotin-labeled proteins, while the secreted proteome was analyzed by shotgun proteomics. In addition, we investigated *in vivo* vascular accessible PDAC biomarkers, by carrying out whole body perfusions of mice bearing orthotopic pancreatic tumors using a reactive biotin ester solution. Biomarker candidates of interest were selected after the proteomics discovery experiment and underwent thorough validations, using complementary antibody based (immunofluorescence, Western Blot) and antibody independent (single reaction monitoring, RT-qPCR) techniques. The validation experiments were performed both *in vitro* – using the PDAC cell culture model and *in vivo* – using tumor xenografts developed in immunodeficient mice.

More than 2500 proteins were identified after completing the cell surface proteome analysis, and over 1700 proteins could be reported in the conditioned cell culture medium. We selected the two pan-PDAC biomarker candidates, protocadherin-1 (PCDH1) and lipocalin-2 (LCN2), as well as the two exocrine-like biomarker candidates, cadherin-17 (CDH17) and galectin-4 (LGALS4), for further validations.

In vitro immunofluorescence confirmed the exclusive presence of PCDH1 and LCN2 on the surface of primary patient-matched PDAC cell lines, which correlated with elevated levels of the corresponding mRNAs. The expression of the two exocrine-like protein biomarker candidates was restricted to the predicted subtype, as proven by immunofluorescence and single reaction monitoring. The secretion of the two exocrine-like proteins in the cell culture medium was additionally confirmed by single reaction monitoring. The regulation of the two exocrine-like PDAC biomarkers was at least partially due to the upregulation of the

corresponding mRNAs. Additionally, we could report that CDH17 and LGALS4 colocalize on the surface of cultured exocrine-like PDAC cells. *In vivo* immunofluorescence analyses corroborated the presence of PCDH1 and LCN2 on the surface of pancreatic cancer cells and in the tumor environment, while being completely absent in healthy human pancreata. Exocrine-like tumor xenografts expressed CDH17 and LGALS4 at high levels, as evaluated by immunofluorescence. However, a limited activation of the two putative exocrine-like biomarkers could be detected *in vivo* in some of the classical pancreatic tumors. Protein expression patterns correlated with reported levels of the respective mRNAs for all the investigated biomarker candidates, evaluated in both EpCam+ epithelial cells isolated from orthotopic xenografts and whole tumor tissue.

Moreover, LCN2 and LGALS4 could be detected in the course of a proof-of-principle *in vivo* perfusion experiment using mice harboring orthotopic tumors. Therefore, we could conclude that the two proteins are accessible from the blood stream and could represent candidates for novel PDAC targeted therapies.

Taken together, our study identifies novel pan-PDAC as well as subtype-specific protein biomarkers candidates, which can be used in clinical applications for patient stratification. Additionally, the proteins were secreted by pancreatic cancer cells, and could therefore represent the basis for developing novel non-invasive diagnostic tools. Importantly, the validated protein biomarker candidates were also accessible from the vascular system, making them eligible targets for novel antibody based therapies of PDAC.

1.2 Zusammenfassung

Das duktale Pankreasadenokarzinom weist unter allen Krebserkrankungen eine der höchsten Mortalitätsraten auf, die mittlere Überlebenrate nach der Diagnose beträgt nur 6 Monate. Die Standardbehandlung bestehend aus operativer Entfernung des Tumors gefolgt von adjuvanter Chemotherapie kann lediglich bei 20% der neu diagnostizierten Patienten angewandt werden und zeigt zudem nur geringen Nutzen für die Patienten. In den vergangenen zwei Jahrzehnten konnte die Therapie trotz intensiver Bemühungen sowohl seitens akademischer Forschungsstellen als auch seitens der pharmazeutischen Industrie nicht signifikant verbessert werden. Duktales Pankreasadenokarzinom wird derzeit bei allen Patienten einheitlich behandelt. Vor kurzem wurden, basierend auf Genexpressionsprofilen von durch Lasermikrodissektion aus humanem Tumorgewebe generierten Krebszellen sowie kommerziell erhältlichen Zelllinien, drei molekulare Subtypen des duktalen Pankreasadenokarzinoms (klassisch, exokrin-ähnlich und quasimesenchymal) beschrieben. Diese neu definierten molekularen Subtypen wurden mit unterschiedlichen Überlebensraten der Patienten assoziiert. Darüber hinaus können diese vermutlich zur Vorhersage des Behandlungserfolges mit verschiedenen Chemotherapeutika dienen. Jedoch gibt es noch keine klinischen Marker für die Stratifizierung von Patienten nach diesen molekularen Subtypen.

Das Ziel dieser Studie war die Identifizierung und Validierung neuer Pan- und Subtyp-spezifischer Protein-Biomarkerkandidaten des duktale Pankreasadenokarzinoms, welche zur Entwicklung neuer Diagnostika und zielgerichteter Therapien genutzt werden können. Anhand von zwölf aus Patientenmaterial abgeleiteten, primären Zelllinien, die in serum-freiem Medium kultiviert wurden, konnten mittels LC-MALDI-MS neue Zelloberflächen- und sekretierte Biomarkerkandidaten idenfiziert werden. Dabei dienten zwei kommerziell erhältliche, gesunde Pankreas-Zelllinien als Kontrollgruppe. Die Analyse des Zelloberflächenproteoms erfolgte durch *in vitro*-Biotinylierung der Zellen mit darauffolgender Affinitäts-basierter Anreicherung der kovalent markierten Proteine auf Streptavidin, wohingegen das sekretierte Proteom durch einen Shotgun-Ansatz untersucht wurde. Darüberhinaus wurden vaskulär erreichbare Biomarker durch *in vivo*-Perfusion von orthotop pankreastumortragenden Mäusen mit einem reaktiven Biotin-Ester Derivat markiert und analysiert. Aus den Proteom-Daten wurden daraufhin interessante Biomarkerkandidaten ausgewählt und mittels komplementären, auf Antikörpern basierenden Techniken (Immunfluoreszenz, Western Blot) sowie Antikörper-unabhängigen Techniken (Single Reaction Monitoring, RT-qPCR) validiert. Die Validierung erfolgte sowohl *in vitro* auf Zellkulturmodellen als auch *in vivo* mittels Tumor-Xenotransplantaten in immundefizienten Mäusen.

Durch die Analyse des Zelloberflächenproteoms konnten mehr als 2500 Proteine identifiziert werden, im konditionierten Zellkulturmedium wurden über 1700 Proteine entdeckt. Zur weiteren Validierung wurden die zwei Pan-Biomarkerkandidaten Protocadherin-1 (PCDH1) und Lipocalin-2 (LCN2) sowie die zwei exokrin-ähnlichen Biomarkerkandidaten Cadherin-17 (CDH17) und Galectin-4 (LGALS4) ausgewählt.

Durch Immunfluoreszenz wurde das spezifische Vorhandensein von PCDH1 und LCN2 auf der Zelloberfläche von Zelllinien, die von humanem duktalen Pankreasadenokarzinomen abgeleitet wurden, bestätigt, was darüberhinaus mit erhöhten Konzentrationen der korrepondierenden mRNAs korreliert. Wie durch Immunfluoreszenz und single reaction

monitoring gezeigt wurde, erwies sich die Expression der zwei exokrin-ähnlichen Biomarkerkandidaten wie vorhergesagt als beschränkt auf den exokrin-ähnlichen Subtyp. Die Sekretion der zwei exokrin-ähnlichen Subtyp-spezifischen Proteine in das Zellkulturmedium wurde darüber hinaus durch single reaction monitoring bestätigt. Die Regulation der zwei exokrin-ähnlichen Biomarker korrelierte dabei mit den erhöhten Konzentrationen der korrespondierenden mRNAs. Darüber hinaus konnte die Kolokalisierung von CDH17 und LGALS4 auf der Oberfläche von dem exokrin-ähnlichen Subtyp zugeordeten Zellen gezeigt werden. Die *in vivo* Immunfluoreszenz-Analysen konnten die Präsenz von PCDH1 und LCN2 auf der Oberfläche von Pankreastumorzellen und in der Umgebung des Tumors bestätigen, wohingegen die Biomarker in gesundem, humanen Pankreasgewebe nicht nachgewiesen werden konnten. Exokrin-ähnliche zugeordnete Xenotransplantate überexprimierten CDH17 und LGALS4 ebenfalls stark, wie durch Immunfluoreszenz gezeigt werden konnte. Jedoch konnte eine geringe Aktivierung der zwei mutmaßlich exokrin-ähnlichen Subtyp-spezifischen Biomarker *in vivo* auch im klassischen Subtyp detektiert werden. Die Proteinexpression korreliert für alle Biomarkerkandidaten stark mit den erhaltenen Konzentrationen der korrespondierenden mRNAs, welche in EpCam positiven Epithelzellen aus orthotopen Xenotransplantaten und Tumorgewebe evaluiert wurden.

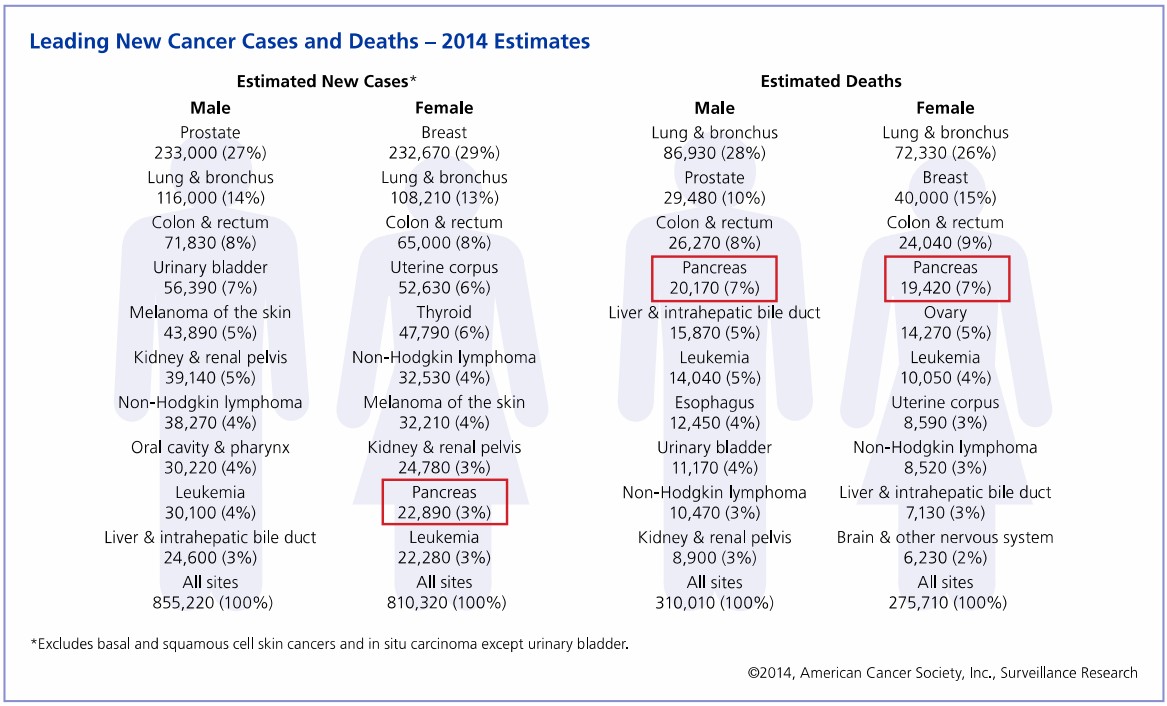
Darüber hinaus konnten LCN2 und LGALS4 in einem *in vivo*-Perfusionsexperiment mit orthotop tumortragenden Mäusen detektiert werden. Daraus lässt sich folgern, dass die zwei Protein-Biomarker vaskulär zugänglich und somit vielversprechende Kandidaten für neue zielgerichteten Therapien darstellen.

In dieser Studie wurden neue Pan- und Subtyp-spezifische Protein-Biomarkerkandidaten für das duktale Pankreasadenokarzinom identifiziert, die klinisch zur Stratifizierung von Patienten genutzt werden können. Darüber hinaus werden diese Proteine von Pankreastumorzellen sekretiert und können daher auch als Basis für die Entwicklung neuartiger, nicht invasiver Diagnostika dienen. Die validierten Biomarkerkandidaten sind zudem durch das Blutgefäßsystem zugänglich, was sie zu geeigneten Zielstrukturen für neue Antikörper-basierte Therapien des duktalen Pankreasadenokarzinoms macht.

2. Introduction

2.1 Pancreatic ductal adenocarcinoma

Despite significant advancements made in the field of oncology, pancreatic ductal adenocarcinoma (PDAC) continues to represent one of the most lethal forms of cancer. Accounting for less than 5% of cancer cases, PDAC represents the fourth most common cause of cancer-related deaths in the United States (**Figure 1)** [1]. In Europe, in 2012 pancreatic cancer was evaluated to be the fifth most common type cancer, with approximately 104'000 new cases reported per year. The high mortality rate reported for all patients made it the fourth cause of cancer related death in Europe [2]. The majority of patients succumb to the disease within 6 months from the time point of diagnosis, and the five year overall survival is a mere 5% [3, 4]. One of the main reasons for this devastating prognosis is that most patients, due to lack of early symptoms, present with an already advanced (i.e. metastatic disease) – and early pre-cancerous lesions are usually non-detectable using standard imaging techniques [5]. Only a limited number of patients (~30%) present with resectable, localized pancreatic cancer. Surgical removal of the primary tumor, however, increases the five year survival rate to only 10-20% [5]. PDAC is the most common type of pancreatic cancer, accounting for approximately 90% of pancreatic malignancies. Interestingly, other types of pancreatic cancers, such as neuroendocrine tumors and even the less common cystic neoplasms (less than 1% of cases), exhibit better prognoses and improved responses to chemotherapy compared to PDAC [6].



Leading New Cancer Cases and Deaths – 2014 Estimates

Estimated New Cases*		Estimated Deaths	
Male	**Female**	**Male**	**Female**
Prostate 233,000 (27%)	Breast 232,670 (29%)	Lung & bronchus 86,930 (28%)	Lung & bronchus 72,330 (26%)
Lung & bronchus 116,000 (14%)	Lung & bronchus 108,210 (13%)	Prostate 29,480 (10%)	Breast 40,000 (15%)
Colon & rectum 71,830 (8%)	Colon & rectum 65,000 (8%)	Colon & rectum 26,270 (8%)	Colon & rectum 24,040 (9%)
Urinary bladder 56,390 (7%)	Uterine corpus 52,630 (6%)	Pancreas 20,170 (7%)	Pancreas 19,420 (7%)
Melanoma of the skin 43,890 (5%)	Thyroid 47,790 (6%)	Liver & intrahepatic bile duct 15,870 (5%)	Ovary 14,270 (5%)
Kidney & renal pelvis 39,140 (5%)	Non-Hodgkin lymphoma 32,530 (4%)	Leukemia 14,040 (5%)	Leukemia 10,050 (4%)
Non-Hodgkin lymphoma 38,270 (4%)	Melanoma of the skin 32,210 (4%)	Esophagus 12,450 (4%)	Uterine corpus 8,590 (3%)
Oral cavity & pharynx 30,220 (4%)	Kidney & renal pelvis 24,780 (3%)	Urinary bladder 11,170 (4%)	Non-Hodgkin lymphoma 8,520 (3%)
Leukemia 30,100 (4%)	Pancreas 22,890 (3%)	Non-Hodgkin lymphoma 10,470 (3%)	Liver & intrahepatic bile duct 7,130 (3%)
Liver & intrahepatic bile duct 24,600 (3%)	Leukemia 22,280 (3%)	Kidney & renal pelvis 8,900 (3%)	Brain & other nervous system 6,230 (2%)
All sites 855,220 (100%)	All sites 810,320 (100%)	All sites 310,010 (100%)	All sites 275,710 (100%)

*Excludes basal and squamous cell skin cancers and in situ carcinoma except urinary bladder.

Figure 1: Predicted cancer statistics – estimated morbidities and cancer related deaths in the US for 2014; derived from the American Cancer Society.

Efforts to develop more efficient therapies have been mostly unsuccessful. No major breakthroughs have been made since gemcitabine was approved as the standard

chemotherapeutic agent for PDAC in 1997, when an only moderate overall survival could be reported, compared to 5-fluorouracil (5-FU) therapy [7]. The most recent advancement has been the introduction of the Epidermal Growth Factor Receptor (EGFR) inhibitor Erlotinib in parallel with gemcitabine treatment, extending the median survival by only14 days [8].

2.1.1 Molecular and genetic aspects of PDAC

It is speculated that early detection and a better understanding of the particular molecular mechanisms driving pancreatic ductal adenocarcinoma could help thwart the current therapeutic challenges. Risk groups have been identified, based on environmental and genetic factors. Smoking and alcohol consumption have been linked to an increased risk of developing PDAC [4]. Diabetes patients are also more likely to be diagnosed with PDAC, although it is not yet clear whether their increased body index represents a risk factor – or early symptom of the malignancy [4]. The description of familial cases of PDAC has also suggested the presence of genetic predisposition factors. Members of families with a history of developing pancreatic cancer have a six-fold increased risk of developing the disease themselves, while living in families with sporadic PDAC cases also poses a twofold increased risk of being diagnosed with the same ailment [4]. Mutations in BRACA2, PALB2 (partner of interaction for BRACA2), SERPIN B12, ATX, MLH1, as well as ARF/INK4A germ line mutations have been listed as putative genetic risk factors for PDAC familial cases [4, 9].

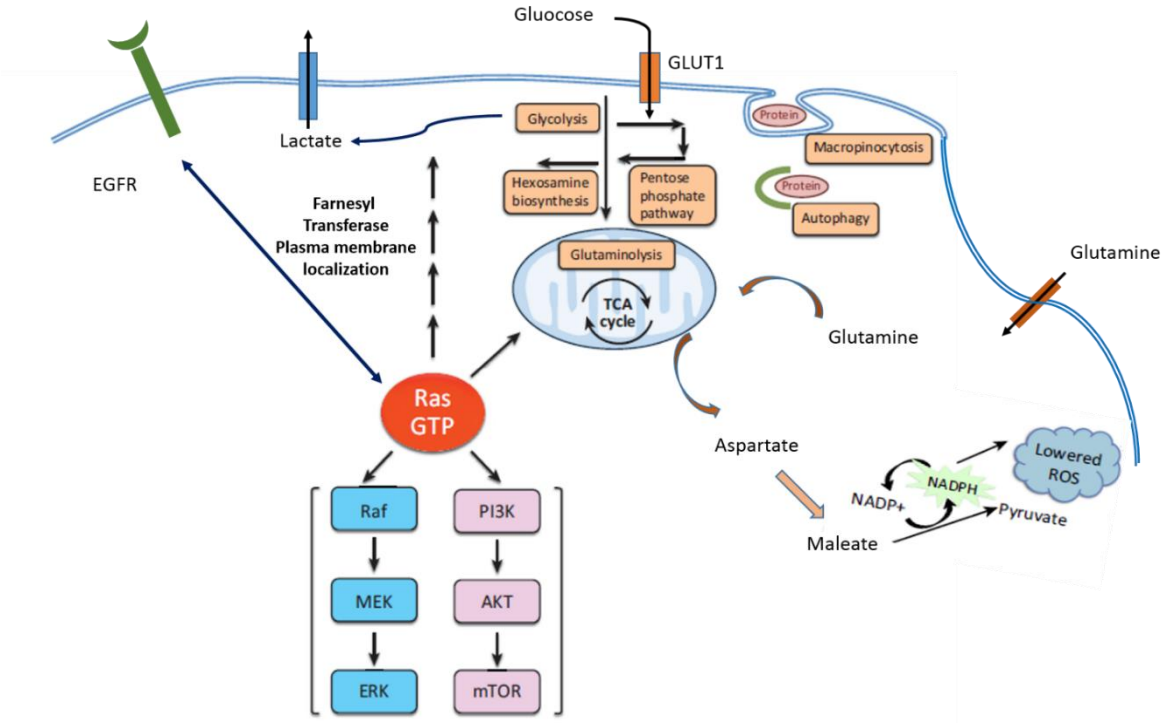


Figure 2: KRAS driven metabolic reprogramming in PDAC: activated KRAS has been shown to facilitate glucose and glutamine via the upregulation of specific transporters. Phosphorylated glucose is also shunted into the non-oxidative pentose phosphate pathway, leading to increased nucleic acid production. Oncogenic KRAS has been involved in glutamine reprogramming, leading among others to increased maleate production, responsible for balancing the intracellular ROS levels. KRAS also influences autophagy in pancreatic tumor cells. Plasma membrane localization of Ras is regulated by specific farnesyl transferases. Adapted from Bryant *et al*. [10].

The main genetic event driving PDAC progression, metabolism and proliferation appears to be KRAS mutation, described in approximately 95% of identified cases [10]. This small GTPase can be activated via a series of activating mutations, the most common ones being three point mutations, affecting the residues G12, G13 and Q61. These genetic activator mutations can either lead to the formation of steric hindrance, preventing GTPase activating proteins from binding to KRAS (the G12D amino acid substitution, occurring in 98% of PDAC cases), or the mutated amino acid prevents the correct coordination of a water molecule needed for the hydrolysis of bound GTP (Q61 substitutions) [10, 11]. Interestingly, KRAS mutation appears to be an early event in the progression of pancreatic cancer, since it has already been reported in over 90% of pancreatic intraepithelial neoplasia (PanIN) pre-cancerous lesions [5, 10, 11]. KRAS mutation is subsequently followed by inactivation of cyclin-dependent kinase inhibitor 2A, p53 and SMAD4 [10, 11]. Oncogenic KRAS is involved in a plethora of signaling pathways responsible for increased proliferation and inhibition of apoptosis, altered metabolism and microenvironment, promotion of inflammation leading to metastasis and evasion of immune response (**Figure 2**) [10, 12].

2.1.2 Extracellular matrix characteristics in PDAC

One particular characteristic of PDAC is the highly fibrotic tumor microenvironment, created by a dense extracellular matrix (ECM), leading to increased interstitial fluid pressure and the subsequent collapse of many tumor-associated capillaries [8, 13]. Up to 80% of PDAC tumors are necrotic, with little vasculature and a highly expressed stromal compartment. The tumor-associated ECM is rich in collagen and glycosaminoglycans, among which hyaluronic acid is one of the major ingredients [8, 13]. Responsible for the matrix deposition are activated pancreatic stellate cells, which represent myofibroblast cells normally associated with the exocrine pancreas [14, 15]. These essential niche components typically switch between inactive (containing vitamin A lipid droplets) and active state (when they start expressing α-SMA), and can be activated by a plethora of factors, including transforming growth factor-β1 (TGF-β1), tumor necrosis factor α (TNFα), interleukins such as IL1, IL6, oxidative stress, etc. [14, 15]. In the active form, pancreatic stellate cells are responsible for the massive production and remodeling of ECM components, while they also release cytokines, chemokines and adhesion molecules. Moreover, they can perform phagocytosis, similar to liver Kupffer-cells [15].

Although it may seem counter-intuitive, the abundance of hyaluronic acid in the tumor microenvironment and the increased interstitial fluid pressure – both leading to the formation of extended hypoxic areas in the pancreatic tumors, appear to benefit the progression of PDAC rather than inhibiting it [8]. One reason might be that hyaluronic acid and other ECM components can bind surface receptors and lead to the activation of signaling pathways responsible for cell proliferation and survival of cancer cells [13]. It has recently become acknowledged that hypoxia and ECM signaling play important roles in cancer metastasis, and intra-tumoral hypoxia appears to be a pathway for metastasis rather than an impediment [16]. Several speculations exist regarding the mechanisms employed by pancreatic cancer cells to tackle with the hypoxic and nutrient depleted environment: they could activate alternative metabolic pathways, increase their glucose intake, increase autophagy (see also **Figure 2**), etc. [10, 13].

The hypo-perfusion of PDAC tumors also impends the diffusion of chemotherapeutic compounds, explaining the clinical failure of treatment agents that showed positive results in early *in vitro* and animal based *in vivo* tests [8]. Treatment with stabilized hyaluronidases has been shown to lead to increased perfusion of PDAC mouse tumors and improve response to chemotherapy. Simply treating tumor bearing mice with hyaluronidase prolonged the overall survival of the animals [8], which could be explained by improved infiltration of activated immune system components, or could be due to disruption of paracrine signaling in response to alterations of the tumor-associated ECM. Despite concerns that increasing tumor perfusion might lead to increased metastasis, no such observation could be reported in the initial experiments [8].

However, not all experimental data support the idea that stromal ablation is required for improved PDAC treatment. A recent report by Özdemir and colleagues pointed out that cancer-associated α-SMA positive myofibroblast depletion lead to the development of more invasive, non-differentiated tumors, with enhanced hypoxia, EMT and decreased survival of tumor-bearing mice compared to animals whose stromal compartment was intact [17]. Depletion of α-SMA positive cells correlated with decreased collagen I expression and general tumor ECM remodeling. The report suggested that the tumor-associated stromal compartment might represent a protective response of the organism rather than a tumor survival mechanism, and cautioned against stromal targeting therapies in PDAC until additional experiments are performed [17].

2.1.3 Precursor lesions in PDAC

Evaluation of pancreatic tissue collected from patients suffering from chronic pancreatitis, or PDAC patients who were eligible for surgical removal of the primary tumor, revealed the presence of histological abnormalities, representing early dysplastic lesions in the progression of pancreatic cancer. These observation come in agreement with observations that other adenocarcinomas also go through an adenoma to carcinoma progression, a stepwise process of accumulating genetic and morphologic abnormalities [5, 18, 19]. Three types of precursor lesions have been described, based on their morphological characteristics: pancreatic intraepithelial neoplasms (PanIN), and two type of cystic lesions – intraductal papillary mucinous neoplasms (IPMN) and mucinous cystic neoplasms (MCN) [5, 9]. A schematic representation of the three premalignant lesions is represented in **Figure 3**.

PanINs represent the most common premalignant lesion of the pancreas [9, 19]. Usually asymptomatic, PanINs contain columnar cells which lose their cuboidal architecture and increase production of mucins, while exhibiting various levels of cytological and genetic abnormalities [5, 18]. When discovered in patients suffering from chronic pancreatitis, they are believed to signal a high risk of progression to carcinoma, but unfortunately the lesions can only be detected through invasive biopsy procedures, as they are invisible by cross-sectional imaging or endoscopic ultrasound [5, 18]. Interestingly, PanINs have been described in as much as 30% of autopsy specimens collected from elderly patients [9].

Based on the cytological criteria, PanINs can be classified in three grades: low grade PanIN-1A (flat), PanIN-1B (papillary), followed by intermediate grade PanIN-2 (displaying mild cytological atypia) and finally high grade PanIN-3 lesions, also described as *'carcinoma in situ'*, characterized by severe tissue abnormalities, although still confined by the basement

membrane [5]. KRAS mutation is an early event in the development of PanINs, activating mutations being already present in stage 1 lesions, and are followed, as the cellular atypia progresses, by loss of tumor suppressor genes such as p16, p53, SMAD4 [9]. Telomere shortening is also an early event and leads to the increased expression of telomerase by the dysplastic cells [5, 18]. Mouse models expressing mutated KRAS (G12D substitution) under the control of PDX1 transcription factor promoter (essential for the maturation of pancreatic and β cells), conditionally knocked out for p53, develop the full range of PanIN lesions, culminating with aggressive PDAC. Therefore, mouse models confirmed the hypothesis that PanINs represent precursor lesions of pancreatic cancer [11, 19].

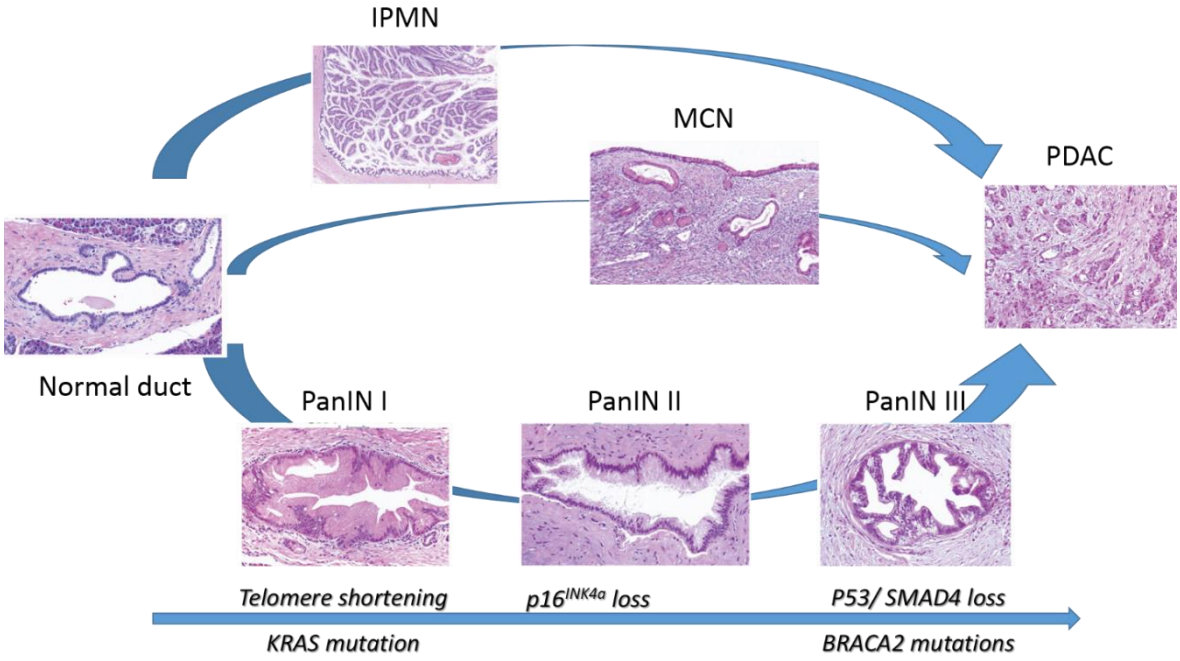


Figure 3: Precancerous lesions and the sequential accumulation of genetic abnormalities leading to the formation of PDAC. Adapted from Hezel *et al*. [9].

Of interest is the fact that morphological and gene expression changes specific for epithelial to mesenchymal transition (EMT) have been described early in the progression of PanIN lesions, raising concerns that metastatic spread in PDAC can represent an early event [11]. The results were confirmed using two mouse models harboring activated KRAS G12D mutations under the conditional control of two different promoters – the above mentioned PDX1, and the Mist1 transcription factor promoter, required for complete differentiation of acinar cells in mice [11]. Early accumulation of genetic and epigenetic abnormalities is accompanied by increased expression of atypical protein markers, which can be later found in fully developed PDAC and metastatic lesions, such as aberrantly glycosylated MUC1 [20], or secreted protein lipocalin-2 (LCN2) [21]. This raises the prospect of using protein biomarkers for detection of early pre-cancerous lesions, with profound clinical applications.

IPMNs are invasive carcinoma precursors, showing an increased incidence in recent years. IPMNs have been described as tumors of the duct epithelium, characterized by papillary epithelial proliferation, increased mucin production and cystic dilatation of affected ducts [5]. According to their site of origin, IPMNs are divided into main duct (MD-IPMN) or branch duct (BD-IPMN); when both duct types are involved, the lesions are termed “mixed duct type”. Based on the histological properties, IPMNs can be subdivided into intestinal, pancreato-

biliary, oncocytic and gastric, with the first three types usually originating from the main duct, while the gastric subtype arises typically from branch ducts [5]. From the listed IPMN subtypes, the intestinal subtype is one of the most often described pre-malignant lesion, being characterized by tall columnar cells, elongated nuclei and amphophilic cytoplasm [5]. At a molecular level, many of the described genetic abnormalities are similar to those already identified in PanIN lesions. Additionally, a common mutation affects the GNAS locus [5]. Unlike the PanIN lesions, IPMNs developing from the branch duct can be detected by MRI, and early treatment for these aggressive lesions is therefore possible [5, 19].

MCNs represent the most uncommon type of pancreatic precancerous lesion, more often developing in women, which are usually located in the pancreatic body or tail [5]. The lesions are asymptomatic and usually discovered by accident; prognosis is usually good, but surgical removal is advised for all patients. If untreated or undiagnosed, MCNs can grow very large, and include calcifications. The presence of mural nodules signals a high risk of malignancy. Under the microscope, the cells underlying MCN lesions appear columnar and express abundant mucins, like all the previously described precursor lesions. KRAS mutations and inactivation of tumor suppression genes p16, p53 and SMAD4 are a common event [5, 18]. A characteristic of MCN is the presence of ovarian like stroma associated with the atypical epithelial structures, expressing progesterone and estrogen receptors [9]. Therefore, the stroma can undergo luteinization similar to that described in the normal ovarian stroma. Despite the low risk of malignancy, diagnosed MCNs are typically surgically removed [5].

2.1.4 Molecular subtypes in PDAC

Despite the increased understanding of the molecular mechanisms driving pancreatic cancer, novel therapies designed for PDAC continue to fail in clinical trials, and even response to standard therapy is extremely heterogeneous. One of the major problems is that PDAC has been long viewed as a homogeneous disease [8].

This monodimensional approach has been challenged by Collisson *et al.*, when the analysis of microdissected pancreatic cancer cells and commercially available cell lines revealed differences in their transcriptomic profiles, not described previously [3]. Based on the gene expression level, they were able to describe three molecular subtypes, differing in their response to chemotherapy and overall survival [3]. The three molecular subtypes were termed by the authors: classical, exocrine-like and quasi-mesenchymal. Furthermore, they discovered that the molecular subclassification can be achieved by performing gene enrichment analysis using only 62 gene signatures, which they designated as a PDA (pancreatic ductal adenocarcinoma assigner) [3].

Based on the gene expression profiles, they observed that the classical subtype expresses high levels of epithelial-associated and adhesion molecules, the quasi-mesenchymal (QM) subtype is enriched in mesenchymal associated genes, while the exocrine-like subtype continues to express digestive enzyme genes [3]. In their initial study, patients harboring the classical subtype showed better overall survival than the other two subtypes, with the QM patients presenting the worst prognostic; the molecular subclassification of the patients correlated better with the overall survival than previously used factors, such as tumor size, degree of differentiation, etc. [3].

The stratification of patients according to these three molecular subtypes proved to be beneficial for therapy selection as well. The classical subtype appeared to be more dependent on KRAS signaling than the other two described subtypes. It comes therefore as no surprise that the classical subtype responded best to EGFR inhibitory therapy with Erlotinib [3]. This addiction of the classical subtype to KRAS signaling could allow patients harboring the classical subtype to benefit from KRAS targeted therapies, which might one day become clinically available, considering the increased interest manifested in their development [3, 7, 22]. By contrast, the QM subtype responded best to gemcitabine treatment [3].

The exocrine-like subtype response to chemotherapy could not be evaluated *in vitro* by Collisson *et al*., due to its complete absence from commercially available serum grown cell cultures [3]. Initial concerns that the exocrine-like subtype might represent an artifact introduced by the statistical analysis were refuted after additional *in vivo* analysis. Moreover, Dr. Christian Eisen (HI-STEM , DKFZ, Heidelberg, Germany) confirmed the existence of this subtype *in vitro* after developing serum-free cell culture methods (Noll, Eisen *et al*., manuscript in preparation). Furthermore, Eisen and collaborators established that the exocrine-like subtype pancreatic cancer cells exhibit increased resistance to several chemotherapeutic agents used for the treatment of PDAC, and this resistance could be linked to activation of detoxification pathways typically expressed by hepatocytes (Noll, Eisen *et al*., manuscript under preparation). Although at first disheartening, the proof that exocrine-like PDAC patients would not benefit from standard chemotherapies spells out the importance of developing optimal stratification strategies of PDAC patients, in parallel with the development of novel, targeted therapeutical strategies.

2.1.5 PDAC therapies – standard of care and novel approaches

Surgery continues to represent the best treatment, despite the fact that the overall survival of patients with resectable tumors does not raise above 20% [5]. The surgical approach is dependent on the location of the primary tumor. Patients with tumors situated in the head of the pancreas will typically undergo a Kausch-Whipple pancreatoduodenectomy. In contrast, pancreatectomy – together with hilar and spleen nodes removal is performed in patients with tumors present in the body or tail of the pancreas [23].

Neoadjuvant therapies have been proposed in pancreatic cancer, as a means to shrink the primary tumor and allow surgical removal. Neoadjuvant therapy can be administered as soon as the tumor is diagnosed [24]. Examples of neoadjuvant therapies investigated in PDAC include: radiotherapy, chemotherapeutical agents – 5-FU, gemcitabine, cisplatin, used as monotherapies, or in combination [24]. A potential drawback of opting for neoadjuvant therapies is the fact that they delay surgical intervention [7, 25], thus potentially allowing the pancreatic tumor to advance to more aggressive and metastatic levels.

Adjuvant therapies are administered to patients after undergoing surgical resection, and include radiotherapy, chemotherapy – or a combination of both [24]. Historically, 5FU has been the chemotherapeutic of choice, until in 1997 gemcitabine was proven to improve the overall survival (albeit moderately) and the five year survival rate [7]. Gemcitabine is a nucleoside analogue designed to trigger apoptosis in fast dividing cells by stalling DNA replication after its incorporation in place of a normal cytidine [26]. Novel therapies have been attempted in combination with gemcitabine, but despite early promising results *in vitro*

and in early clinical trials, they all failed to reach the goal of at least significantly improving patient survival [27].

With such a dismal prognosis, even for patients diagnosed with resectable PDAC, efforts have been made recently to develop targeted therapies against pancreatic cancer. The tyrosine kinase inhibitor Erlotinib has been the only targeted therapy in combination with gemcitabine to receive FDA approval for the first line treatment of PDAC patients, after a small (about two weeks), but statistically significant increase in overall survival could be observed for the treatment arm receiving the combined therapy compared to gemcitabine treatment alone [7]. An alternative approach has been to target EGFR with the chimeric monoclonal antibody Cetuximab, but despite promising initial results, stage II and III trials failed [27].

One desired target for PDAC therapies is KRAS, since activator mutations are a common and early event in pancreatic cancer. However, the development of targeted therapies against KRAS are not trivial, and early attempts have failed in clinical trials. One such example is the use of farnesyltransferase inhibitors, aiming to disrupt the membrane association of KRAS [7, 22]. However, an additional mechanism of KRAS membrane association, through geranyl-geranylation can circumvent the farnesyl transferase inhibitors. Therapies aiming to inhibit both farnesylation and geranylation are prohibited by the anticipation of high toxicities [7]. Novel strategies are being currently considered, targeting specific pockets formed in G12C KRAS mutants [28], but the efficiency of these approaches remains to be tested in clinical trials.

Vascular Endothelial Growth Factor (VEGF) inhibitors have also represented an area of interest for PDAC, since it is overexpressed in 93% of the patients [7]. However, clinical responses have not been encouraging, as both antibody based therapies with Bevacizumab and VEGF receptor inhibitors Sorafenib and Axitinib have failed to bring significant improvements in the overall survival when combined with gemcitabine and Erlotinib [7].

Some strategies aim to target the PDAC associated stroma. Examples include the use of Sonic Hedgehog (Shh) inhibitors (NCT01064622), and the engineered hyaluronidase PEGPH20 (NCT01839487) [7], which recently entered phase II clinical trials.

Antibody-based therapies have also been considered, in combination with pre-existing chemotherapeutics or other targeted approaches. Emerging passive immunotherapies are targeting proteins overexpressed by PDAC, independent of their function in promoting carcinogenesis. One such example is the single chain Fv murine antibody SS1P recognizing mesothelin (NCT00006981), a glycosylphosphatidylinositol anchored protein overexpressed in 90-100% of PDAC cases [6, 7]. Immune checkpoints blocking therapies aiming to reverse tumor associated immune suppression are also currently tested in PDAC [29, 30]. The anti-CTLA4 antibody Ipilimumab, administered in combination with allogeneic pancreatic cancer cells transfected with GM-CSF, have revealed a slight increase in overall survival, compared to Ipilimumab treatment alone (NCT00836407).

2.2 Biomarker discovery and proteomics

2.2.1 Molecular biomarkers – definition and identification

According to the American National Institute of Health (NIH), a biomarker is defined as: “a characteristic used to measure and evaluate objectively normal biological processes, pathogenic processes, or pharmacological responses to therapeutic intervention” [31, 32]. Therefore, biomarkers represent valuable tools for early disease detection and differential diagnostic, providing prognostic and predictive information, while also allowing the monitoring of high-risk populations. Additionally, they also support the development and selection of targeted therapeutics [31].

Biomarkers can be classified according to different criteria. Based on their general characteristic, we can divide them in biomarkers for imaging and molecular biomarkers. Molecular biomarkers have biophysical properties allowing their detection and quantification, and include nucleic acid based biomarkers (DNA, RNA), peptides, proteins, lipids as well as small metabolites [32, 33]. Based on their utility, biomarkers can also be classified in diagnostic biomarkers, disease staging biomarkers, disease prognostic biomarkers, or biomarkers for monitoring of therapy response. Finally, the NIH ‘Biomarkers and Surrogate Endpoint Working Group’ has defined three types of biomarkers – type 0 (markers of the natural history of a certain disease, correlating with clinical indices), type I (markers measuring drug activity) and type II (also known as surrogate markers) [33]. A surrogate marker is expected to be a reliable substitute for a morbid event end-point, and must act as both an epidemiologic marker as well as a therapeutic responder [34]. This subchapter will focus on molecular biomarkers and experimental screening procedures employed for their detection.

Genomic approaches for biomarker investigation are typically evaluating differences in gene expression profiles or variation in the genetic sequence which can be associated with a particular pathological condition. For example, DNA-microarrays and next generation sequencing techniques allow the investigator to identify gene mutations, copy number variations and single nucleotide polymorphisms (SNP) [32, 35]. The field of transcriptomics, analyzing differences in RNA transcripts expression over time, and between healthy and diseased cells, has developed an array of sophisticated methods, widely used in biomarker discovery experiments. Several RNA species can be investigated, including mRNAs, non coding RNAs and small RNAs [36]. A series of techniques have been developed over the years and are being continuously optimized for the evaluation of mRNA levels such as: Northern Blot, serial analysis of gene expression (SAGE), RNA-seq; additionally, DNA-microarrays can as well be used in transcriptomics experiments [35-37]. RNA-seq has the advantage that it does not require predefined probes, while the coverage of the whole transcriptome of an organism can be achieved. Moreover, it can also be used to distinguish between alternative-splicing products in various pathological or developmental conditions [35, 36]. Genomic approaches have the advantage of being suited for high-throughput investigations, while the overall cost of such approaches has decreased significantly over the last decade [38]. However, genomic experiments have several caveats. For example, single cell RNA sequencing has been associated with increased noise when estimating expression levels of low copy mRNA species, while at the same time being accompanied by the introduction of a 3’ end amplification bias [35]. Moreover, the quality control guidelines for many genomics methods are not clearly defined. For this reason, additional

validation techniques following transcriptomics-based discovery experiments, such as RT-qPCR, are often mandatory [36]. But the most important draw-back of genomic approaches is the fact that variation in mRNA copy numbers or genetic mutations cannot always accurately predict the actual final gene product levels in the cell/tissue, nor can they predict the complex post-translational modifications (PTMs), alternative splicing, or allosteric interactions of the respective encoded proteins [39-41]. Nonetheless, the recent publication of the human proteome draft by Wilhelm *et al.* suggested that, at least for healthy tissues, the translation rate might represent a constant characteristic of a given transcript, implying that the mRNA levels can often be employed for predicting the abundance of the final protein product [42]. Of importance, some nucleic acid species can potentially be used as surrogate biomarkers, and they can be easily detected in biological fluids. For example, microRNAs (miRNAs) represent small endogenous non-coding RNAs, responsible for the post-transcriptional regulation of target genes, playing important roles in many physiological and pathological processes [43]. Several reports have indicated dysregulations of circulating miRNAs in the blood of patients suffering from different disorders, and miRNAs are currently intensely investigated as cancer prediction biomarkers [44].

Proteomics approaches for biomarker discovery include 2D-PAGE based methods, mass spectrometric identification, antibody microarrays and tissue microarrays. Differential Gel Electrophoresis (also known as 2D-DIGE) has been for almost a quarter of a century the work-horse in proteomics [45]. However, gel based methods are laborious, require high amounts of starting material, and have been plagued by problems associated with the solubility of hydrophobic trans-membrane proteins [46, 47]. Antibody based proteomics has traditionally been widely used, due to its rapid, sensitive, and high-throughput options [48]. However, antibody based techniques such as Enzyme-linked immunosorbent assay (ELISA), protein micro-arrays and immunohistochemistry (IHC) are dependent on the availability of highly specific antibodies, with the ability to recognize splice isoforms or PTMs [48]. In addition, ELISA and tissue micro array (TMA) based identifications are difficult to multiplex [39, 48]. Protein micro-array/protein chips represents a high-throughput method, developed on a technology similar to the DNA micro-array chips. Protein chips can be grouped into analytical arrays, functional proteomics micro-arrays and reverse phase protein micro-array [49]. The difference between these array types is dictated by the nature of the macromolecule immobilized on the chip's surface. Analytical arrays contain trapped antibodies, affibodies (small engineered affinity proteins) or aptamers (single stranded oligonucleotides which bind their target with high affinity and specificity), which will capture the corresponding proteins against which they were raised and detection can be achieved using ELISA type methods [31, 49]. Functional micro-arrays contain immobilized whole proteins which can be used for the investigation of protein-protein/protein-DNA or RNA/protein-phospholipid or protein-small molecule interactions. Finally, in reverse phase protein micro-arrays, complex protein extracts (such as cell lysates) are immobilized and detected by the use of an antibody [31, 49, 50]. Protein arrays suffer from the same limitation earlier mentioned for ELISA and TMA antibody based methods. Additionally, protein chips need to immobilize proteins on solid surfaces without leading to loss of tertiary and quaternary structures, and the subsequent recovery of the corresponding proteins from the chip for further analysis is difficult [49]. Moreover, considering the wide dynamic range of the proteome [51], it is difficult to construct protein arrays with a sufficient capacity for a complete proteome analysis.

Mass spectrometry has become the method of choice for the majority of proteomics based biomarker discovery projects. The development of the two stable soft ionization techniques – Electro Spray Ionization (ESI) and Matrix-Assisted Laser Desorbtion/Ionization (MALDI) – allowing the concurrent ionization and transfer of the biological analytes into the gas phase without extensive degradation, represented one of the major instrumentation advancements, leading to the development of modern day proteomics [45, 52]. Efforts made to improve the quality of digitally available databases for spectra identification, and the recent publication of the first drafts of the human proteome allow for advanced MS based proteomics biomarker identification [53] Mass spectrometric based biomarker identification has the advantage that it can process complex biological samples, without the requirement of introducing any pre-analysis hypothesis, thereby eliminating any prespecification associated biases. However, the complexity of the samples analyzed, in particular the wide dynamic range of mammalian proteomes, affects the sensitivity of the method and leads to stochastic sampling [51, 54]. Instrumentation advancements over the last decade have also facilitated MS-based discovery experiments. Significant improvements in the acquisition rate translated in an increased number of identified proteins per minute. For example, a recent study by Hebert *et al.* reported the identification of over 4000 yeast proteins based on a 1.3 hours chromatographic-MS method, at an impressive identification rate of 67 proteins per minute [55]. A summary of modern MS methods will be presented in the following chapter.

Recently, a novel technique termed mass cytometry has been developed combining the advantages of fluorescence assisted cell sorting (FACS) and mass spectrometry. Briefly, the method utilizes antibodies labeled with lanthanide atoms (absent from biological samples), which recognize and bind their specific target, after which the cell is vaporized in a CyTOF analyzer and the corresponding atom is detected [56]. The method has been successfully used to detect differences in cell-surface protein expression under different treatment conditions and has the advantage of eliminating spectral overlap (a common problem in FACS sorting). On the down side, the sample is vaporized during the analysis, therefore no additional evaluations can be performed [56]. Additionally, the method does not address the problems associated with the use of antibodies, plaguing all antibody based detection techniques.

The field of lipidomics investigates alterations in lipid compositions, biosynthesis or downstream metabolism. Lipids are complex macromolecules, exhibiting a high structural diversity, which can include neutral molecules (triglycerides, sterols), polar lipids (glycerophospholipids), and small signaling molecules, such as eicosanoides, comprising numerous isomers [36]. One of the most widely use lipidomics marker in the clinic is cholesterol, which, depending on its association with low-density or high-density lipoproteins, represents a prognosis marker for patients at risk of developing atherosclerosis [57]. Currently, lipidomics discovery experiments are mostly performed using MS based methods [36]. However, the coverage of the lipidome by mass spectrometry is limited by the scarcity of data bases and spectral libraries. The development of novel lipid based bioinformatics resources and software tools (e.g. the LIPID MAPS lipidomics gateway), is one step forward in the direction of improved lipidome discovery and quantification experiments [36].

Metabolomics is the field analyzing the whole metabolome under a given set of physiological, environmental, and/or clinical conditions [36]. An exact definition of the

metabolome is more difficult to pin down, however, it can be generally referred to as the "quantitative element of all the low molecular weight molecules present in a particular physiological and developmental state" [36]. Historically, two major techniques have been employed for identification experiments in the field: MS and NMR (nuclear magnetic resonance) [58]. MS based metabolomics techniques can analyze a large number of unrelated metabolites, with a wide range of chemical properties. Sample prefractionation by chromatographic methods is a prerequisite, and current platforms allow high resolution screening and quantification for a variety of metabolites. One limitation of LC-MS based metabolomics is the scarcity of tools for data annotations, and the fact that the high chemical variety of different metabolites makes it (at least at the moment) impossible to measure the entire metabolome of one sample [36]. An alternative approach has been to use MS based proteomics to investigate the PTMs and general expression patterns of proteins involved in known metabolic pathways [59]. NMR techniques have the advantage that they do not require any sample prefractionation prior to analysis. Additionally, it is a nondestructive, noninvasive technique, providing detailed information of the metabolite's molecular structure, based on atom-centered nuclear properties and interactions [58]. However, the interpretation of NMR spectra is not trivial. For clinical applications, kit based technologies have also been developed [36].

2.2.2 Short introduction to proteomics and mass spectrometry

Proteomics represents the large scale systematic study of protein structure and function in an organism, tissue, cell, organelle or specific pathway [54]. Several techniques can be employed for the study of proteins, however, currently mass spectrometry represents the most common methodological approach.

All mass spectrometers use the analysis of a mass-to-charge ratio for the identification of proteins. The most important components of a mass spectrometer are the ion source, the mass analyzer and the data processing electronics [52].

The polarity, volatility and thermal instability of proteins and peptides hindered for a long time their non-destructive transfer into the gas phase. Thus, the use of mass spectrometry in proteomics has been limited by the development of soft ionization techniques, compatible with these analytes. However, the development of two stable ionization techniques: ESI and MALDI in the late 1980's has allowed for a more stable ionization of macromolecules and the subsequent development of MS based proteomics methods [45, 52, 60].

In MALDI, analytes are co-crystalized with a matrix (selected based on the purpose of the experiment). The matrix will absorb the energy of the laser and transfer it to the analyte, while the rapid pulsed laser heating leads to the desorption of the matrix and analyte ions into gas phase [52]. The development of robotic systems capable of performing automated mixture of the matrix of choice and analytes eluted from a chromatography column has simplified MALDI sample preparation, initially viewed as too laborious [45]. MALDI generated ions are usually single charged [52]. Several variations of matrix associated ionization techniques have been developed, including Surface-assisted laser desorption/ionization mass spectrometry (SALDI) and Desorption/ionization on silicon (DIOS), using porous graphite, and silicon, respectively, to increase the tolerance of MALDI MS methods to detergents and salts [52]. Another variation of the technique – termed Surface-Enhanced Laser Desorption/Ionization (SELDI), allowed evaluation of biochemical

activity and subsequent enrichment of a particular subset of the analyzed sample. In SELDI the proteins/peptides of interest are first deposited on a surface modified with a chemical functionality, allowing the selective binding of distinct target analytes. The corresponding matrix is added after non-bound peptides/proteins are washed away [61].

On the other hand, in ESI, an electrically charged spray is created by applying a high voltage between an emitter at the end of a separation pipeline (e.g. a capillary chromatography column) and the inlet of the mass spectrometer [52]. ESI generated ions are usually multiply charged [52].

Mass analyzers sort the analyte ions based of their mass-to-charge (m/z) ratio. Different mass analyzers have been built, employing different separation strategies. Ion trap, ion cyclotron resonance and orbitrap instruments analyze the resonance frequency of ions. Quadrupole systems use the stability given by m/z in an oscillating electric field to selectively stabilize the paths of distinct analyte ions. Finally, time of flight analyzers use the flight time of ions accelerated under the same potential to separate the different analytes. Hybrid mass analyzers have been built, increasing the versatility of the analysis [52]. ESI systems can be attached to a variety of mass analyzers and are often preferred due to the possibility of coupling them directly (online systems) to a liquid chromatographic separation system. MALDI systems are often coupled with time of flight (TOF) analyzers, and can only be used offline, which increases the overall analysis time. However, MALDI based analysis is highly stable and has the advantage of allowing long term storage and subsequent re-analysis of a sample of interest [45].

Two types of distinct proteomic approaches can be used for protein identification: bottom-up approach, and top down proteomics [52].

Bottom-up approaches are currently the most popular tactic employed in proteomics, when dealing with complex samples. For this method, also known as shotgun proteomics, proteins are digested prior to performing the analysis and the resulting peptide masses and sequences represent the basis for protein identification, similarly to the DNA shotgun sequencing [52]. Peptides generated by enzymatic digestion are later dissociated into smaller fragments, using different dissociation methods, such as collision induced dissociation (CID) or electron transfer dissociation (ETD). The subsequent tandem mass spectrometric data (MS^2 or MS/MS) are compared to *in silico* generated fragmentation patterns available in public databases [52, 62]. The bottom-up approach has the advantage of being highly sensitive, fully automatable and fractionation techniques are easily applicable for peptides. Drawbacks of the method include the possible loss of PTMs, ambiguity of the origin of redundant peptide sequences and limited sequence coverage, depending on the ionization method and the pre-fractionation/enrichment methods applied [45, 52].

Top-down proteomic approaches are less commonly used, but have recently attracted more interest, after the development of highly sensitive mass analyzers. In top down proteomics, whole proteins are analyzed and tandem mass spectrometry is performed after using alternative fragmentation methods, such as ETD and electron capture dissociation (ECD). Advantages of top-down proteomics include a better sequence coverage, identification of specific isoforms and easier detection of complex PTMs. However, front end separation techniques are more challenging for intact protein mixtures, and the method is dependent on the development of high mass accuracy instruments, such as Fourier Transformation MS or Orbitrap instruments [52, 63].

The majority of tandem MS analyses are performed in a data dependent acquisition mode, meaning that the data collected after the first MS analysis is used for the selection of precursor ions that will be fragmented and are subject to a second round of MS analysis, following a dissociation step [64]. Lately, the lab of Rudi Aebersold, has been developing a data independent acquisition method, termed SWATH MS, in which all precursors contained in a predetermined window are fragmented and the resulting spectra are recorded [48, 64].The method promises to allow improved quantification for most of the proteins in a sample [64], but complicates the subsequent bioinformatics analysis.

MS based proteomics can be used for the identification, and quantification of protein expression between complex biological samples, evaluation of PTMs and associated metabolic pathways, evaluation of protein-protein interactions, etc [41, 60]. Depending on the aim of the experiment, the general shotgun proteomics workflow includes sample selection, the preparation and optimization of whole protein extracts (which includes cell/tissue lysis, depletion of abundant proteins), enzymatic digestion, sample fractionation and finally the MS and MS^2 analysis [52, 60]. A schematic representation of a typical shotgun MS proteomic workflow is represented in **Figure 4.** The previously mentioned steps are common for both discovery and targeted proteomics, which will be discussed in greater detail in the following chapters.

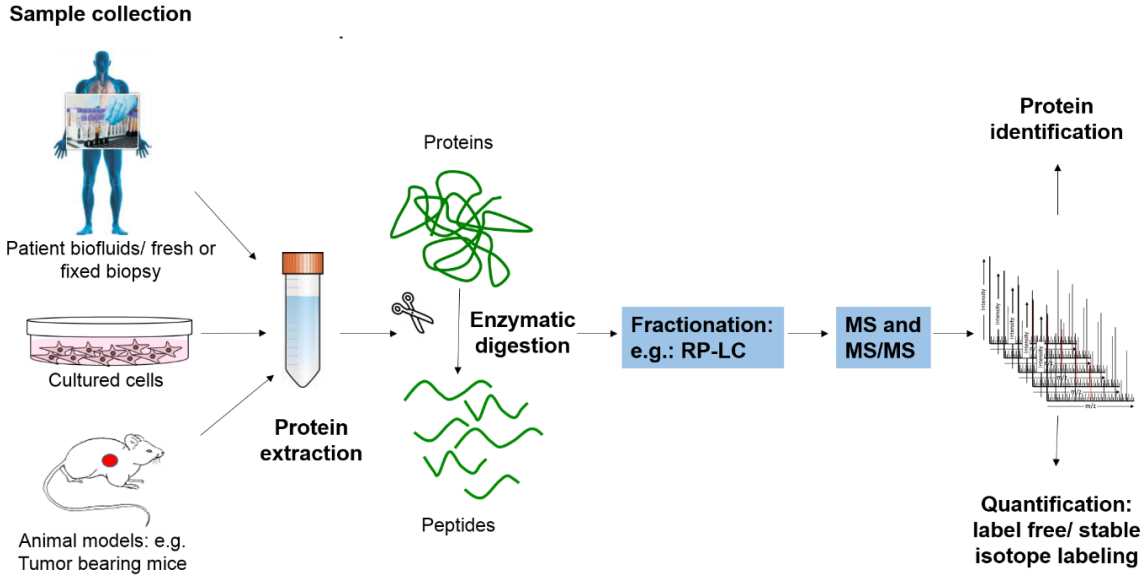


Figure 4: Schematic illustration of the shotgun proteomics workflow. For simplification purposes, isotope labeling techniques used for the quantification steps are omitted. RP-LC = reverse phase liquid chromatography.

Due to the high dynamic range of mammalian proteome, sample pre-fractionation is required in order to allow the mass spectrometric identification [45]. Some of the most commonly used prefractionation methods include reverse-phase chromatography and strong cation exchange chromatography. The two chromatographic methods can also be combined in a two dimensional prefractionation step (also known as multi-dimensional protein identification technology or MudPIT). Additionally, fractionation methods can enable the enrichment of a proteome subset, such as immobilized metal affinity chromatography (IMAC) for phosphorylated peptides, or lectin affinity chromatography for the enrichment of glycosylated proteins [47, 52, 62].

Finally, mass spectrometry based applications allow for the subsequent comparison and quantification of proteins from different samples. In order to obtain a snap shot of protein concentration characterizing different biological states, an experimenter can choose between label-based and label-free quantification methods [45, 52]. Label based methods make use of the ability of mass analyzers to distinguish between peptides containing different isotopic species. Heavy isotopes can be incorporated in peptide populations using metabolic, chemical or enzymatic methods, and a minimum of two distinct populations (labeled and unlabeled) can be analyzed in parallel. Relative quantification is done by comparing the intensity of co-eluting peaks, distinguished by their mass difference [45, 52]. Chemically labeling techniques commonly used are isobaric labeling for relative and absolute quantification (iTRAQ), isotope coded protein labelling (ICPL), etc [45]. Enzymatic labeling usually leads to the incorporation of ^{18}O isotopes in digested peptides, and is currently less applied [52]. Metabolic labeling is today one of the quantification methods of choice, with the development of SILAC (stable isotope labeling by amino acids in cell culture) and super-SILAC (mix of SILAC labeled cell lines) by the group of Mathias Mann, allowing the quantification of complex biological samples, for which metabolic labeling would not be an option [65]. Label free quantification methods have required the development of proper bioinformatics tools, but represents an appealing alternative to isotope labeling, since they are cheaper and less laborious to use. For better quantification and peak accuracy measurement, added internal standards are recommended prior to MS analysis [45]. It is important to mention that the minimization of processing steps is recommendable for all relative and absolute quantification methods, as they can introduce experimental artifacts [45]. Also, the optimization of the enzymatic digestion is essential for bottom-up quantification experiments, and the protein to enzyme ratio, incubation time and temperature are just a few of the parameters that should be optimized in order to obtain accurate quantitative proteomic results [66].

In conclusion, MS based proteomic approaches are versatile identification and quantification methodologies, although the high dynamic range of the eukaryote and especially mammalian proteome does not currently allow for the accurate simultaneous identification and quantification of all protein species. The experimenter must therefore optimize individual proteomic workflows, according to the aim of the experiment [54].

2.2.3 Clinical proteomics in cancer diagnostic and therapy – functional and methodological requirements

The field of clinical proteomics investigates the differences between healthy and diseased tissues/fluids, thus identifying and quantifying protein differences between the samples [54]. Consequently, clinical proteomics can help shed light on the pathology of the disease, enables better stratification of patients, while at the same time offering novel targets for the development of targeted therapies, thus aiding the development of customized/personalized medicine [48, 54].

Clinical proteomics has the advantage that it can identify and quantify a large number of proteins in complex biological samples, without requiring any pre-specified theory; however the broad and unbiased approach comes with the cost of reduced sensitivity and reproducibility, the latter due to stochastic sampling [39, 54]. Current proteomic techniques

have been grouped into discovery and targeted methods, the latter being employed in the validation phase of a biomarker discovery experiment [48].

In order for a cancer protein biomarker to be approved for clinical use, it needs to fulfill several requirements. Diagnostic markers need to be released by the tumor into the general circulation at amounts high enough to allow detection, should be specific to the tissue of origin and should be detected in early stages of the disease [32]. Additionally, biomarkers developed for targeted therapies need to be accessible from the blood stream – and can include antigens expressed on the surface of cancer cells or the associated stroma, the extracellular matrix, or markers expressed by the tumor neo-vasculature [67]. When validating putative biomarkers it has to be analyzed whether their levels can be altered by non-cancerous pathologies [32]. Therefore, experimental procedures employed for the identification and subsequent validation of a protein cancer biomarker need to be rigorous and reproducible, in order to successfully develop and implement the discovery into clinical practice.

The general protocols for biomarker discovery and validation include three important phases: the preanalytical phase (including sample collection), the analytical phase (the actual assay) and the postanalytical phase (data interpretation and subsequent validation strategies) [32]. Several precautions need to be taken in each phase in order to avoid investing time and resources in the development of an improper biomarker. During the preanalytical phase, proper controls need to be established and the cohort size should be sufficient to allow proper statistical prediction of putative biomarkers and avoid model overfitting. Sample handling and storage needs to be standardized and thoroughly planned, to avoid experimental artifacts, which could result from improper handling (such as proteins or other putative markers released by lysing blood cells) [32]. The proper selection and optimization of the proteomics protocol is crucial for the analytical phase. Sample pre-fractionation is often essential for the detection of cancer biomarkers, otherwise their signal could be eclipsed by the most abundant proteins present in the raw sample [41]. The final mass spectrometric analysis needs to be performed on an instrument with an appropriate sensitivity and dynamic range [31, 48]. Data interrogation and statistical analysis performed in the post analytical phase needs to be rigorous and an adequate number of clinical samples of high quality need to be available for proper validation [32]. Sample prefiltering (using correction methods, or false discovery rates) are often used to reduce the size of the data set, but could lead to overly optimistic predictions of the efficiency of a putative biomarker [68]. Finally, proper validation techniques, highly reproducible and accurately quantifying the putative biomarker, should be employed and will be discussed in greater detail in a later subchapter.

So far, the majority of the FDA approved biomarkers for cancer diagnosis include protein markers developed in the last decades of the 20th century such as: carcinoembriogenic antigen (CEA) for colorectal cancer, carbohydrate antigen 19-9 (CA 19-9) for PDAC, prostate specific antigen (PSA) for prostate cancer, carbohydrate antigen 125 (CA125) for ovarian cancer and human chorionic gonadotropin for seminoma [69]. However, each of the enumerated biomarkers has been reported to have limitations: either their levels are also elevated in benign diseases (e.g. PSA is upregulated in other benign prostate disorders), or their sensitivity is suboptimal [70]. Several other biomarkers have been considered as putative diagnostic or metastasis markers. Examples include mucine-1 (MUC1) (elevated and abnormally glycosylated in many cancers, including PDAC) [20], HER-2 (proposed as

a marker for invasive breast cancer, currently used only as a predictive factor for Trastuzumab treatment), human epididimus protein 4 (HE4), etc. [69, 70]. Some of the novel diagnostic secreted biomarkers have been reported to have similar sensitivities to already existing clinical approved markers, which initially discouraged their adoption into clinical practice. However, for those biomarkers exhibiting an increase specificity compared to the already approved markers, testing a combination of both old and more recent biomarkers has been suggested. Recently, a combination of CA125 and HE4 has received FDA approval to be used for distinguishing between benign and malignant pelvic masses [70]. Thus, combinational tests should also be considered when developing novel soluble diagnostic biomarkers for the clinic.

Combining a set of biomarkers is especially important, since it is difficult to identify proteins showing absolute specificity to a given tumor tissue, with the exception of seldom PTMs (e.g.: pancreatic ribonuclease in PDAC and kallikrein in ovarian cancer) [32]. Cancer specific protein isoforms, recapitulating splice patterns specific for embryonic development, have also been described in malignancies and other pathologies (e.g. rheumatoid arthritis, endometriosis, atherosclerosis). For example, splice isoforms of tenascin-C and fribronectin are specifically expressed in the extracellular matrix of several tumors at the sub-endothelial level [67]. Once a vascular accessible protein biomarker has been properly validated, it can be used for the development of targeted therapies, such as antibody based therapy. Cancer specific antigens can be targeted with a monoclonal antibody in native state or armed (conjugated with cytotoxic drugs, cytokines, radionucleotides). Different antibody formats can also be investigated (e.g.: diabodies, small immune proteins, full IgGs). The choice for the optimal delivery agent is influenced by its uptake in the normal and diseased tissue and by the clearance profile [67]. An alternative to the use of antibodies as delivery agents has been the development of high affinity ligands-based targeted cytotoxics [71].

In the next subchapters we will discuss in greater detail the specific sample preparation methods reported in literature for secreted and cell surface protein discovery and quantification. Finally, the last subchapter will be dedicated to commonly used validation methods and recent developments in the field of targeted proteomics.

2.2.4 Using mass spectrometry to identify membrane associated protein biomarkers

The membrane proteome has been defined as the entire complement of membrane proteins present in a cell under specific conditions at a specific time [72]. Proteins associated with the cell and organelle membranes play important roles in cell-cell communication, signal transduction and the transport of metabolites, ions and other solutes [47].

Membrane proteins can be classified based on different criteria. Based on their level of association with the plasma membrane, they can be divided into membrane spanning and peripheral membrane proteins. Based on their secondary structure, they have been classified into α-helix and β-barrel containing structures. And finally, based on the number of transmembrane helixes, membrane proteins can be classified into single pass proteins (further subclassified into type I, i.e. secreted proteins containing one transmembrane helix, and type II, i.e. non secreted) and multipass proteins [72].

It is approximated that 20-30% of all open reading frames in the human genome encode for membrane proteins [47]. However, due the amphipathic nature of transmembrane and membrane associated proteins, it has been notoriously difficult to study them using standard LC-MS methodologies [47, 72]. Membrane proteins represent attractive targets for biomarker and drug discovery, and they account for more than 60% of all approved therapeutic drug targets. For this reason, efforts are being made to optimize experimental methods that can allow the selective enrichment of membrane fractions and facilitate their proteomic analysis [72, 73].

Initial analysis of membrane proteins using 2D-PAGE have indicated that the method is not suitable for the investigation of the membrane proteome, due to solubility issues of these hydrophobic proteins in the buffers required for the first dimension of separation [47, 72]. For this reason, the majority of modern, large scale membrane proteomics experiments opt for gel free approaches [72]. The membrane enrichments methods summarized in this chapter will therefore refer to protocols tested and optimized for LC-MS based identification and quantification techniques.

The most simple enrichment methods for membrane fractions are ultracentrifugation and gradient centrifugation. Unspecific, weakly bound proteins can be removed by including a sodium carbonate wash [47].

Targeted enrichment strategies have also been developed, including – chemical labeling reagents (e.g. reactive ester derivatives of biotin), coating with colloidal silica particles, proteolysis approaches, or glycoprotein affinity enrichment strategies [47, 73, 74]. A schematic representation of the top three most commonly used plasma membrane proteome enrichment strategies for proteomic analysis are represented in **Figure 5**.

The covalent labeling of cell surface proteins with biotin can be achieved by the use of chemical reagents containing biotin coupled to a reactive ester, which can covalently modify either protein associated amino groups (the case of N-hydroxysuccinimide esters, **Figure 5**), sulfhydryl residues (N-iodoacetyl or maleimide) and aldehydes (hydrazide groups). After the chemically labeling of the membrane proteome, affinity streptavidin pull-down and stringent washes, ensure the enrichment of the membrane fraction and the subsequent removal of unspecifically bound proteins [47, 73, 75]. The spacer separating the biotin residue and the functional reactive group can be optimized in order to increase the hydrophilic nature of the biotinylation reagent (thus limiting cellular uptake of the biotin ester), or to allow a cleavable linker region (e.g.: sulfo-NHS-SS-biotin) [47, 73].

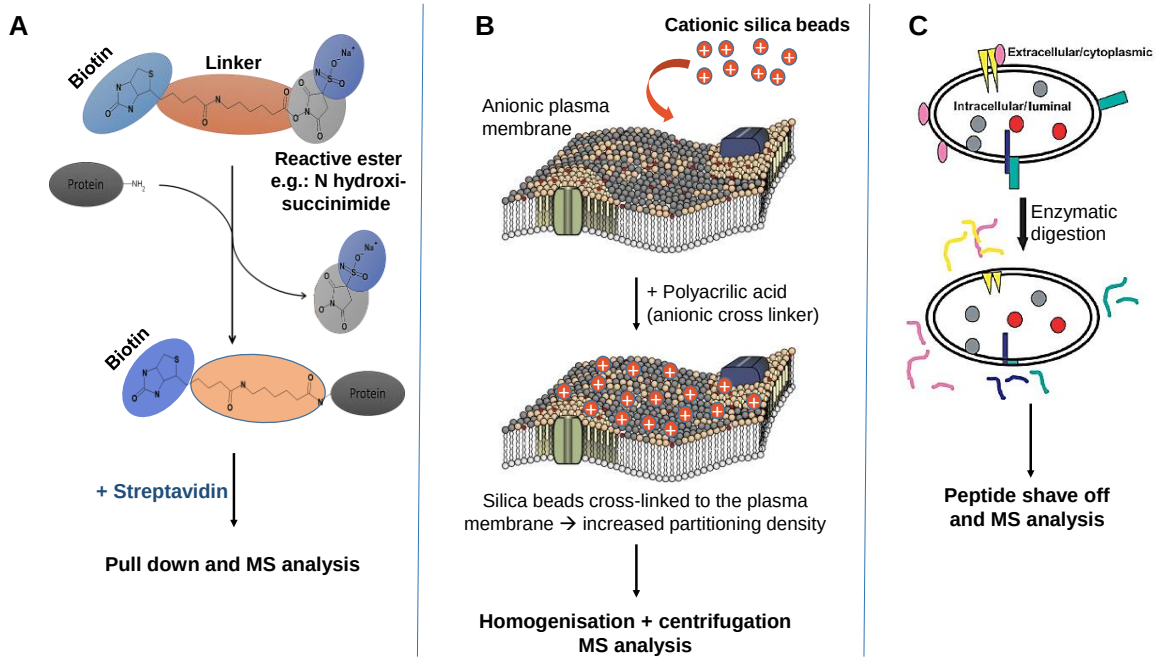


Figure 5: Targeted enrichment strategies for the analysis of plasma membrane proteome. **A**: Reactive ester derivatives of biotin are used for the tagging of proteins expressed on the cell surface and/or in the ECM, followed by streptavidin pull down and MS analysis. In this example, an N- hydroxysuccinimide ester was represented, reacting with primary amino groups present either at the N-terminal end of the proteins or in the amino acid side chain of lysine. Curtesy of Dr. C. Rösli **B**: Cationic silica beads can be applied to enrich the plasma membrane fraction, by taking advantage of the anionic nature of the cell surface. Anionic cross-linkers enable the formation of membrane fractions with increased partitioning density, which can be isolated by centrifugation. Adapted from Kornhuber *et al.* [76] **C**: Enzymatic digestion of intact cells can be used to obtain extracellular peptides. Adapted from Wu *et al.* [74].

Particle enrichment strategies rely on the anionic nature of the plasma membrane, which allows the electrostatic attachment of cationic colloidal silica beads. The addition of an anionic cross linker (e.g. polyacrilic acid) leads to the covalent association of the silica beads to the intact cell membrane (**Figure 5**). This covalent modification increases the partitioning density and facilitates sample collection by centrifugation. Following cell lysis, the desired membrane fraction can be pelleted from the crude extract, using a simple centrifugation technique [47].

Another method to investigate transmembrane proteins consists in the use of short duration enzymatic treatment on intact cells, thus 'shaving' the exposed extracellular epitopes. The method circumvents problems associated with low soluble hydrophobic transmembrane originating peptides, but is dependent on membrane integrity (**Figure 5**) [47]. Wu and collaborators developed a method combining sequential enzymatic digestion that allows in parallel the characterization of the protein's membrane topology. Intact cells are incubated first with the protease of choice and after the collection of extracellular shaved peptides, the plasma membrane integrity is compromised by homogenization at high pH, followed by a second enzymatic digestion [74].

Given that a large number of extracellular protein domains are glycosylated, lectin pull-down can be used to enrich cell surface glycoproteins. Deglycosylation with peptide-N4-(N-acetyl-beta-glucosaminyl) asparagine amidase (also known as PNGase F), or other specialized

enzymes can help improve the protein coverage following MS analysis [66, 77]. Additional use of glycoprotein enrichment methods in biomarker discoveries will be addressed in the following subchapter, concerning secreted protein biomarker discovery.

Better protein sequence coverage for transmembrane proteins can be achieved by using solvents and/or detergents for increased solubilization of highly hydrophobic proteins. Historically, methanol has been the solvent of choice for protocols aiming to improve membrane proteome solubilization and subsequent identification. Solvent volatility is compatible with downstream mass spectrometric identification, permitting easy removal prior to analysis. The disadvantage of such methods is that enzymatic digestion efficiency is reduced and it requires constant addition of fresh proteases during the peptide generation step [47]. Using detergents to improve the solubilization of hydrophobic proteins has been the most widely adopted method, but their implementation is restricted by incompatibilities with the downstream chromatographic and mass spectrometric experiments. There is a large variety of detergents which can be used for proteomic protocols, including both denaturating (e.g. SDS) and non-denaturating reagents (e.g. Triton X-100, used when protein functionality and protein-protein interactions need to be preserved) [47]. A recent improvement (addressing the compatibility of MS and detergent) has been the development of acid-labile surfactants – such as sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxyl]-1-propanesulfonate, commercialized by Waters under the name RapiGest, which can be used to improve protein solubility and are easily removed prior to LC-MS analysis by solution acidification and centrifugation [47, 78]. Several groups have reported improved membrane proteome recovery after using RapiGest. Nevertheless, the acid hydrolysis of the surfactant leads to the formation of hydrophobic residues, which must be removed by centrifugation in order to avoid interference with the mass spectrometric analysis and the pelleting step can result in co-precipitation of hydrophobic peptides of interest [47, 78].

Another obstacle for membrane proteomics identification is the fact that many plasma membrane associated proteins are not optimally digested by trypsin. The problem can either arise from the existence of large spanning hydrophobic sequences, or from the abundance of PTMs, such as glycosylations, masking the protease's substrate [47, 66]. Improved digestion of transmembrane domains can be achieved by using distinct proteases such as proteinase K or Lys-C, or cyanogen bromide (which cleaves peptide bonds at the C-terminus of methionine residues) [47]. Masked trypsin cleavage sites can be better exposed by using techniques such as filter aided sample preparation (FASP), or enzyme-immobilized reactors, both allowing a more rapid and efficient sample preparation, with an increased tolerance for high urea and detergent concentrations in the sample to be digested [47, 66].

2.2.5 Using mass spectrometry to identify secreted protein biomarkers

The secretome can be defined as the subproteome including all proteins released by a cell, a tissue, or an organism through different secretion mechanisms [46]. The secretome of a certain cell/organ includes cytokines, chemokines, hormones, other paracrine signaling molecules, but also extracellular matrix proteins and proteins shed from the cell surface [79, 80]. Typically, proteins can be secreted via classical mechanisms, involving the ER/Golgi pathway, via non-classical pathways (for e.g. involving caspase 1), exosomes, or be simply

shed from the cell surface by extra-cellular proteases [46, 81]. It is speculated that approximately 10% of the human genome encodes secreted proteins [46]

The secretome of a particular cell/tissue is extremely dynamic and can suffer significant changes during organ development, pathological transformation and in response to environmental stimuli, including organ or tissue damage [80]. Tapping into the pool of differentially secreted proteins in response to disease onset and progression is desirable for the field of biomarker discovery, with an increased interest being manifested in the area of cancer biomarker research. Sources for cancer secretome analysis are plasma, conditioned cell culture medium, and tumor proximal liquids – such as ovarian cancer associated ascites, pancreatic juice collected from tumor bearing patients, etc. [80]. Secreted protein biomarkers are often found in very low concentrations and require concentration steps prior to their analysis, such as solvent precipitation or ultrafiltration, which unfortunately lead to partial loss of the proteins contained in the supernatant [82]. Each source has distinct advantages and disadvantages for MS-based identification and validation methods.

Plasma and serum are easy to collect through non-invasive procedures and carry potential markers able to elucidate organ confined events or to evaluate the overall state of health [46, 83]. Unfortunately, the plasma sample's appeal is diminished by the presence of highly abundant proteins – such as serum albumin and transferrin, which mask the signal of less abundant proteins. Depletion of serum abundant proteins is however critical, since it can lead to the pull down of some of the putative biomarkers, but it is often necessary, as the detection range of most mass spectrometers is inferior to the dynamic range of serum proteins [54, 83]. Biomarker candidate concentrations in plasma are typically in the range of low ng/mL. Targeted MS approaches, such as selected reaction monitoring (SRM) of unfractionated plasma can only detect proteins in the low µg/mL range, making depletion steps mandatory, even for sensitive targeted MS approaches [48]. The storage and processing of blood samples, as well as the age, sex and ethnicity of biofluid donors can also significantly affect protein identification experiments, and accurate controls must be established in order to avoid experimental artifacts [32].

The collection of tumor proximal fluids diminishes the problem of abundant proteins eclipsing low abundant biomarkers, but it represents an invasive procedure, and limited blood contamination can still occur (for e.g. during fine needle aspiration) [46, 80]. The problem of serum contamination when collecting tissue fluids can be partially circumvented by using capillary ultrafiltration probes, implanted into various tissues, which can collect secreted and ECM proteins. Once ultra-filtration probes are implanted in the tissue of interest, a vacuum is applied to a semi-permeable filter, leading to the collection of tissue fluids [80]. Pancreatic juice has been identified as a rich source of putative PDAC associated biomarkers. It is secreted by the exocrine pancreas and represents a bicarbonate based solution (produced by the ductal cells) containing digestive enzymes (secreted by the acinar cells) [46, 84]. The presence of secretory enzymes complicates proteomic analysis and rigorous procedures need to be employed for sample collection and processing to prevent the proteolytic degradation of the juice protein components [84].

Conditioned cell culture medium is often a preferred source for secreted biomarker evaluation, since it represents a stabilized, easy to handle model and it includes the possibility of comparing primary and metastatic cell lines with different aggressiveness or metastatic behavior [46, 81]. However, cell lines are an artificial isolated system, depleted

for e.g. of niche stimuli. The dependency of cultured cells on added mammalian serum complicates proteomic discovery, similarly to the way abundant serum proteins hinder the identification of low abundant proteins in the blood. Two strategies are typically employed to deal with the added serum contamination: either pre-fractionation procedures are added to the proteomic work flow – which increases the number of experimental steps and can compromise the accuracy of protein quantification, or cells are cultured in serum deprived medium prior to collection – leading to cell death and the artificial alteration of the secretome profile by apoptotic and pre-apoptotic cells [45, 79, 85]. Attempting to address the problem of secretome contamination with serum proteins, Eichelbaum *et al.* developed a method of labeling newly synthesized proteins with azide-containing amino acids, thus allowing the use of added serum for cell cultures without affecting the subsequent identification and quantification of secreted proteins. Newly synthesized secreted proteins can be captured from complex cell culture media by using click chemistry, eliminating the need for laborious concentration and pre-fractionation techniques [79].

Villarreal *et al*. have analyzed to which extent serum starvation affects the quality of cell culture derived secretomes. By evaluating the secretome profile of various breast cancer cell lines grown in serum free medium for various amounts of time, they were able to establish a list of proteins associated with increased apoptosis of cultured cells, thus establishing an internal control. The apoptosis reporter protein list included mostly tRNA synthetases and chaperones [81]. Additionally, the group investigated if some of the intracellular annotated proteins identified in the conditioned medium were released in the extracellular space after cell lysis, or if they were truly secreted via an unconventional pathway specific to cancer cells. To their surprise, upon performing IHC on breast cancer tissue samples, they observed that a number of the nuclear/cytoplasmic proteins relocated in the neoplastic tissue and could therefore be detected in the extra-cellular environment [81]. Hence, they cautioned against discarding typically intra-cellular proteins from the subsequent cancer biomarker validation steps, and acknowledged the need for further investigations of unconventional secretion pathways and their role in mammalian cell biology [81]. The group also expressed interest in proteins secreted via exosomes, since it has been shown that tumor released exosomes represent more than a simple source of cancer biomarkers [81, 82]. In the particular case of melanoma, exosomes alone have been shown to prime their future metastatic site, ensuring the hospitability of the new niche to the invading cancer cells [82].

CA19-9 is the most widely used FDA approved serum biomarker for PDAC. The protein, a sialylated Lewis A antigen, is normally present on the surface of cells, but can also be secreted. However, CA19-9 is also elevated in benign diseases such as chronic pancreatitis and obstruction of the biliary duct [84]. Genetic variations in some individuals resulting in the inability to shed CA-19-9 antigen additionally complicate its use as a PDAC biomarker [46, 86]. For these reasons, several attempts have been made by the proteomic community to identify novel protein biomarker candidates secreted in pancreatic juice or cell culture medium that can lead to improvements in the diagnostic and disease monitoring capacity [46, 84]. This subchapter will not detail individual experiments and proteins of interest selected by the authors; an overview of these approaches can be found in the reviews of Pavlou *et al.* [46] and Makridakis *et al.* [85]. Very few attempts have been made to discover novel PDAC markers in the serum of cancer bearing patients or patients diagnosed with preneoplastic lesions. One such interesting discovery experiment has been performed by Nie *et al*. The authors selectively enriched circulating glycoproteins, using a lectin resin, and

proposed a panel of three protein markers, namely α-1-antichymotrypsin, thrombospondin-1,and haptoglobin, to be used in combination with CA19-9, which could distinguish between cancer, diabetes, obstructive jaundice and chronic pancreatitis patients [77].

2.2.6 Validation methods and targeted proteomics

Once the postanalytical data interpretation has been completed and a number of protein biomarker candidates has been selected, the next step of the biomarker development pipeline is the preclinical validation.

Multiple techniques have been employed over the years. Historically, antibody based techniques such as ELISA, IHC and western blot (WB) have been preferred as convenient, rapid, sensitive and high-throughput solutions, despite the fact that they are difficult to multiplex [48]. However, the availability of suitable antibodies is often a limiting factor for the development of such validation method. The cross reactivity of commercially available antibodies needs to be taken into consideration when the aim is to quantify proteins with high sequence similarities between species (when animal xenografts samples are used), or between members of a certain protein family [48]. IHC validation is complicated by the obligatory sample fixation protocols, which additional to being time consuming, can also affect the availability of the antigen of interest. Moreover, IHC can only be used for semiquantitative evaluations of protein abundances [31, 48]. Novel technological advances have been developed in order to increase the specificity and sensitivity of antibody based validation methods. A novel immunoassay supported by Erenna^{TM} Immunoassay Systems (Singulex, Alameda, CA, USA) utilizes photon fluorescence detection and paramagnetic microparticles to increase precision and sensitivity, being able to quantify and report concentrations below 1 pg/mL [31]. Another new development is the digital ELISA (Quanterix Corporation, Cambridge MA, USA), which can detect single enzyme-linked immunocomplexes on beads packed in arrays consisting of femtoliter wells, and can detect proteins present at sub-fM concentrations, thus being three orders of magnitude more sensitive than a conventional ELISA [31]. New affinity based techniques which are replacing antibodies with affibodies and aptamers, are also being considered [31, 54].

RT-qPCR is also commonly used in the validation phase of proteomics based biomarker discovery experiments, despite it mostly being regarded as a transcriptomic tool. The method interrogates whether the changes in protein expression levels observed between samples can be correlated with mRNA levels, or if a post-translational mechanism is more likely to be responsible for the recorded variations. Briefly, total RNA extracts need to be first reverse-transcribed, using oligo-dT, random, or specific primers. The resulting cDNA is interrogated for the amount of a specific mRNA of origin, based on the fluorescent signal detected in the sample. The fluorescence is either generated by a TaqMan hydrolysis probe, or most commonly by the SYBR green DNA intercalating agent [87]. The moment when the overall fluorescence detected in the sample surpasses the statistically defined background noise is called the threshold cycle (Ct), and it is an indicator of the amount of cDNA molecules present in the sample. The higher the initial number, the quicker can the Ct be reached [87]. Specific primers can be designed for the genes of interest, including the possibility to recognize species specific mRNA reverse transcribed sequences. Therefore, RT-qPCR can be used for the relative quantification of the genes of interest also in xenografted tissues, making it possible to exclude the signal originating from the host's

stromal cells present in the tissue sample/contaminating the pre-sorted cells. For TaqMan probe detection protocols, an additional step for the selection of the oligonucleotide probe needs to be included [87, 88]. However, when designing a pair of primers for RT-qPCR, one needs to consider putative signals originating from potential pseudogenes, sometimes transcribed, although usually not translated into a functional proteomic entity [89, 90]. Properly designed RT-qPCR primers could also distinguish between splice isoforms [89]. However, as previously mentioned, gene expression experiments cannot always predict accurate protein levels, nor can they offer information on the wealth of PTMs associated with a specific protein marker or accompanying pathologic transformations [40].

Mass spectrometric approaches represent an appealing alternative to antibody and other affinity compound based validation strategies. Targeted MS methods have been developed, aiming to allow multiplexed analysis, validation and quantification of specific protein biomarkers, while circumventing the traditional limitations of shot-gun discovery proteomics [48].

The main caveat of traditional MS discovery designed instruments is their limited sensitivity when analyzing complex samples [31]. First, problems can arise from stochastic sampling, which is an intensity dependent effect. Secondly, ion suppression effects can arise depending on the presence of compounds in the sample altering the efficiency of gas ion formation. Finally, informative ion species can be excluded from the analysis, due to their elution within a predefined dynamic exclusion filtering window, designed to avoid resampling of the same ions [48, 54, 91]. Novel MS strategies, such as SRM (also known as multiple reaction monitoring or MRM), parallel reaction monitoring (PRM), or accurate inclusion mass screening (AIMS) have been developed to increase selectivity and improve MS quantification [48].

SRM is a targeted proteomic approach, usually conducted on a triple quadrupole instrument, in which the first and the third quadrupoles act as mass filter, while the second quadrupole performs the CID fragmentation. In SRM, a specific peptide (the precursor) is selected in the first mass filter, fragmented, and a single fragment ion is filtered by the third mass analyzer. The number of fragment ions is counted by the instruments detector, allowing for precise quantification of the precursor peptide and therefore its corresponding protein of origin [48, 54]. A schematic representation of shotgun vs SRM proteomics is outlined in **Figure 6**. The pair of precursor and fragment ions selected during the analysis are referred to as transitions [48]. Alternatively, all product ions of a certain peptide can be analyzed in parallel, thus increasing the confidence rate for the identified peptide and providing a higher tolerance for co-isolated background peptides. This technique has been termed PRM [54, 92]. Proteotypic peptides as well as species specific peptides can be selected as precursors for the transitions, thus allowing for the accurate quantification of the protein of origin in complex samples and samples isolated from xenografts models [93]. SRM's sensitivity also allows the detection of highly similar peptide sequences, such as somatically mutated and disease associated proteins [39]. However, optimization of transitions is not trivial and implementing the method requires significant technical knowledge, which can be a deterrent for clinical applicability [54]. Nonetheless, SRM has been proven to be a highly reproducible quantification method across different laboratories and it is currently used in the clinics for evaluating the concentration of small molecules [48].

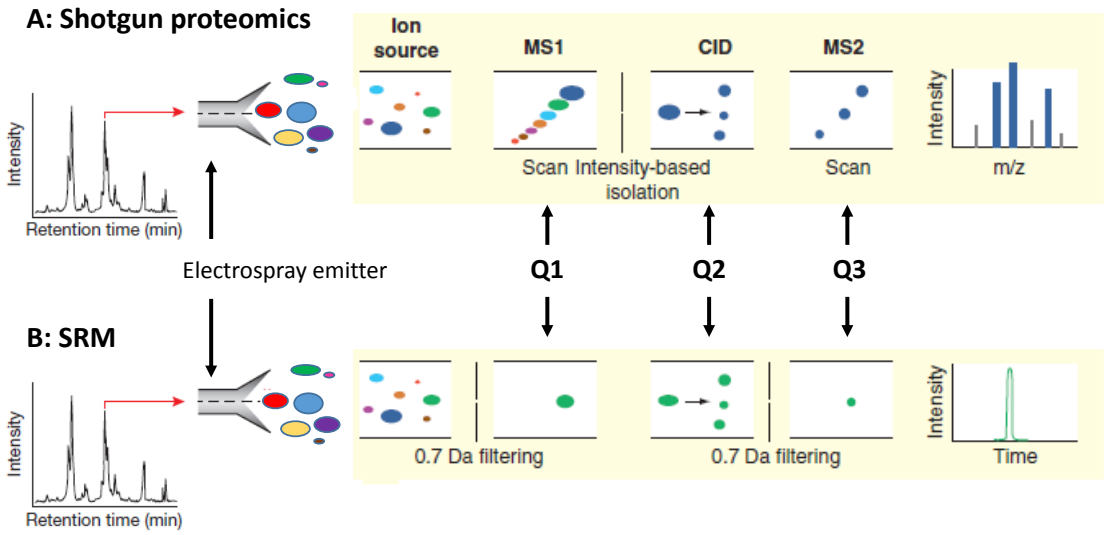


Figure 6: Comparison of the MS principles in shotgun proteomics and SRM analysis – schematic representation for ESI-quadrupole systems. Enzymatically digested peptides are loaded on a chromatographic column for prefractionation, and eluted. **A**: In shotgun proteomics, all eluting peptides are screened in the first quadrupole (Q1), and a full MS spectrum is created. Peptides with the highest MS intensity are selected for dissociation (Q2) and all fragment ions are screened in the second mass filter – Q3, thus obtaining the MS^2 spectra which will be searched against available data bases. **B**: In SRM, Q1 selects only one prespecified precursor ion, with a defined m/*z*, and the peptide is fragmented in the second quadrupole. Q3 acts as a mass filter for a single fragment ion and the respective intensity is recorded. Further, relative or absolute quantification is achieved by comparing peak intensity between different samples or internal standards, respectively. Modified from Gilette *et al.* and Liu *et al.* [48, 54].

For a better quantification, SRM analysis is often performed in the presence of stable isotope labeled peptides corresponding to the precursor peptides to be analyzed. These internal standards allow a reliable evaluation of protein level differences in distinct samples. However the spike in of stable isotope labeled peptides cannot predict the effect of sample prefractionation protocols on the availability of the precursor peptides for the MS analysis, nor can it predict the efficiency of the enzymatic digestion [48]. Thus, for absolute quantification, the use of synthetic labeled whole proteins, undergoing the same preanalytical procedures, is recommended [39, 54].

The complexity of the sample needs to be considered, since it affects the method's limit of detection. Plasma often needs to be depleted for the most abundant proteins, using immunoaffinity depletion columns prior to the MS analysis, despite the risk of losing lower abundant proteins in the process [48]. The detection limit for unfractionated plasma is in the range of 0.3 – 1 µg/mL and therefore above the required level for distinguishing potential protein biomarkers. The detection boundary can be decreased to 1 – 10 ng/mL in fractionated plasma [48]. When analyzing unfractionated whole cell lysates, SRM allowed the detection of proteins present in as low as 7500 copies per cell [39].

Sample handling is also highly important and hemolysis should be avoided, since it can negatively affect the mass spectrometric subsequent analysis [93]. SRM can also be used in combination with antibody based enrichments protocols, allowing for a better limit of detection in complex biological samples. Protein or peptide enrichment can be carried out

prior to the MS based analysis, using high affinity antibodies. The sensitivity of the antibody is not as important as its affinity, since the quadrupole filtering can distinguish between the isolated species. However, the method is limited by the existing of commercially available antibodies with high affinity for the peptide or protein of interest [48].

2.3 Current knowledge about novel PDAC biomarker candidates

The following subchapters will summarize important information about the four main putative PDAC biomarker candidates of interest, identified in the current study. A short overview is presented in **Table 1**.

Table 1: Short overview of the top four novel putative PDAC markers

Protein name	Phylogenetic distribution	Cellular localization	Organ distribution (adult humans)	Function
Cadherin-17	Vertebrates and invertebrates	Plasma membrane	Small and large intestine	Cell-cell interaction. Cell adhesion
Galectin-4	Animals and plants	Cytoplasmic + secreted	Digestive tract and associated glands, excepting liver and pancreas	Carbohydrate domain recognition. Protein-protein interaction
Protocadherin-1	Vertebrates and invertebrates	Plasma membrane	Central nervous system, airway epithelial cells	Cell-cell interaction
Lipocalin-2	Eukaryotic organisms, mostly vertebrates	Secreted	Bone marrow, spleen, tonsils, respiratory + digestive track, kidney, prostate, uterus	Iron transport. Innate immune response

2.3.1 Cadherin-17

Cadherins represent a family of calcium dependent adhesion proteins, typically involved in cell recognition, cell signaling, morphogenesis, angiogenesis, migration and tumorigenesis [94, 95]. All family members contain a series of extra-cellular amino terminal domains, termed cadherin repeats; most cadherins are trans-membrane proteins, including a carboxi-terminal cytoplasmic domain, varying in size and composition significantly between different cadherin families, responsible for intra-cellular signaling [94-96]. The cadherin repeat consist typically of approximately 110 amino acids arranged in seven β-strands and two short α helixes, resulting in a β barrel structure similar to that of an immunoglobulin domain [97]. In the presence of calcium ions, the extra-cellular domain is stabilized and adopts a rod like structure; the binding of ions takes place at the junction between cadherin domains (also known as ectodomains) [98].

The cadherin family members can be classified in at least six sub-families, based on the composition of their extra and intra-cellular domains [96]:

a) Classical cadherins: containing five extra-cellular cadherin domains, and a cytoplasmatic tail of 120-150 amino acids
b) Desmosomal cadherins
c) Seven domains cadherins: containing seven extra-cellular cadherin repeats and a very short cytoplasmatic tail, consisting of approximately 20 amino acids
d) Seven transmembrane cadherins
e) Truncated cadherins: containing no transmembrane domain, instead GPI anchored to the cell membrane
f) Protocadherins (which will be discussed in greater detail in a following subchapter)

Cadherin-17 (CDH17) belongs to the sub-family of seven domains cadherin, as it contains seven extra-cellular cadherin repeats, and it was first described as a proton dependent peptide transporter in the intestinal CaCo2 cell line [96, 97, 99].

Phylogenetic analysis have suggested that the first two cadherin domains of CDH17 resulted as a duplication of the first two ectodomains described in classical cadherins [97]. In support of the duplication theory comes also the observation that the calcium binding motif between cadherin domains two and three is incomplete [97]. The cytoplasmic domain of CDH17 is very short and no direct intra-cellular partners of interactions could be described so far [96, 100].

Cadherins expression is highly regulated during embryogenesis and adulthood [101]. In mouse and human, CDH17 has been reported at the baso-lateral level of enterocytes and goblet cells of the intestine, while in adult rat it is also present in hepatocytes, which is the reason why the initial name for the protein was Liver-Intestine Cadherin [96, 100]. Some authors have reported occasional scattered weak expression in ductal cells of the healthy pancreas [95, 96].

Classical cadherins are known to mediate cell adhesion via their cytoplasmatic tail, which interacts with β-catenin, plakoglobin and α-catenin [94]. However, in the case of CDH17 the adhesion mechanism is unknown. Not only that the 20 amino acids intracellular domain is insufficient for direct β-catenin interaction, but a GPI anchored engineered form of the protein was capable of mediating cell-cell adhesion to the same extent as the native CDH17 [96, 100]. It has been however suggested that CDH17 could dimerize with E-cadherin, although this type of interactions are more likely to occur during embryogenesis, since the two cadherins are located at different junctions in the adult cell [96, 97]. It has been suggested that the interaction with E-cadherin can lead to the release of β-catenin from the cytoplasm, thus potentially implicating CDH17 in Wnt signaling [96]. Takamura *et al.* reported galectin-3 as a potential partner of interaction for CDH17, using a pancreatic ductal adenocarcinoma cell line [95]. The interaction would occur via the extra-cellular domains of CDH17, which have been shown to be subject to N-glycosylation [95, 96]; they could show that washing the cells with a lactose solution lead to a decrease in association between the two proteins (evaluated via co-immunoprecipitation – co-IP) in a dose dependent manner, but it was not affected upon washes with a sucrose solution [95]. Of interest is the fact that galectin-3 has also been described to be involved in Wnt signaling, based on its ability to interact with β-catenin in the nucleus and axin in the cytoplasm [96]. Recently, Bartolomé *et al*. described α2β1 integrin as an interaction partner for CDH17, based on co-IP experiments and mass spectrometric determination of the pulled down proteins [102].

Interaction between the two adhesion molecules lead to activation of focal adhesion kinase (FAK) in a metastasising colorectal cancer cell line, and initiated Ras signaling. Interestingly, silencing of either the α2 or the β1 integrins resulted in loss of colocalization of the remaining chain with CDH17, leading to the conclusion that both α and β subunits were necessary for the proper interaction to occur. Moreover, CDH17 knock down could be correlated with a decrease conformational activation of β1 integrin [102]. Thus it appears as if CDH17 might mediate cell adhesion exclusively via key interactions of its extracellular domain.

The CDH17 gene expression has been shown to be modulated by two major transcription factors – HNF1α and CDX2, based on chromatin-immunoprecipitation experiments and deletion/mutation analysis, using hepatocellular carcinoma cell lines [99].

A summary of CDH17 mediated signaling and transcriptional regulation is presented in **Figure 7.**

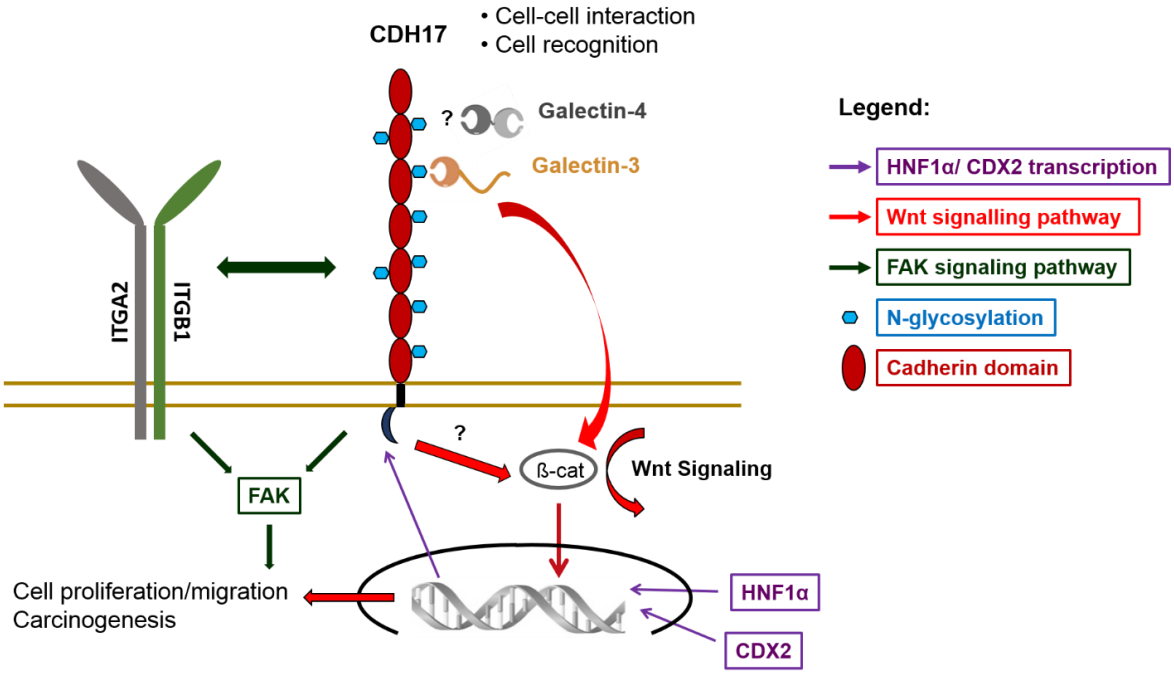


Figure 7: Schematic representation of signaling associated pathways and transcriptomic regulation of CDH17. The seven extracellular domains of the protein are predicted to be subject to N-glycosylation, and the glycosyl chains are expected to mediate interaction with members of the galectin family, such as galectin-3. Galectin-4 and CDH17 co-expression has been reported in literature, however their putative interaction has not been tested. Wnt signaling (indicated by red arrows) and activation of FAK pathway via integrin β1 – CDH17 interaction (indicated by green arrows) are the main signaling pathways in which CDH17 has been shown to participate so far. Purple arrows indicate transcriptional events. ITGA2 = integrin α2; ITGB1 = integrin β1.

CDH17 expression has been reported to be disregulated in malignant tissues, with different prognosis depending on the carcinoma type – in hepatocellular carcinoma, CDH17 overexpression has been associated with bad prognosis [96, 103], while in PDAC it has been linked to improved prognosis, enhanced differentiation of the tumor and increased survival [96, 104].

In the case of gastric carcinoma, CDH17 together with LGALS4 have been reported as markers for a molecular subtype (termed by the authors G-Intestinal), defined by better

response to 5-FU therapy, the corresponding group of patients showing a moderate survival benefit compared to patients who did not express the two markers (G-diffuse subtype) [105].

In hepatocellular carcinoma, a splice variant of CDH17 has been described, where exon 7 is excluded from the mature mRNA. This mutation seems to be specific to hepatocellular carcinoma and is associated with shorter survival, increased tumor recurrence and tumor venous infiltration [96]. In one study, knockdown of CDH17 in highly metastatic hepatocellular carcinoma cell lines lead to reduced proliferation, and was associated with cytoplasmic relocalization of β-catenin and a decrease in cyclin D1 levels, linking it potentially to Wnt signaling [103]. In another study, CDH17 was also shown to cross-talk with Met signaling [96].

On the other hand, in the case of intrahepatic cholangiocarcinoma, the second most common type of liver cancer, CDH17 seems to have a protective role. Takamura *et al*. down regulated CDH17 and observed an increased expression of Metal Responsive Transcription Factor 1 (MTF1) and its downstream target Placental Growth Factor (PIGF) [106]. PIGF has been shown to be essential for angiogenesis and plasma extravasation in ischemia, inflammation and cancer [106]. In addition, MTF1 loss in Ras-transformed cells lead to reduced tumor growth, increased matrix deposition and reduced angiogenesis [106].

In conclusion, CDH17 is an intriguing adhesion molecule, potentially linked to Wnt and Ras signaling [95, 96, 102], whose role in cancer progression appears highly dependent on the tissue of origin [106].

2.3.2 Galectin-4

Galectins represent a class of animal lectins that typically bind β-galactose containing glycoconjugates [107, 108].They were initially named S-type lectins, due to the fact that they seemed to require reducing conditions in order to maintain their lectin-binding activity, thus indicating that a free cysteine residue was vital for the biological function of the protein. However, galectin-4 activity is independent of the existence of reducing conditions in the environment [108]. All galectins contain a carbon recognition domain (CRD), of usually around 130 amino acids arranged in a β-sandwich completely lacking an α-helix, domain responsible for the recognition of and interaction with specific sacharides [108].

The galectin family has been subclassified in three major groups [108, 109]:

a) Prototypical galectins: contain a single CRD, often forming dimmers
b) Chimeric galectins: more common in invertebrates, contain a single CRD and a large Pro-Gly-Tyr rich domain, which is often very sensitive to metalloproteinases. The only vertebrate galectin described in this class is galectin 3
c) Tandem repeat galectins: they contain at least 2 CRDs, linked together by an amino-acid chain of 5-50 amino acids long. The linker peptides are often sensitive to proteolytic activity. This class includes galectin-4.

Galectin-4 (LGALS4) is a 323 amino acid long protein (approximately 32 kDa), containing two CRDs (consisting each of approximately 130 amino acids), linked by a 34 amino acid sequence similar to the Pro-Gly repeats described for galectin 3 [110]. The two CRDs of LGALS4 display only 40% sequence similarity, and although they bind lactose with equal affinities, they exhibit different affinities for larger sacharides. This had led scientists to

hypothesize that it can act as a cross-linker and potential stabilizer of complex structures [110].

In development and adulthood, LGALS4 is expressed at high levels in the digestive tract from mouth to rectum, but is not expressed in accessory digestive glands such as liver and pancreas [110]. It can be found either intracellulary, associated with adherens junctions, microvilli or lipid rafts [107, 110], or in the extracellular environment, where it is expected to act as a cross-linker for various glycoproteins [110]. No classical secretion signal peptide could be identified for any member of the galectin family, and therefore it can only be suspected that their export is done via a non-canonical secretion pathway [108, 111]. The exact mechanism is unknown, but it is believed to have evolved in order to avoid interaction between the galectins and the glycosylated proteins found in the Golgi classical secretion pathway, especially since the CRD is already active prior to externalization [107, 108, 111].

The CRD seems to be essential only for the extracellular functions of galectins [107]. In the intestine, LGALS4 forms soluble high molecular weight complexes with brush border enzymes in lipid rafts [107, 112]. Upon LGALS4 depletion, proteins normally localized at the apical level were trapped intracellulary, thus indicating a sorting defect caused by the absence of the lectin [107, 112]. Sulfatides with long chain hydroxylated fatty acids, which are typically enriched in lipid rafts, were described as high affinity ligands for LGALS4 [107, 113], leading to the conclusion that that the lectin is important for the apical delivery of proteins upon interaction with the sulfatides clustered on the lipid rafts [107, 113]. LGALS4 has been shown to also play a role in the development of inflammatory bowel disease, since it stimulates $CD4^+$ T cells to produce the pro-inflammatory cytokine IL6 [107, 109]. The IL6 production is induced by a PKCθ signaling pathway, through the immune synapse [107, 109]. The simple administration of a LGALS4 antibody to mice developing intestinal inflammation could lead to suppression of disease progression [107].

Two c-Rel binding sites could be identified at the promoter region of LGALS4, leading to suggestions that LGALS4 could be induced in inflammatory gut disease as part of the innate immunity response triggered by NF-kB [110, 114]. The promoter region of LGALS4 has also been suggested to be under control of Myo-D transcription factor, as well as multiple transcription factors belonging to the Hepatocyte Nuclear Factor 3/HNF3/fork transcription factor family, such as HNF3β, HNF4, HFH2 [110]. Thus, described putative transcription binding sites upstream of the LGALS4 promoter suggest that the gene is controlled by regulators of epithelial development, differentiation, as well as malignant transformation [110].

In fact, LGALS4 has been found to be associated with a number of malignancies. In colon cancer, a loss of supra-nuclear LGALS4, concomitant with increased cytoplasmatic expression could be described during cancer progression [110]. In one study, increased expression of LGALS4 in colon carcinoma was correlated with poor prognosis, although in early colon carcinogenesis, LGALS4 mRNA expression has been reported to decrease compared to the expression observed in the surrounding normal tissue [115]. Another example of abnormal expression has been described in breast cancer: although the healthy breast tissue is devoid of LGALS4, abnormal areas of benign tumors and especially carcinomas show high levels of the protein [110].

The exact function of LGALS4 in human malignancies has however been insufficiently investigated, with reports sometimes reaching contradictory conclusions for distinct malignancies. A study using ovarian and lung carcinoma derived cell lines has reported that

poorly differentiated epithelial cancer cells not capable of forming polarized layers express less LGALS4, tend to migrate more and express higher levels of galectin-1, typically associated with tissues of mesenchymal origin [110]. Another study investigated the localization of LGALS4 in the colon metastasis derived cell line T84. The investigators observed accumulation of the lectin of interest at the cell-substrate interaction for freshly seeded cells, at the basal membrane for confluent cells, and at the leading edge of lamellipodia in the case of sub-confluent cells [115, 116]. They therefore concluded, that at least in the case of colon carcinoma, LGALS4 plays a role in cell-substrate interactions and may be important to cell migration and metastasis [115]. Of great importance was the report showing that LGALS4 may help cells survive in nutrient depleted environments. Thus MDCK cells transfected with LGALS4 were able to survive in serum starved medium for 7-8 weeks, in stark difference to the mock transfected cells who became apoptotic after one week in the same conditions [110]. Considering that for PDAC the microenvironment is reported to be severely hypoxic and under-perfused [13], LGALS4 overexpression might therefore represent a survival advantage for these cancer cells.

In 2011, Tan *et al*. reported that co-expression of CDH17 and LGALS4 can be correlated with better survival prognosis for gastric cancer patients, as well as increased responses to 5-FU and oxaliplatin, despite showing augmented resistance to cisplatin treatments [105]. No attempts to functionally link the two proteins has been made. However, previous studies have identified 'stalked' brush border enzymes – containing only a small cytoplasmatic tail and large, heavily glycosylated extracellular domains, similar to the general organization of CDH17 – to be partners of interaction for LGALS4 [96, 112]. It would be of interest to investigate if the two proteins interact *in vivo* (as mentioned is **Figure 7**) and if this interaction plays a role in cancer progression and response to chemotherapy.

To sum up, LGALS4 is a versatile carbohydrate binding protein, involved in inflammatory response and cancer progression. LGALS4 activation in cancer tissues can often be associated with improved overall survival. However, its upregulation seems to give a survival advantage for cancer cells living in nutrient deprived media [110].

2.3.3 Protocadherin-1

Protocadherins represent the largest subfamily of the cadherin family, briefly discussed earlier. The term protocadherin was introduced by Sano *et al*., when they identified a subclass of proteins containing cadherin ectodomains, holding different features from those of classical cadherins. The new class of proteins was highly expressed in the brain and their presence in a wide range of vertebrate and invertebrate species inspired the name protocadherin, from the Greek 'protos'= first [117, 118].

Based on their molecular and genetic features, protocadherins have been subdivided into several clusters: α-, β-, γ- protocadherins, flamingo cadherins, large protocadherins (including FAT cadherins) and nonclustered protocadherins δ and ε [117, 119].

Protocadherin-1 (PCDH1) belongs to the δ protocadherin subfamily, which has been further subdivided in δ1 and δ2 protocadherins [119, 120]. Based on the latter subclassification, PCDH1 is a δ1 protocadherin, first identified under the name of PC42 in human brain [118]. δ1 protocadherins contain seven cadherin repeat domains, and three intracellular conserved motifs (CM), termed CM1, CM2 and CM3; CM3 is specific for δ1 protocadherins,

located adjacent to the CM1 domain, and seems to be responsible for interaction with Protein Phosphatase 1α (PP1α) [119, 120]. PCDH1 has been shown to be subject to alternative splicing, with at least two major isoforms often reported in literature: isoform 1 expresses a short cytoplasmatic tail, while isoform 2 has a long cytoplasmatic region [118, 119]. Several additional splice isoforms have been described as well by Koning *et al*., their inclusion/exclusion from the mature mRNA being most likely dictated by the differentiation state of the investigated cells; several alternative transcription initiation sites have also been suggested [119]. The exact function of the different splice isoforms has been so far insufficiently investigated.

Protocadherins are currently regarded as proteins expressed mostly in the central nervous system, developmentally regulated and their expression tends to be higher in adult compared to fetal and neonatal tissues [117, 118]. Additionally to the brain localization, first reported for PCDH1 [118], the protein has also been reported to be expressed by airway epithelial cells, at the apical border, and in macrophages [121].

Very little is known about the function of PCDH1. Protocadherins in general have been shown to both promote, and inhibit cell adhesion, but the exact mechanisms of adhesion have not been explained [117]. PCDH1 is known to express a PDZ domain binding site at the C terminal end of the cytoplasmic domain. PDZ is a structural domain found in many different signaling proteins of many regna, responsible for helping transmembrane proteins anchor to the cytoskeleton and hold together signaling complexes [119]. Their defined expression by distinct subpopulations of functionally connected neurons as well as their localization at the synaptic junctions, lead researchers to presume that protocadherins could be involved in modulation of synaptic transmission, synaptogenesis and synaptic plasticity, all important for learning and memory formation [117, 120]. Protein isoforms expressing the full CM3 cytoplasmatic domains include a RRVTF sequence responsible for interaction with Protein Phosphatase 1α [120]. PP1α has been involved in regulation of synaptic plasticity, but also in lung development [120, 121]. Other putative interaction partners for δ1 protocadherins include TAF1/set (involved in cell cycle regulation) and Frizzeled 7 receptor [120, 122]. PCDH1 has been identified as a susceptibility gene for bronchial hyperresponsiveness and asthma [121], and the expression level is increased in highly differentiated bronchial cells. Additionally, the expression of splice isoforms may also be dependent on the differentiation status of the bronchial cells [119]. PCDH1 has been suggested to be involved in epithelial repair, since the mRNA and protein levels were upregulated in a keratinocyte wound repair model [119].

References about protocadherins expression and involvement in cancer have been scarce. Protocadherin-7, a δ1 protocadherin family member showing many structural similarities to PCDH1 has been reported to be involved in non-small-cell lung cancer [119]. To our knowledge, no report has to date ever linked PCDH1 to PDAC.

2.3.4 Lipocalin-2

The lipocalin protein family belongs to the structural superfamily of calycins (together with avidins and Fatty Acids Binding Proteins), and consist of over 50 small molecules, usually involved in binding and transporting small hydrophobic molecules (such as prostaglandins, retinol, hormones, etc) [21, 123-125]. Despite the low level of structural similarities between family members (~20%), the three dimensional structure is common for all affiliated

proteins, consisting of a single eight-stranded, continuously hydrogen bounded anti-parallel β-barrel, which forms an enclosing cavity, responsible for binding specific molecules [124, 125].

Lipocalin-2 (LCN2), also known as neutrophil gelatinase (NGAL – since it was first isolated from human activated neutrophils), or siderocalin (due to its iron binding capacity) is a 25 kDa protein (in its monomeric form), which can also be found as a disulphide linked homodimer (46 kDa in size) or a disulphide linked heterodimer with matrix metalloproteinase 9 (MMP9) (135 kDa) [124, 126]. In healthy tissues, LCN2 has been reported in the bone marrow, in the spleen and tonsils, as well as in organs prone to exposure to microorganisms such as epithelia of the respiratory and digestive tracks [127]. It can also be found in low levels in the kidney, prostate and uterine mucosa [124].

LCN2 is a versatile protein, performing a variety of functions, depending on the tissue, developmental stage and health status. It is involved in the innate immune response, as it is released by neutrophils at a site of infection and inflammation, where it binds iron and is involved therefore in the antibacterial iron depleting strategy [124]. Its iron binding and transport capacity has also been shown to contribute to the activation/repression of iron responsive genes and cell growth, after its cytoplasmic internalization followed by iron unloading in acidic endosomes [124]. Lipocalin has also been reported to be a regulator of apoptosis, through its iron binding and transporting function – it can block the induction of pro-apoptotic protein Bim and prevent pro-caspase 9 activation. Alternatively though, LCN2 can bind several siderophores in the cytoplasm, reducing intracellular iron concentration and leading to Bim expression [124, 128]. Moreover, LCN2 can protect against acute ischemic injury and its expression level reflects the extent of tissue damage, making it one of the most promising markers for nephrologic pathological states [124, 126]. However, neutrophil gelatinase has been reported to be upregulated in several inflammatory conditions, including appendicitis, inflammatory bowel disease, diverticulitis, and it seems to be under transcriptional regulation of NF-kB [126]. LCN2 is also involved in differentiation of primordial cells during embryogenesis, its expression being highest before the onset of cellular differentiation [21]. A schematic overview of the LCN2 involvement in signaling pathways relevant for tumor progression is represented in **Figure 8**.

In human malignancies, LCN2 is often upregulated compared to the normal tissue, but controversially it tends to be downregulated in metastatic tumors compared to the primary tumor [124]. Both pro- and anti-neoplastic functions have been described in literature for the protein, depending on the type of tumor and the differentiation status of the neoplastic cells studied. For example, in the case of esophageal cancer, LCN2 interaction with MMP9 leads to a protected, more active gelatinase, favoring cancer cell invasion [124, 126]. However, in K-Ras mutated colon cancers and pancreatic ductal adenocarcinomas, or Ras transformed breast cancer cells, LCN2 over-expression lead to decreased migration and metastasis. This effect could be due to decrease of Ras dependent phosphorylation and subsequent inactivation of E-Cadherin, or in the case of PDAC, by altering FAK phosphorylation at tyrosine level 397 (critical for the kinase's activation) and subsequent decrease of migration and metastasis of pancreatic cancer cells [126]. Its function in cancer progression is therefore controversial and may be highly dependent on the tissue of origin.

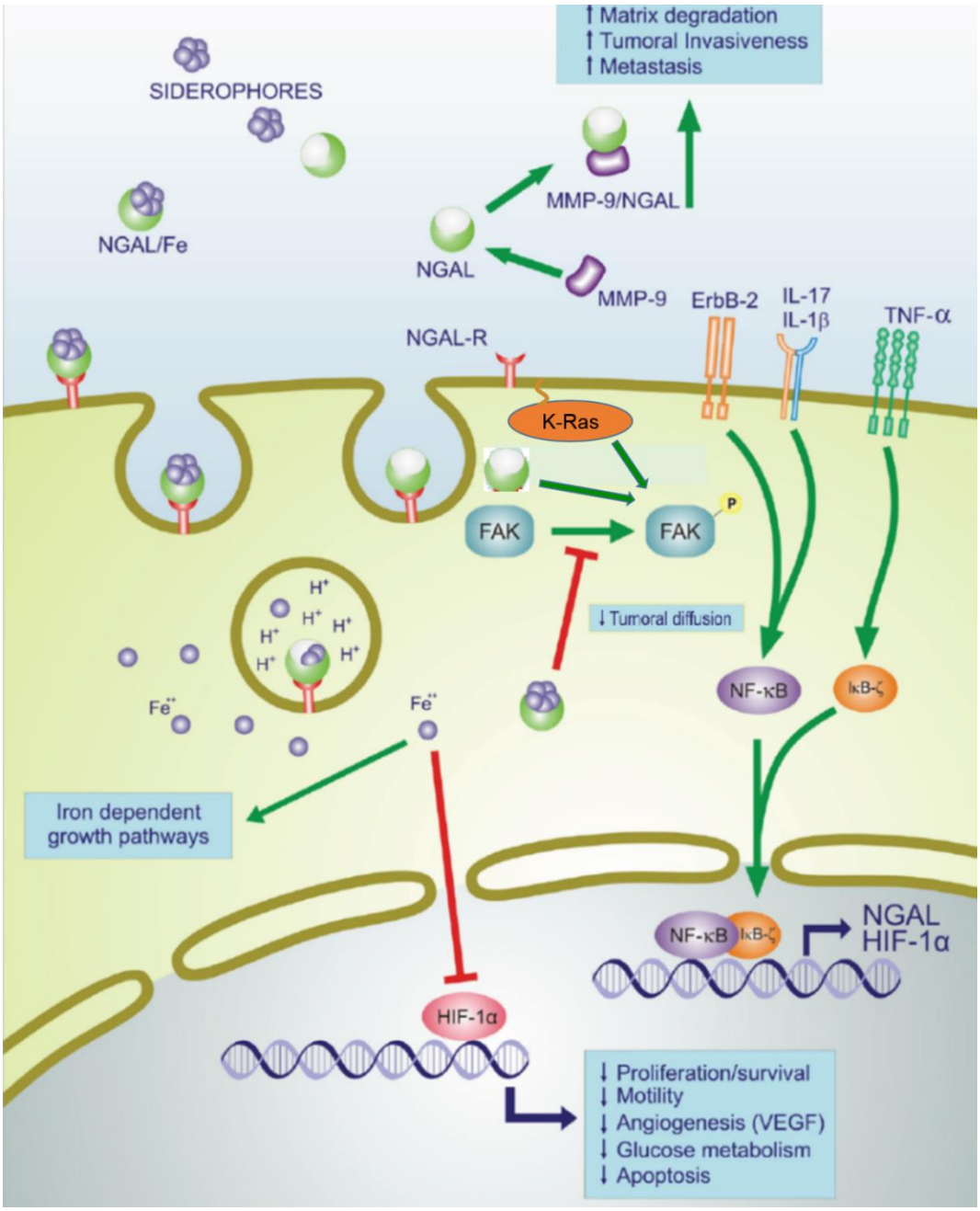


Figure 8: LCN2 (also known as NGAL) involvement in survival, motility, proliferation and angiogenesis. New acronyms presented in the picture include: NGAL-R = NGAL receptor; v-Erb-B2 Avian Erythroblastic Leukemia Viral Oncogene Homolog = ErbB-2/HER2; IL-17: interleukin-17; IL-1β: interleukin-1 beta; NF-κB: nuclear factor kappa B cells; IκBζ: Inhibitor of NF-κB zeta subunit; HIF-1α = hypoxia inducible factor-1α. Adapted from Candido *et al.* [124].

In the particular case of PDAC, LCN2 has been reported to be significantly upregulated at both mRNA and protein level, compared to chronic pancreatitis samples [21]. It could also be detected in the blood of PDAC patients, using ELISA or Radioimmunoassay (IRA). For this reason, several groups studied its potential use as a diagnostic marker, suggesting that despite having similar sensitivity to the only so far approved PDAC marker, CA19-9, it might show increased specificity [129, 130]. Of interest is the report of Moniaux *et al*., showing that LCN2 is upregulated early in the progression of pre-cancerous lesions: PanIN- I lesions already show high expression of the protein, and this comes in agreement with earlier

studies showing LCN2 upregulation in IPMNs [21]. High LCN2 levels were however associated with early dysplastic lesions as well as better differentiated tumor cells. It is therefore believed that LCN2 plays a role in the initiation of pancreatic cell transdifferentiation [21, 131]. Several groups have reported that down-regulation of LCN2 in PDAC cell lines usually leads to reduced attachment, invasion and decreased tumor growth, despite having no effect on tumor cell proliferation [126, 128]. However some reports suggested that LCN2 upregulation lead to downregulation of VEGF and HUVEC tube formation [126], while others have indicated an increased expression of both VEGF and HIF-1α *in vitro* and *in* vivo, as well as increased vascularization *in vivo* [128]. Whether or not LCN2 promotes chemoresistance is still controversial, and seems to depend on the cell line used, as well as the length of the cytotoxicity experiment [126, 128]. LCN2 over-expression has also been associated with increased E-Cadherin expression and decreased vimentin presence in pancreatic cancer, indicating that it might play a role in preventing EMT [131].

In conclusion, LCN2 has been already reported as a promising marker for pancreatic cancer, with potential diagnostic applications, although it does not optimally distinguish between pre-malignant lesions and PDAC. Clinical applications have already been suggested, and Xu *et al*. conducted early experiments, assessing the applicability of an oncolytic virus harboring the LCN2 gene for PDAC treatment [132].

3. Aim of the study

Pancreatic ductal adenocarcinomas remain the solid tumors with the worst overall survival and attempts to develop novel therapies for the disease have been without results. As for other solid tumors, the stratification of patients into subgroups with tumors featuring similar molecular characteristics might allow to better evaluate the treatment options. Recently, three molecular subtypes have been proposed for PDAC, based in transcriptomic profiling. These molecular subtypes predict patient overall survival and response to therapy. However, to date no clinical biomarkers distinguishing between the three molecular subtypes have developed.

The aim of our current study was the identification and validation of novel pan-PDAC as well as subtype-specific protein biomarker candidates, which can be employed in the development of novel clinical applications, such as diagnostics and/or targeted PDAC therapies.

For this purpose, we set out to investigate novel cell surface as well as novel secreted protein biomarker candidates using a mass spectrometric approach employing an *in vitro* system recapitulating all three PDAC molecular subtypes. In addition, we investigated vascular accessible biomarkers using an *in vivo* mouse xenogaft system. Promising biomarker candidates were selected and further validated using both antibody based and antibody independent methods.

We show the identification and validation of two novel pan-PDAC as well as two novel exocrine-like subtype specific protein biomarker candidates. The new biomarker candidates can be employed for patient stratification and the development of alternative antibody-based therapies in PDAC, as well as the development of new non-invasive diagnostic tools.

4. Materials and methods

4.1 Materials

Cell lines

Cell line	Origin	Medium
HPDE	Dr.C. Eisen, HI-STEM/DKFZ; first isolated by Furukawa et al [133]	Keratinocyte serum free + 50 µg/mL bovine pituitary extract
HPNE	ATCC (CRL-4023)	DMEM + M3 + 5% FBS
PACO cell lines	Dr.C. Eisen, HI-STEM/DKFZ	Cancer Stem Cell medium

Mouse strains

NOD.Cg-*Prkdc*^{scid} *Il2rg*^{tm1Wjl}/SzJ JAX (NSG) mice were obtained from the Jackson Laboratory (Bar Harbor, USA) and bred in the DKFZ animal facilities. Mice were kept under pathogen free conditions and males ages 10-12 weeks were used for the generation of orthotopic and metastasis tumor models. All animal handling and experiments followed German laws and were previously approved by the governmental review board of Baden-Württemberg, Germany. Animal experiments were performed under the TVA number G25/12.

Cell culture and xenograft work products

Product	Company	Ordering number
0.2 µm syringe filter	Life sciences	PN4612
100 µm cell strainer	BD	352360
15 mL conical falcon tubes	TPP	Z707724
20 mL Luer Lock syringes	Terumo	SS-20L1
40 µm cell strainer	BD	352340
50 mL conical falcon tubes	Greiner	T2318
50 mL reagent reservoir, sterile	Corning	4870
70 µm cell strainer	BD	352350
75 cm^2 cell culture flasks polystyrene, vented cap	TPP	90076
75 cm^2 primaria cell culture flasks	BD	353810
Acepromazin (Vetranquil)	Ceva	233A6
Cell culture system Inserts 1,0 µm	Falcon	353102

Cryotube 1.8 mL sterile	Nunc	375418
Eppendorf tubes 0.5, 1.5, 2mL safe lock	Eppendorf	13625, 12682, 12776
Falcon Companion-Plattes 6-well	Falcon	353502
Gentle MACS C tube	Miltenyi	130-093-237
Ketamin (Ketavet)	Parke-Davis	51913
LD column	Miltenyi	130 - 0 42-9 01
MACS M tube	Miltenyi	130-093-236
Microvette 500K3E	Sarstedt	20.1341
Nalgene freezing container Mr Frosty	Bunc	51000001
Nunc MaxiSorb Plates	Fischer Scietific	44-2404-21
Primaria cell culture plate 6well	BD	353846
Serological pipettes 2, 5, 10, 25, 50 mL sterile	BD	3565-0.7/-29/-30/-35/-50
TipOne 1000, 200, 20, 10µL filter tips	StarLab	14262, 12704, 12705, 12706
Vacuum filtration system 500 mL 0.2 µm pore size	TPP	99500
Xylazin (Rompun)	Bayer	02169592

Cell culture media and media components

Product	Company	Ordering number
Accutase	Life technologies	A11105
Advanced DMEM/F12	Life Technologies	12634010
Ascorbic Acid	Sigma	A4544
Basic FGF	Peprotech	100-18B-1000
Bovine Serum Albumin	Sigma	A2153
Bromphenol Blue Sodium Salt	Sigma	B8026
CO_2 independent medium	Invitrogen	18045088

Cryostore CS10	Sigma	C2874
DMEM Wo glucose, L-glutamine, phenol red, sodium pyruvate and sodium bicarbonate	Sigma	D5030
DMEM/F12	Life Technologies	21041025
EGF	Peprotech	AF-100-15-1000
Fetal Bovine Serum	Life Technologies	10270
Glucose 45% solution (sterile)	Sigma	G8769
Heparine sodium cell culture tested	Sigma	H3149
HEPES	Life Technologies	15630056
IGF-R3	Sigma	I1271
Insulin recombinant chain	Sigma	I9278
Iron Chloride anhydrous	Sigma	451649
Keratinocyte serum free medium (Kit) with L-Glutamine, EGF and BPE	Invitrogen	17005075
L-Glutathione	Sigma	G-6013
Lipid MIX	Sigma	L-0288
M3 Base F Culture Media	Incell Corp	M300F-500
N2 supplement	Invitrogen	17502048
DPBS 1X	Sigma	P5368
Progesterone	Sigma	P8783
Puromycin	Sigma	P8833

Putrescine dihrydrochloride (cell culture tested)	Sigma	P5780
Sodium bicarbonate Cell culture tested	Sigma	S5761
Sodium pyruvate	Sigma	S8636
Sucrose	Sigma	S0389
Trace Elements A, B, C	Cellgro	MT-99-182-Clc; MT-99-175-CIc; MT-99-176-Cic
Trypsin/EDTA solution 1X	Sigma	T3924
Water (RNase, DNase free, sterile, for cell culture)	Gibco	10977-035
β-Mercaptoethanol	Life Technologies	31350010

Chemical and biological reagents

Product	Company	Ordering number
1-Butanol	Chemosolute	2513.1000
2-Propanon	Sigma	33539
Acetone	Sigma	32201
Acetonitrile ULC/MS purity	Biosolve	01204101
ACK lysis buffer	Lonza	10-548E
ACQUITY UPLC PST C18 nanoAcquity 10K psi column	Waters	186003545
Agarose	GE Healthcare	17-0554-02
Bicine	Sigma	B3876

BM Blue POD Substrate Soluble	Roche	11 484 281 001
Bovine Serum Albumin	Sigma	A2153
C18 OMIX tips 10-100 µL	Agilent	A57003100K
Calcium Chloride	Sigma	C4901
CHCA matrix	Proteochem	P9100
Chlorophorm	ROTH	7331.2
DAKO fluorescent mounting medium	DAKO	S3023
DAKO Glycergel mounting medium	DAKO	C0563
DAKO pen	DAKO	S2002
DAPI dihidrochloride	Sigma	D9542
Dextrane powder 40000 MW	US Biological	D6030
Diisopropyl ether	AppliChem	A4322,1000
DNase (RNase free)	Qiagen	14325
Donkey Serum	Th Geyer	BW/S2170/000100
DTl- Dithiothreitol	Sigma	D9779
EDTA	Sigma	036099
EDTA tetrasodium salt	Sigma	27261
Ethanol absolute	Sigma	32205
Ficoll Paque Plus	GE Healthcare	17-1440-02
Formalin solution 10%	Sigma	HT50-1-2

Fungizone (Amphotericin B)	Life Technologies	15290-018
Glycine	Sigma	G8898
Goat Serum	Sigma	G6767
Hematoxylin	Sigma	51275
Histogel	Thermo	HG 4000-012
Hydrochloric acid 37%	Sigma	30721
Iodoacetamide	Sigma	I1149
L-Cysteine	Sigma	30089
Matrigel	BD	354234
Methanol	Sigma	34860
Midori Green	Nippon Genetics	MG04
Milk (Skim) Powder	Sigma	70166
NP-40	Sigma	74385
NuPAGE 4-12% Bis-Tris gells; 1.5 mm, 10 wells	Life Technologies	NP0335
NuPAGE 4-12% Bis-Tris gells; 1.5 mm, 15 wells	Life Technologies	NP0336
NuPAGE MOPS running buffer	Life Technologies	NP0001
OCT embedding medium	TissueTek	4583
PBS tablets	GIBCO	18912-014
pH9 Antigen retrieval buffer	DAKO	S236784
Protease inhibitor EDTA free tablets	Roche	04-693132 001

PVDF membrane 0.45µm pore size Immobilon P	Millipore	IPVH00010
Quick Load 100bp DNA ladder	Bio Labs	NO467G
RapiGest Surfactant	Waters	186001861
RNA from total human pancreas	Ambion	AM7954 (Lot no: 1011005)
ROCK inhibitor Y-27632	Selleck	S1049
Sodium chloride	Sigma	31434
Sodium citrate tribasic dihydrate	Sigma	C8532
Sodium Dodecyl Sulphate	Sigma	L6026
Sodium hydroxide	Sigma	S5881
Streptavidin-Sepharose High performance	GE Healthcare	175113-01
Sulfo-NHS-LC-Biotin	Proteochem	B2103
Sulfuric acid 95-98%	Sigma	258105
SYBR green mix	Life technologies	4367659
TCEP	BioVision	1202
Trifluoroacetic acid	ProteoChem	LC6203
Triton X-100	Sigma	X100
Trizma Base	Sigma	T1503
Trypsin Porcine sequencing grade	Promega	V511A
Tween-20	Sigma	P137-9

Ultrafree centrifugal filters PVDF	Millipore	UFC30SV00
Urea	Sigma	33247
Water ULC/MS purity	Biosolve	23214 1B1
Xylol	Sigma	33817
β-Mercaptoethanol	Sigma	63689

Antibodies

Primary antibodies

Antibody	Clone/host	Ordering number	Company
CD31 mouse	rat	550274	BD
CD326 Epcam human microbeads	not specified/mouse	130-061-101	Miltenyi Biotech
CDH17	3H2/mouse	SAB1403654-100UG	Sigma
GAPDH	2D4A7/mouse	ab37187	Abcam
LCN2	MM0458-2G32/rat	NBP2-11778	Novus Biological
LGALS4	1E8/mouse	SAB1401230-100UG	Sigma
PCDH1	5D5/mouse	ab55504	Abcam

Secondary antibodies and staining reagents

Antibody	Clone/Lot no	Ordering number	Company
Donkey anti-rat Alexa-488 labeled	Polyclonal/110472	Jackson	712-546-153
Donkey anti-goat Alexa-647 labeled	Polyclonal/111792	Jackson	705-606-147
Donkey F(ab')2 anti-mouse Alexa-488 labeled, preabsorbed	Polyclonal/GR126425-1	Abcam	ab150101

Goat anti-mouse Biotin labeled pre-adsorbed	Polyclonal/GR36816	Abcam	ab98725
Goat anti-mouse HRP labeled	Polyclonal/895492A	Invitrogen	62-6520
Goat anti-mouse-IgG2a	Polyclonal/GR11071-7	Abcam	ab98694
Goat anti-rat Alexa-594 labeled	Polyclonal/716827	Invitrogen	A11007
Rat anti-mouse-IgG1	Monoclonal/clone SB77e (IgG2b)	Abcam	ab99601
Goat anti-mouse FcR block	Polyclonal/108592	115-007-003	Jackson
Streaptavidin Alexa-488 labeled	--	S32354	Invitrogen
Streptavidin-HRP	--	RPN1231V	GE Healthcare

Isotype controls

Antibody	Clone	Ordering number	Company
Mouse IgG1 isotype biotinylated	P3	13-4714-85	eBioscience
Mouse IgG1 isotype unconjugated	P3.6.2.8.1	14-4714-82	eBioscience
Mouse IgG2a isotype unconjugated	eBM2a	16-4724-85	eBioscience
Rat IgG2a isotype unconjugated	RTK2758	400502	BioLegend

Peptides

Peptide	Molecular weight	Source	Application
TVFDEAIR	951.06	PSL Peptide	MALDI MS Quantification
TGVFDEAIRTVGF	1411.72	PSL Peptide	MALDI MS Quantification
CLEHMYHDLGLVRDF	1846.87	PSL Peptide	MALDI MS Quantification
EEQPSTPAPKVEQQEEILC	2155.02	PSL Peptide	MALDI MS Quantification
VVFNTLQGGK	1071.61	JPT	Galectin 4 – MRM
VGSSGDIALHINPR	1446.78	JPT	Galectin 4 – MRM
DAYVFYAVAK	1155.60	JPT	CDH17 – MRM
AENPEPLVFGVK	1308.71	JPT	CDH17 – MRM

Real Time qPCR primers

Target Gene	Primer sequence 5' – 3'	Amplicon size (bp)	Accesion number
CDH17	FW: AGGCACACAGAGTTTGAGGAGAGG RV: GTGGGTATCCCAGTCTGGTG	286	NM_004063.3
GAPDH	FW: ATGGCCTTCCGTGTCCCCACTG RV: GTGGGTGTCGCTGTTGAAGTCAG	183	NM_001256799.1
LCN2	FW: CAGCAGAACTTCCAGGACAAC RV: TTGCGGGTCTTTGTCTTCTC	93	NM_005564.3
LGALS4	FW: ACCCGCCTGTGCCATATTTC RV: TGGGTTGTGGGTGATCTTCTT	236	NM_006149.3
PCDH1	FW: GCTCCATCCCCAGGCCACG RV: GGGCACCCACCTCTAGCTTG	140	NM_002587.4
RPL13a	FW: AAGTACCAGGCAGTGACAG RV: CCTGTTTCCGTAGCCTCATG	80	NM_012423

Kits

Kit name	Company	Ordering number
BCA Protein assay kit	Pierce	23227
DAB Peroxidase substrate kit	Vecto Labs	VC-SK-4100-KI01
ECL Prime Western Blotting detection reagent	GE Healthcare	RPN2232
MACS Tumor dissociation kit, human	Miltenyi Biotech	130-095-929
ProteoSpin Abundant Serum Protein Depletion	NORGEN BIOTEK CORP	17300
QIAshredder (250)	QIAGEN	79656
Rneasy Mini Kit (250)	QIAGEN	74106
SuperScript VILO cDNA Synthesis Kit	Invitrogen	11754-050
Vectastain ABC reagent	Biozol diagnostica	VEC-PK-4000

Laboratory equipment/tools

Equipment	Company
Axioplan Wide Field Microscope	Zeiss
CanoScan 5600F scanner	Alternate
Centrifuge AvantiJ-26XP	Beckmann Coulter
Criotome Microm HM 525	Thermo
Discofix Luer Lock 3SC 10 cm	B Braun
Eppendorf centrifuge 5424	Eppendorf
Eppendorf centrifuge 5424 R with cooling	Eppendorf
Eppendorf centrifuge 5810 R with cooling	Eppendorf
Film developing machine Classic E.O.S	AGFA

Flow hood 1300 series A2 Class II	Thermo
Heidelberger prolongation	B Braun
Heracell 240i CO_2 Incubator	Thermo Scientific
Histo star embedding machine	Thermo
Homogenizor T10 S	IKA
Insulin syringes 1mL 29G	BD
Intas photo imager	INTAS
LSM 700 Confocal Laser Scanning Microscope	Zeiss
MALDI Spotter/micro fraction collector, SunCollect	SunChrom
MALDI TOF/TOF 5800	ABSciex
Microtome Microm HM355S	Thermo
Mupid One electrophoresis unit	Nippon Genetics
Nano Drop ND1000	Thermo
NanoACQUITY UPLC	Waters
nanoAcquity UPLC column	Waters
Operating table (with heating)	Hugo Sachs Elektronik – Harvard Apparatus
pH-meter 211	Hanna Instruments
QTrap 6500	ABSciex
Rotamax 120	Heidolph
Sonicator W-250D	Branson

SpectraMax M5 Plate reader	Molecular Devices
SpeedVac RVC 2-25 CD plus	Christ
Surgical clamps	Fine Science Tools
Syringe pump	Landgraf Laborsysteme
T3000 Thermocycler	Biometra
Thermomixer comfort	Eppendorf
Tuberculin syringes ½ CC 27G	Becton Dickinson and Company
Tumbler	Cole Parmer
UltraCentrifuge Optima L-90K	Beckman Coulter
Venofix 25G Luer Lock	B Braun
Vi-Cell XR Cell Viability Analyzer	Beckman Coulter
Viia7 Real Time PCR system	Thermo
Waterbath	Julabo
Western Blot CL-X Posure Films	Thermo
Xcell II Blot Module	Invitrogen
Xcell SureLoc Mini-Cell Electrophoresis System	Invitrogen

Software tools

Tool	Version	Source
GraphPad Prism	6	www.graphpad.com
Microsoft Office	2013	http://office.microsoft.com
MS_QBAT	1.0	Developed in house (A. Kerner and Dr. C. Rösli, HI-STEM)
NetPrimer – premier biosoft	2014	http://www.premierbiosoft.com/netprimer
ProteinPilot	4.0/4.5	http://www.absciex.com
Viia7	1.0	https://www.lifetechnologies.com
WebGestalt	2014	http://bioinfo.vanderbilt.edu/webgestalt/
ZEN lite	2011	http://www.zeiss.com

Solutions and media formulation

Cancer Stem Cell medium:

500 mL Advanced DMEM/F12

5 mL N_2 supplement

0.25 mL Trace elements A

0.5 mL Trace elements B

0.5 mL Trace elements C

1 mL Lipid-MIX 1

2 mM L-Glutamine

0.6% glucose (added to final concentration)

5 mM HEPES

2 µg/mL heparine

100 µM β-Mercapto ethanol

1 µg/mL GSH

50 ng/mL hBasic FGF

20 ng/mL hEGF

10 ng/mL IGFR3

+ 0.06% BSA (final concentration excluding BSA from the Advanced DMEM/F12)

+25 mL H_2O

Cancer Stem Cell medium, BSA/Transferrin depleted:

500 mL DMEM/F12

0.25 mL Trace elements A

0.5 mL Trace elements B

0.5 mL Trace elements C

1 mL Lipid-MIX 1

2 mM L-Glutamine

0.6% glucose (added to final concentration)

5 mM HEPES

2 µg/mL heparine

1 mM Sodium pyruvate

10 mg/L Insulin

10.01 mM Putrescine

0.002 mM Progesterone

0.00863 mM Ascorbic Acid

0.05 mM Iron Chloride

100 µM β-Mercapto ethanol

1 µg/mL GSH
50 ng/mL hBasic FGF
20 ng/mL hEGF
10 ng/mL IGFR3
+25 mL H_2O

DMEM + M3 medium:
75% DMEM without glucose
25% M3 medium
1.5 mM L-glutamine
1.125 g/L $NaHCO_3$
5.5 mM D-glucose
10 ng/mL human recombinant EGF
750 ng/mL puromycin
5% fetal bovine serum

Keratinocyte medium:
Keratinocyte serum free medium
50 µg/mL bovine pituitary extract
5 ng/mL EGF (Gibco)

CO_2 independent medium:
CO_2 independent medium
2 mM L-Glutamine
1% BSA

2mM PEB/10mM PEB:
2 mM EDTA/10 mM EDTA
1x PBS
1% BSA

Wash buffer A:
1% NP-40
0.1% SDS
1x PBS

Wash buffer B:

2 M NaCl

0.1%SDS

1x PBS

Trypsin digestion buffer (TDB):

50 mM Tris-HCl

1 mM $CaCl_2$

pH 8.0

Cell lysis buffer:

10 mM EDTA

2% NP-40

0.2% SDS

1x PBS

1x Protease inhibitor added before use

RIPA (RadioImmunoPrecipitaion) lysis buffer:

50 mM Tris-HCl pH 8.0

150 mM NaCl

1% NP-40

0.5% sodium deoxycholate

0.1% SDS

pH 8.0

Tissue lysis buffer:

2% (wt/v) SDS

50 mM Tris-HCl

10 mM EDTA

1x PBS

pH 6.8

1x Protease inhibitor added before use

Antigen retrieval buffer pH6:

10 mM Trisodium citrate

0.05% Tween 20

pH 6.0

Anesthetic mixture (surgeries):

10 mg/mL Ketamin (Ketavet)

1.6 mg/mL Xylazin (Rompun)

Prepared in water

Anesthetic mixture (terminal perfusions):

20 mg/mL Ketamin (Ketavet)

2 mg/mL Xylazin (Rompun)

0.3 mg/mL Acepromazin (Vetranquil)

Prepared in 0.9% NaCl solution

SDS Loading buffer (5x):

208 µM Tris, pH = 6,8

33% Glycine

5% SDS

0,06% Bromphenolblau

DNA loading buffer (6X):

1.168 M sucrose

3.6 mM Bromphenol Blue sodium salt

In water

Western Blot running buffer:

25 mM Bicine

25 mM Bis-Tris

1 mM EDTA

50 nM Chlorobutanol

10% Methanol

In water

TBE buffer:

0.089 M Tris

0.089 M H_3BO_3

20 mM EDTA pH8

14 mM NaOH

Final pH 8.3

4.2 Methods

4.2.1 Cell culture methods

Culturing of PACO cells

A total of twelve primary pancreatic cancer cell lines were obtained from Dr. Christian Eisen (HI-STEM/DKFZ). Cells were grown on Primaria 75 cm^2 flasks in serum free CSC medium developed and optimized by Dr. Martin Sprick and Dr. Christian Eisen, until ~80% confluence was reached.

For the subculturing of the PACO cells, the medium was aspirated, washed the flask with PBS, and 5 mL of Accutase were added per 75 cm^2 flask. Cells completely detach from the substrate in ~15 minutes, at which point the enzymatic activity was stopped using CO_2 independent medium (double amount compared to Accutase volume) and the cell suspension was transferred in a falcon tube. Cells were centrifuged (450*xg*/4°C/5 minutes) and the pellet was resuspended in 1mL CSC medium. The standard split rate for the PACO cells was 1:10; for the classical subtype cell lines, sometimes a split rate of 1:20 was necessary, due to their fast growth rate.

Freezing of PACO cells

For cryopreservation purposes, the cells were detached and span down as previously mentioned. Then, the pellet was resuspended in Crystor CS10 and aliquoted into cryovials. The vials were placed in a Mr. Frosty and transferred to -80°C for 1-2 days, prior to being long term stored in liquid nitrogen.

Culturing of the control cell lines

The two control cell lines were grown in the medium recommended by the provider. For the HPNE cells, the growth medium was DMEM-M3 supplemented with EGF and 5% FBS. For the HPDE cell line, the recommended medium was Keratynocyte medium supplemented with bovine pituitary extract and EGF [133]. All control cell lines were grown in TPP 75 cm^2 flasks.

For the subculturing of the control cell lines, cells were detached using 3 mL trypsin-EDTA/T75 flask. HPNE cells needed 5 minutes for complete detachment, while the HPDE cell line required 15 minutes. Trypsin was subsequently inactivated using the specific cell line medium, added in double amount compared to the volume of the detachment enzyme. Cells were later span down using the above described centrifuge parameters and the pellet was resuspended in 1mL of the respective growth medium. Split ratios were 1:10 for the HPNE cell line, and 1:6 to 1:10 for the HPDE cell line.

Freezing of control cell lines

For cryopreservation, cells were detached and centrifuged as previously mentioned, and the resulting pellet was resuspended in Crystor CS10 (for the HPDE cell line) or 95%

DMEM-M3 medium + 5% DMSO cryo-medium (for the HPNE cell line). Alliquoting, freezing and storage was performed identically as for the PACO cell lines.

Counting of cells

Cell numbers were dermined using the ViaCell counter. For this purpose, the cells were detached (depending on the subsequent experiment, using 10mM PEB buffer or Accutase), pellet and resuspended in 1 mL PBS. A 1:10 dilution was performed in a Beckman Counter 4 mL sample cup (final volume 500 µL) and the cells were counted by the ViaCell counter.

4.2.2 Generation of orthotopic pancreatic mouse tumors

NSG male mice were anesthetized using the anesthetic mixture for surgeries (administered at a dose of 10 µL/g), injected intra-peritoneal. The abdominal cavity was opened and a mixture of 10^6 viable PACO cells and Matrigel (2 mg/mL final concentration) was injected in the head of the pancreas (injection volume 30 µL). The peritoneum was sutured and the skin was clipped and disinfected, after which mice were kept under observation for 24 hours.

Tumor growth was monitored by regular palpation of the mice. The animals were sacrificed when the size of the tumor reached ~ 1 cm in diameter. Generally, classical derived tumors required two months to develop, exocrine derived tumors three to four months, while QM tumors required six to nine months.

4.2.3 Discovery of novel protein biomarkers in vitro

4.2.3.1 Discovery of novel cell surface biomarkers in vitro

Biotinylation of cells in vitro

Cells were grown in their recommended medium, until 70-80% confluence was reached. At this point, the medium was removed, cells were washed twice with PBS and a 0.5 mg/mL Sulfo-NHS-LC biotin solution in PBS was added to the flask (total biotin ester amount: 2 mg). The flasks were incubated on a shaker for 10 minutes at room temperature, and the excess biotin ester was inactivated using 100 µL of a 1 M Tris-HCl buffer. The solution was removed, cells were washed once with PBS and were later detached using a 10 mM EDTA, 1% BSA in PBS solution. After ~20 minutes, the detached cells were resuspended and centrifuged.

Cell lysis

The resulting pellet of cells was washed once with PBS, and the cell number was determined. Cells were lysed using the cell lysis buffer. Briefly, cells were incubated in the buffer for 30 minutes on ice, vortexed and later sonicated (10% intensity, 3 x 30 seconds). Cell debris were pellet by centrifuging full speed for 20 minutes, and the supernatant was stored at -20°C. Protein concentration of the total lysates was determined using the BCA kit (Pierce), according to the instructions of the provider.

Enrichment of cell surface proteins and on resin tryptic digestion

250 µg of total protein lysate were loaded onto a sepharose-streptavidin column, pre-equilibrated in wash buffer A. SDS was added to the sample to a final concentration of 2%. The biotinylated fraction was captured for two hours, tumbling at room temperature and unspecific binders were removed by performing a series of stringent washes: two washes with wash buffer A (1% NP-40, 0.1% SDS in PBS), two washes with 10% SDS in PBS, two washes with the high salt buffer B (2 M NaCl, 0.1% SDS in PBS) and ten washes in the trypsin digestion buffer (TDB) (50 mM Tris HCl, 1 mM $CaCl_2$ pH8).

The sepharose-streptavidin slurry containing the captured biotinylated fraction was digested using 800 ng trypsin (Promega) dissolved in TDB (final reaction volume 210 µL) for ten hours at 37°C.

4.2.3.2 Discovery of novel secreted biomarkers in vitro

Collection of supernatants containing secreted proteins

Cells were grown in their standard culture conditions until they reached ~60% confluence.

At this point, the medium was removed, cells were washed twice with PBS and incubated with the CSC medium depleted of BSA and transferrin. The depleted medium was formulated to match the normal CSC medium, while completely excluding BSA and transferrin. After 48 hours, the medium was collected and underwent a series of centrifugations for the removal of cell debris (5 minutes/500*xg*/4°C), of intracellular organelles (20 minutes/20'000*xg*/4°C) and finally microvesicles (2 hours/100'000*xg*/4°C). The resulting cleared supernatant was aliquoted and stored at -20°C.

In solution digest of secreted proteins

Protein concentration of the supernatant fraction collected after the 100'000*xg* centrifugation step was determined using the BCA method. Due to high dilution of proteins in the supernatant and interactions from components present in the culture medium with the BCA reagent, proteins were precipitated prior to concentration evaluation using the chloroform-methanol method.

Briefly, a sample aliquot was precipitated by adding four volumes of methanol, one volume of chloroform and three volumes of water (vortexed after each step). The mix was centrifuged at 15'000*xg* for 2 minutes, and the aqueous phase was discarded. The proteins located in the interface were precipitated by adding four volumes of methanol, and were subsequently pelleted by centrifugation. Finally, the pellet was dried using the SpeedVac. The precipitate was solubilized directly in the final BCA reagent volume suggested by the provider, and protein concentration determination was achieved by the standard method recommended by the manufacturer.

For the in solution digest, 15 µg protein were precipitated using the above described chloroform-methanol method and resuspended directly in trypsin digestion buffer. Trypsin was added in a 1:50 ratio (trypsin: total protein) and enzymatic digestion was performed for 10 hours at 37°C.

4.2.4 Evaluation of in vivo vascular accessible protein markers

Whole body perfusion using reactive biotin esters

Orthotopic pancreatic tumors were grown in NSG mice, according to the protocol mentioned above.

The whole body perfusion was performed as described by Rösli et al [134]. When the palpated tumor reached ~ 1 cm in diameter, mice were anesthetized using a mixture of Ketamin, Xylazin and Acepromazin. After confirming loss of pain sensitivity in the anesthetized animals, the abdominal cavity was opened, followed by the opening of the thoracic cavity. A needle was inserted in the left ventricle of the heart, connected to an automated pumping system containing 1 mg/mL reactive biotin ester solution in 10% dextrane-PBS. A cut was made in the right atrium, to allow evacuation of blood and pumped liquid, and the solution was pumped at a rate of 1.1 mL/minute for 10 minutes. The thoracic and abdominal cavity were sprinkled with a 50 mM Tris-HCl solution prepared in PBS pH 7.4 to prevent biotinylation due to the evacuating unreacted ester from the circulatory system. After ten minutes, the biotin ester solution was replaced with a quenching solution, containing 50 mM Tris in dextrane-PBS, and the mouse circulatory system was washed for an additional ten minutes, using the same flow rate. All solutions were kept at 40° C prior to use, and all operations were performed on a heated surgery table, in order to prevent capillary contractions as a result of hypothermia.

Pancreatic tumors and any visible metastases were collected and snap frozen for proteomic analysis, or embedded in OCT for evaluation of biotinylation/analysis of expressed protein markers. Healthy liver and in the case of healthy controls normal pancreas samples were also collected.

Evaluation of perfusion success using immunofluorescence on cryosections

Pieces of biotinylated organs were snap frozen in OCT medium and cut in 8 µm thick sections using a Microm HM 525 cryotome.

Cuts were fixed in acetone at -20°C for eight minutes, and later delineated with a DAKO silicon pen. Blocking was achieved using 20% goat serum in PBS for one hour at room temperature. Incubation with the primary antibody (rat anti-mouse CD31) was performed for one hour at room temperature. After rinsing the cuts with PBS, incubation with secondary antibody Alexa-594 goat anti-rat and additionally Streptavidin-Alexa488 (for the biotin detection) was carried out for one hour at room temperature. Negative controls were incubated only with the secondary goat anti-rat antibody.

Cuts were thoroughly washed in PBS, mounted with DAKO fluorescent medium and stored at 4°C in the dark. Pictures were acquired and processed as described in the validation section below.

Tissue lysis of biotinylated organs

Collected tissues were weighted and minced carefully. Upon transferring the tissue in a 15 mL tube, 20 µL of tissue lysis buffer was added per mg sample, and the tissue was

homogenized 3 x 2 minutes (for pancreas and liver samples) or 5 x 2 minutes (for tumor samples), using an TS10 tissue homogenizer at full power.

Homogenized samples were sonicated (Vibra-cell) 21 x 5 sec at 25% intensity (5 sec off pause on ice) and heated to 95°C for 20 minutes. Tissue debris were removed by centrifugation at 21'000*xg*, for 20 min, at room temperature. The supernatant was aliquoted and stored at -20°C.

Evaluation of biotinylation level using ELISA

A standard biotinylation curve was designed, using biotinylated and native BSA. Briefly, a total of 6.65 µg BSA (corresponding to 10^{-5} M) were pipetted in each well, by summing native and biotinylated protein. The concentration range of the curve varied between 0 to 2.5 µg biotinylated protein.

Total protein concentration was determined using the BCA kit, as recommended by the manufacturer. Based on the determined value, samples were diluted to allow coating of 10^{-5} M antigen (the average protein mass was considered to be 53 kDa). Samples were pipetted on a Nunc maxisorp plate, in parallel with the biotinylation standard curve, and allowed to bind overnight, at 4°C. The following morning, the wells were washed and blocked using 2% BSA in PBS for one hour at room temperature. All staining reagent dilutions were performed in the blocking solution. 100 µL streptavidin-HRP (diluted 1:1000) were added to the wells and incubated for one hour at room temperature. Wells were thoroughly washed using PBS containing 0.1% Tween and PBS and 100 µL BM Blue POD substrate (Roche) were added for detection. The samples were incubated with the substrate for a maximum of 20 minutes, depending on the intensity of the blue color observed, and the reaction was stopped with 70 µL of 1 M H_2SO_4.

The plates were analyzed using the SpectraMax M5 plate reader, by measuring the absorbance at 450 nm and subtracting the absorbance observed at 690 nm.

Streptavidin enrichment of vascular accessible biotinylated proteins, alkylation of cysteines and tryptic digestion

The equivalent of 50 µg biotinylated protein was used for subsequent enrichment on a streptavidin column and tryptic digestion.

Samples were diluted in buffer A containing 2% SDS and 2 mM TCEP and allowed to bind for two hours at room temperature, while tumbling. Alkylation was performed by adding 100 µL iodoacetamide (50 mM in PBS, pH 7.4) and incubating for 30 minutes in the dark. The reaction was stopped by adding an excess of cysteine (300 µL of 100 mM cysteine solution in water) and incubating for an additional 15 minutes in the dark.

A delipidation step was performed by adding 400 µl wash buffer A and 800 µl butanol-DIPE (di-isopropyl ether) solution (40:60 [v:v]). Samples were incubated for 30 minutes at room temperature under constant agitation and the solvent phase was later removed.

Unbound proteins were removed by washing two times with wash buffer A and ten times with TDB. Tryptic digestion was performed as previously described for the *in vitro* discovery of biomarkers.

4.2.5 Mass spectrometric analysis

Purification of tryptic peptides

Following the tryptic digestion, the peptide containing aqueous fraction was acidified with trifluoroacetic acid (TFA) to a final concentration of 0.1%. The acidification of the solution is required for optimum purification and de-salting of peptides using C18 packed mini-columns (OMIX tips, Agilent).

The C18 columns were activated by washing two times with a 50% acetonitrile (ACN) solution, and equilibrated by washing two times with 0.1% TFA solution. Peptides were captured on the C18 column by pipetting up and down twenty times. The column was washed three times with the equilibration buffer and peptides were afterwards eluted in a 75% ACN, 0.1% TFA solution. The peptides were ultimately dried in a SpeedVac and stored at -20°C.

Fractionation of the tryptic peptide mix using reverse phase UPLC and spotting of the eluted fraction

The dried tryptic peptides were resuspended in 20 µL 5% ACN, 0.1% TFA and loaded on an Acquity UPLC BEH130 C18 column (Waters), 75 µm x 250 mm with 1.7 µm particles and an average pore size of 142 Å. 8 µL of sample were loaded directly on the analytical column and eluted using an ACN gradient. Buffer B (ACN containing 0.1% TFA) concentration was varied as follows: buffer B concentration was increased from 5% to 11% in 0.33 minutes, maintained for another 0.67 minutes, then increased to 14% in 2.66 minutes. Elution was performed using two consecutive linear gradients, from 14% to 30% buffer B for 64 minutes, and from 30% to 40% for 13.34 minutes. The column was washed by first linearly increasing the concentration of buffer B to 85% in nine minutes, maintaining it for an additional five minutes, after which the concentration of buffer B was brought back to base line 5% in two minutes. The column was equilibrated with the standard 5% buffer B for an additional 13 minutes. Total running time was 110 minutes, the constant flow rate was 350 nL/minute and the temperature of the column was set to 35°C.

The eluted fractions were mixed with α-Cyano-4-hydroxycinnamic acid (CHCA) matrix solution prepared in 80% ACN containing a mixture of four standard peptides spiked in and spotted onto a MALDI plate by an automated spotter, depositing a spot every 4 seconds, in a total of 1200 fractions (Sunchrom). The concentration of the spiked in standard peptides in the final CHCA matrix mixture is summarized in **Table 2**:

Table 2: Concentration of internal standard peptides in final matrix mixture for MALDI analysis

Peptide sequence	Concentration (fmol/µL)
CLEHMYHDLGLVRDF	107.2
EEQPSTPAPKVEQQEEILC	214.3
TGVFDEAIRTVGF	53.6
TVFDEAIR	26.8

MALDI MS and MS/MS analysis

The samples were analyzed by the AbSciex MALDI TOF/TOF 5800. The mass spectrometer was operated in positive ion reflector mode. The number of laser shots per spot was set to 2000 and the acquisition rate was 400Hz. The velocity of the laser was 1000 µm/second and the intensity of the laser was optimized prior to each analysis. Data were acquired in DDA (Data Dependent Acquisition). The TOF/TOF Series Explorer software calculated the peaks on interest, and MS^2 was performed using the Dynamic Exit processing method, with a maximum of 35 MS^2 spectra analyzed per spot, weakest precursor first. The spectra acquisition would be completed after a maximum number of 3000 laser shots per precursor, or until the quality of the spectra was evaluated by the software as high. The acquisition rate was set to 1000 Hz and the velocity of the laser was 1200 µm/second.

For the *in vitro* samples, the MS/MS spectra were processed using the ProteinPilot software version 4.0, or version 4.5, and the search against the human data base was performed using the Paragon algorithm version 4.0.0.0 or version 4.5.0.0. The Homo sapiens reference proteome data set was downloaded from UniProt on 16th September 2011, and the search data base included 71020 proteins. For the *in vivo* samples, searches were performed against a mix human and mouse data base. The reference proteome data set was also downloaded from UniProt, on the 15th of May 2013, and the combined data base contained 243244 proteins, including reverse sequences. For the *in vitro* samples, ProteinPilot searches were performed allowing for missed cleavages and no cysteine modification were defined. The search mode was set to Thorough ID, with ID focus on biological modifications and amino acid substitutions. A detected protein threshold of 1.3 was chosen and peptides with a confidence interval greater than 95% were accepted. For the *in vivo* samples, the definitions for the search algorithm were identical to the ones previously mentioned, with one exception: cysteine modifications – namely alkylation, were specified for the search algorithm.

4.2.6 Validation of protein biomarkers

4.2.6.1 Antibody based validations

4.2.6.1.1 Western Blot

Total cell lysates obtained as previously described were incubated with the reducing buffer for 10 minutes at 95°C and loaded on a precasted NuPAGE Bis-Tris 4-12% gradient gel (Novex). Samples were ran in MOPS SDS running buffer for one hour, 200 V constant.

Immunoblotting was performed using PVDF membranes (activated prior to transfer in methanol) according to the manufacturer's instructions, for 1.5 hours at 30 V. Membranes were blocked with 4% milk-PBS solution over night at 4°C. Staining with the primary antibody was performed in 2% milk-PBS solution for one hour at room temperature, followed by incubation with a horse radish peroxidase (HRP) labeled secondary antibody diluted in 2% milk solution for another hour. Membranes were developed using the ECL development kit, according to the instructions of the provider.

Dilutions used for the primary antibodies were: PCDH1 1:2000, GAPDH 1:10000. The secondary goat anti-mouse – HRP labeled antibody was diluted 1:2000.

4.2.6.1.2 Immunocytofluorescence on methanol fixed cells

First validation steps were performed *in vitro* using the cultured primary PDAC and control cells. For this purpose, cells were grown on special cell inserts for 6 well plates, representing an adherent porous membrane (pore size 1 µm). Cells were seeded at a density of 5 x 10^5 cells per well and were allowed to grow for 48 hours (with the exception of PACO14 and PACO16 cell lines, requiring 96 hours to reach the desired confluence). Two milliliters of medium were added both on top, and in the six well plates holding the inserts, to insure sufficient amount of nutrients for the growing cells, as well as proper humectation of the membranes on both sides. Cells were later fixed with ice cold methanol for 10 minutes at -20°C and washed with PBS. The membranes were cut in smaller pieces and then stored in 0.1 M glycine in PBS at 4°C for 24 hours.

The cut membrane inserts were later stained following the standard protocol. Inserts were incubated in 0.2% Triton X solution in PBS for 15 minutes for membrane permeabilization. The fixed cells were blocked with a 10% serum solution in PBS, depending on the nature of the secondary antibody for 30 minutes at 37°C. The primary and secondary antibodies were diluted in 3% BSA solution in PBS. Incubation with the primary antibody was performed for one hour, at 37°C, membranes were washed three times with PBS and incubated with the secondary antibody for another hour at 37°C. DAPI staining was performed in parallel with the secondary antibody incubation. After the final washing step, the cut membrane inserts were mounted using Fluorescence mounting medium (DAKO) and stored at 4°C in the dark awaiting visualization. Isotype controls were performed for each specific antigen.

The individual dilutions used in the experiment were: CDH17 antibody 1:50; LGALS4 antibody 1:30; PCDH1 antibody 1:20. Isotype controls were diluted to the same concentration as that of the corresponding primary antibodies. The dilutions for the secondary antibodies were: Goat anti-mouse biotin labeled (Abcam) 1:200 incubated with Streptavidin-Alexa488 (Invitrogen) 1:200 (for CDH17 detection); Donkey anti-mouse Alexa488 labeled (Jackson) 1:250 (for the LGALS4 and PCDH1 primary antibodies). DAPI dilution was 1:100.

Fluorescence staining and localization of the antigen of interest was achieved using the LSM 700 confocal fluorescence microscope (Zeiss) and pictures were later processed using the ZEN lite 2012 software. Scale bars were added and contrast for individual channels were optimized identically for all samples stained for one specific antigen.

4.2.6.1.3 Immunocytofluorescence on formalin fixed paraffin embedded cells

For those antibodies that did not stain optimally when cells were fixed with methanol, immunofluorescence was performed on formalin fixed cells embedded in Histogel, and finally embedded in paraffin. For this purpose, cells were detached non-enzymatically using a 10mM EDTA, 1% BSA solution in PBS, and were later fixed in formalin for 10 minutes at room temperature. Fixed cells were resuspended in 300 µL Histogel (pre-warmed to 60°C), allowed to cool and stored in formalin until paraffin embedding was performed. Stainings

were performed using the protocol for immunohistofluorescence of formalin fixed, paraffin embedded mouse derived tumors, described below.

Antibody dilutions were: LCN2 1:150; donkey anti-rat 1:250; DAPI 1:100. Pictures were acquired using the LSM700 (Zeiss) and processed as previously described.

4.2.6.1.4 Testing colocalization of putative exocrine-like biomarkers using confocal microscopy

In order to investigate if the two putative exocrine-like protein biomarkers interact, their localization was assessed using confocal laser microscopy after co-staining methanol fixed PACO cell lines. Isotype specific unlabeled secondary antibodies, raised in different hosts, were employed for the recognition of the primary antibodies previously used, taking advantage of the fact that they were of different isotypes. The final detection was achieved using fluorescently labeled tertiary antibodies, distinguishing the different species from which the secondary antibodies have been obtained.

Briefly, the cells were grown on cell inserts and fixed as previously described. After permeabilizing the cells and blocking with 10% donkey serum + 10% goat serum (prepared in PBS), the cells were incubated with a mixture of CDH17 antibody (diluted 1:75) and galectin 4 (diluted 1:30) for one hour at 37°C. After washing the inserts, the cells were incubated (one hour/37°C) with a mixture of unlabeled secondary antibodies distinguishing between the isotypes of the primary antibodies. The secondary rat anti-mouse IgG1 antibody (recognizing the CDH17 antibody) was diluted 1:200, while the goat anti-mouse IgG2a antibody (recognizing the LGALS4 antibody) was diluted 1:400. Cells were washed with PBS and incubated with the final tertiary antibodies using the following dilutions: donkey anti-rat Alexa-488 labeled (1:200) and donkey anti-goat Alexa-647 labeled (1:200). DAPI was added to the tertiary antibody mix to a final dilution of 1:100 and the cells were incubated for one final hour at 37°C. Mounting and image acquisition/processing were performed as previously mentioned.

Appropriate controls were also stained in parallel, evaluating the specificity of the secondary and tertiary antibodies. For this purpose, either individual primary antibodies were excluded from the analysis (while using both secondary and both tertiary antibodies), or, each or both secondary anti-isotype antibodies were omitted, while using both primary and tertiary antibodies. A scheme of the appropriate staining controls is listed in **Table 3**.

Table 3: Staining controls for colocalization evaluation experiments. All samples were incubated with the same mixture of tertiary antibodies, while the primary (first row) or secondary antibodies (second row) were selectively excluded to assess the specificity of the secondary and respective tertiary immunoglobulins. Secondary and tertiary antibodies were color coded in the table representation. Legend: Ab = antibody

Indirect IF strategy	Individual stainings + control stainings		
Primary Ab **+** **Secondary Ab** **+** **Donkey anti-rat & Donkey anti-mouse tertiary Ab**	CDH17 +LGALS4 + anti-IgG1 & anti-IgG2a + Tertiary Ab	CDH17 + anti-IgG1 & anti-IgG2a + Tertiary Ab	LGALS4 + anti-IgG1 & anti-IgG2a + Tertiary Ab
	CDH17 +LGALS4 + anti-IgG1 + Tertiary Ab	CDH17 +LGALS4 + anti-IgG2a + Tertiary Ab	CDH17 +LGALS4 + No secondary Ab + Tertiary Ab

4.2.6.1.5 Immunohistofluorescence

Orthotopic PACO derived pancreatic tumors were fixed in 10% formalin for 48 hours at 4°C and embedded in paraffin. For staining, the tissues were rehydrated and antigen retrieval was achieved by boiling the tissue in pH 6 citric buffer (for LGALS4 and PCDH1 antigens) or pH9 retrieval buffer (DAKO) (for CDH17 and LCN2 antigens) for 15 minutes. After rinsing the slides, blocking was performed using a 20% donkey serum solution in PBS for one hour at room temperature, followed by an additional blocking step using FcR fragments goat anti-mouse (Jackson) diluted 1:10 in 12% BSA prepared in PBS. Incubation with the primary antibody was carried out over night at 4°C, while the secondary antibody incubation was performed for one hour at room temperature. All antibody dilutions were performed in 12% BSA solution. Slides were rinsed and mounted using the Fluorescence mounting medium (DAKO). Isotype controls were performed at concentrations matching to those of the antibodies of interest.

Individual dilutions used in the experiments were: CDH17 1:750; LGALS4 1:600; LCN2 1:600; PCDH1 1:400. Secondary antibodies were diluted as follows: donkey anti-mouse – Alexa488 coupled 1:750 for CDH17 stainings, 1:250 for LGALS4 and PCDH1 stainings; donkey anti-rat Alexa488 coupled 1:300. DAPI dilution was 1:100.

Picture acquisition and processing was performed as previously described.

4.2.6.2 Antibody independent validations

4.2.6.2.1 MRM validations

Multiple reaction monitoring was used to quantify and validate the presence of the markers of interest in both whole cell lysates and secreted fraction.

For validation/quantification of proteins in PACO and control cell lines, cells were lysed directly on the flask using the RIPA buffer, and the total cell lysates were clarified by centrifuging 20'000*xg* for 20 minutes. The protein concentration was determined using the BCA method, and 100 µg proteins were precipitated using the chloroform – methanol method previously described, in order to further increase the protein concentration.

The proteins were resuspended in 0.05% RapiGest solution in TDB for better resolubilisation. RapiGest surfactant allows for improved digestion, without inhibiting enzymatic activity, and can later be removed from the sample by acidification of the solution. Disulfuric bounds were reduced by adding DTT to final concentration of 5 mM and heating the samples to 60°C for 30 minutes. Alkylation was performed by adding iodoacetamide solution to a final concentration of 15 mM and incubating in the dark for 30 minutes. Tryptic digestion was performed overnight with a 50:1 protein: trypsin ratio, at 37°C, under agitation.

RapiGest surfactant was removed by acidifying the sample using TFA to a final concentration of 0.5% and incubation for 45 minutes at 37°C, followed by precipitation of the detergent at 17'000*xg* for 10 minutes. The tryptic peptides present in the aqueous phase were desalted as previously described and the peptides were dried in the SpeedVac and stored at -20°C.

For the validation/quantification of the proteins of interest in the secreted fractions, we used the collected supernatants of cells grown in BSA/transferrin depleted medium, fractionated by ultracentrifugation as detailed earlier. 50 µg of total protein from cleared supernatants were reduced using 5 mM DTT at 60°C/30 minutes, followed by alkylation with 15 mM iodoacetamide at 24°C/30 minutes.

Alkylated proteins were precipitated using acetone, according to the following protocol: five times excess ice cold acetone was added to the protein solution and incubated at -20°C overnight. Proteins were pelleted by centrifuging at 20'000*xg*, for 30 minutes at 4°C. Pellets were washed with 80% ice cold acetone and centrifuged at 20'000*xg* for 5 minutes at 4°C, followed by complete solvent removal and drying of the protein precipitate at room temperature.

Proteins were solubilized in 4 mM Urea in TDB, and digested with trypsin added in a 1:50 (trypsin: protein) ratio for 10 hours/37°C. Digested samples were acidified with TFA (0.1% final concentration) and peptides were desalted using the standard OMIX protocol described earlier.

For the MRM analysis, samples were diluted to a final concentration of 2 µg/mL. Analysis were performed on an ABSciex QTrap 6500 coupled online to a Waters online nanoACQUITY system, in collaboration with Dr.Christoph Rösli and Wiebke Nadler (Biomarker Discovery Group, HI-STEM). Sample fractionation was performed on Atlantis dC18 nanoACQUITY Column or Fortis HPLC column 1.7µ Fortic C18 capillary. Atlantis dC18 columns are silica based reverse phase C18 columns, with particles containing di-functionally bonded C18 ligands, optimized for highly aqueous phases. The column was

15 cm long, with a diameter of 300 µm, a 3 µm particle diameter and 100 Å pore size. In the case of Fortis HPLC column, recommended for their stability over a wide pH range (1-12), the size characteristics were the following: 15 cm length, 200 µm diameter, with 1.7 µm particles. The buffers used for the nanoACQUITY system were: UPLC purity grade water with added 3% DMSO, 0.1% formic acid and 0.01% TFA (defined as buffer A) and ACN UPLC grade purity with DMSO, formic acid and TFA added in the previously mentioned percentages (defined as buffer B). The peptides were eluted using an 117 minutes gradient, with the following specifications: the concentration of buffer B (ACN based) was increased linearly from 3% to 15% over 30 minutes; a second linear gradient increasing the buffer B concentration from 15% to 25% was performed for the following 55 minutes, and a third linear gradient, raising the concentration of buffer B from 25% to 32% was applied for another 25 minutes. The column was washed by increasing the concentration of buffer B to 90% in two minutes and the high ACN concentration was maintained constant for an additional four minutes. Finally, the column was re-equilibrated by returning the buffer B concentration to the base-line 3% within one minute. The flow rate of the mobile phase was 4 µL/minute and the temperature of the column was set to 45°C.

For each protein to be quantified, we optimized three transitions per proteotypic peptide, and we used two precursor peptides per protein. Heavy labeled corresponding peptides were ordered from JPT and spiked in the samples, to a final concentration of 15.6 fmol/sample. The details regarding the mass filter of Q1 and Q3 quadrupoles, the optimized declustering potential, collision energy and the observed retention time of precursor peptides (light and heavy) are listed in **Table 4.** All selected precursor ions were double charged, while all fragment ions analyzed were single charged.

Table 4: Transition of Cadherin-17 and Galectin-4 peptides for SRM

Legend: D P = declustering potentional; CE = collision energy; RT = retention time

Protein	Peptide	Precursor mass	Fragment ion	Fragment mass/charge state	DP (V)	CE (eV)	RT (min)
Cadherin-17	DAYVFYAVAK light	573.795	y6	698.387	72.9	29.5	44.2
			y7	797.455			
			y8	960.518			
	DAYVFYAVAK heavy	577.802	y6	706.401			
			y7	805.469			
			y8	968.533			
	AENPEPLVFGVK light	650.350	y4	450.271	78.5	32.3	44.56
			y7	759.476			
			y9	985.571			
	AENPEPLVFGVK heavy	654.357	y4	458.285			
			y7	767.490			
			y9	993.585			
Galectin-4	VVFNTLQGGK light	531.800	y6	603.346	69.9	28	29.12
			y7	717.388			
			y8	864.457			
	VVFNTLQGGK heavy	535.807	y6	611.360			
			y7	725.403			
			y8	872.471			
	VGSSGDIALHINPR light	718.386	y5	636.357	83.5	34.7	30.53
			y6	749.441			
			y7	820.478			
	VGSSGDIALHINPR heavy	723.390	y5	646.365			
			y6	759.449			
			y7	830.487			

4.2.6.2.2 Gene expression analysis

RNA isolation from cultured cells and total tumor tissue

In order to evaluate the gene expression levels for the markers of interest, we performed first an *in vitro* analysis. For this purpose, cells were seeded on 6 well Primaria plates at a density of 5 x 10^5 cells/well and allowed to grow for 48 hours. Cells were lysed using the RLT Lysis buffer provided by the RNeasy Mini Kit (Qiagen) containing β-mercaptoethanol and passed through the QIAshredder spin column (Qiagen), according to the instructions of the manufacturer. RNA was isolated according to the recommended protocol of the provider and DNase treatment was also performed (Qiagen). RNA concentration and purity were evaluated using the NanoDrop analyzer.

Gene expression levels were also evaluated *in vivo*. Therefore, 30 mg of pancreatic tumor tissue derived from the mouse orthotopic xenografts were placed into a MACS M Tube (Miltenyi) together with the aforementioned lysis buffer and the tissue was dissociated using a Gentle MACS dissociator (Miltenyi). RNA extraction followed the previously mentioned protocol.

EpCAM+ cell isolation and lysis for RNA extraction

Because the total tumor tissue contains epithelial human cancer cells together with mouse stromal cell, we performed an enrichment of EpCAM+ human cells and isolated RNA from the enriched epithelial fraction for qPCR analysis. Tumor tissue was cut into small pieces, resuspended in CO_2 independent medium and enzymes were added from the MACS tumor dissociation kit together with ROCK inhibitor and fungizone. The mix was placed in a MACS C Tube (Miltenyi) and fixed in a MACSMix rotator for 45 minutes at 37°C, with regular vortexing. The enzymatic cocktail was inactivated with fetal calf serum (FCS) and the dissociated cells were filtered through a 40 µm mesh. Cell debris and dead cells were removed using a ficol gradient centrifugation. Erythrocytes were removed by incubating the pellet of cells with ACK lysis buffer for 1 minute. The remaining single cells were incubated with anti-human EpCAM beads (Miltenyi) for 30 minutes at 4°C. The cell suspension was loaded on an LD column (Miltenyi) placed in a MACS separator, and after discarding the flow-through representing the stromal fraction, the EpCAM positive fraction was recovered from the column with the use of the plunger provided by the manufacturer, in a 2mM EDTA – PBS buffer containing 1% BSA. Cells were lysed using the QIAshredder kit and RNA was isolated as previously described.

qPCR reaction

Human specific primers were designed using the Netprimer software. In order to assess the specificity of the selected primers, the melting curves of each individual pair of designed primers were investigated for the presence of more than one product of amplification. Additionally, qPCR products were run on a 2% agarose gel (containing Midori Green for DNA band detection) in TBE buffer. Thus, it was confirmed that only one band can be detected following amplification, and the size of the amplicon was also assessed and

confirmed to match the predicted length. Band size was evaluated with the help of a 100 bp DNA ladder (Bio Labs).

One microgram of total RNA extracted was reverse transcribed using the SuperScript VILO cDNA kit according to the instructions of the manufacturer. Reverse transcription reaction was performed for one hour at 42°C. 10 ng of total cDNA were used for the qPCR reaction, and the final concentration of the forward and reverse specific primers was 10 µM. The qPCR reaction was performed on the Viia7 Thermo, using the 2x concentrated SYBR green master mix (Applied Biosystems). Final reaction volume was 10 µL and the reaction was performed in the following conditions: the reaction mix was heated for 2 minutes at 50°C, followed by the initial denaturation performed at 95°C for 10 minutes. The denaturation and annealing plus extension cycles were carried out at 95°C for 15 seconds and 60°C for 1 minute respectively, and repeated 40 times. Melting curves were generated and triplicate threshold cycles (Ct) values for each sample were averaged and compared to the Ct average of the housekeeping genes for each sample. mRNA levels were determined by relative quantification using the Viia7 RUO software version 1.0, and reported to two housekeeping genes: glyceraldehyde 3-phosphate (GAPDH) (primers designed to be human specific and the amplicon to span between different exons) and 60S ribosomal protein L13a (RPL13a). *In vivo* qPCR reactions were reported to total human pancreas RNA available from Ambion, which was processed prior to quantification identically to the total RNA extracts obtained from cell culture and mouse xenografts.

5. Results

We set out to uncover novel cell surface and secreted protein markers in PDAC, which could be later used in clinical application, such as the development of novel diagnostic tests or novel targeted therapies. For this purpose, we employed for our discovery and validation experiments twelve patient matched primary pancreatic cancer cell lines (PACO cells) developed by Dr.Christian Eisen, DKFZ. Among the developed cell lines, three were present in duplicate, as biological replicates developed from the same patient. The subclassification of the cell lines into the individual subtypes was done using the gene set enrichment analysis (GESA) using the pancreatic ductal adenocarcinoma assigner as recommended by Collisson et al [3] (analysis performed by Dr. Christian Eisen). However, two cell lines originating from the same patient – PACO19 and PACO20 were initially classified as belonging to different subtypes – classical and respectively QM. Using both transcriptomic data (Dr.Elisa Espinet and Elisa Noll, HI-STEM, Metastasis Initiating Cells group) and whole cell proteome data (Wiebke Nadler, HI-STEM, Biomarker Discovery group), we observed that the PACO20 cell line clustered together with PACO19 and the two other classical cell lines – PACO2 and PACO17 (data not shown). For this reason, we decided to list PACO20 as a classical cell line in our validation experiments. The list of PACO cell lines and their subclassification is shown in **Table 5**. Cell lines derived from the same patient are: PACO3 and PACO10 (exocrine-like subtype); PACO7 and PACO8 (QM PDAC). For microscopy based validation experiments we employed two additional exocrine-like cell lines: PACO25 and PACO26, which were not part of the cell lines used for the mass spectrometric discovery phase. Novel exocrine-like cell lines were provided by Dr. Elisa Espinet (HI-STEM).

Table 5: Subclassification of PACO cell lines

Molecular subtype	Cell line	Observations
Exocrine-like	PACO3	Derived from the same patient. With medium length cell cycle
	PACO10	
	PACO14	Growing in three dimensional islands. Very slow cycling
	PACO18	Slow cycling
Classical	PACO2	Fast cycling
	PACO17	Fast cycling
	PACO19	Derived from the same patient; fast cycling
	PACO20	
Quasimesenchymal	PACO7	Derived from the same patient. With medium length cell cycle
	PACO8	
	PACO9	With medium length cell cycle
	PACO16	Very slow cycling

Two control cell lines, derived from healthy pancreatic tissue, were used as healthy controls in both the discovery and *in vitro* validation steps of the project. The two control cell lines were HPNE (Human Pancreatic Nestin Expressing Cells) and HPDE (Human Pancreatic Ductal Epithelial Cells). HPNE cells were immortalized using hTERT polymerase, while HPDE cells were immortalized by transfection with the E6E7 gene [133].

5.1 Identification of novel protein biomarkers using mass spectrometry

The project focused on the *in vitro* discovery of novel pancreatic cancer biomarker candidates located on the cell surface of tumor cells or secreted by the tumor cells into the local environment. Specific enrichment protocols were optimized for each sub-proteome of interest. Additionally, we started investigating vascular accessible biomarkers associated with orthotopic pancreatic tumors developed in immunodeficient NSG mice, by performing whole body perfusion with a reactive biotin ester and enriching the biotinylated fraction prior to mass spectrometric analysis.

5.1.1 Identification of novel cell surface protein biomarkers for PDAC using in vitro cultured primary, patient-matched cell lines

The cell surface proteome of PDAC primary cells was analyzed after chemically enriching the plasma membrane associated proteins, upon using a reactive biotin ester – in this particular case Sulfo-NHS-LC Biotin. Amino groups (including amino terminal aminoacids or lysine residues) of proteins accessible to the reagent were covalently labeled upon reacting with the N-hydroxysuccinimide ester. Cells were incubated with the biotinylation reagent for 10 minutes at room temperature, and the unreacted ester was quenched using a Tris-HCl solution (for details see *Materials and Methods*). Afterwards, cells were detached using non-enzymatic methods, and lysed. Biotinylated proteins, present in 250 µg total protein, were enriched on a streptavidin resin and digested with trypsin, as described in the methods chapter. In parallel, negative controls have been performed for each sample, by using the same total protein amount and performing identical enrichment, digestion and fractionation protocols on cell lysates without prior biotin labeling. A minimum of four biotinylated samples and two negative controls were analyzed per cell line. The complex peptide mixture was fractionated using a reverse phase UPLC column, and the eluting fractions were mixed with CHCA resin and spotted on a MALDI plate, prior to MS analysis on a MALDI TOF/TOF 5800 instrument.

After each MS run, we created 2D peptide maps, which enable an overview of the complexity of the sample to be analyzed, as well as the quality of the chromatographic fractionation. Spotted samples for which the 2D peptide maps revealed tailed elutions, contaminations or significant delays in the start of elution, were excluded from further MS^2 analysis. Thus, the created 2D peptide maps served as a first level quality control for the UPLC fractionated samples. We selected two representative maps: one for the biotinylated fraction (**Figure 9A**) and one for an unbiotinylated negative control (**Figure 9B**). The y axis indicates the fraction number in which the peptide was detected, whereas the x axis indicates the m/*z* value of each peptide. The red arrow points to one of the constant signal of the spiked in peptide having a mass of ~1411 Da, which was used for the normalization of the peptide maps. The two selected maps indicate the general high level quality of the reverse phase UPLC fractionation performed. As it can be observed from **Figure 9**, the majority of the eluted peptides can be identified in only few fractions, and no significant elution tails could be reported for the majority of the fractionated peptides. Exceptions could be observed for very abundant peptides, such as the ones derived from streptavidin digestion (delineated in red). The number of peptides in the negative controls (**Figure 9B**)

is, as expected, significantly lower compared to the number detected in biotinylated samples. Biotin carboxilazes, present in both biotinylated samples and negative controls, are enzymes employing biotin as cofactor for their catalytic activity [135]. These proteins are therefore enriched upon streptavidin pull down, thus explaining the presence of positive signals (additional to the streptavidin tryptic peptides) in the 2D peptide maps of the negative controls.

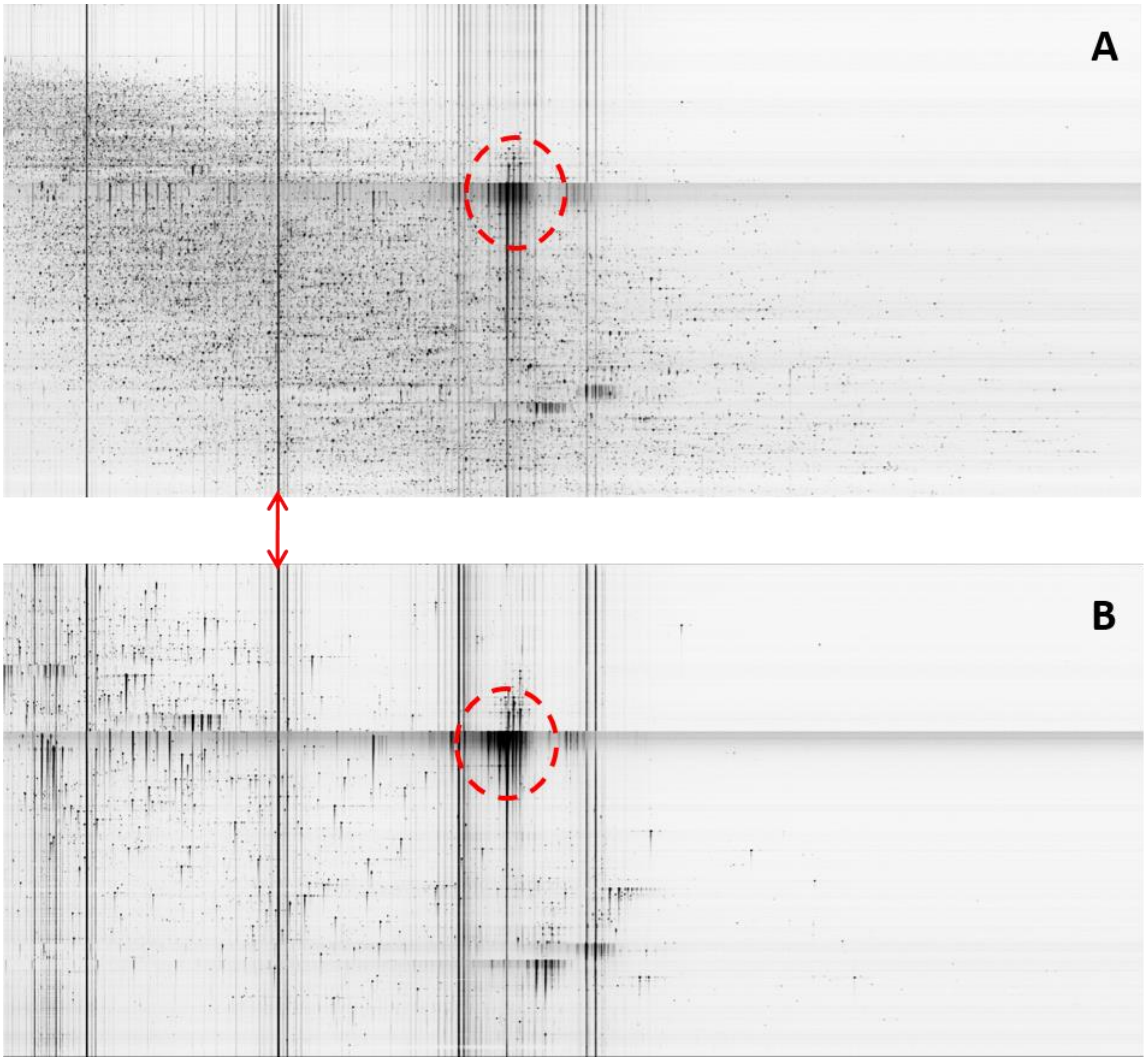


Figure 9: 2D peptide maps of cell surface proteome samples. A: PACO3 biotinylated sample – run 6. B: HPNE non-biotinylated negative control – run 4. Abundant streptavidin derived peptides are outlined in red.

Tandem MS was performed and the resulting spectra were used for protein identification using the ProteinPilot software, by employing the Paragon algorithm and searching against a non-redundant human proteome data base, downloaded from UniProt. The resulting protein lists were manually curated in order to evaluate the presence of putative isoforms. Proteins identified at least once with two peptides were included in the final cell surface proteome list.

In total, we analyzed 108 biotinylated samples, and 46 negative controls. 16 healthy control samples were compared to 30 exocrine-like, 33 classical, and 29 QM samples. The final total list of proteins identified (excluding one hit wonders, as previously mentioned), included

2567 proteins. 1295 of them were present in all cell lines analyzed, while 860 were detected exclusively in one or multiple PDAC derived cell lines and were completely absent from the healthy controls. 200 proteins could be identified in the negative (unbiotinylated) control, consisting mostly of biotin carboxylases and abundant intracellular proteins – such as proteins of the cytoskeleton, ribosomal proteins, etc (**Supplementary Information** – digital format). Subtype specific proteins could be identified. Additionally, some of the identified proteins were present in two or more PDAC molecular subtypes. The exact number of proteins belonging to each subtype is represented in the Venn diagram displayed in **Figure 10**.

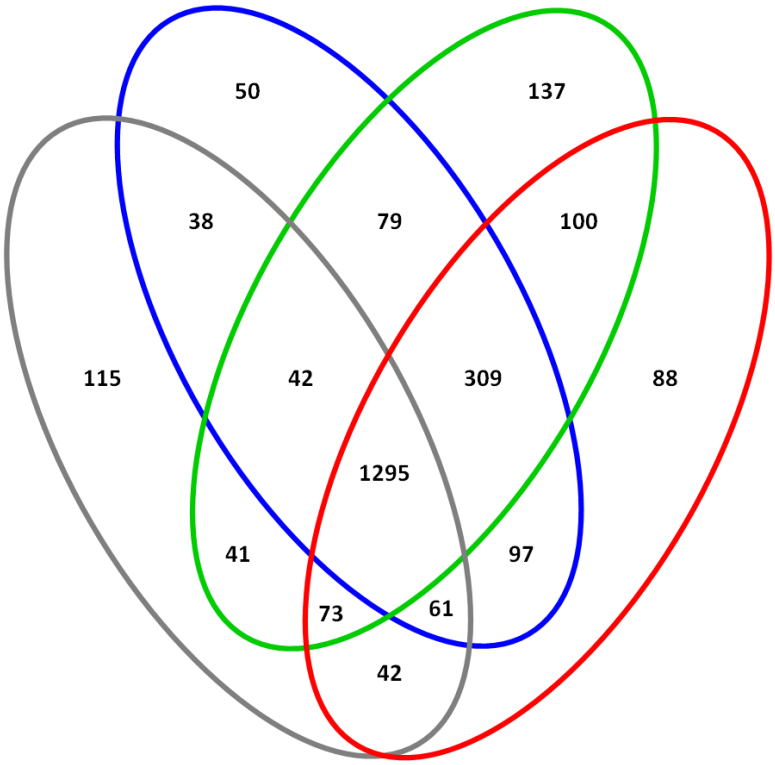


Figure 10: Venn diagram representing the number of proteins identified after performing the cell surface proteome analysis. The following color code was used to distinguish between the different subtypes analyzed: **grey**: healthy control; **blue**: classical subtype PDAC; **green**: exocrine-like PDAC; **red**: QM PDAC.

In order to evaluate the localization of the identified proteins, we performed a gene ontology analysis, using the WebGestalt software. As shown in **Figure 11**, a significant number of proteins identified by MALDI MS after the enrichment of the cell surface proteome were localized at the plasma membrane (more than 800 proteins) or are expected to participate in the formation of anchoring junctions (96 proteins). More than 1300 of the proteins were expected to be associated with membranous fractions (both intracellular and extracellular). With regard to proteins predicted to reside in intracellular membranes, 389 were annotated to the ER membrane fraction.

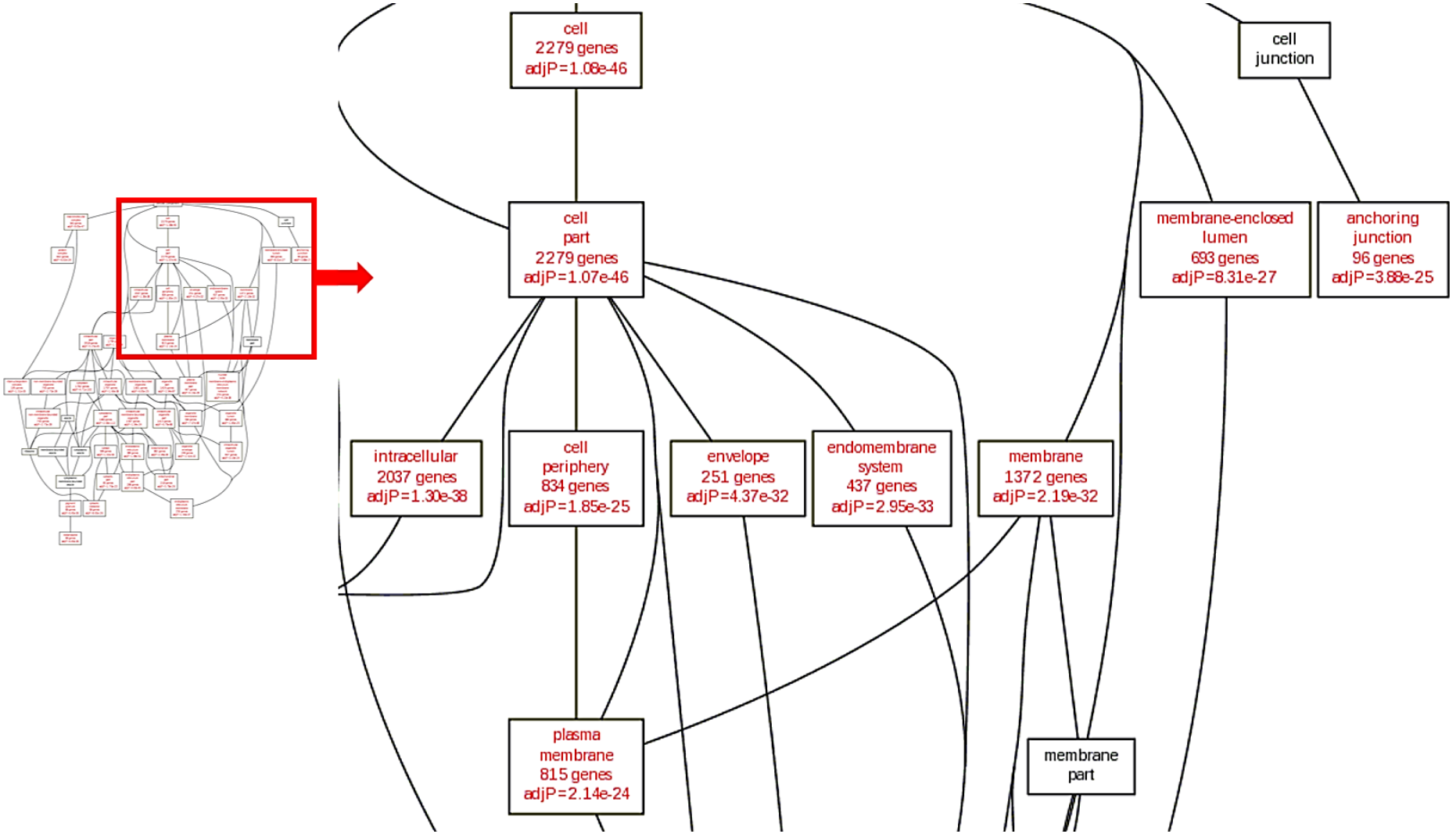


Figure 11: Gene ontology (GO) analysis of the complete cell surface proteome: close up.

Furthermore, the GO analysis allowed investigating the involvement of the identified proteins with disease associated pathways. The list was enriched for proteins involved in adhesion, neoplasm invasiveness and metastasis, as well as the development of epithelial cancers (data not shown). We additionally investigated which disease associated pathways are enriched in each individual subtype, and we could observe the same pattern as previously described for the complete cell surface proteome list (data not shown).

When analyzing the complete cell surface proteome list, we could identify multiple proteins present in all the analyzed PACO cell lines, but in none of the controls, thus representing putative pan-PDAC markers. A short list of the proteins identified with the highest number of peptides in all the PDAC derived cell lines is shown in **Table 6**. The maximum number of peptides identified for each molecular subtype and for the healthy controls is indicated in the last four columns.

As it can be observed from the list, protocadherin-1 (PCDH1) represented one of the top proteins identified in the PDAC cell surface proteome analysis, being detected in all PACO cell lines analyzed, and in almost every biological replicate performed (the absence of PCDH1 identification in some of the biological replicates can be explained by stochastic sampling). Moreover, PCDH1 could not be identified in any of the healthy control cell lines. Other important hits, previously described in literature as being associated with PDAC are: mesothelin [6], mucin-1 (MUC1) [20], carcinoembryonic antigen-related cell adhesion molecule-5 and -6 (CEACAM5 and CEACAM6) [136], and putative cancer stem cell marker prominin-1 (also known as CD133) [137]. Novel putative pan-PDAC markers include PCDH1, Single IgIL-1-related receptor (involved in Toll receptor signaling [138]), podocalyxin (particularly enriched in QM samples), lipocalin-2 (also known as Neutrophil gelatinase-associated lipocalin) and prominin-2. Out of the list, we selected two putative pan-PDAC markers for further validations: PCDH1 and LCN2.

We were also interested in identifying subtype specific markers, and for this reason we inquired the cell surface proteome list for proteins present exclusively, or significantly enriched (based on the maximum number of peptides identified), in one molecular subtype. In order for a marker to be considered as being putative subtype specific, the respective protein had to be identified at least once in all of the cell lines belonging to that particular subtype.

To our surprise, the exocrine-like cell lines stood out, expressing several subtype specific proteins (**Table 7**). As indicated in **Table 7**, the top hits for the exocrine-like subtype were cadherin-17 (CDH17) and galectin-4 (LGALS4), identified in every cell line belonging to the exocrine-like subtype, and we selected both markers for further validations. Not only did we identify the two proteins with the highest number of peptides, but they were also present in more than 95% of all analyzed samples (including biological replicates) from cell lines of the exocrine-like subtype. The exocrine-like enriched group included additional putative biomarkers of interest which we did not pursue further, such as: deleted in malignant brain tumors 1 protein (previously described as being enriched in PDAC [41]), firbroblast growth factor receptor 4, vilin-1, etc.

For the classical and the QM subtypes we could not identify any protein expressed solely or significantly enriched in the individual subtypes. A short list of proteins most abundant in classical and QM PACO cells is available in **Table 8**. Some of the proteins identified in classical PDAC with the highest number of peptides could only be detected in on cell line alone: PACO2. Two of the proteins identified with the highest number of peptides in the QM

samples – Zinc transporter ZIP4 and Ephrin type-A receptor 3, were unfortunately only present in two of the QM assigned cell lines. For this reason, we did not select any putative marker for further validations.

Several proteins appeared enriched in one or two distinct molecular subtypes: for example mucin-16, present in classical and QM cells, but completely absent in exocrine-like PACOs, or metalloreductase STEAP1, enriched in the classical subtype, but also present in exocrine-like cells. STEAP1 could be identified in all cell lines belonging to the classical and exocrine-like subtype at least once, and appeared to be slightly enriched in classical cells. We considered STEAP1 for future validation experiments, which will not be presented in this report.

The complete list of proteins identified in the cell surface proteome and corresponding secretome is available in the **Supplementary Materials**, in digital format.

Table 6: Putative pan-PDAC markers identified exclusively in PACO cell lines. The maximum number of peptides identified for each subtype is represented in the right columns (in brackets – number of cell lines in which the protein was identified/out of the number of cell lines belonging to the respective molecular subtype). Proteins are listed dependent on the percentage of PACO samples (including biological replicates) positive for the respective protein

Protein Name	Gene Name	% positive PACO	Max Ctrl	Max Exocrine	Max Classical	Max QM
Protocadherin-1	PCDH1	97.83	0	22 (4/4)	15 (4/4)	22 (4/4)
Mucin-1	MUC1	91.3	0	19 (4/4)	11 (4/4)	15 (4/4)
Probable G-protein coupled receptor 110	GPR110	89.13	0	5 (4/4)	6 (4/4)	11 (3/4)
Single Ig IL-1-related receptor	SIGIRR	83.7	0	8 (4/4)	10 (4/4)	7 (4/4)
Carcinoembryonic antigen-related cell adhesion molecule 6	CEACAM6	82.61	0	7 (4/4)	7 (4/4)	6 (3/4)
Neutrophil gelatinase-associated lipocalin	LCN2	81.52	0	10 (4/4)	9 (4/4)	11 (3/4)
Carcinoembryonic antigen-related cell adhesion molecule 5	CEACAM5	79.35	0	12 (4/4)	16 (4/4)	10 (3/4)
Periplakin	PPL	78.26	0	21 (4/4)	15 (4/4)	22 (4/4)
Sodium- and chloride-dependent neutral and basic amino acid transporter B(0+)	SLC6A14	77.17	0	6 (4/4)	7 (4/4)	4 (3/4)
Canalicular multispecific organic anion transporter 2	ABCC3	72.83	0	17 (4/4)	8 (3/4)	16 (4/4)
Cadherin EGF LAG seven-pass G-type receptor 1	CELSR1	69.57	0	20 (4/4)	13 (4/4)	7 (4/4)
Mesothelin	MSLN	65.22	0	6 (2/4)	8 (3/4)	7 (4/4)
Poliovirus receptor-related protein 4	PVRL4	65.22	0	6 (4/4)	5 (4/4)	9 (3/4)
Amyloid-like protein 2	APLP2	61.96	0	5 (4/4)	3 (4/4)	5 (4/4)
Prominin-1	PROM1	58.7	0	19 (4/4)	4 (3/4)	4 (2/4)
Epiplakin	EPPK1	58.7	0	44 (4/4)	11 (3/4)	12 (2/4)
Low-density lipoprotein receptor-related protein 8	LRP8	56.52	0	7 (3/4)	5 (4/4)	7 (4/4)
Tumor necrosis factor receptor superfamily member 21	TNFRSF21	56.52	0	5 (2/4)	5 (4/4)	7 (3/4)
Intercellular adhesion molecule 2	ICAM2	45.65	0	3 (3/4)	5 (4/4)	3 (3/4)
Prominin-2	PROM2	44.57	0	18 (4/4)	10 (3/4)	6 (2/4)
Envoplakin	EVPL	42.39	0	4 (2/4)	12 (4/4)	10 (2/4)
Podocalyxin	PODXL	39.13	0	3 (1/4)	4 (2/4)	10 (3/4)

Table 7: List of proteins enriched in the exocrine-like subtype. The percentage of samples positive for a given protein is included in parallel with the maximum number of peptides identified for each molecular subtype. In brackets: number of cell lines in which the protein was identified/out of the number of cell lines belonging to the respective molecular subtype

		% positive samples			Maximum number of peptides identified			
Protein Name	**Gene Name**	**% Exocrine**	**% Classical**	**% QM**	**Max Ctrl**	**Max Exocrine**	**Max Classical**	**Max QM**
Cadherin-17	CDH17	100.00	0.00	0.00	0	46 (4/4)	0	0
Galectin-4	LGALS4	96.67	3.03	0.00	0	12 (4/4)	1	0
Claudin-2	CLDN2	66.67	0.00	0.00	0	2 (4/4)	0	0
Endothelial cell-selective adhesion molecule	ESAM	56.67	51.52	0.00	0	4 (3/4)	3	0
Semaphorin-4G	SEMA4G	56.67	27.27	0.00	0	7 (3/4)	3	0
Ly6/PLAUR domain-containing protein 2	LYPD2	46.67	0.00	0.00	0	3 (2/4)	0	0
Receptor-type tyrosine-protein phosphatase H	PTPRH	46.67	0.00	0.00	0	7 (3/4)	0	0
Lysozyme C	LYZ	40.00	12.12	0.00	0	2 (4/4)	1	0
Protein KIAA1199	KIAA1199	40.00	9.09	0.00	0	5 (4/4)	1	0
Alpha-2A adrenergic receptor	ADRA2A	36.67	0.00	0.00	0	4 (2/4)	0	0
Cadherin-12	CDH12	36.67	0.00	0.00	0	6 (1/4)	0	0
Fibroblast growth factor receptor 4	FGFR4	33.33	0.00	0.00	0	4 (2/4)	0	0
Glutathione peroxidase 2	GPX2	33.33	0.00	0.00	0	6 (4/4)	0	0
Protein phosphatase 1 regulatory subunit 1B	PPP1R1B	33.33	0.00	0.00	0	2 (3/4)	0	0
Deleted in malignant brain tumors 1 protein	DMBT1	30.00	0.00	0.00	0	3 (2/4)	0	0
Probable G-protein coupled receptor 125	GPR125	26.67	0.00	0.00	0	3 (2/4)	0	0
Villin-1	VIL1	23.33	0.00	0.00	0	5 (2/4)	0	0

Table 8: List of proteins enriched in the classical and QM subtypes. The percentage of samples positive for a given protein is included in parallel with the maximum number of peptides identified for each molecular subtype In brackets: number of cell lines in which the protein was identified/out of the number of cell lines belonging to the respective molecular subtype

		% positive samples			Maximum number of peptides identified			
Protein Name	**Gene Name**	**% Exocrine**	**% Classical**	**% QM**	**Max Ctrl**	**Max Exocrine**	**Max Classical**	**Max QM**
Protocadherin Fat 2	FAT2	0.00	24.24	0.00	0	0	12 (1/4)	0
Contactin-1	CNTN1	0.00	24.24	0.00	0	0	8 (2/4)	0
Interferon-induced GTP-binding protein Mx2	MX2	0.00	24.24	0.00	0	0	4 (3/4)	0
MICAL-like protein 1	MICALL1	0.00	24.24	0.00	0	0	3 (1/4)	0
Pirin	PIR	0.00	18.18	0.00	0	0	3 (1/4)	0
Matrix metalloproteinase-28	MMP28	0.00	15.15	0.00	0	0	3 (1/4)	0
Metalloreductase STEAP1	STEAP1	33.33	69.70	3.45	2 (1/2)	4 (3/4)	5 (4/4)	1 (1/4)
Mucin-16	Q8WXI7	0.00	75.76	58.62	0	0	25 (3/4)	14 (3/4)
Chloride intracellular channel protein 3	CLIC3	0.00	0.00	48.28	0	0	0	4 (3/4)
Zinc transporter ZIP4	SLC39A4	0.00	0.00	44.83	0	0	0	3 (2/4)
Syndecan-2	SDC2	0.00	0.00	24.14	0	0	0	4 (1/4)
Ephrin type-A receptor 3	EPHA3	0.00	0.00	20.69	0	0	0	10 (1/4)
Latrophilin-3	LPHN3	0.00	0.00	20.69	0	0	0	4 (1/4)
Keratin-81-like protein	--	0.00	0.00	20.69	0	0	0	3 (2/4)
Protein Wnt-11	WNT11	0.00	0.00	17.24	0	0	0	2 (1/4)
Toll-like receptor 5	TLR5	0.00	0.00	10.34	0	0	0	2 (1/4)

5.1.2 Identification of novel secreted protein biomarkers for PDAC in vitro

Putative novel diagnostic cancer biomarkers are expected to be proteins either secreted by the tumor cells into their microenvironment, or enzymatically shaved off plasma membrane components. In order to investigate putative novel PDAC diagnostic biomarkers, we employed the conditioned cell culture medium of the twelve PACO cell lines. The CSC medium for the PDAC derived cell lines has the advantage that it does not contain any added serum, which normally hinders mass spectrometric protein identification. However, the CSC PACO medium did contain high amounts of albumin and transferrin, which could have masked the signal of less abundant peptides originating from putative protein biomarkers. For this reason, together with Dr. Martin Sprick (HI-STEM/DKFZ), we optimized a novel CSC based medium, completely devoid of albumin and transferrin. Iron was supplemented to the new medium, in order to satisfy the metabolic needs of the growing cells. Since the new medium was not depleted in nutrients or growth factors (supplemented separately), we expected that the new formulation would not significantly impact the metabolism, phenotype and secretion pattern of the investigated cells.

All cell lines were grown for 48 hours in the new chemically defined CSC albumin/transferrin depleted medium (termed secretome CSC medium), including the two healthy control references: HPDE and HPNE. The medium was later collected and underwent a series of centrifugation steps, meant to remove organelles released in the growth medium by apoptotic cells and microvesicles (for the full protocol see *Materials and Methods*). A minimum of three technical replicates were performed for each sample. We did not observe any changes in the cell morphology and division rate of the PACO and HPNE cells, which was to be expected, since the secretome CSC medium formulation is highly similar to their original cell culture medium. However, for the HPDE cell line (grown in a Keratinocyte cell medium), slight morphological changes could be observed, and the cells appeared more elongated than the counterparts grown in their native medium after the 48 hours incubation period (**Figure 12**).

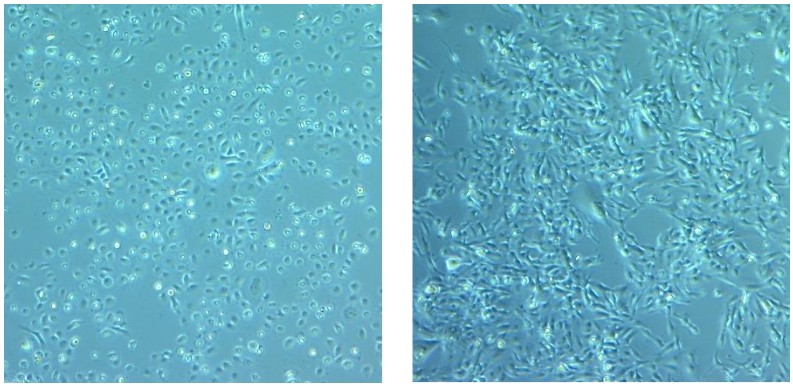


Figure 12: Microscopy pictures of HPDE cell line grown in regular Keratinocyte culture medium and in secretome optimized CSC medium are shown on the left and on the right side respectively.

Secreted proteins were precipitated and digested with trypsin, and the peptide mixture was fractionated by reverse phase UPLC. Two representative 2D peptide maps (for HPNE and PACO19 collected supernatants) are shown in **Figure 13**. We observed once again that the majority of peptides were eluted in only a few fractions.

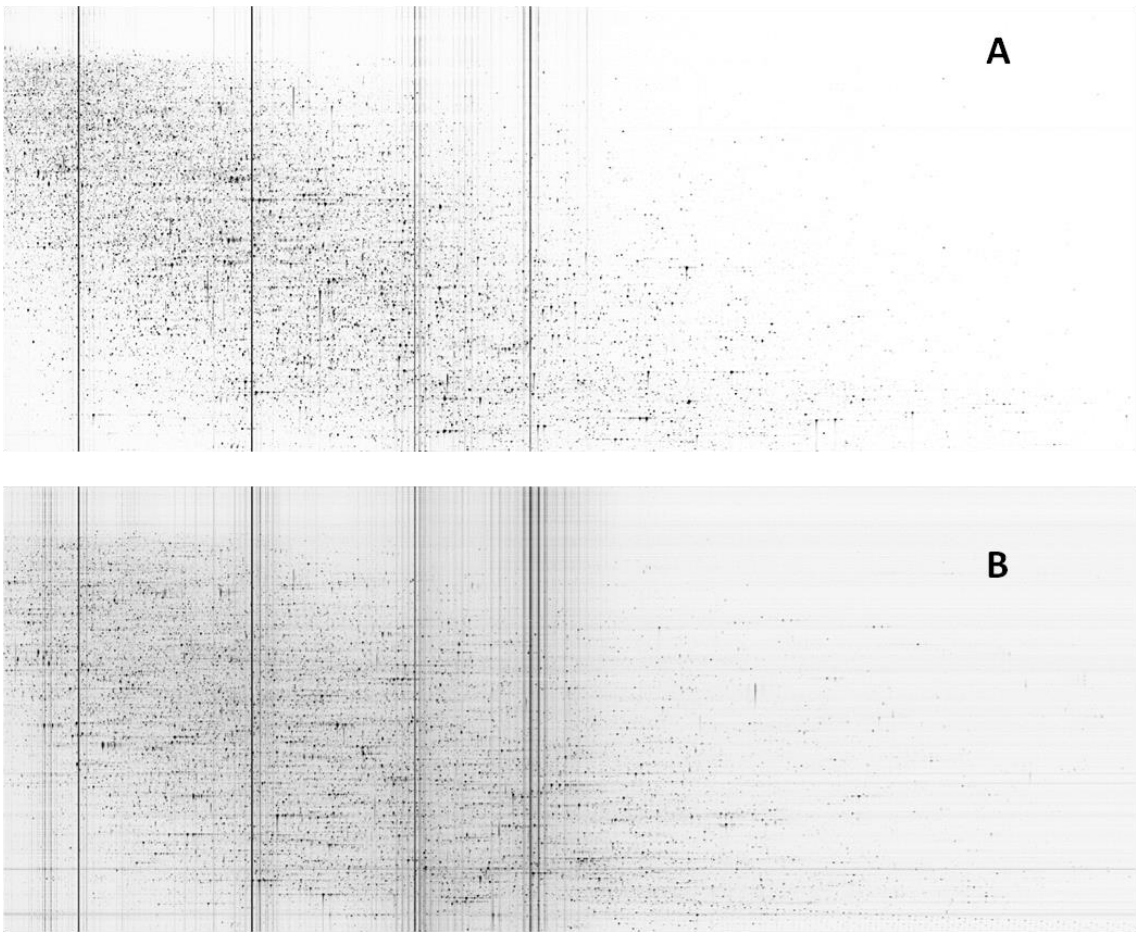


Figure 13: 2D peptide maps of secretome samples. A: HPNE – run 2. B: PACO19 – run 4.

We analyzed a total of 47 supernatant samples, including 6 healthy controls, 13 exocrine-like, 12 classical, and 16 QM PDAC samples. After manually curating the list of proteins identified (as previously mentioned for the cell surface proteome), we could confirm the presence of 1781 proteins in the secreted fractions, identified at least once with a minimum of two peptides. 786 of the identified proteins were common for the control and PDAC cell lines, while 175 detected proteins were identified in all PACO subtypes, but were absent in the healthy control cell line secretome. The number of secreted proteins identified and their distribution among the different subtypes is summarized in the following Venn diagram (**Figure 14**).

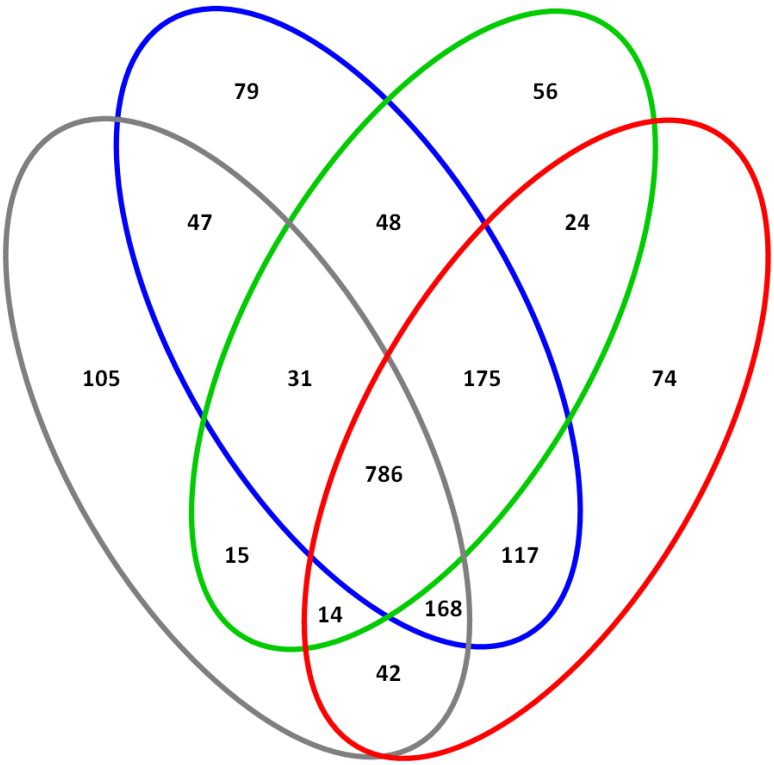


Figure 14: Venn diagram representing the number of proteins identified in the secretome analysis. The following color code was used to distinguish between the different subtypes analyzed: **grey**: healthy control; **blue**: classical subtype PDAC; **green**: exocrine-like PDAC; **red**: QM PDAC.

We investigated the predicted cellular localization of the secreted proteins identified, by performing gene ontology analysis. **Figure 15a** shows a close up of the GO enrichment analysis results, using the complete secretome list for investigation. A clear enrichment of proteins annotated to the GO term “extracellular region” and “extracellular matrix” can be observed (more than 500 proteins in total). Over 100 of the discovered proteins were predicted to be located on the cell surface. Some of the intracellular proteins identified were described to be enriched in membrane bound and cytoplasmic vesicles. When analyzing the involvement of identified secreted proteins in disease associated pathways, we observed similar enrichments to the ones previously described for the cell surface proteome, with adhesion, neoplasm invasiveness and metastasis representing some of the most relevant disease associated pathways (data not shown).

Interestingly, the secretome of individual subtypes revealed some relevant differences, compared to the whole secretome. For example, the secretome of the exocrine-like cells expressed a subset of proteins enriched in cell junctions, additionally to the already predicted extracellular and cell surface proteins (**Figure 15b**, close-up). In contrast, the classical and the QM subtype’s secretome were enriched in proteins expressed at the basement membrane level. For simplification purposes, we only represented close ups of the classical secretome GO analysis (**Figure 15c**).

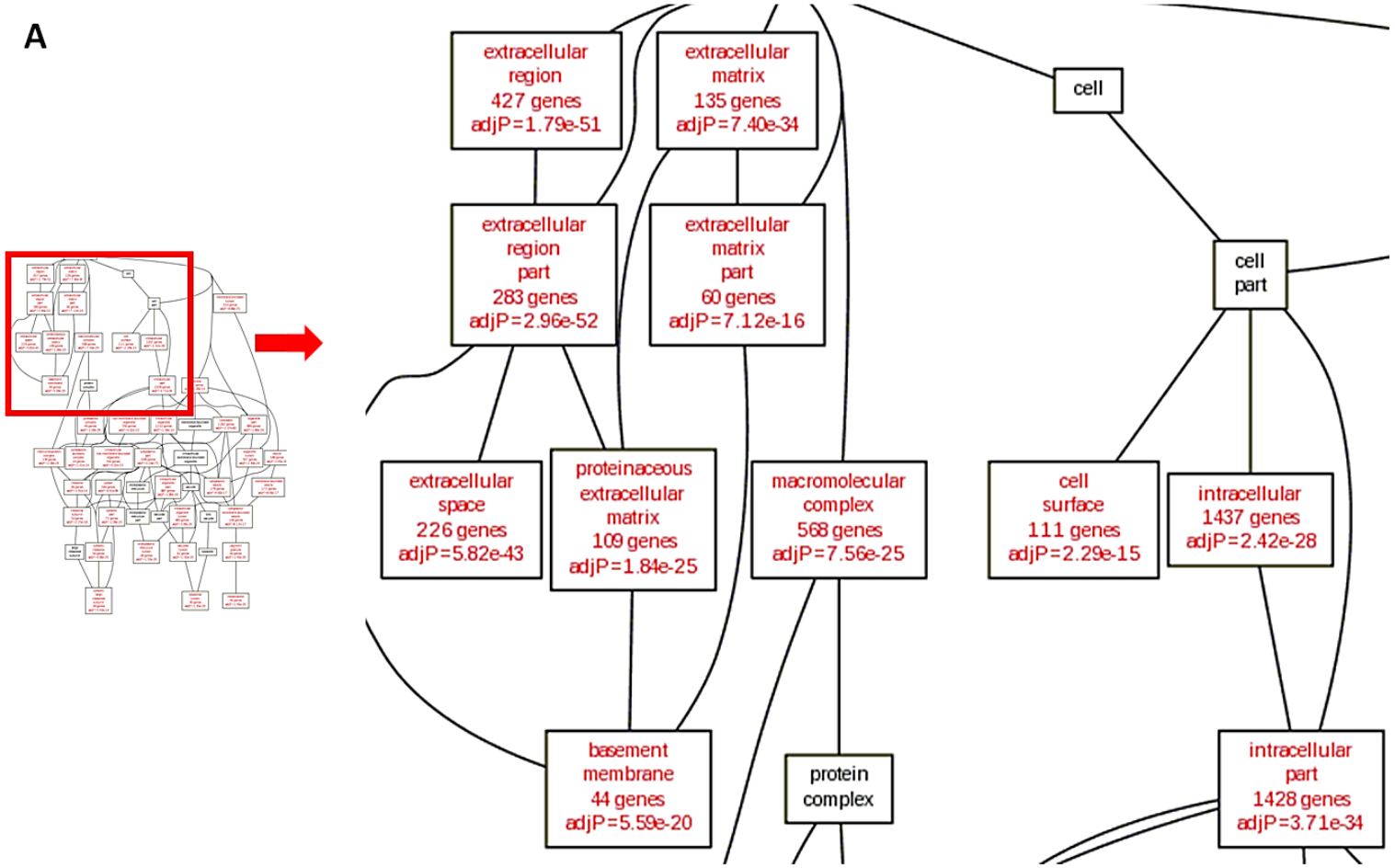


Figure 15 (a): Gene ontology analysis of the complete *in vitro* secretome: close up.

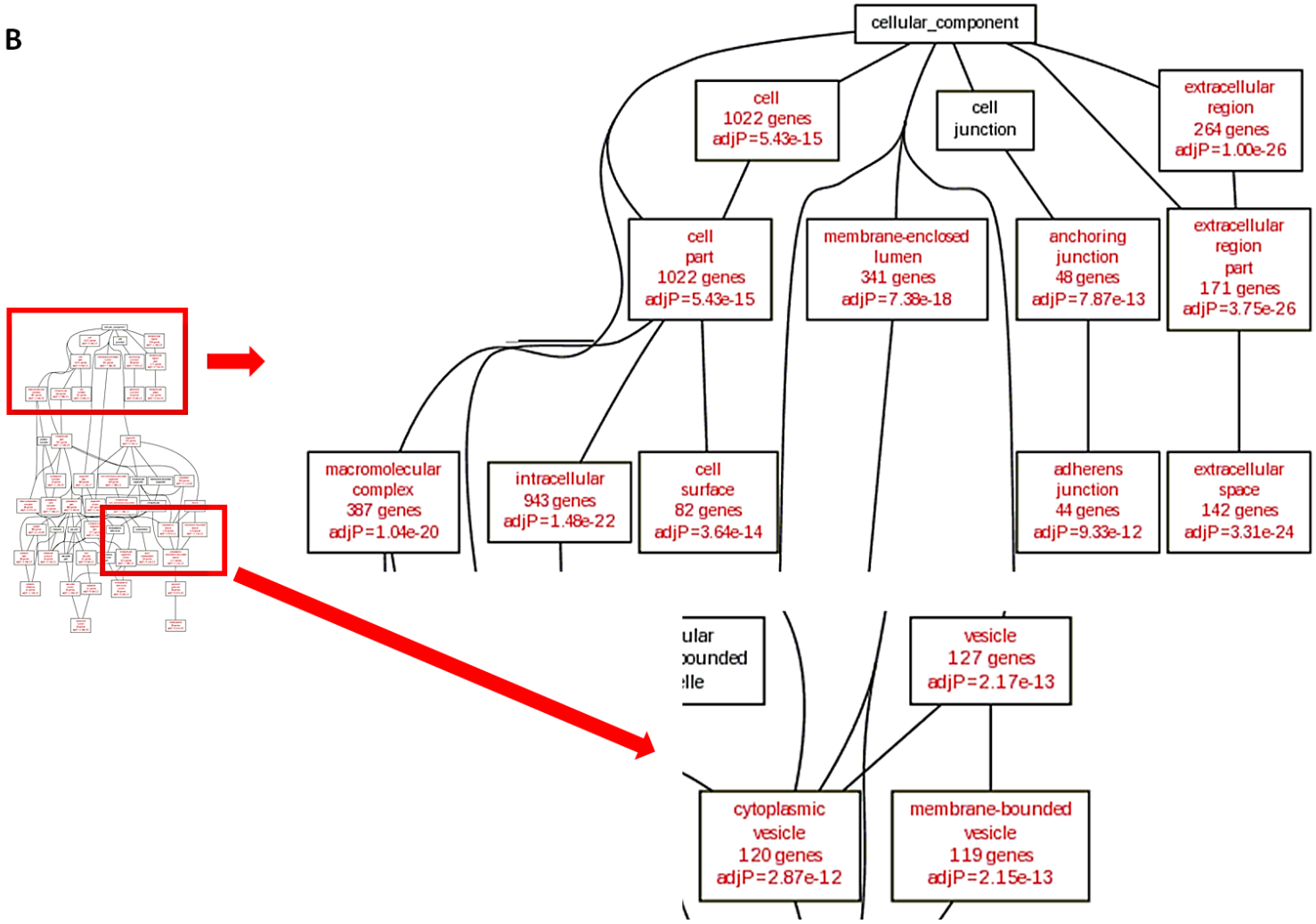


Figure 15 (b): Gene ontology analysis of the exocrine-like *in vitro* secretome: close up.

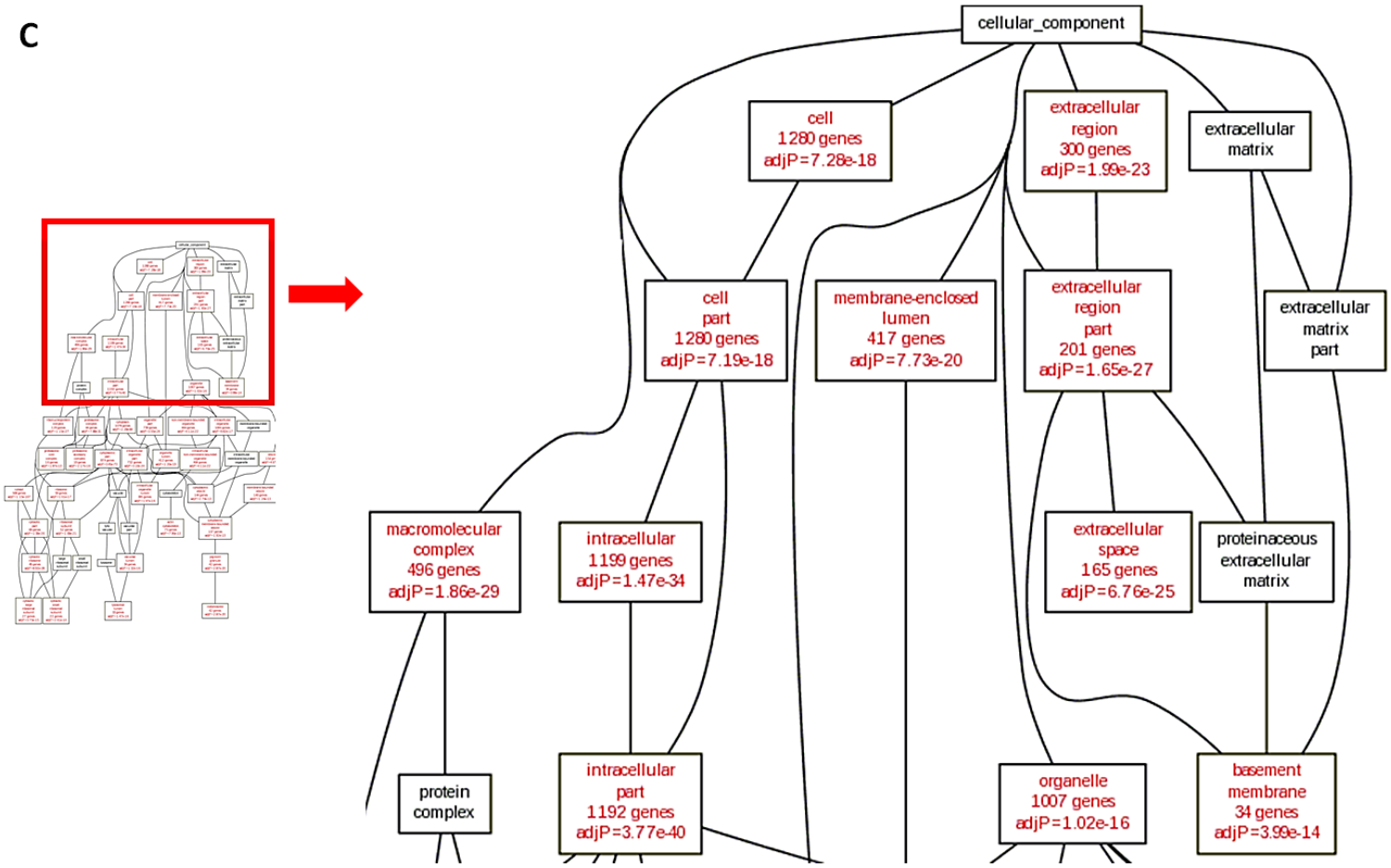


Figure 15 (c): Gene ontology analysis of the classical *in vitro* secretome: close up.

We screened the complete secretome list for proteins associated exclusively with the PACO cell lines, but absent in the healthy control secretome. A short list of the most important hits is listed in **Table 9**. Several of the top hits identified in the cell surface proteome could also be detected in the secreted fraction. LCN2, one of the putative pan-PDAC markers detected during the cell surface analysis, was identified in the secreted fraction of all PACO cell lines as well, with a high number of peptides. Mesothelin, PCDH1, MUC1, CEACAM6 were also identified in the secretome of PDAC cancer cell lines. Among the secreted proteins present throughout the whole molecular subtype repertoire, we could discern remodeling agents of the ECM, such as matrilysin (MMP7) (appearing to be enriched in the exocrine-like subtype based on the maximum number of peptides identified and the percentage of positive cells) and ADAM8. Interestingly, trypsin-2 was secreted by exocrine-like and QM PACO cell lines.

Several of the previously mentioned exocrine-like cell surface proteins could be detected in the secreted fraction as well, including the two putative biomarkers we had selected for further validations (**Table 10**). Among the top hits of exocrine-like specifically secreted proteins we noticed several proteins involved in mediating cell adhesion (such as cadherins-6, -12 and -17). Interestingly, the pair ligand-receptor fibroblast growth factor 19 (FGF19) and fibroblast growth factor receptor 4 (FGFR4) were both secreted by exocrine-like cells.

A short list of proteins enriched in the supernatant of classical and QM cell lines is presented in **Table 11**. Among the proteins highly expressed in the classical subtype are: contactin-1 (previously observed on the cell surface proteome of PACO2 cells), C-X-C motif chemokine 14 and 16, GMCSF receptor subunit alpha. The top hit for QM specific secreted proteins was matrix-remodeling-associated protein 5, detected in half of the samples analyzed.

Table 9: Putative pan-PDAC secreted markers identified exclusively in PACO cell lines. The maximum number of peptides identified for each subtype is indicated in the right columns (in brackets: number of cell lines in which the protein was identified/out of the number of cell lines belonging to the respective molecular subtype)

Protein Name	Gene Name	% positive PACO	max ctrl	max exocrine	max classical	max QM
Neutrophil gelatinase-associated lipocalin	LCN2	100.00	0	87 (4/4)	25 (4/4)	51 (4/4)
Retinoic acid receptor responder protein 1	RARRES1	87.80	0	18 (4/4)	9 (4/4)	11 (4/4)
Retinal dehydrogenase 1	ALDH1A1	80.49	0	16 (3/4)	9 (4/4)	8 (4/4)
Mesothelin	MSLN	80.49	0	19 (3/4)	10 (4/4)	7 (4/4)
Insulin-like growth factor-binding protein 3	IGFBP3	75.61	0	6 (3/4)	11 (3/4)	12 (4/4)
Na(+)/H(+) exchange regulatory cofactor NHE-RF1	SLC9A3R1	73.17	0	3 (4/4)	13 (3/4)	8 (4/4)
Semaphorin-3B	SEMA3B	70.73	0	14 (3/4)	5 (2/4)	19 (4/4)
Protein-glutamine gamma-glutamyltransferase 2	TGM2	70.73	0	3 (2/4)	22 (3/4)	39 (4/4)
Calcyphosin	CAPS	65.85	0	9 (3/4)	5 (2/4)	7 (4/4)
Pterin-4-alpha-carbinolamine dehydratase	PCBD1	65.85	0	6 (4/4)	2 (4/4)	4 (4/4)
Matrilysin	MMP7	63.41	0	51 (4/4)	5 (4/4)	5 (1/4)
Protocadherin-1	PCDH1	60.98	0	6 (2/4)	11 (4/4)	6 (3/4)
Disintegrin and metalloproteinase domain-containing protein 8	ADAM8	58.54	0	2 (2/4)	3 (4/4)	2 (4/4)
CD2-associated protein	CD2AP	58.54	0	9 (2/4)	9 (3/4)	3 (3/4)
Mucin-5AC (Fragments)	MUC5AC	58.54	0	15 (4/4)	18 (3/4)	2 (2/4)
Periplakin	PPL	58.54	0	11 (2/4)	13 (3/4)	17 (4/4)
Protocadherin Fat 1	FAT1	56.10	0	10 (4/4)	14 (2/4)	7 (3/4)
Carcinoembryonic antigen-related cell adhesion molecule 6	CEACAM6	46.34	0	11 (3/4)	6 (4/4)	2 (1/4)
Mucin-1	MUC1	34.15	0	8 (2/4)	5 (3/4)	2 (3/4)
Trypsin-2	PRSS2	26.83	0	10 (2/4)	7 (3/4)	0 (0/4)

Table 10: Putative exocrine-like secreted markers. The maximum number of peptides identified for each subtype is indicated in the right columns. Proteins are listed according to the percentage of samples in which the protein was identified (largest to smallest). In brackets: number of cell lines in which the protein was identified/out of the number of cell lines belonging to the exocrine-like molecular subtype

		% positive samples			Maximum number of peptides identified			
Protein Name	**Gene Name**	**% Exocrine**	**% Classical**	**% QM**	**max ctrl**	**max exocrine**	**max classical**	**max QM**
Annexin A4	ANXA4	61.54	0.00	0.00	0	4 (4/4)	0	0
Cathepsin E	CTSE	61.54	0.00	0.00	0	4 (3/4)	0	0
Regenerating islet-derived protein 4	REG4	61.54	0.00	0.00	0	9 (4/4)	0	0
Galectin-4	LGALS4	46.15	0.00	0.00	0	8 (3/4)	0	0
Fibroblast growth factor 19	FGF19	38.46	0.00	0.00	0	13 (2/4)	0	0
Fibroblast growth factor receptor 4	FGFR4	38.46	0.00	0.00	0	4 (3/4)	0	0
Kallistatin	SERPINA4	38.46	0.00	0.00	0	4 (3/4)	0	0
Nephronectin	NPNT	30.77	0.00	0.00	0	4 (2/4)	0	0
Complement C5	C5	23.08	0.00	0.00	0	16 (1/4)	0	0
Cell adhesion molecule 4	CADM4	23.08	0.00	0.00	0	6 (2/4)	0	0
Cadherin-6	CDH6	23.08	0.00	0.00	0	10 (1/4)	0	0
Interleukin-1 receptor accessory protein	IL1RAP	23.08	0.00	0.00	0	3 (1/4)	0	0
Adseverin	SCIN	23.08	0.00	0.00	0	3 (2/4)	0	0
Angiotensinogen	AGT	15.38	0.00	0.00	0	3 (1/4)	0	0
Alkaline phosphatase, placental-like	ALPPL2	15.38	0.00	0.00	0	10 (1/4)	0	0
Cadherin-12	CDH12	15.38	0.00	0.00	0	2 (1/4)	0	0
Cadherin-17	CDH17	15.38	0.00	0.00	0	4 (1/4)	0	0
Macrophage metalloelastase	MMP12	15.38	0.00	0.00	0	4 (1/4)	0	0
Neurotensin/neuromedin N	NTS	15.38	0.00	0.00	0	4 (1/4)	0	0

Table 11: Selected proteins secreted by the classical and QM cell lines. The maximum number of peptides identified for each subtype is indicated in the right columns. Proteins are listed according to the percentage of samples in which the protein was identified (largest to smallest) In brackets: number of cell lines in which the protein was identified/out of the number of cell lines belonging to the exocrine-like molecular subtype

		% positive samples			Maximum number of peptides identified			
Protein Name	**Gene Name**	**% Exocrine**	**% Classical**	**% QM**	**max ctrl**	**max exocrine**	**max classical**	**max QM**
Polypeptide N-acetylgalactosaminyltransferase 6	GALNT6	0.00	58.33	0.00	0	0	9 (4/4)	0
Catalase	CAT	0.00	41.67	0.00	0	0	5 (3/4)	0
Contactin-1	CNTN1	0.00	33.33	0.00	0	0	5 (1/4)	0
Hematopoietic lineage cell-specific protein	HCLS1	0.00	33.33	0.00	0	0	2 (1/4)	0
Caprin-1	CAPRIN1	0.00	25.00	0.00	0	0	5 (2/4)	0
Carboxypeptidase D	CPD	0.00	25.00	0.00	0	0	2 (2/4)	0
Granulocyte-macrophage colony-stimulating factor receptor subunit alpha	CSF2RA	0.00	25.00	0.00	0	0	2 (1/4)	0
C-X-C motif chemokine 14	CXCL14	0.00	25.00	0.00	0	0	2 (1/4)	0
C-X-C motif chemokine 16	CXCL16	0.00	25.00	0.00	0	0	3 (1/4)	0
Matrix-remodeling-associated protein 5	MXRA5	0.00	0.00	50.00	0	0	0	5 (2/4)
Chloride intracellular channel protein 3	CLIC3	0.00	0.00	43.75	0	0	0	8 (2/4)
Cellular retinoic acid-binding protein 2	CRABP2	0.00	0.00	43.75	0	0	0	6 (2/4)
Protein kinase C-binding protein NELL2	NELL2	0.00	0.00	43.75	0	0	0	2 (2/4)
Insulin-like growth factor-binding protein 5	IGFBP5	0.00	0.00	31.25	0	0	0	2 (2/4)
Selenium-binding protein 1	SELENBP1	0.00	0.00	31.25	0	0	0	3 (2/4)
Testican-2	SPOCK2	0.00	0.00	31.25	0	0	0	4 (2/4)

5.1.3 Overview of total protein list detected on the cell surface proteome and secretome of PDAC cell lines versus healthy controls

After completing the *in vitro* analysis of both proteomic subsets, we decided to combine the two lists of identified proteins and analyze them together. The lists were again curated manually and only proteins identified at least once with two peptides in either the cell surface proteome or the secretome were considered for the final list. In total, we identified 3288 proteins, out of which 1450 were common for both fractions (**Figure 16**). Only 546 proteins were selectively identified in the supernatant of cultured cells, while 1292 could be detected only in the enriched plasma membrane fraction.

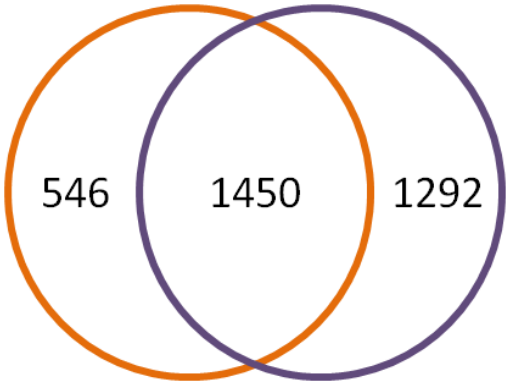


Figure 16: Venn diagram illustrating the total number of proteins identified at least once with two peptides in the cell surface and secreted proteome of cultured PDAC primary cells and healthy controls.

The unified list of both proteomes is available in digital format in the **Supplementary Materials**.

5.1.4 In vivo biotinylation of tumor xenograft mouse models for the identification of novel subtype specific biomarker candidates

In parallel we aimed to investigate protein markers associated with orthotopic pancreatic tumors, including proteins expressed on the surface of cancer cells, on the plasma membrane of stromal cells, as well as proteins present in the tumor associated ECM. For this purpose, 10^6 PACO cells were injected in the pancreas of immunodeficient NSG male mice. When the tumor reached ~1 cm in diameter (evaluated by weekly palpations), mice were anesthetized, the thoracic cavity was opened and the mice received a full body terminal perfusion with a reactive biotin ester prepared in dextrane-PBS solution. The perfusion protocol, performed in optimal temperature conditions to prevent vasoconstriction, leads to the covalent biotin labeling of proteins accessible from the blood stream, including vasculature associated proteins, as well as cell surface and ECM proteins expressed and secreted by the cancer cells or the tumor associated stroma. *In vivo* biotinylation is a powerful technique, as it allows the investigation of the tumor and the associated niche proteome, while at the same time offering potential valuable candidates for future targeted therapies.

One cell line per molecular subtype was selected for the generation of orthotopic mouse tumors: PACO3 for exocrine-like PDAC, PACO17 for classical PDAC and finally PACO8 for QM PDAC (a minimum of 8 mice per experiment). Male mice were selected in an attempt to limit the influence of hormone levels variations (such as the ones associated with the female oestrous cycle) on the development of primary tumors. In total we perfused 33 tumor bearing mice, and nine healthy NSG male mice. For each experimental group we included a minimum of two negative controls – animals that were not perfused with the biotin ester. For PACO8 tumor bearing mice and the healthy control groups, we also performed a number of perfusions using the vector solution alone (dextrane-PBS) as additional negative controls. The number of mice used for the orthotopic tumor perfusion experiments are listed in **Table 12**.

Table 12: Overview of the number of animals used for the *in vivo* proteomic discovery experiments

Cell line	Type of xenograft	Number of mice	Number of non-biotinylated contrls	Number of biotinylated mice	No of mice optimally perfused
PACO3	Orthotopic	12	3	9	8
PACO8	Orthotopic	13	1+1 Dextrane	10	7
PACO17	Orthotopic	8	2	6	5
Negative ctrl	Healthy NSG	9	2 + 2 Dextrane	5	3

The efficiency of the perfusion experiments was evaluated first visually: the color of optimally perfused organs shifted to a lighter shade as a result of washing out the blood from the vascular system. If no dark clotted spots could be reported in the liver and kidney of perfused animals, we expected an overall good biotinylation quality. If however large clots could be detected, we labeled the organs as only partially perfused. An example of a successfully perfused mouse is presented in **Figure 17**.

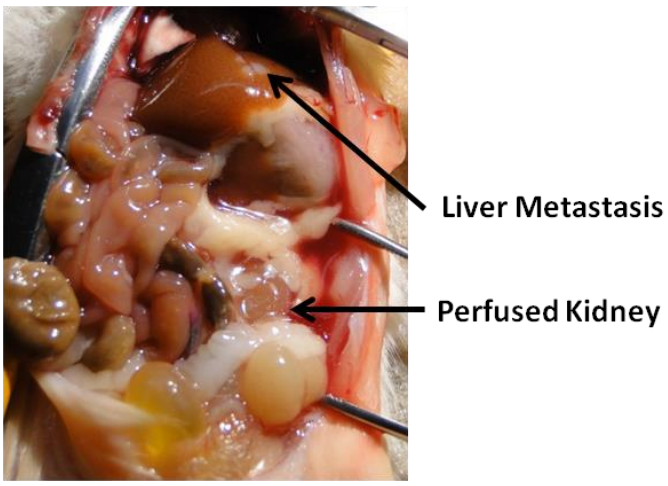


Figure 17: Mouse after receiving terminal whole body perfusion. Tissue discoloration can be observed for liver and kidney, indicating complete organ perfusion. The mouse had been injected intra-splenic with the PACO17 cell line and developed liver malignant lesions (metastases model).

The primary pancreatic tumor (for tumor bearing mice), or total pancreatic tissue was collected and snap frozen in liquid nitrogen. Some tumor bearing mice developed in the incubation time liver metastases, which were also collected for further proteomic analysis. Liver tissue was collected as controls for the assessment of the whole body perfusion's success. Small tissue samples were additionally embedded in OCT cryomedium and used for further experiments, evaluating the biotinylation efficiency. Frozen samples were stored at -80°C. Mice received number codes, which included the number of the PACO cell line from which they were derived, the acronym O (for orthotopic model) and a two number code indicating the identifier of the cage and the mouse number within the cage.

5.1.4.1 Evaluation of perfusion success using immunofluorescence

Prior to the proteomic analysis, a quality control check evaluating the biotinylation efficiency was performed, using the collected frozen OCT embedded tissues. Biotinylated areas were revealed by using a streptavidin-A488 conjugate, while the CD31 positive endothelial cells were visualized by means of secondary immunofluorescence (for details see *Materials and Methods*). For each sample we performed negative staining controls, from which the streptavidin conjugate and the primary anti-CD31 antibody were omitted.

We evaluated the perfusion efficiency on PACO3 and PACO17 derived pancreatic tumors (**Figure 18**).

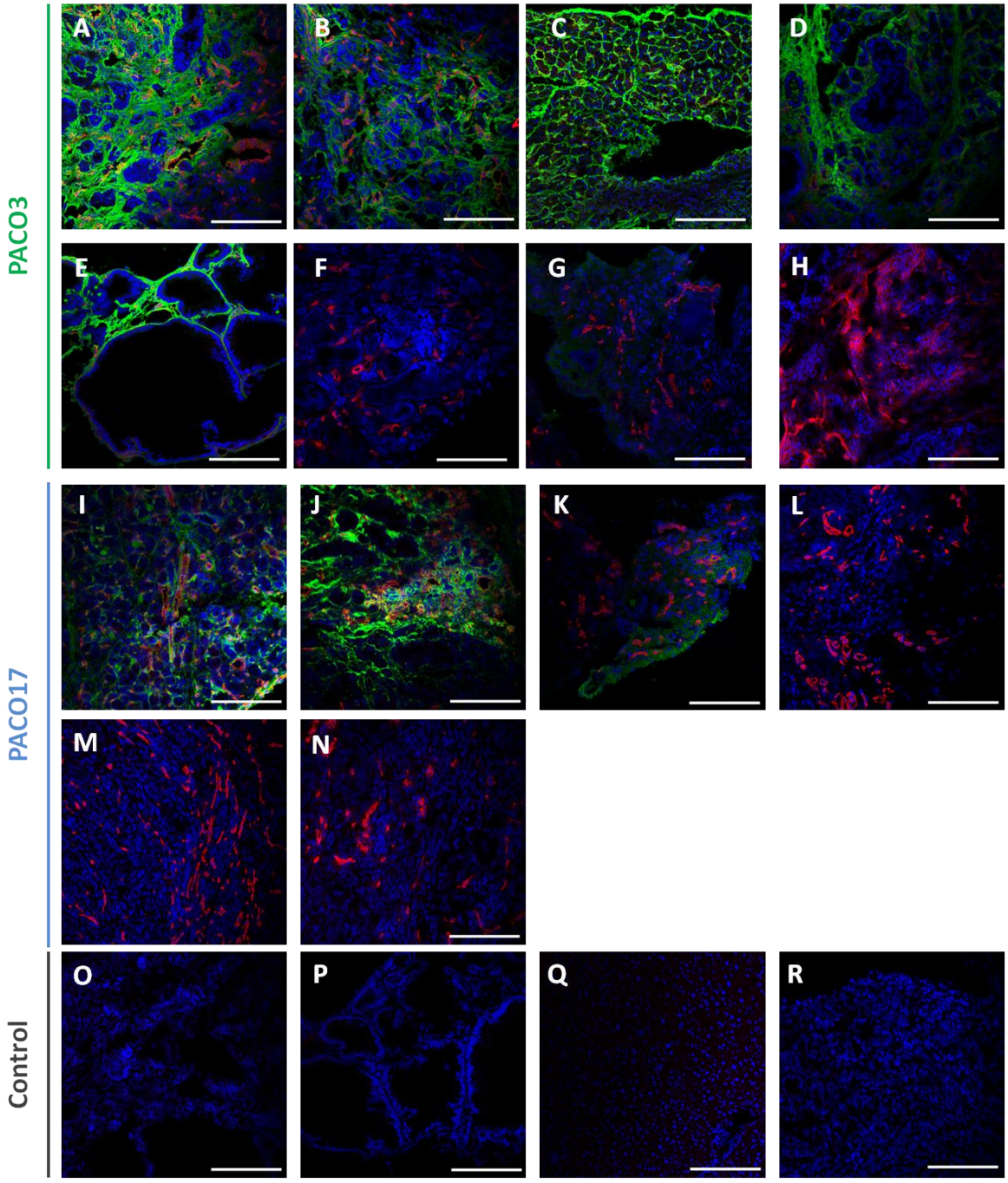


Figure 18: Orthotopic PACO derived pancreatic tumors: evaluation of biotinylation efficiency on frozen tissue. In **green**: biotin staining. In **red**: CD31 staining. **A**: PACO3.O.1.1 biotinylated tumor; **B**: PACO3.O.1.2 biotinylated tumor; **C**: PACO3.O.2.2 biotinylated tumor; **D**: PACO3.O.2.3 biotinylated tumor; **E**: PACO3.O.2.3 liver metastases – biotinylated; **F**: PACO3.O.1.4 suboptimal perfusion; **G**: PACO3.O.2.1 suboptimal perfusion; **H**: PACO3.O.1.3 non biotinylated control; **I**: PACO17.O.2.2 biotinylated tumor; **J**: PACO17.O.2.4 biotinylated tumor; **K**: PACO17.O.2.3 biotinylated, tumor with liquid cyst; **L**: PACO17.O.2.5 biotinylated, tumor with liquid cyst; **M**: PACO17.O.1.1 non biotinylated control; **N**: PACO17.O.2.1 non biotinylated control; **O**: PACO3.O.1.1 staining negative control; **P**: PACO3.O.2.3 liver metastases stainings negative control; **Q**: PACO17.O.2.2 staining negative control; **R**: PACO17.O.2.4 staining negative control. (Scale bars: 100 µm).

Figure 18 includes representative immunofluorescence stainings of the analyzed samples. As it can be observed, tumors which had been estimated as optimally perfused exhibited a strong biotin signal (**Figure 18 A → E** for PACO3 tumors; **I** and **J** for PACO17 tumors). The biotinylation positive signal was not restricted to areas in the immediate vicinity of the CD31 positive cells, thus indicating a good level of penetration for the reactive biotin ester used for the experiment. The spontaneous liver metastasis collected from one of the PACO3 tumor bearing mice was also optimally perfused (**Figure 18 E**). For the mice with suspected suboptimal perfusion, the biotin signal was indeed very low or absent, independent on the CD31 staining (**Figure 18 F → G** PACO3 tumors). In the case of PACO17, two of the collected primary tumors harbored a large liquid cyst at the center of the cancerous lesions, and to our surprise, despite what appeared to be a good body perfusion, the biotinylation signal recorded within the tumor was either very low (**Figure 18 K**), or absent (**Figure 18 L**). In nonbiotinylated control samples we could not detect any biotin signal (**Figure 18 H** and **M → N**), thus confirming the specificity of our detection system. Negative staining controls (**Figure 18 O → R**) could confirm the specificity of our primary and secondary immunofluorescence detection system. We also performed IF stainings on biotinylated liver samples, for additional confirmation of the whole body perfusion efficiency. As expected, we could confirm a strong biotin signal in those liver samples collected from mice where a high biotin signal was also detected in the orthotopic pancreatic tumors, thus confirming the efficiency of the *in vivo* perfusion (data not shown).

5.1.4.2 Evaluation and quantification of perfusion success using ELISA

One of the first demands for optimal discovery and quantitative proteomics experiments is the proper normalization of the amount of starting material for each sample analyzed. As already discussed, the efficiency of the whole body perfusion of tumor bearing mice can vary between distinct animals, depending on the proficiency of the experimenter, the existence of biological impediments (a hypoperfused tissue area, a tendency of the animal to clot abnormally due to liver damage), the maintenance of constant physical parameters during the perfusion (temperature, homogeneous solutions), etc.

In order to optimize the starting material for the mass spectrometric analysis, the amount of biotinylated protein needed to be determined in relation to the total protein concentration. For this reason we performed ELISA analyses, using a biotinylated albumin-based standard curve for quantification as described in *Materials and Methods*. Whole tumors where homogenized, and the total protein concentration was determined. 10^{-5} M total protein lysates were used for ELISA determination, as previously mentioned. We selected a number of tumors derived from each molecular subtype for this quantification experiment, based on the post perfusion evaluation of the biotinylation efficiency. Two classical (17.O.2.2 and 17.O.2.3), three exocrine-like (3.O.1.2, 3.O.2.2 and 3.O.2.3) and four QM (8.O.1.3, 8.O.1.4, 8.O.2.2 and 8.O.2.3) PDAC derived tumors were evaluated for the amount of total biotinylated proteins and the detected levels are presented in **Figure 19**.

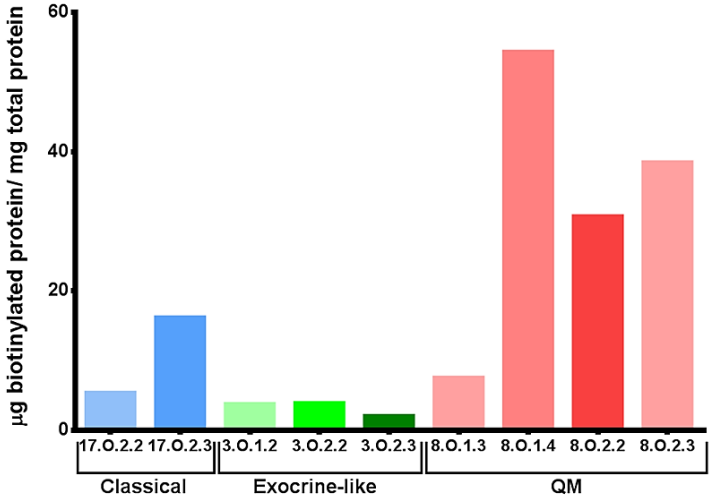


Figure 19: ELISA based evaluation of the biotinylation efficiency in PACO derived primary tumor xenografts.

Figure 19 indicates that tumors derived from different cell lines might be perfused with different efficiencies. PACO3 exocrine-like tumors for example showed the lowest level of biotinylation (~5 µg/mg total protein). By contrast, PACO8 QM derived tumors showed the highest level of biotinylation, with more than 20 µg/mg total protein. PACO17 tumors had intermediate biotinylation levels.

5.1.4.3 Identification of novel protein biomarkers in vivo using mass spectrometry

50 µg total biotinylated protein were enriched on a streptavidin column and analyzed as mentioned in the materials and methods chapter. Unlike the previous *in vitro* based discovery experiments, cysteines were alkylated and the enriched fraction was also delipidated using an organic solvent mixture, as decided upon optimization experiments.

For an initial proof of principle experiment, we selected two classical derived primary tumors (17.O.2.2 and 17.O.2.3) and one exocrine-like primary tumor (3.O.1.2). MS^2 spectra obtained from each sample were searched against a mixed human and mouse data base. The assignment of one protein to the human or mouse proteome was dependent upon the identification of at least one species specific proteotypic peptide, as determined by the Paragon algorithm.

In total, we identified 530 proteins in the three analyzed samples, out of which 177 were designated by the Paragon algorithm to be of human origin. Selected top hits and other interesting proteins identified are listed in **Table 13**.

Some of the putative PDAC markers we had selected for future validations, based on the *in vitro* discovery experiments could also be detected *in vivo*, such as: LCN2 (present in all three analyzed samples) and LGALS4. Surprisingly however, we identified LGALS4 in only one sample, belonging to the classical subtype PDAC, namely PACO17 derived tumor 17.O.2.3. Although the number of samples analyzed does not allow any generalized conclusions, it was reassuring to rediscover some of the top hits derived from the *in vitro* discovery experiment in the xenografted tumor material. Some proteins appeared to be

present only in the two classical PDAC tumors (such as integrin α-6), or only in the exocrine-like sample (pancreatic alpha-amylase). The mouse form of the digestive enzyme pancreatic alpha-amylase could also be detected, this time in all three analyzed tumors. Some of the top mouse specific proteins included ECM proteins (various collagen chains, integrins – see also **Supplementary Material**). Complement proteins secreted by the mouse-derived stromal cells were also identified. This observation is in line with literature data, indicating that complement system proteins can be also produced by the pancreas [139]. Based on these preliminary data, it is too early to draw a conclusion whether the mouse tumor associated proteome features components exclusively associated with distinct PDAC molecular subtypes. Examples of such putative subtype associated proteins deduced upon a first *in vivo* analysis could be represented by: galectin-1 (identified in the two classical tumors), ferritin heavy chain and moesin (present in the stromal fraction of the PACO3 exocrine-like tumor). Further analyses are required to distinguish if these proteins are truly associated with distinct molecular subtypes, or only with the particular cell lines we took for analysis.

Table 13 Selected proteins identified in the course of the *in vivo* proteomic discovery experiments, using orthotopic pancreatic tumors grown in immunocompromised mice. Both human and mouse specific proteins are listed

				Max number of peptides identified		
Protein Name	**Gene Name**	**SwissProt**	**Species**	**PACO17 17.O.2.2**	**PACO17 17.O.2.3**	**PACO3 3.O.1.2**
Collagen alpha-1(I) chain	Col1a1	P02452	HUMAN		28	46
Mucin-5AC (Fragments)	MUC5AC	P98088	HUMAN	8	2	
Pancreatic alpha-amylase	AMY2A	P04746	HUMAN			6
Integrin alpha-6	Itga6	P23229	HUMAN	5	2	
Collagen alpha-1(IV) chain	Col4a1	P02462	HUMAN		4	1
Laminin subunit beta-1	LAMB1	P07942	HUMAN	3		2
Laminin subunit gamma-1	Lamc1	P11047	HUMAN	3	2	
Galectin-4	LGALS4	P56470	HUMAN	3		
Integrin alpha-5	Itga5	P08648	HUMAN		3	
Neutrophil gelatinase-associated lipocalin	LCN2	P80188	HUMAN	1	1	2
Integrin alpha-V	Itgav	P06756	HUMAN			3
Plectin	PLEC	Q15149	HUMAN		2	2
Pancreatic alpha-amylase	Amy2	P00688	MOUSE	7	4	16
Complement C3	C3	P01027	MOUSE	50	35	69
Collagen alpha-1(I) chain	Col1a1	P11087	MOUSE	45	55	114
Collagen alpha-2(I) chain	Col1a2	Q01149	MOUSE	40	55	105
Collagen alpha-1(III) chain	Col3a1	P08121	MOUSE	21	28	61
Ferritin heavy chain	Fth1	P09528	MOUSE			3
Galectin-1	Lgals1	P16045	MOUSE	2	1	
Complement C2	C2	P21180	MOUSE	4	1	

Additionally, we also performed gene ontology enrichment analysis on the human specific, as well as the mouse specific protein fractions identified. An enrichment for proteins located in the extracellular compartment, basement membrane as well as basal lamina components could indeed be confirmed (**Figure 20**). However, for the human fraction, we identified several proteins with predicted intracellular localization.

Gene ontology analysis of the mouse specific list confirmed an enrichment of extracellular and secreted proteins upon the biotinylation and subsequent streptavidin pull down experiments (**Figure 21**). ECM components, basement membrane associated proteins as well as plasma constituents (such as plasma lipoprotein particles) were also enriched in our list of mouse specific proteins.

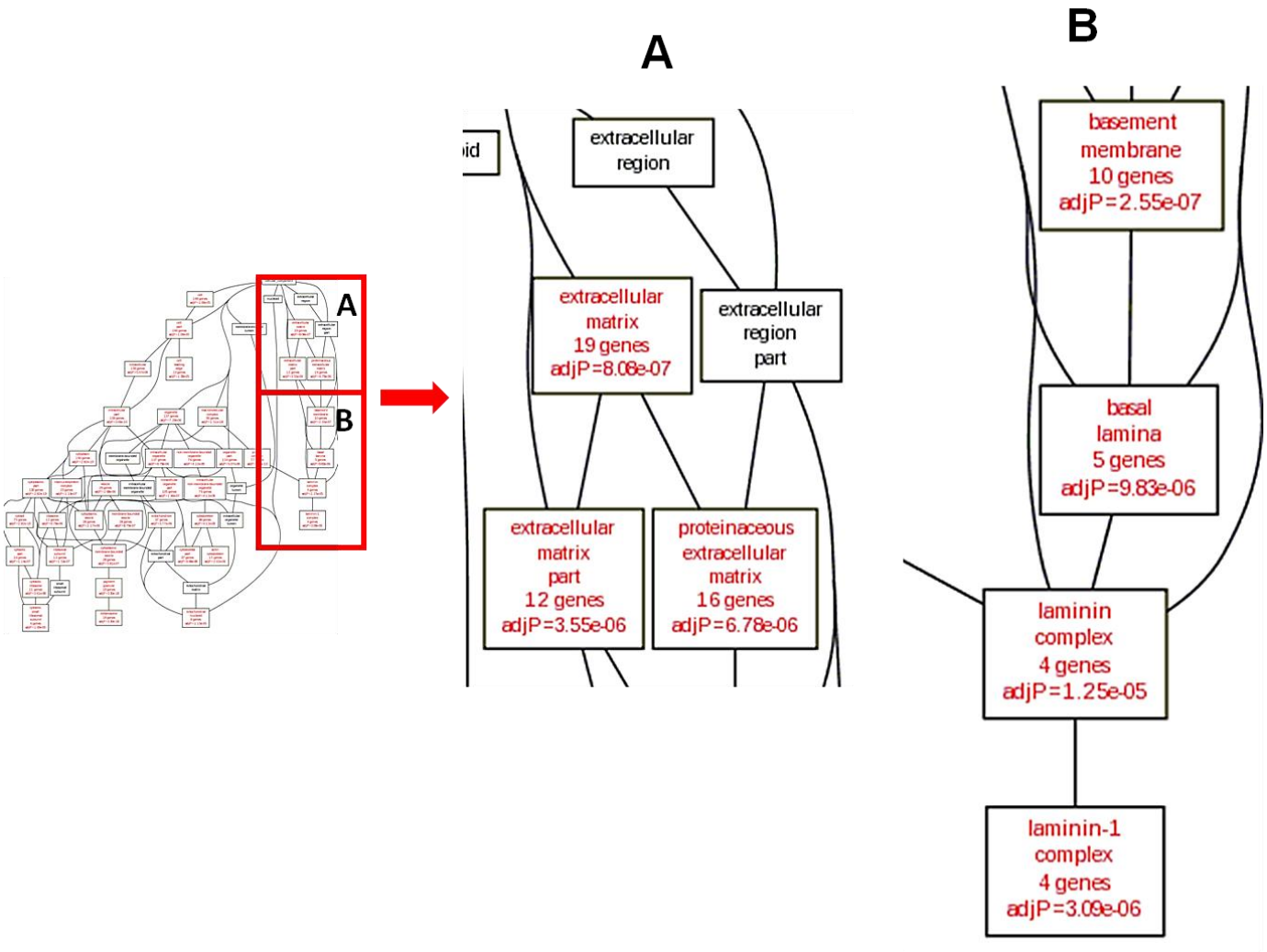


Figure 20: GO analysis of human proteins identified in the NSG mouse tumor xengrafts, upon performing *in vivo* biotinylation.

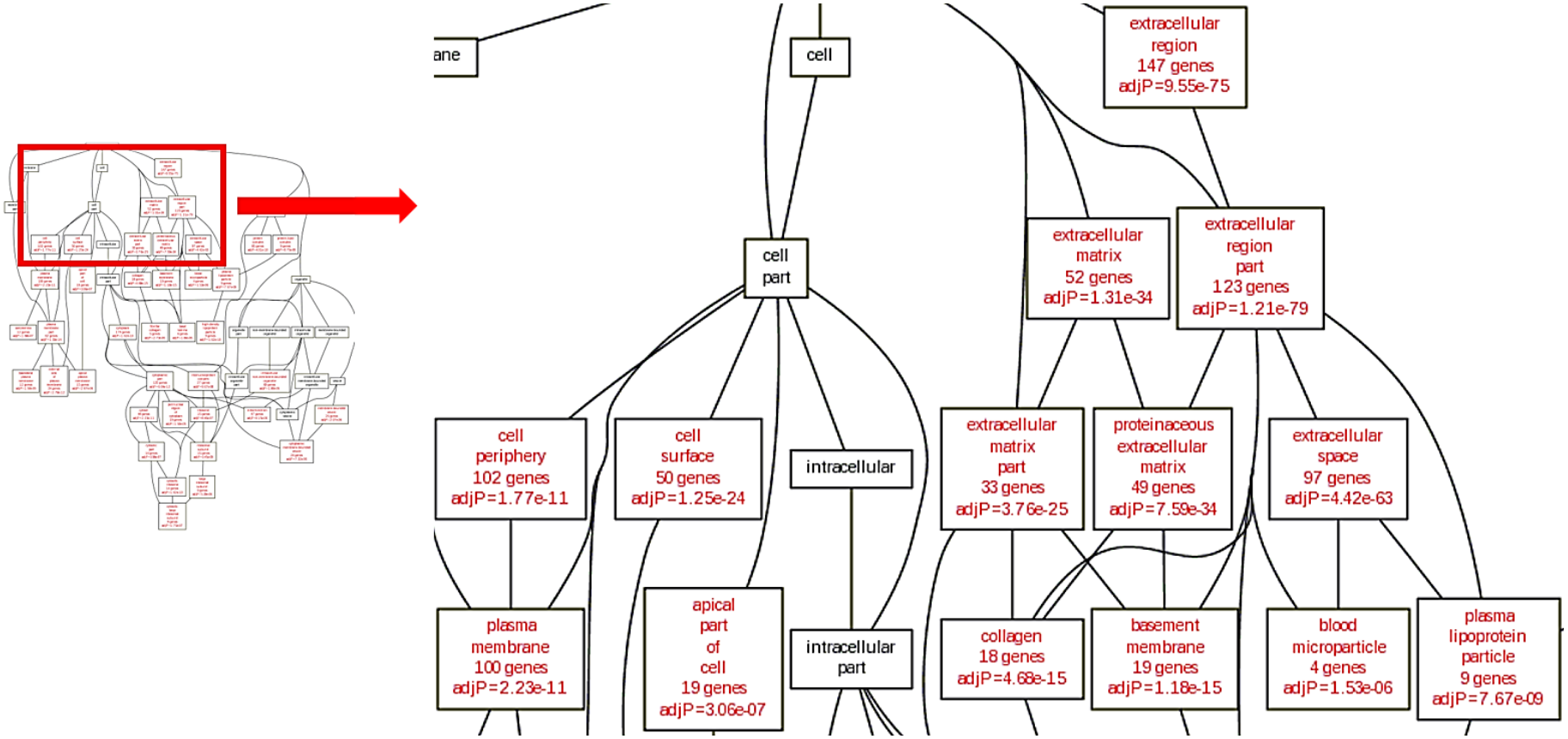


Figure 21: GO analysis of mouse proteins identified in the NSG mouse tumor xengrafts, upon performing *in vivo* biotinylation.

5.2 Validation of detected pan-PDAC and exocrine-like PDAC protein biomarker candidates

Based on the list of cell surface proteins identified in the course of the large-scale proteomics investigation on *in vitro* biotinylated cell lines, we selected two exocrine-like (CDH17 and LGALS4) and two pan-PDAC (PCDH1 and LCN2) protein marker candidates for further validations. The selection of the four candidates was based on the abundance of the proteins in the cell lines analyzed. We additionally evaluated their putative biomarker potential after consulting the available literature data about their localization in healthy and disease tissues, their known interactions and their novelty for the field of PDAC and biomarker discovery.

For all proteins of interest we performed thorough validations including antibody based (Western Blotting, Immunofluorescence) and antibody independent (SRM, RT-qPCR) methods. After first validating the expression patter of the four markers of interest *in vitro* (on cell lines), we investigated their presence in orthotopic xenografts developed in immunodeficient mice and patient material.

5.2.1 In vitro antibody based methods

Antibody based techniques used for validations are versatile and allow the evaluation of cellular localization for the antigen of interest, potential interactions between different antigens, or enable the distinction between different protein isoforms [31, 67]. Additionally, antibody based detection methods are commonly employed in the clinic, for the evaluation of markers circulating in the blood (ELISA), or for investigating protein expression patterns in fixed tissue samples (IHC) [31].

5.2.1.1 In vitro validations

5.2.1.1 In vitro validations using immunocytofluorescence

We first investigated if the four proteins of interest are present exclusively on the subtype(s) of interest, and if they are localized on the cell surface of the cancer cells. For this reason, we performed immunocytofluorescence using cells grown on special cell inserts, as described in the methods chapter. With regard to the putative pan-marker LCN2, we performed the initial validations on EDTA detached cells, fixed in formalin and embedded in histogel, since the antibody of choice did not properly recognized the antigen after fixing the cells with methanol.

For the two putative exocrine-like markers, IF stainings of methanol fixed cells showed a clear, exclusive expression of both CDH17 and LGALS4 on the exocrine-like cell lines: PACO3, PACO10, PACO14 and PACO18.

In the case of **CDH17**, a clear and strong membrane staining could be observed on PACO10 and PACO14 cell lines (**Figure 22**). PACO3 and PACO18 cells expressed CDH17 at lower levels, and in the case of PACO18 the staining pattern suggested the presence of some of the protein of interest in the cytoplasm of the cells. No distinguishable membrane staining

could be identified in classical or QM PACO cell lines, nor in the healthy and isotype controls. Potential back-ground staining could be reported in the case of some cell lines such as PACO19, PACO7 and the control cell line HPDE. We therefore performed additional experiments, in order to evaluate if the cytoplasmic signal reported in some of the non-exocrine-like cell lines could be attributed to background noise or atypical cellular localization. The complete panel of isotype control staining can be found in **Supplementary Figure 1**.

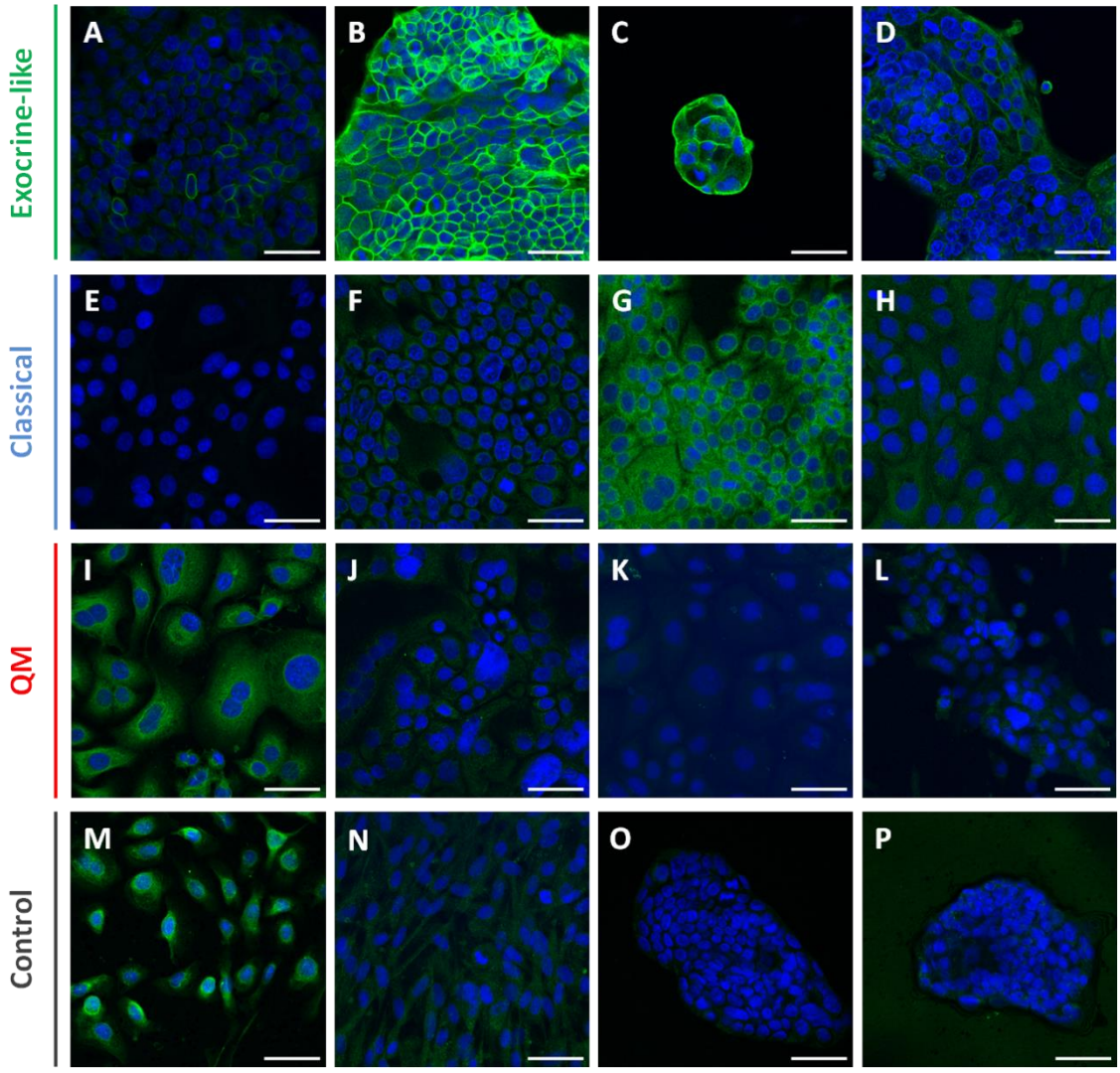


Figure 22: *In vitro* IF validation of putative exocrine-like marker CDH17 on PACO and control cell lines. **A**: PACO3; **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO2; **F**: PACO17; **G**: PACO19; **H**: PACO20; **I**: PACO7; **J**: PACO8; **K**: PACO9; **L**: PACO16; **M**: HPDE; **N**: HPNE; **O**: PACO10 isotype control; **P**: PACO14 isotype control (Scale bars: 50µm).

LGALS4 immunocytofluorescence staining revealed a strong honey-comb pattern staining (indicating membrane localization of the protein) in all cell lines belonging to the exocrine-like subtype (**Figure 23)**. In accordance with literature data, we could also observe cytoplasmic staining in these cell lines, and some putative nuclear localization in the case of cell lines PACO10 and PACO3 (data not shown). But for the exocrine-like subtype, no relevant staining for LGALS4 could be reported, with the exception of weak cytoplasmic signals present in classical cell lines PACO17 and PACO20. However, based on the IF data

alone we cannot distinguish if the observed signal represents actual cytoplasmic staining, or simply a higher background. Additional isotype control stainings are available in **Supplementary Figure 2**.

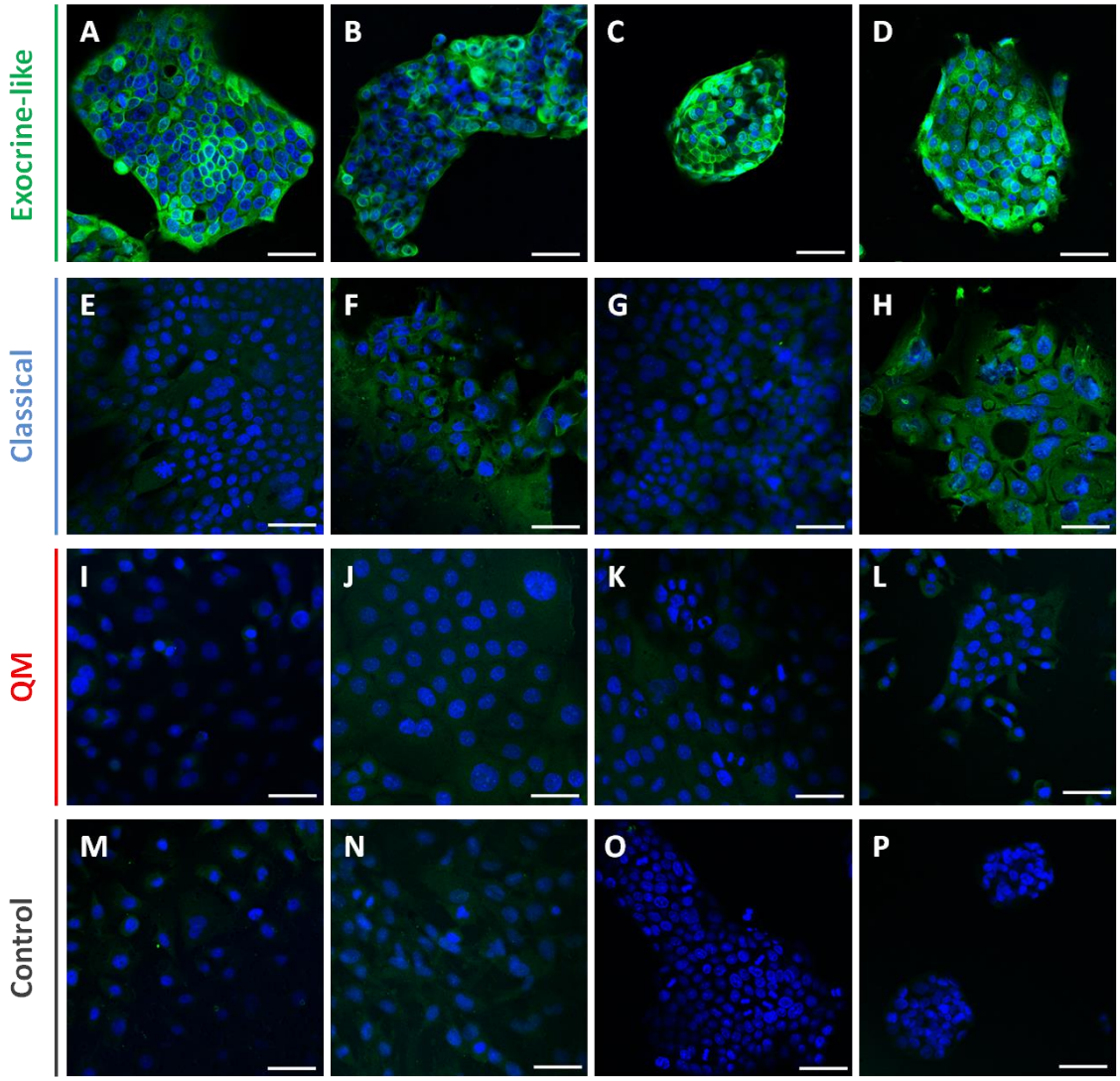


Figure 23: *In vitro* IF validation of putative exocrine-like marker LGALS4 on PACO and control cell lines. **A**: PACO3; **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO2; **F**: PACO17; **G**: PACO19; **H**: PACO20; **I**: PACO7; **J**: PACO8; **K**: PACO9; **L**: PACO16; **M**: HPDE; **N**: HPNE **O**: PACO3 isotype control; **P**: PACO14 isotype control (Scale bars: 50µm).

For putative pan-PDAC marker **LCN2** we performed stainings using EDTA detached histogel-embedded cells, since our antibody of choice did not stain methanol fixed cells. As shown in **Figure 24**, all tested PACO cell lines were positive for LCN2 staining, although the level of expression as well as the preferential localization of the protein varied greatly between cell lines, and even between clones isolated from the same patient. A stronger staining could be reported for cell lines PACO3, PACO8, PACO9, PACO14, PACO17, PACO19 and PACO20, where a membrane localization of the signal could also be deduced. PACO10 and PACO18 displayed a less intense signal, although it seemed to be mostly localized on the cell surface of the proteins, or associated with the membrane. PACO2 and PACO7 cells displayed weak staining, which appeared to be mostly localized in the

cytoplasm. It is interesting to observe great differences in staining intensity between cell lines PACO3 and PACO10, isolated from the same patient, and an even more striking difference between genetically identical cell lines PACO7 (displaying very low staining) and PACO8. Despite the fact that we could not detect any LCN2 expression in control cell lines after performing MS analysis of membrane fraction enriched samples, our IF stainings revealed a positive staining in the cytoplasm of HPNE cell lines. However, in line with our proteomics data, no membrane staining could be observed. The complete set of negative controls (samples incubated with the secondary antibody only) can be found in **Supplementary Figure 3**.

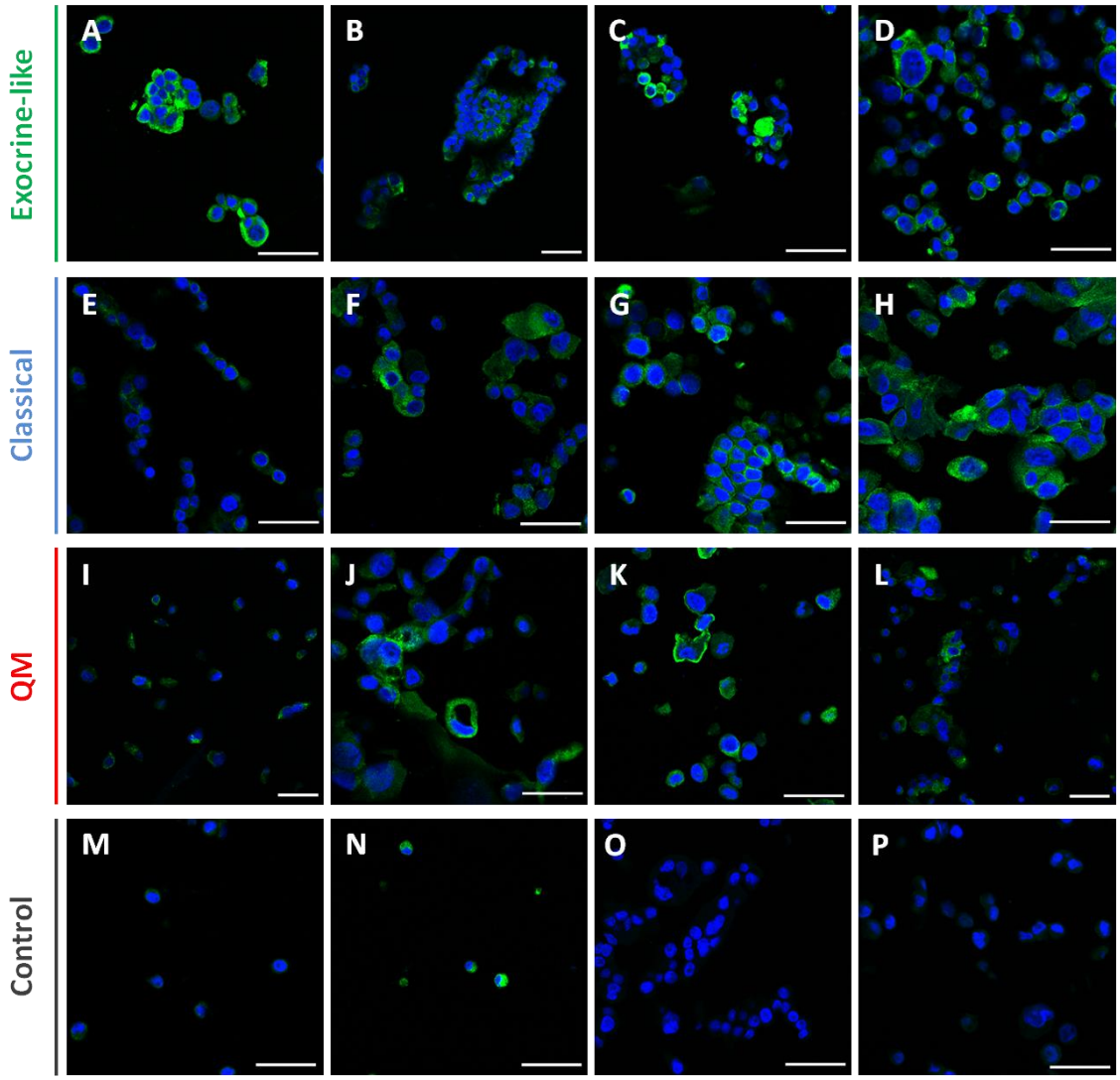


Figure 24: *In vitro* IF validation of putative pan-PDAC marker LCN2 on histogel embedded PACO and control cell lines. **A**: PACO3; **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO2; **F**: PACO17; **G**: PACO19; **H**: PACO20; **I**: PACO7; **J**: PACO8; **K**: PACO9; **L**: PACO16; **M**: HPDE; **N**: HPNE **O**: PACO2 negative control; **P**: PACO9 negative control (Scale bars: 50µm).

In vitro fluorescent staining of the putative pan-PDAC marker **PCDH1** showed variable levels of intensity in the investigated PACO cell lines (**Figure 25**). We could report the strongest stainings in QM cell lines, with PACO8 standing out with the most intense membrane signal observed among the investigated cell lines. Classical cell lines showed a

moderate staining, while the lowest staining intensity could be reported in exocrine-like cell lines, particularly PACO14 and PACO18. Besides weak cytoplasmic staining, a very clear sharp membrane staining could be easily detected in all PDAC cell lines, including the ones with the lowest overall level of staining. No clear membrane staining could be reported in the two control cell lines. The corresponding isotype controls confirmed the detected signal not to be the result of unspecific background (pictures **O** and **P**, **Figure 25** and **Supplementary Figure 4**).

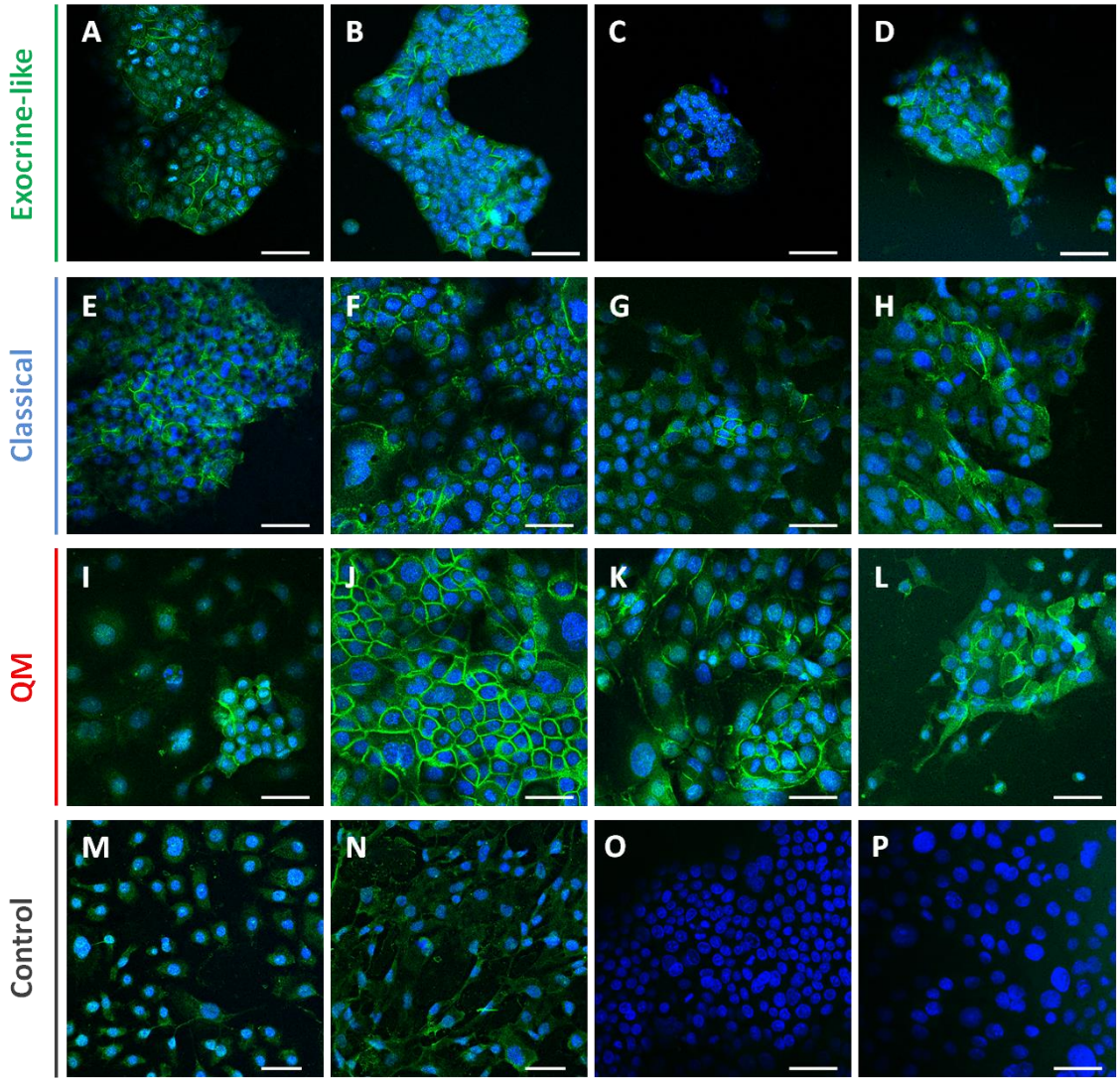


Figure 25: *In vitro* IF validation of putative pan-PDAC marker PCDH1 in PACO and healthy control cell lines. **A**: PACO3; **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO2; **F**: PACO17; **G**: PACO19; **H**: PACO20; **I**: PACO7; **J**: PACO8; **K**: PACO9; **L**: PACO16; **M**: HPDE; **N**: HPNE; **O**: PACO2 isotype control; **P**: PACO8 isotype control (Scale bars: 50µm).

5.2.1.1.2 Assessment of potential co-localization of exocrine-like protein biomarker candidates using confocal laser microscopy

In order to evaluate if the two putative exocrine-like protein markers could be interacting partners, we performed co-staining of cells grown on cell inserts and evaluated their localization using confocal laser microscopy. For this purpose, we took advantage of the fact that the primary mouse monoclonal antibodies raised against CDH17 and LGALS4 were of different isotypes. By using secondary antibodies raised against distinct mouse IgG isotypes, we were able to detect simultaneously the two antigens of interest. Fluorescent detection was achieved after using Alexa-dye labeled tertiary antibodies, recognizing the species of origin of the secondary antibodies.

We investigated the potential co-localization of CDH17 and LGALS4 on two exocrine-like cell lines (PACO10 and PACO25), one classical cell line – PACO2, and the QM cell line PACO9. New exocrine-like cell lines PACO25 and PACO26 were obtained from Dr.Elisa Espinet (HI-STEM; Metastasis Initiating Cells group).

In the case of the two exocrine-like cell lines, we observed strong membrane staining for CDH17 and cytoplasmatic as well as potential membrane staining for LGALS4 (**Figure 26**). In the classical and QM cell lines only weak background staining could be reported. Negative controls in which both secondary antibodies were excluded from the experiment were also performed in parallel, and revealed no unspecific background originating from the tertiary fluorescently labeled antibodies. After performing confocal laser microscopy of co-stained cells, we could detect areas were the two antigens co-localize. The co-localization seemed to occur, as expected, at the cell surface level and not intracellular.

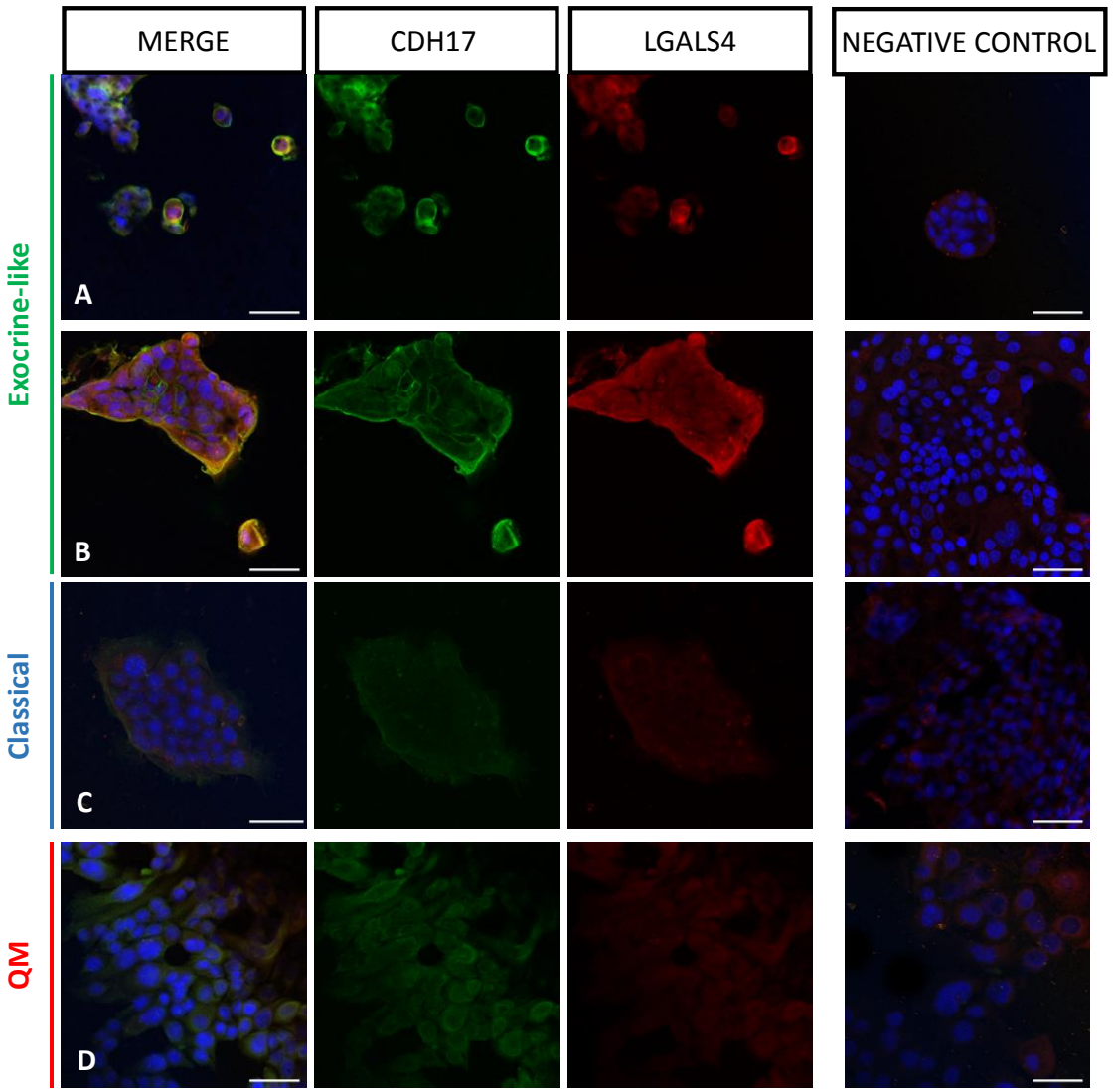


Figure 26: CDH17 and LGALS4 co-staining *in vitro*. From left to right: Alexa-488 and Alexa-647 channel merge; Alexa-488 channel (CDH17); Alexa-647 channel (LGALS4); negative control (no secondary antibodies) Alexa-488 and Alexa-647 channel merge. From top to bottom: PACO10 (**A**); PACO25 (**B**); PACO2 (**C**); PACO9 (**D**). (Scale bars: 50 µm).

In order to evaluate the specificity of the secondary antibodies, we performed control stainings, excluding one of the secondary antibodies. Similarly, we checked the specificity of the tertiary antibody by selectively excluding one tertiary antibodie and investigating if the corresponding fluorescent signal could still be detected. No cross reactivity of the secondary or tertiary antibodies could be detected after performing the above mentioned controls (**Figure 27**).

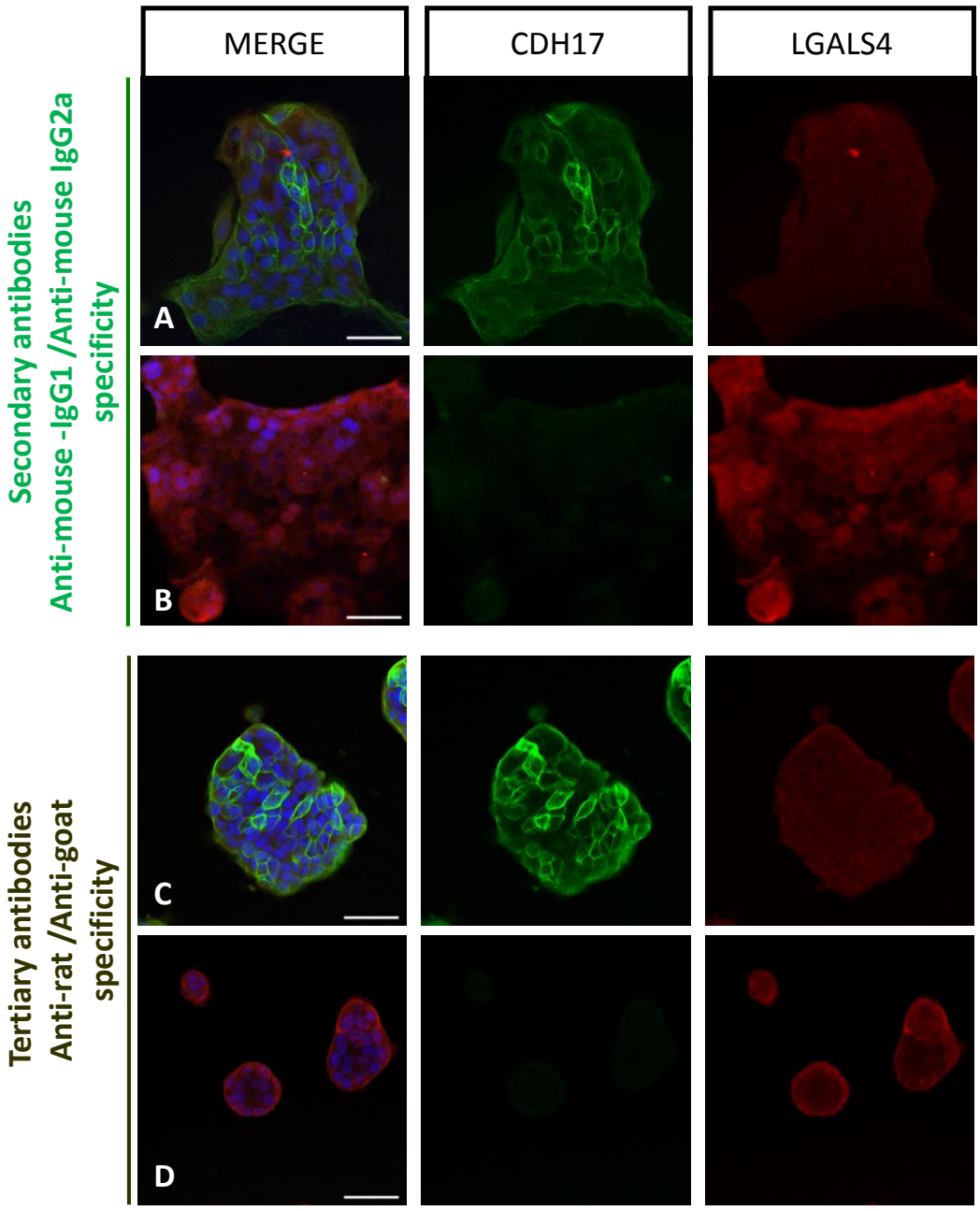


Figure 27: CDH17 and LGALS4 co-staining *in vitro* – control stainings assessing cross-reactivity of the secondary anti-mouse IgG1/anti-mouse IgG2a and tertiary anti-rat/anti-goat antibodies. **A**: PACO25 stained with CDH17 and LGALS4 primary antibodies + secondary rat anti-IgG1 (detecting CDH17 primary antibody); **B**: PACO25 stained with CDH17 and LGALS4 primary antibodies + secondary goat anti-IgG2a (detecting LGALS4 primary antibody). **C**: PACO10 stained with CDH17 and LGALS4 primary antibodies + secondary isotype specific antibodies + tertiary donkey anti-rat (detecting CDH17); **D**: PACO10 stained with CDH17 and LGALS4 primary antibodies + secondary isotype specific antibodies + tertiary donkey anti-goat (detecting LGALS4). (Scale bars: 50 µm).

5.2.1.1.3 Assessment of isoform distribution of pan-PDAC biomarker candidate PCDH1 using Western Blotting

Multiple isoforms have been reported in literature for PCDH1, although insufficient data exist about their distinct functions [119]. We wanted to investigate if PCDH1 expressed by PACO cell lines is represented predominantly by a single isoform, or if we can distinguish a subtype specific pattern of isoform expression. The monoclonal mouse anti-PCDH1 (clone 5D5) antibody employed was raised against a synthetic peptide, corresponding to the N-terminal domain of PCDH1. Therefore, the epitope recognized by the monoclonal antibody is common for both isoform 1 and isoform 2 of PCDH1. Although the 5D5 anti-PCDH1 antibody cannot directly discriminate between the two major isoforms of the protein, molecular weight differences reported between the two splice forms can in the end be used to discern signals originating from different isofoms.

Isoform 2 of PCDH1 contains an additional C-terminal domain, consisting of approximately 200 aminoacids. After analyzing the peptide sequences identified by the Paragon algorithm with 99% confidence in our cell surface proteomics identification experiment, we were able to confirm the presence of peptides spanning the unique C-terminal domain of PCDH1 isoform 2. **Figure 28** shows the alignment of isoforms 1 and 2 for PCDH1, performed using the ClustalW alignment tool (for simplification purposes, we selected for representation only the C-terminal sequences starting at position 900 for both isoforms). Identified peptides, common for both isoforms are marked in green, whereas isoform 2 specific peptides are highlighted in yellow. Considering the complete list of PCDH1 peptides, identified with a 99% confidence in all the samples analyzed during the cell surface proteome identification experiment, we could confirm a protein coverage of approximately ~28.5% for isoform 2 of PCDH1. The complete alignment of the two isoform sequences is presented in **Supplementary Materials**.

```
sp|Q08174|PCDH1_HUMAN        PSGKASKGNKSKGKKSKSPKPVKPVEDEDEAGLQKSLKFNLMSDAPGDSP
sp|Q08174-2|PCDH1_HUMAN      PSGKASKGNKSKGKKSKSPKPVKPVEDEDEAGLQKSLKFNLMSDAPGDSP
                             **************************************************

sp|Q08174|PCDH1_HUMAN        RIHLPLNYPPGSPDLGRHYRSNSPLPSIQLQPQSPSASKKHQVVQDLPPA
sp|Q08174-2|PCDH1_HUMAN      RIHLPLNYPPGSPDLGRHYRSNSPLPSIQLQPQSPSASKKHQVVQDLPPA
                             **************************************************

sp|Q08174|PCDH1_HUMAN        NTFVGTGDTTSTGSEQYSDYSYRTNPPKYPSKQVGQPFQLSTPQPLPHPY
sp|Q08174-2|PCDH1_HUMAN      NTFVGTGDTTSTGSEQYSDYSYRTNPPKYPSKQLPHRRVTFSATSQAQEL
                             **********************************: :     :. . .:

sp|Q08174|PCDH1_HUMAN        HGAIWTEVWE----------------------------------------
sp|Q08174-2|PCDH1_HUMAN      QDPSQHSYYDSGLEESETPSSKSSSGPRLGPLALPEDHYERTTPDGSIGE
                             :..   . ::

sp|Q08174|PCDH1_HUMAN        --------------------------------------------------
sp|Q08174-2|PCDH1_HUMAN      MEHPENDLRPLPDVAMTGTCTRECSEFGHSDTCWMPGQSSPSRRTKSSAL


sp|Q08174|PCDH1_HUMAN        --------------------------------------------------
sp|Q08174-2|PCDH1_HUMAN      KLSTFVPYQDRGGQEPAGAGSPSPPEDRNTKTAPVRLLPSYSAFSHSSHD


sp|Q08174|PCDH1_HUMAN        -------------------------------------
sp|Q08174-2|PCDH1_HUMAN      SCKDSATLEEIPLTQTSDFPPAATPASAQTAKREIYL
```

Figure 28: Alignment of isoform 1 (higher lane) and isoform 2 (lower lane) of PCDH1 performed using ClustalW tool. In green: peptides identified common to both isoforms. In yellow: isoform 2 specific peptides identified. For simplification purposes, the presented alignment commences at position 900 for both isoforms.

To further evaluate PCDH1 isoform distribution in the PACO cell lines, we performed WB analysis using whole cell lysates, as described in the methods section. GAPDH was employed as loading control.

As presented in **Figure 29**, we detected in most PACO cell lines a single band with an approximate molecular weight of 135 kDa, corresponding most likely to isoform 2 of PCDH1. These observations were in line with the MS/MS data, which allowed the identification of peptides specific to isoform 2 alone in several of the cell lines analyzed (data not shown). In the case of PACO20, an additional band could also be detected, at a higher molecular weight. We could not determine if the additional band represents a hyper glycosylated form, or another splice isoform of the protein. No band corresponding to PCDH1 protein could be detected in either of the healthy control cell lines.

According to the WB experiment, the highest PCDH1 signal could be detected in PACO20 cell line. High amounts of the protein could also be detected in PACO8, PACO9 and cells belonging to the classical subtype, mostly confirming the observations of the *in vitro* IF experiments (see **Figure 25**). Surprisingly, we could not detect any PCDH1 signal in the cell line PACO16.

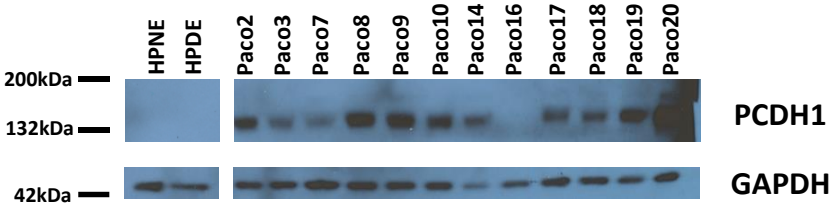


Figure 29: Western Blot analysis of PCDH1 expression in PACO and control cell lines using whole cell lysates.

5.2.1.2 In vivo immunohistofluorescence validations using mouse derived xenografts

In order to evaluate if the markers present *in vitro* on the surface of the pancreatic cancer cells maintain their pattern of expression *in vivo*, we performed IF experiments using pancreatic tumor xenografts grown in immunocompromised NSG mice. As previously described, PACO cells were injected in the head of the pancreas of male mice and the tumors were allowed to grow until they reached ~1 cm in diameter. Mice were sacrificed and pancreatic tumors, as well as any growing liver or lung metastasis were preserved in formalin and latter embedded in paraffin. In addition, we received three novel xenografts from Dr. Elisa Espinet (HI-STEM, Metastasis Initiating Cells group), derived from the novel exocrine-like cell lines PACO25, PACO26 and PACO28. In parallel we also analyzed the pancreas of healthy NSG male mice and human healthy pancreas samples obtained from patients undergoing tumor resection (material provided by Dr. Elisa Espinet).

When injected into NSG mice, PACO cells developed tumors with distinct histologies, depending on the molecular subtype to which they are annotated. Exocrine-like and classical xenografts often exhibited ductal like structures; whereas QM derived tumors usually display a poorly differentiated histology. One representative xenograft for each molecular subtype is shown in **Figure 30** – slides were stained only with haematoxylin, and were selected from the panel of isotype controls performed for IHC experiments on paraffin embedded orthotopic tumors developed in NSG mice.

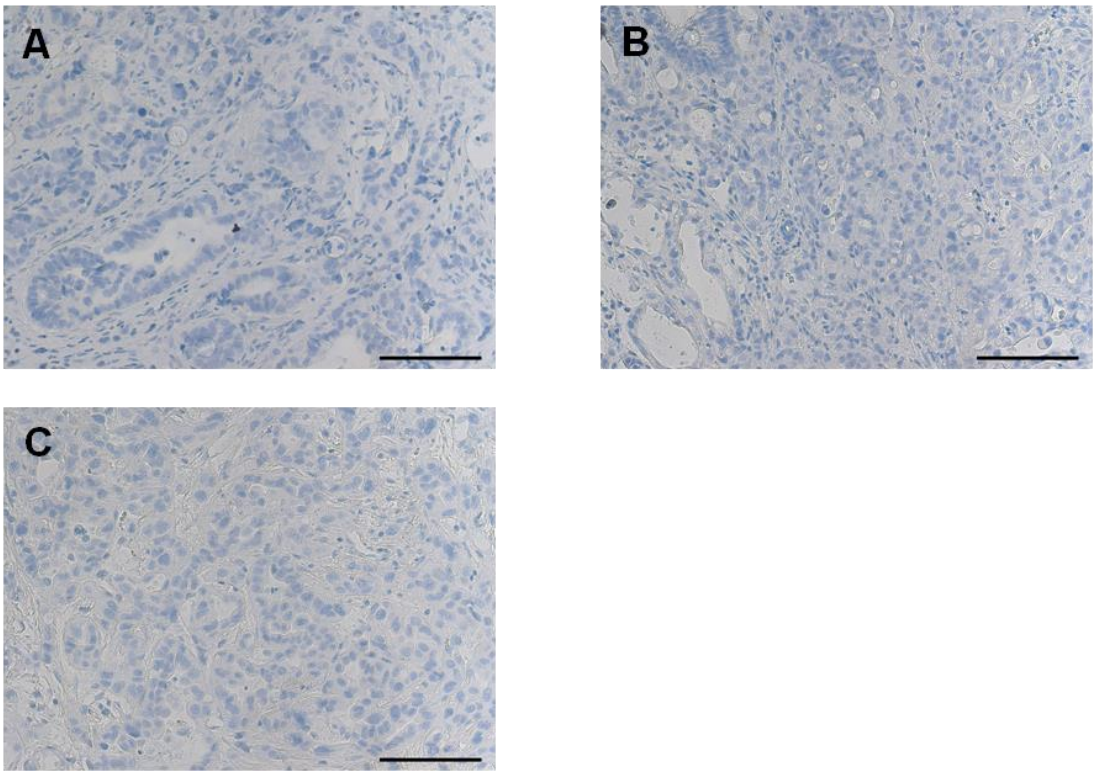


Figure 30: Haematoxylin staining of mouse derived PDAC tumors, revealing histological differences between the different molecular subtypes. **A**: PACO3 derived tumor (exocrine-like); **B**: PACO19 derived tumor (classical); **C**: PACO9 derived tumor (QM). (Scale bars: 100µm).

CDH17 expression *in vivo* confirmed most of the initial observations made i*n vitro*. Primary tumors derived from exocrine-like PACO cell lines showed as expected a clear membrane staining. These observations were also true for the novel tested PACO25, PACO26 and PACO28 cell line derived xenografts. However, PACO26 expressed lower levels of the exocrine-like putative biomarker. As it could also be reported *in vitro*, PACO18 CDH17 staining remained weak (**Figure 31**). Additionally, we could report CDH17 expression in some of the classical derived pancreatic tumors, with PACO19 showing the strongest expression, including a clear membrane staining. QM derived xenografts remained negative for the exocrine-like marker, despite showing a significant back ground, which could also be reported in the isotype controls (**Supplementary Figure 5**). We could not detect any expression for CDH17 neither in the two healthy human controls, nor in healthy mouse NSG pancreata. The signal contrast for the Alexa-488 channel was processed, after picture acquisition, using the same parameters for all samples, in the ZEN lite 2012 software, for better visualization. Due to significant back-ground staining, typical for the healthy mouse pancreas, the contrast of the acquired picture for the healthy NSG pancreata staining was not modified.

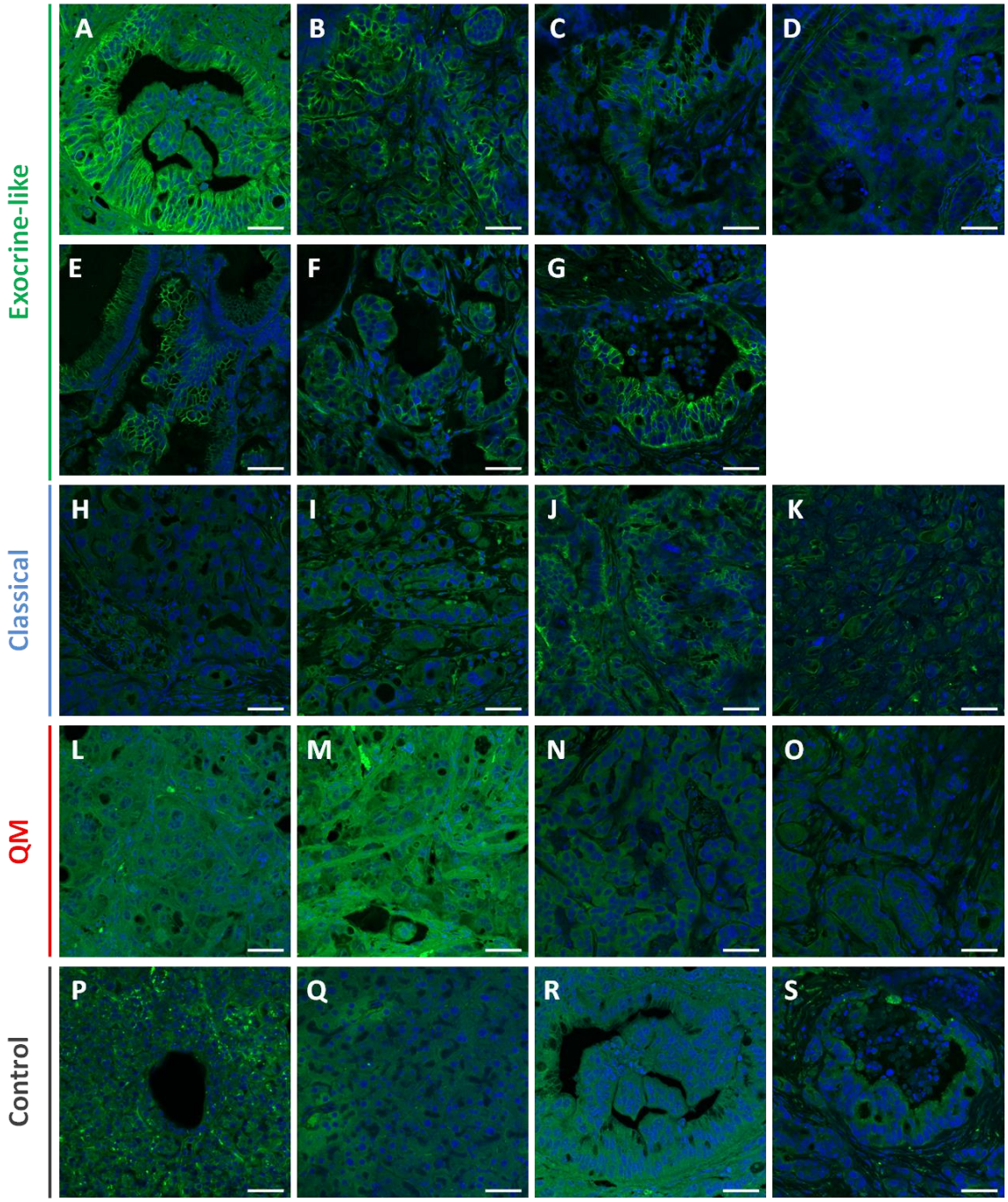


Figure 31: CDH17 staining *in vivo* using PACO derived pancreatic tumors in NSG mice and comparison to healthy pancreas isolated from human and mouse. **A**: PACO3, **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO25; **F** PACO26; **G**: PACO 28; **H**: PACO2; **I**: PACO17; **J**: PACO19; **K**: PACO20; **L**: PACO7; **M**: PACO8; **N**: PACO9; **O**: PACO16; **P**: human healthy pancreas patient # 4984; **Q**: healthy mouse NSG pancreas; **R**: PACO3 isotype control; **S**: PACO28 isotype control. (Scale bars: 50 µm).

We also investigated if CDH17 protein levels are up- or down-regulated in different metastatic lesions. For this reason, we employed PACO10 and PACO20 spontaneous lung and liver metastases, and compared the intensity of the staining to that of the primary tumor (**Figure 32**). As it can be observed in **Figure 32** for the PACO10 derived tumors we could observe an increase in signal intensity in the metastasis compared to the primary tumor, indicating a stronger expression of CDH17 in the liver metastasis. A similar observation

could be made for PACO3 spontaneous liver metastases, when the tissue was embedded in OCT and stained for CDH17 (data not shown). However, in the PACO10 lung metastasis, CDH17 expression seemed to be completely absent, when compared to the isotype control (both positive and isotype staining had a high background). For the classical cell line PACO20, we could as well observe a significant increase in the intensity of CDH17 staining in the metastases, with a clear membrane pattern for lung metastases. Unfortunately, no liver metastasis was available for comparison staining. Additional experiments will be required to asses if CDH17 is typically up-regulated in metastatic cells, and if the organ harboring the secondary cancer and the molecular subtype dictate the expression pattern.

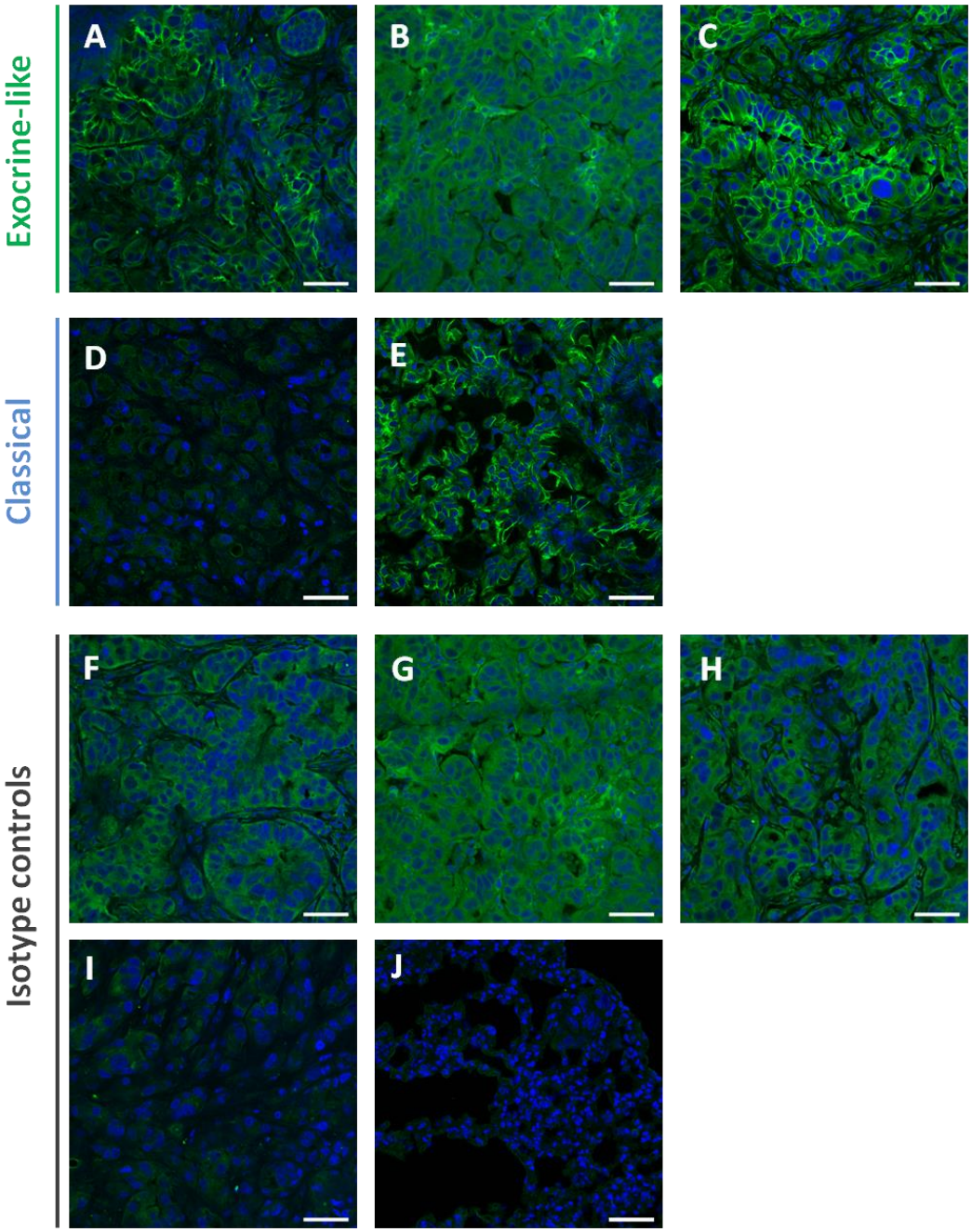


Figure 32: CDH17 expression in primary tumor and metastasis. **A**: PACO10 tumor pancreas; **B**: PACO10 lung metastasis; **C**: PACO10 liver metastasis; **D**: PACO20 tumor pancreas; **E**: PACO20 lung metastasis. The lower panel of pictures (**F** → **J**) includes the isotype controls corresponding to the **A** → **E** positive stainings. (Scale bars: 50 µm).

LGALS4 IF *in vivo* revealed a strong staining in all exocrine-like derived xenografts, with the signal localizing both in the cytoplasm and at the plasma membrane (**Figure 33**). In the case of PACO14 and PACO25 originating tumors, the cytoplasmatic staining appeared less intense, while most of the protein was either localized at the plasma membrane, or associated with the cell surface. As reported for CDH17, we could also observe an *in vivo* activation of LGALS4 in some xenografts derived from classical cell lines – PACO17 showed a low expression, while the LGALS4 levels of PACO19 appeared to be comparable to the ones observed for the exocrine-like pancreatic tumors. QM derived cancers were mostly negative for LGALS4, with the exception of PACO16, where a moderate LGALS4 expression could be detected. The pancreas of healthy NSG mice was negative for LGALS4, while we could detect a moderate cytoplasmic and potentially membrane staining in the healthy human pancreas sample obtained from resected pancreata. No unspecific signal could be detected in the appropriate isotype control stainings (**Supplementary Figure 6**).

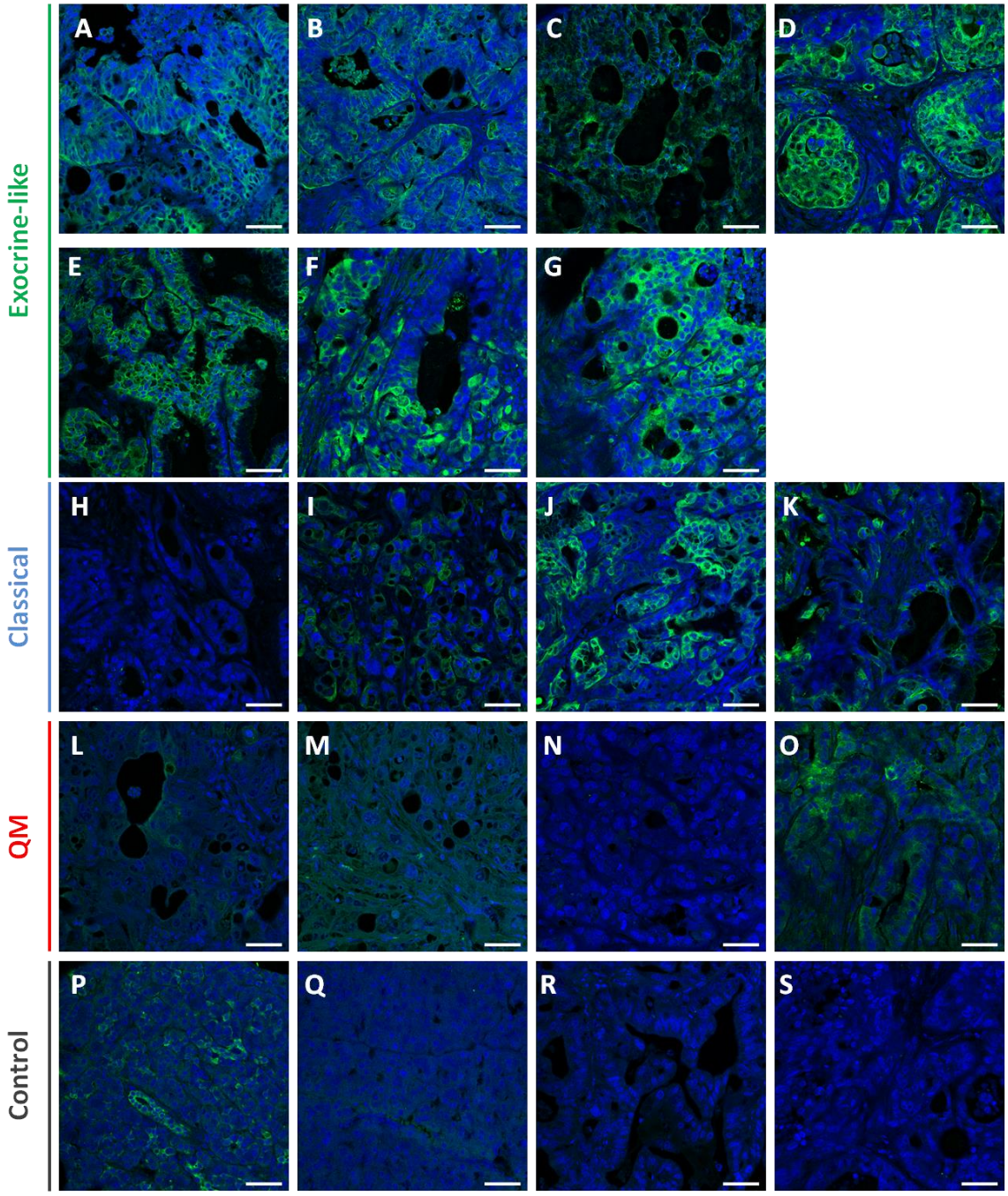


Figure 33: LGALS4 in vivo staining using PACO derived pancreatic tumors in NSG mice and comparison to healthy pancreas isolated from human and mouse **A**: PACO3, **B** PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO25; **F** PACO26; **G**: PACO 28; **H**: PACO2; **I**: PACO17; **J**: PACO19; **K**: PACO20; **L**: PACO7; **M**: PACO8; **N**: PACO9; **O**: PACO16; **P**: human healthy pancreas patient # 4984; **Q**: healthy mouse NSG pancreas; **R**: PACO10 isotype control; **S**: PACO18 isotype control. (Scale bars: 50 µm).

When investigating the presence of LGALS4 in PACO10 (exocrine-like) and PACO20 (classical) spontaneously derived metastases (**Figure 34**), we could report similar patterns as earlier described for CDH17. LGALS4 staining was absent in lung metastases derived from the exocrine-like cell line PACO10. In the liver metastases however, the staining intensity was similar to that detected in the primary tumor and the likely originated from the

cytoplasm and the membrane of metastatic tumor cells. PACO20 primary tumor staining for LGALS4 was negligible. A strong activation of LGALS4 protein expression could be reported in the lung metastases, with the majority of the fluorescent signal originating from the cytoplasm (**Figure 34**). Isotype control staining confirmed that the observed signals were not the result of unspecific antibody binding.

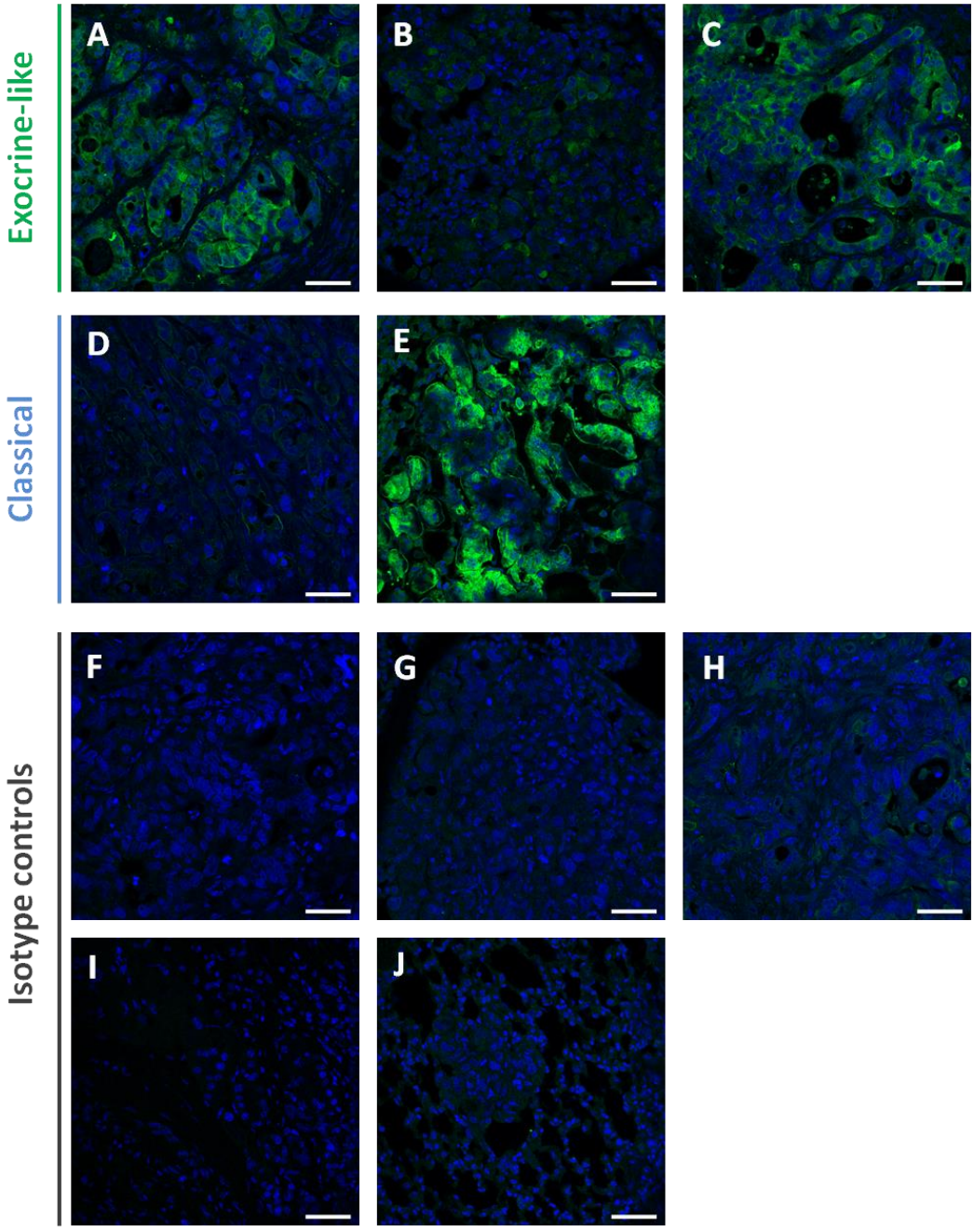


Figure 34: LGALS4 expression in primary tumors and metastases: **A**: PACO10 tumor pancreas; **B**: PACO10 lung metastasis; **C**: PACO10 liver metastasis; **D**: PACO20, tumor pancreas; **E**: PACO20 lung metastasis. The lower panel of pictures (**F** → **J**) includes the isotype controls corresponding to the **A** → **E** positive stainings. (Scale bars: 50 µm).

For the putative pan-PDAC marker **LCN2**, *in vivo* IF results indicated that most pancreatic cancer cell lines continue to produce the protein at high levels after injection into NSG mouse pancreas (**Figure 35**). Some of the strongest expression could be detected for exocrine-like xenografts, as well as tumors derived from classical cell lines PACO19 and PACO20. Lower expression could be reported for the exocrine-like tumor PACO26 and the primary cancer derived from the PACO8 QM cell line. No clear signal could be detected for PACO2, PACO17 (classical PDAC) and PACO7 (QM PDAC) pancreatic tumors. For PACO2 and PACO7 the results are in agreement with previous *in vitro* validation experiments, where very low levels could be reported for the corresponding samples. LCN2 staining was negative both in two healthy human pancreatic samples (only patient # 4984 is presented in **Figure 35**), and in the pancreata of healthy NSG mice. With regard to the localization of LCN2 in mouse PDAC xenografts, we could observe that most staining appeared to be cytoplasmic and potentially membrane associated. A relatively clear extra-cellular staining could be reported for PACO18 tumors (exocrine-like PDAC). For some xenografts, a stronger LCN2 staining became apparent for ductal like structures, at the apical level (in the case of PACO14, PACO16, PACO25). The complete panel of isotype controls is provided in **Supplementary Figure 7**.

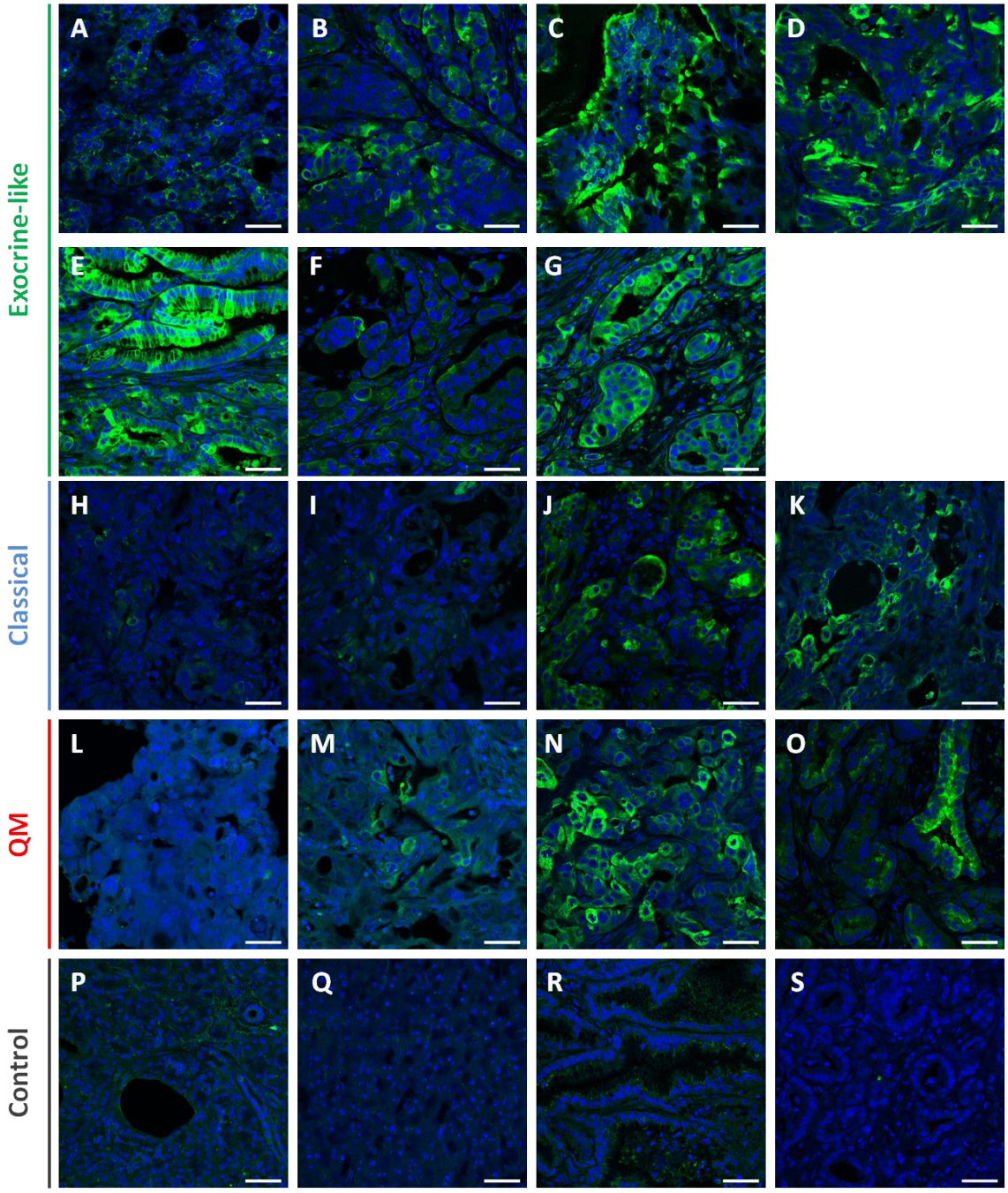


Figure 35: LCN2 *in vivo* staining of PACO derived pancreatic tumors in NSG mice and comparison to healthy pancreas isolated from human and mouse **A**: PACO3, **B** PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO25; **F** PACO26; **G**: PACO 28; **H**: PACO2; **I**: PACO17; **J**: PACO19; **K**: PACO20 **L**: PACO7; **M**: PACO8; **N**: PACO9; **O**: PACO16; **P**: human healthy pancreas patient #4984; **Q**: healthy mouse NSG pancreas; **R**: PACO25 isotype control; **S**: PACO16 isotype control. (Scale bars: 50 µm).

LCN2 expression in metastatic lesions appeared to vary greatly, depending on the originating cell line. For documenting the LCN2 IF in primary and secondary tumors in parallel, the intensity of the 488 nm laser was adjusted for optimal detection to the high fluorescence signal recorded for the PACO10 metastases samples. Thus, a significant upregulation of LCN2 can be clearly observed for both PACO10 derived liver and lung spontaneous metastases when compared to the primary tumor (**Figure 36**). In the PACO20

classical tumors on the other hand, the LCN2 signal disappeared in the lung metastases (metastatic areas are delineated in white). More samples need to be analyzed in order to appreciate if the metastases's expression patterns currently reported can be extrapolated to the whole panel of secondary cancers originating from the same molecular subtype PDAC.

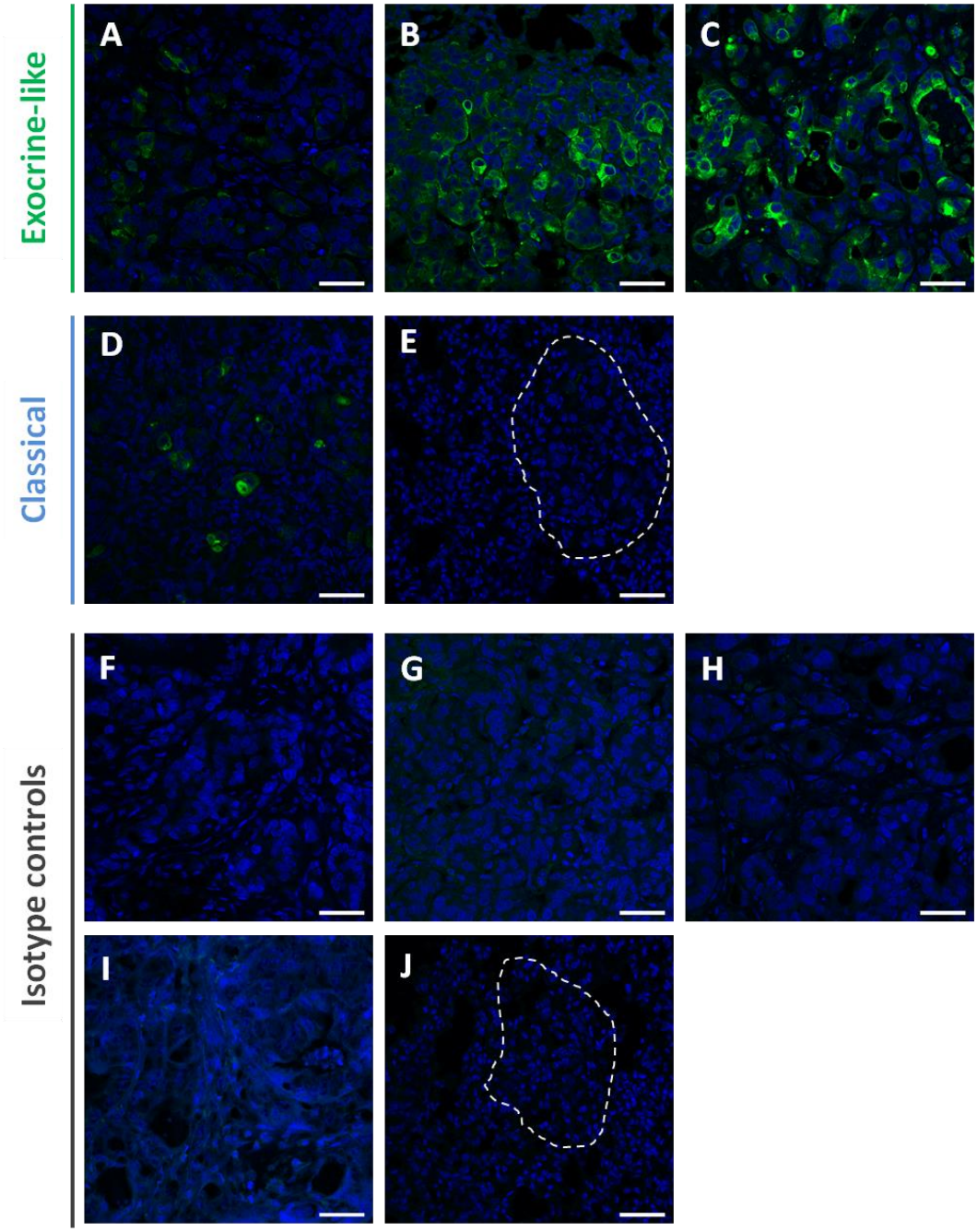


Figure 36: LCN2 expression in primary tumor and metastases: **A**: PACO10 tumor pancreas; **B**: PACO10 lung metastasis; **C**: PACO10 liver metastasis; **D**: PACO20, tumor pancreas; **E**: PACO20 lung metastasis. The lower panel of pictures (**F** → **J**) includes the isotype controls corresponding to the **A** → **E** positive staining. (Scale bars: 50 µm).

PCDH1 *in vivo* staining showed great variance in terms of expression levels in the different samples (**Figure 37**). Some of the highest signals could be detected for exocrine-like xenografts derived from the PACO3, PACO10 and PACO14 cell lines. A strong, likely membrane staining, could be detected for the majority of the analyzed xenografts, clearly exemplified by PACO14, PACO7, PACO8, PACO19 and PACO16 tumors. One of the best examples of strong localized PCDH1 membrane staining can be observed in **Figure 37M** where the positive staining pattern is clearly distinguishable from the isotype control (**Figure 37R**), especially when taking into consideration that the pictures were taken in identical areas of consecutive cuts made from the same paraffin block. The extent of the PCDH1 positive areas varied as well amongst samples: for example, in the case of PACO7 samples, we could detect only small areas of clear membrane staining (data not shown). For PACO10 xenografts on the other hand, the majority of the primary tumor was positive for PCDH1. Despite being one of the highest PCDH1 expressing cell lines *in vitro*, PACO9 showed no detectable staining after transplantation into NSG mice. PACO23 primary tumor appeared also negative for PCDH1 (compared to the staining with the isotype control, available in **Supplementary Figure 8**). PACO26 xenografts showed low staining for PCDH1, which appeared to be localized at the basal and apical level of the cancer cells. Healthy human and mouse pancreata were negative for PCDH1 (albeit a strong background could be detected in the case of the mouse pancreatic tissue).

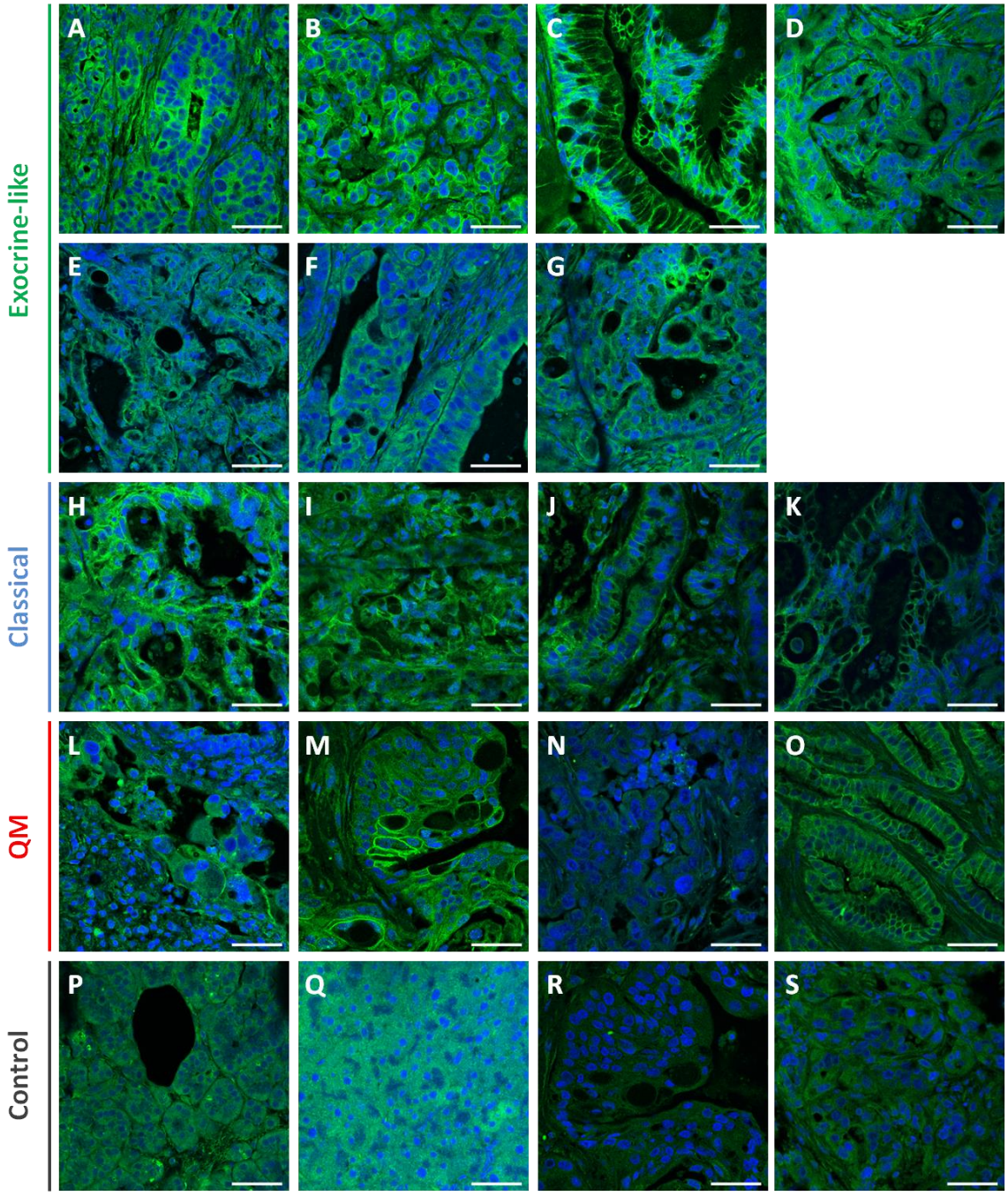


Figure 37: PCDH1 *in vivo* stainings of PACO derived pancreatic tumors in NSG mice and comparison to healthy pancreas isolated from human and mouse **A**: PACO3, **B** PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO25; **F** PACO26; **G**: PACO 28; **H**: PACO2; **I**: PACO17; **J**: PACO19; **K**: PACO20; **L**: PACO7; **M**: PACO8; **N**: PACO9; **O**: PACO16; **P**: human healthy pancreas patient # 4984; **Q**: healthy mouse NSG pancreas; **R**: PACO8 isotype control; **S**: PACO10 isotype control. (Scale bars: 50 µm).

5.2.1.3 In vivo validations on patient material

Original human primary pancreatic tumors from which the patient matched PACO cell lines were derived, was available for testing (courtesy of Dr.Christian Eisen, HI-STEM). We set out to investigate if the expression profile of the markers of interest identified *in vitro* matched that of the original tumor. Unfortunately we did not have original tumor material for all the derived cell lines and in some cases the tumor pieces stored in paraffin were highly necrotic, with very few and sometimes very small ductal-like tumor areas making the validation experiments difficult. Additionally, tumors with a high stromal component are associated with a higher tissue autofluorescence, thereby hindering the IF based validation experiments. At the time at which this report was written, no confocal fluorescent microscope including a spectral unmixing function, for the removal of the autofluorescent unspecific signal, was available for the analysis of the patient stainings. The molecular subtype for each patient primary tumor has been determined by Dr. Christian Eisen (HI-STEM), even for those samples from which no cell line could be stabilized *in vitro*.

CDH17 stainings of the original tumors revealed a clear signal in the respective patient material (PACO14, PACO18) with areas that could be interpreted as membrane restricted stainings. However, high back ground signals could be detected especially in the patient tumors from which the QM tumors had been derived (**Figure 38**). Due to lack of primary material not all the proper negative controls could be performed.

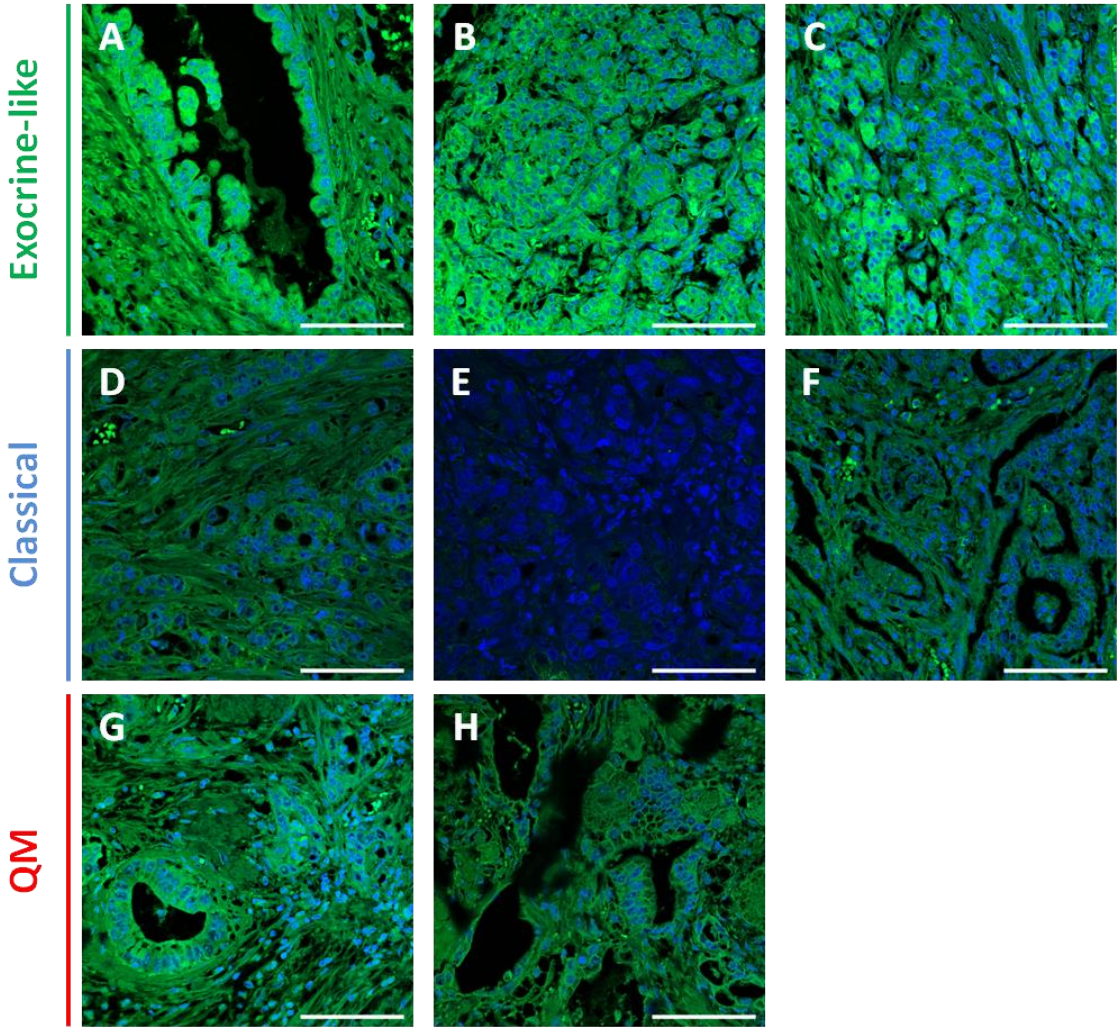


Figure 38: CDH17 stainings on patient original tumors. For simplification purposes, sample names were correlated to the PACO cell lines used in the proteomics investigation. **A**: PACO14; **B**: PACO18; **C**: Putative exocrine patient; **D**: PACO2; **E**: PACO17; **F**: PACO19; **G**: PACO7; **H**: PACO16. (Scale bars: 50 µm).

LGALS4 staining profiles were less obvious for the patient material, compared to those described for its putative partner of interaction. As shown in **Figure 39**, patient samples from which PACO14, PACO19 and PACO16 samples were derived appeared to express medium and high levels of LGALS4, matching the mouse derived xenografts staining profiles previously presented. However, the nonspecific background signal observed in the isotype control staining renders the delineation of a clear conclusion impossible. Especially in the case of the QM patient from which the PACO7 cell line was derived, the background fluorescence was almost identical in intensity to that detected in the stained sample.

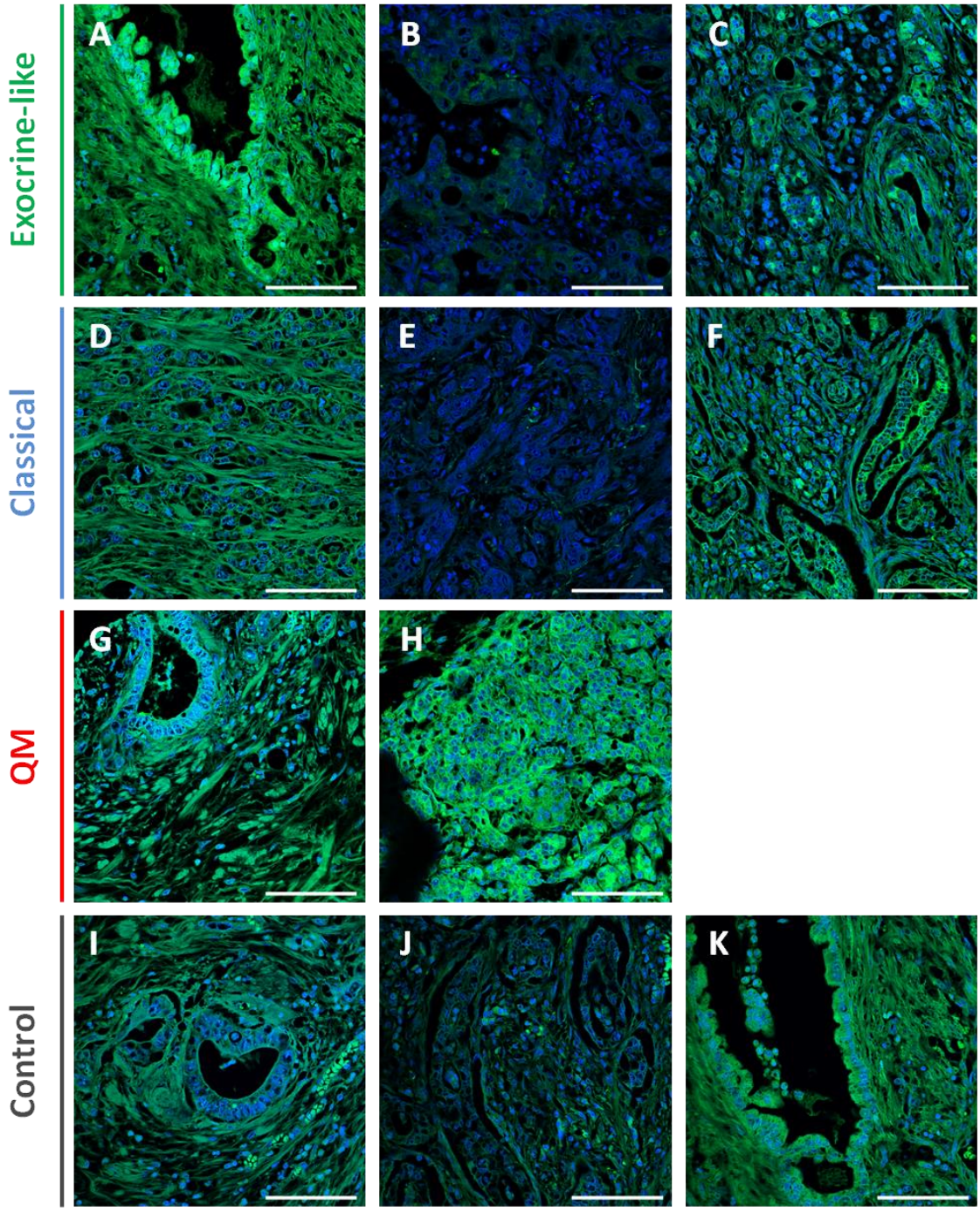


Figure 39: LGALS4 stainings on patient original tumors: Samples names were correlated to the PACO cell lines which had been stabilized in culture. **A**: PACO14; **B**: PACO18; **C**: Putative exocrine patient; **D**: PACO2; **E**: PACO17; **F**: PACO19; **G**: PACO7; **H**: PACO16; **I**: PACO7 isotype control; **J**: PACO19 Isotype control; **K**: PACO14 isotype control. (Scale bars: 50 µm).

In the case of putative pan-PDAC marker **LCN2**, the most intense staining could be observed in patients diagnosed with exocrine-like PDAC (**Figure 40**). A weaker LCN2 signal appeared to be present in the QM patients, while for the classical pancreatic cancerous lesions the back ground intensity did not allow the drawing of any relevant conclusion.

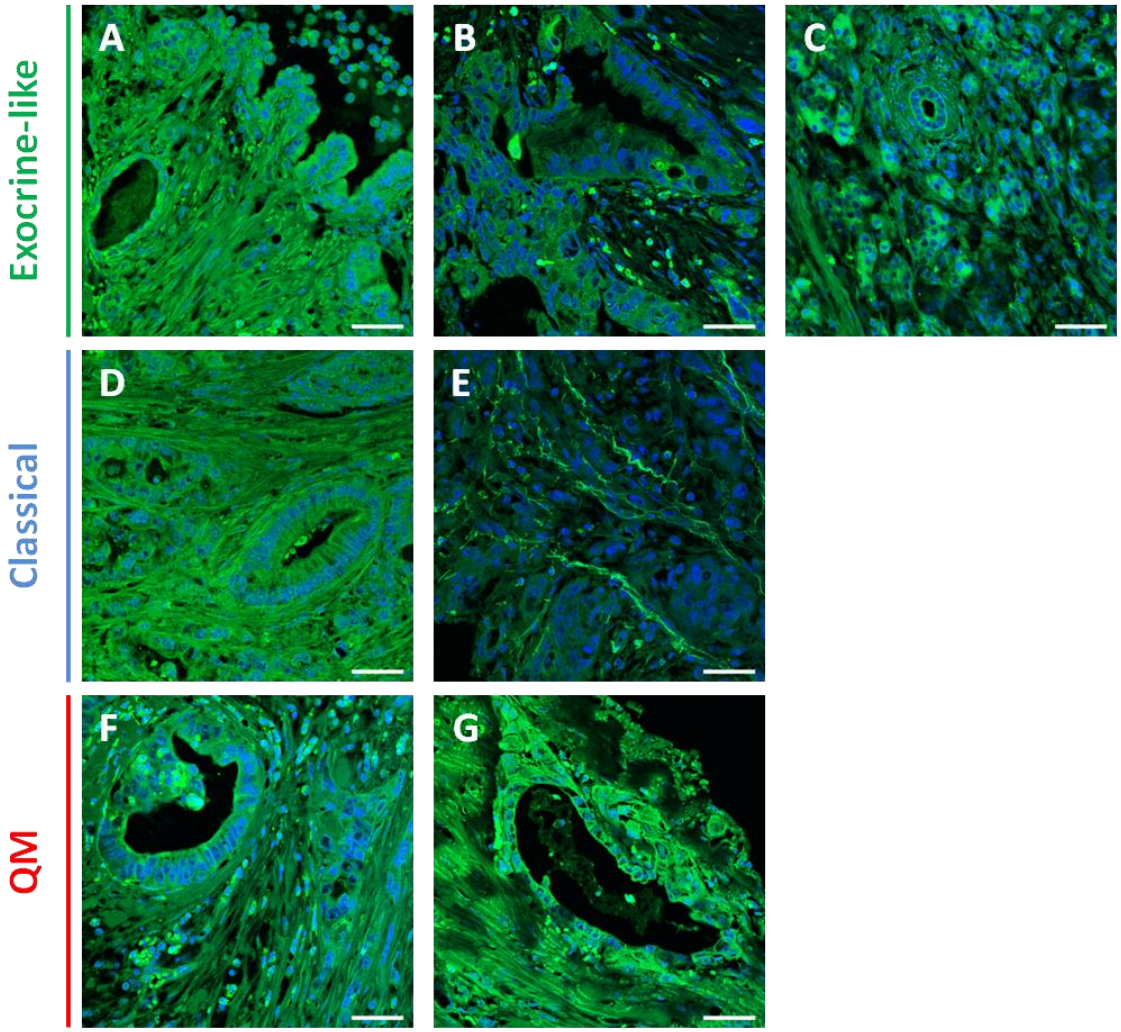


Figure 40: LCN2 stainings on patient original tumors. Samples names were correlated to the PACO cell lines used in the proteomics investigation. **A**: PACO14; **B**: PACO18; **C**: Putative exocrine patient; **D**: PACO2; **E**: PACO17; **F**: PACO7; **G**: PACO16. (Scale bars: 50 µm).

5.2.2 Antibody independent methods:

Pan-PDAC and exocrine-like subtype biomarker candidates were additionally validated by antibody independent methods. First, the markers of interest were quantified by MRM targeted proteomics, using species specific, proteotypic peptides. Second, the gene expression level for the putative biomarkers was evaluated by RT-qPCR, both *in vitro* and *in vivo*, using either total tumor lysate or EpCAM+ tumor cells.

5.2.2.1 In vitro validation and quantification of exocrine-specific biomarkers using MRM

We performed validations and quantifications for the two putative exocrine-like markers *in vitro* using MRM. The method combines the advantages of a high sensitivity with the possibility of performing the analysis of species specific proteotypic peptides [54]. Whole cell lysates prepared using the RIPA buffer and supernatants collected after 48 hour incubation and cleared by ultracentrifugation were digested with trypsin and processed as detailed in the materials and methods chapter. The selection of proteotypic precursor peptides was performed using the in house developed MS_QBAT software (courtesy of Alexander Kerner and Dr. Christoph Rösli, HI-STEM), and the open source software Skyline. For the relative quantification of exocrine-like protein marker candidates, we added heavy labeled peptides corresponding to each of the selected precursor peptides. The raw data were normalized by dividing the relative protein intensity, as determined after the SRM analysis, to the average intensity value of exocrine-like samples. The normalized relative protein expressions are presented in **Figure 41.**

As it can be observed in **Figure 41**, we could detect a clear upregulation of both putative exocrine-like markers CDH17 and LGALS4 in the exocrine-like cell lines compared to the classical, QM and control cell lines when analyzing the whole cell proteome. The expression difference for the two putative exocrine-like markers in the total cell proteome was statistically significant, as determined by the Mann-Whitney two-tailed test (p value < 0.05). An upregulation of secreted CDH17 and LGALS4 could also be reported for the secretome samples, albeit the results were not statistically significant due to the low number of analyzed samples.

When investigating the relative expression levels of CDH17, the highest values could be reported in cell lines PACO3, PACO10 and PACO14, with PACO18 expressing significantly lower levels. The secreted protein levels reflected the profile of the total cell proteome, with PACO10 cell medium representing the most abundant source of secreted/shed CDH17.

LGALS4 expression was the highest in the cell lines PACO18 and PACO14, followed by PACO10 and PACO3. The relative amount of secreted LGALS4 could (as for CDH17) be correlated with the relative expression levels detected in the total cell proteome (**Figure 41**).

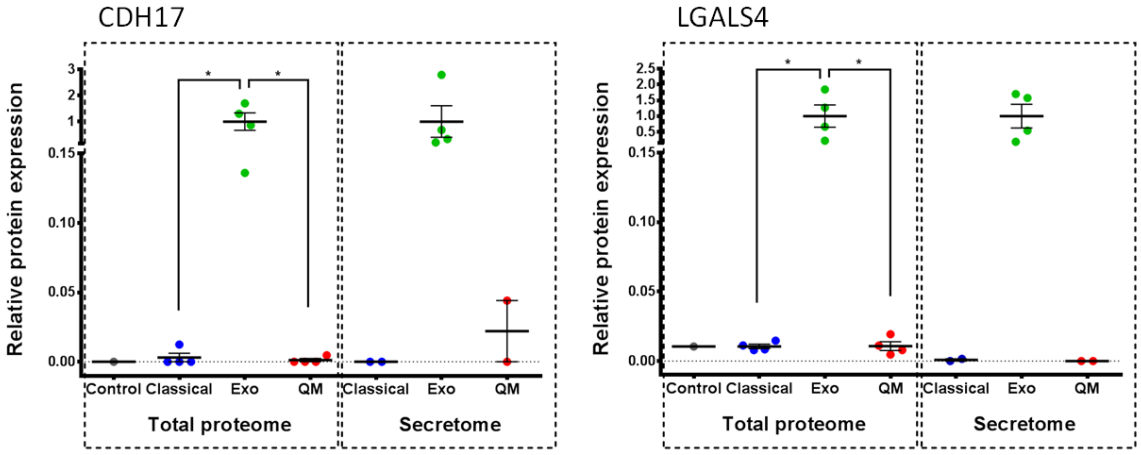


Figure 41: *In vitro* evaluation of protein expression levels for the exocrine-like putative markers CDH17 and LGALS4 by SRM. The analysis was performed on total cell lysates and secretome of cultured PACO and healthy control cell lines. Error bars represent the standard error of mean (SEM), calculated with the Prism software.

5.2.2.2 Evaluation of gene expression levels for the markers of interest using RT-qPCR

After validating the presence of the markers of interest on *in vitro* cultured cells and mouse xenografts, we wanted to evaluate if the upregulation of the putative pan-PDAC and exocrine-like markers was due to changes in gene expression, or a result of a post-transcriptional event. For this purpose, we performed RT-qPCR using both *in vitro* cultured cells, as well as xenograft samples. Primers were designed using the NetPrimer software. After aligning the human and mouse gene, we selected oligonucleotides that bound human specific sequences in the gene to be quantified. The primer pair for the RPL13a human housekeeping gene was provided by Dr. Arnaud Descot (HI-STEM, Metastatic Niches group) and we did not investigate its cross-binding to the mouse equivalent mRNA. Additionally, whenever possible, we designed primers that would amplify multi-exonic sequences, thus allowing the identification of any potential genomic DNA contamination of the RNA sample prior to reverse transcription, based on the size of the resulting amplicon.

5.2.2.2.1 In vitro evaluation of gene expression levels for pan-PDAC and exocrine-like biomarker candidates using RT-qPCR

For the *in vitro* evaluation of the mRNA levels, cells were grown on Primaria 6-well plates, lysed and total RNA was extracted using the RNeasy kit. cDNA amplification was performed using the SuperScript Vilo cDNA kit. 10ng cDNA were used for the RT-qPCR reaction, performed in triplicates, using biological duplicates of each cell line. GAPDH and RPL13a were used as reference genes for the relative quantification performed by the Viia7 1.0 software from Thermo.

Quantification of the gene expression levels for the two putative exocrine-like markers revealed a major upregulation of CDH17 and LGALS4 mRNA levels in the cell lines belonging to the exocrine-like subtype (**Figure 42**), compared to either control, classical or QM PDAC cells. CDH17 mRNA levels were clearly upregulated in all exocrine-like cell lines, compared to all other subtypes and the healthy controls. We could however observe that some QM PACO cell lines expressed higher levels of the LGALS4 mRNA, compared to classical PDAC and control cell lines. Nonetheless, the levels detected in these QM PACO cell lines were inferior to those reported for the exocrine-like cell lines. The increased gene expression reported was statistically significant, as shown by the results of the two tailed Mann-Whitney test (** $p < 0.01$; *** $p < 0.001$).

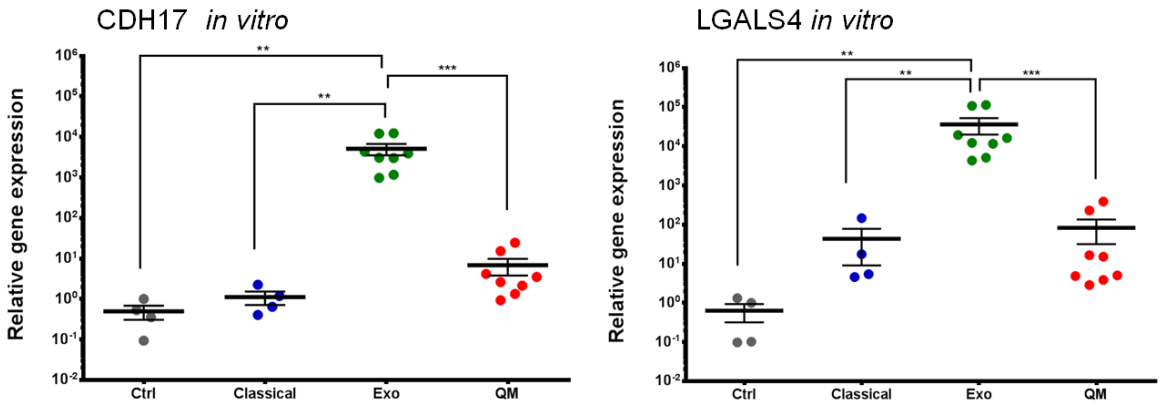


Figure 42: *In vitro* quantification of CDH17 and LGALS4 mRNA levels by RT-qPCR. Statistical relevance was evaluated with the two tailed Mann-Whitney test (** p< 0.01; *** p< 0.001). Error bars represent the standard error of mean (SEM), calculated with the Prism software.

For the pan-PDAC marker candidates investigated we observed a significant upregulation of the corresponding mRNA levels in all PDAC subtypes compared to the healthy controls (**Figure 43**). In both cases, we could report a difference in the mRNA levels of individual healthy control cell lines as well: HPDE cell lines expressed higher mRNA levels of both pan-PDAC marker candidates compared to the HPNE cell line. Also, for LCN2, we could observe a similar level of mRNA expression in the QM cell line PACO7 compared to the control cell line HPDE. However the increase in the gene expression level was significant, as indicated by the two tailed Mann-Whitney test (* p< 0.05; ** p< 0.01).

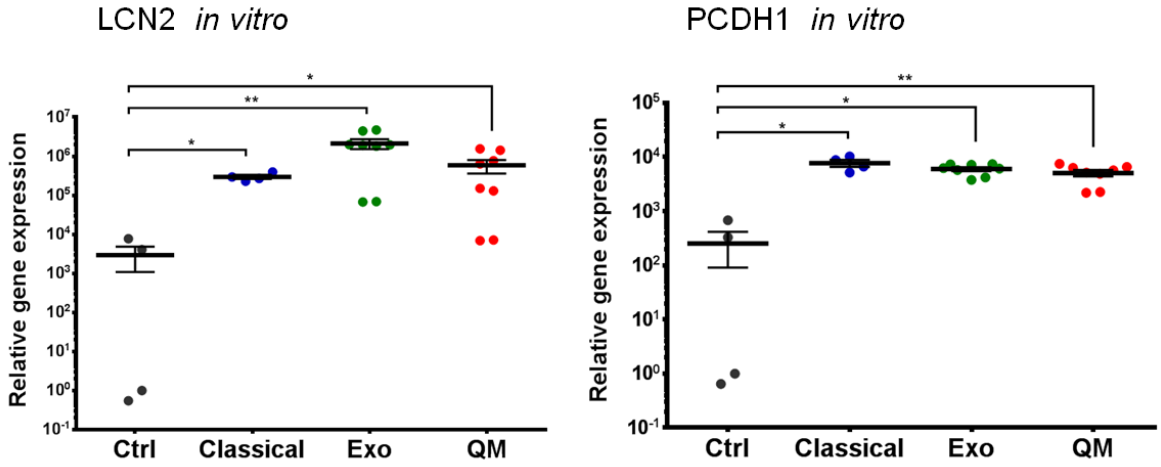


Figure 43: *In vitro* quantification of pan-PDAC marker candidates LCN2 and PCDH1 by RT-qPCR. Statistical relevance was evaluated with the two tailed Mann-Whitney test (** p < 0.01; *** p < 0.001). Error bars represent the standard error of mean (SEM), calculated with the Prism software.

5.2.2.2.2 In vivo evaluation of gene expression levels for pan-PDAC and exocrine-like biomarker candidates using RT-qPCR on total tumor tissue and EpCAM+ cells

We evaluated the gene expression levels of the four biomarker candidates *in vivo* by analyzing both the mRNA levels in EpCAM+ sorted cells and in total tumor tissue. The relative gene expression quantification was performed again using the two house-keeping genes previously mentioned (GAPDH and RPL13a). EpCAM+ cell sorting was performed using anti-human CD326 magnetic beads from Miltenyi. Total tumor tissue was lysed and RNA extraction was performed as detailed in *Materials and Methods*.

In the case of the two putative exocrine-like markers CDH17 and LGALS4, the gene expression levels confirmed the observations of the previous *in vivo* IF stainings. Both genes were up-regulated in classical and exocrine-like xenografts compared to QM derived tumors (**Figure 44**). The up-regulation was comparable in both EpCAM+ isolated cells and in total tumor mRNA.

The highest CDH17 expression could be detected in PACO19 (classical), and exocrine-like PACO14, PACO28 and PACO10 EpCAM+ cells. Total tumor tissue analysis revealed similar CDH17 gene expression patterns as for the EpCAM+ cell analysis, with PACO19, PACO3, PACO10 and PACO28 derived tumors having the highest levels of CDH17 mRNA. When comparing the expression level of CDH17 in classical and exocrine-like pancreatic tumors, only two classical cell lines appeared to express similar mRNA levels to the exocrine-like cells – PACO19 and occasionally PACO2. PACO20 originating tumors (classical subtype) expressed very low CDH17 mRNA levels, similar to many QM xenografts. The *in vivo* gene analysis experiments confirmed PACO18 as a low expressing CDH17 cell line, with PACO10 and the new cell line PACO28 at the opposite end of the spectrum. The commercially available total RNA extracted from healthy human pancreas revealed CDH17 mRNA levels similar to those detected in QM derived tumors.

Overall, the LGALS4 gene expression was comparable in all exocrine-like derived tumors, with a lower relative expression detected only in PACO3 cells. Classical tumors expressed typically inferior levels compared to the above mentioned exocrine-like samples, except for PACO19 derived xenografts. LGALS4 levels were the lowest in QM tumors and in human healthy pancreatic total RNA. However, PACO16 QM derived tumors occasionally appear to up-regulate LGALS4 mRNA, as observed in the course of the *in vivo* experiments (see **Figure 33**).

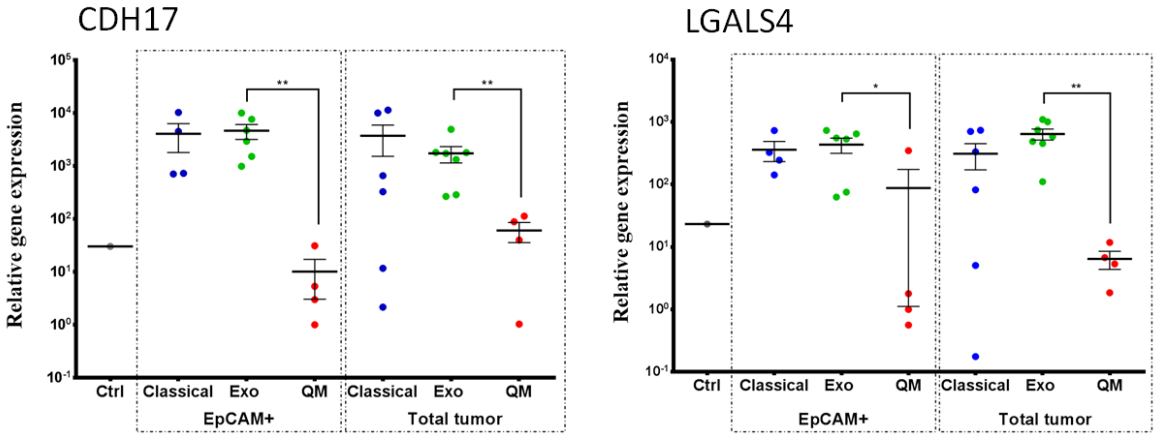


Figure 44: Relative gene expression of CDH17 and LGALS4 exocrine-like marker candidate. The analysis was performed on total RNA extracted from EpCAM+ sorted cells, or from whole tumor tissue. Error bars represent SEM. Statistical relevance was evaluated with the two tailed Mann-Whitney test (* $p < 0.05$; ** $p < 0.01$).

The mRNA levels for the two pan-PDAC marker candidates – LCN2 and PCDH1 – were clearly elevated in all the PACO derived tumor xenografts compared to the healthy pancreatic tissue, as shown in **Figure 45**. The overall expression intensity was similar for all the analyzed tumors and for the corresponding EpCAM+ isolated cells. In the case of LCN2, a moderate increased expression could be detected in the exocrine-like derived EpCAM+ cells, compared to the classical and the QM derived epithelial tumor cells. The highest LCN2 expressing tumors *in vivo* were PACO14 derived xenografts, while the lowest levels could be detected in QM derived xenografts PACO7 and PACO8. PACO2, PACO28 and PACO14 derived tumors were expressing some of the highest PCDH1 mRNA levels.

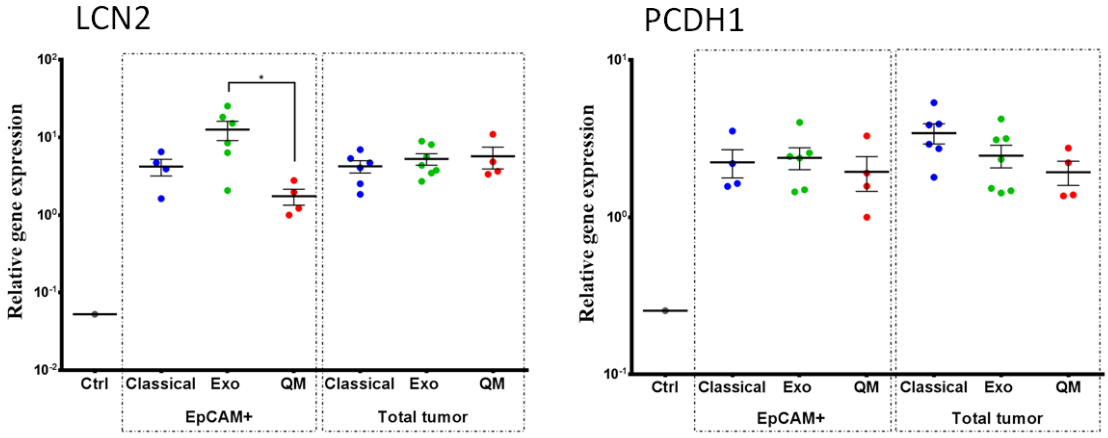


Figure 45: Relative expression of LCN 2 and PCDH1 pan-PDAC marker candidates. The analysis was performed on total RNA extracted from EpCAM+ sorted cells and from whole tumor tissue. Error bars represent SEM. Statistical relevance was evaluated with the two tailed Mann-Whitney test (* $p < 0.05$).

6. Discussions

In summary, the presented study lead to the identification and initial validation of novel cell surface and secreted pan-PDAC as well as subtype specific protein biomarker candidates, which can prime the development of new diagnostic and/or therapeutic clinical applications.

The discovery experiments made use of a panel of patient derived primary pancreatic cancer cell lines (PACOs), grown in a chemically defined cancer stem cell culture medium. These cell lines, developed by Dr. Christian Eisen (HI-STEM, DKFZ), maintain the transcriptomic profile reported in the original patient tumor, even after serial passages, and they recapitulate all three PDAC molecular subtypes described in literature – namely exocrine-like, classical and quasimesenchymal. Additionally, when injected in the pancreas of immunodeficient mice, the cell lines give rise to tumors histologically similar to the original patient tumor.

The initial proteomic discovery experiments focused on the cell surface proteome of pancreatic cancer cells, aiming to identify novel candidates for future antibody based targeted therapies. The plasma membrane associated proteins of cultured cells were enriched after performing *in vitro* biotinylation, followed by streptavidin pull down of biotin-labeled proteins and subsequent mass spectrometric analysis. This proteomic discovery experiment allowed the identification of proteins exclusively expressed on the surface of all/the majority of PACO cell lines, but completely absent in the healthy controls. Furthermore, proteins defining distinct molecular subtypes could also be identified. We selected the two pan-PDAC protein marker candidates PCDH1 and LCN2 for further investigations. Concerning individual molecular subtypes, the exocrine-like PDAC cell lines stood out, expressing several subtype specific proteins, of which we selected two of the top hits (CDH17 and LGALS4) for further validations. No definite classical or quasimesenchymal specific protein biomarkers could be defined after the *in vitro* proteome analysis.

In addition, we investigated the secreted proteome of cultured PDAC and healthy control cell lines using shotgun proteomics. The secretome identification experiments indicated that the four biomarker candidates selected based on the cell surface proteome analysis were as well secreted by the cancer cells in the environment, and could therefore be considered for the development of novel diagnostic tools.

Finally, vascular accessible proteins present on the surface of pancreatic cancer cells and the associated tumor stroma were investigated upon performing whole body perfusion of immunodeficient mice harboring PACO-derived tumor xenografts with a reactive biotin ester. A proof-of-principle experiment confirmed the presence of two of the putative biomarkers identified after the *in vitro* experiments (LGALS4 and LCN2) in the primary pancreatic tumors. In addition, the *in vivo* biotinylation experiments indicated that the two protein biomarkers are accessible from the blood stream, and could represent candidates for novel targeted PDAC therapies, such as antibody-based therapies.

The presence of the four biomarker candidates in the corresponding cell lines was validated first *in vitro*, by employing both antibody based (IF, WB) and antibody independent (SRM, RT-qPCR) methods. IF experiments confirmed the presence of both PCDH1 and LCN2 on the surface of all cultured PDAC cell lines, while they could not be detected at the plasma membrane level of the healthy control cell lines. The initial mass spectrometric identification

revealed the presence of isoform 2 specific peptides for PCDH1 in several of the analyzed PACO cell lines, although the exclusive presence of isoform 2 in the cultured PDAC cell lines could not be clearly confirmed. Additionally, RT-qPCR experiments showed a clear increase in the corresponding mRNA levels of both candidates in all the PDAC cancer cell lines, compared to the control cell lines. Therefore, it appears as if the increased PCDH1 and LCN2 expression in PDAC cells is the result of transcriptional activation of the corresponding genes. Regarding the two subtype specific biomarker candidates CDH17 and LGALS4, IF could confirm their exclusive presence at the plasma membrane level of the corresponding exocrine-like cell lines. LGALS4 was also detected in the cytoplasm of exocrine-like PDAC cells. Moreover, IF experiments suggested that the two exocrine-like biomarker candidates might represent interaction partners, a finding which has been confirmed with appropriate methods. CDH17 and LGALS4 protein levels were also investigated using SRM, which could additionally validate their exclusive presence in the whole cell lysates and in the secretome of exocrine-like cell lines. The corresponding mRNA levels were also upregulated only in those cell lines belonging to the exocrine-like subtype, as detected by RT-qPCR. Therefore, CDH17 and LGALS4 protein expression in exocrine-like PDAC cell lines is likely the result of transcriptional activation of the corresponding genes.

Additionally, we investigated the presence of the four biomarker candidates in pancreatic tumor xenografts developed in immunodeficient mice and in the original patient tumors. *In vivo* IF experiments could confirm the presence of the two pan-PDAC protein biomarker candidates in the extracellular environment and at the plasma membrane level of cancer cells, while being completely absent from healthy human pancreata. The protein expression of PCDH1 and LCN2 in mouse xenografts could be associated with increased levels of the corresponding mRNAs in the tumor xenografts, compared to the healthy control. CDH17 and LGALS4 association with the plasma membrane of exocrine-like cancer cells was also confirmed upon *in vivo* IF analysis. However, in contrast to the *in vitro* validation experiments, we could additionally detect the two biomarker candidates in some of the classical-derived tumor xenografts, although at levels usually inferior to those reported for the exocrine-like tumors. CDH17 was completely absent from healthy human pancreata. On the other hand, some LGALS4 expression could be identified in the normal appearing pancreatic tissue employed for the IF validation experiments. Since the normal human pancreatic tissue was isolated from macroscopically healthy areas of tumor bearing patients (as evaluated at the time of surgery), we could not clearly determine if the LGALS4 present in the selected healthy tissue was not in fact associated with undetected malignant lesions expanding beyond the apparent tumor boarder. Elevated levels of CDH17 and LGALS4 mRNAs could be reported in the exocrine-like and some of the classical tumor xenografts. The four protein marker candidates were also detected in the corresponding original patient tumors, although the quality of the paraffin embedded material prohibited extended validation experiments.

Recently, Lisa Becker, a master student in our lab, generated scFv antibody fragments against the two exocrine-like protein candidates CDH17 and LGALS4 using phage display technology. Initial validation experiments verified the ability of the selected antibodies to recognize the native proteins in ELISA experiments. Furthermore, the scFv fragments could also be employed for fluorescent detection of the corresponding proteins on formalin fixed/paraffin embedded samples. The selected antibodies can be further developed for diagnostic applications and/or novel PDAC immunotherapeutic approaches.

6.1 Mass spectrometric based analysis of the cell surface proteome and secretome of cultured cells reveal novel putative PDAC markers, with the potential to distinguish between molecular subtypes

The study presented in this PhD thesis lead to the discovery and validation of novel cell surface and secreted biomarkers with potential future clinical applications for the diagnosis and treatment of pancreatic cancer. The high mortality and metastasis rate of the PDAC [5, 41], combined with its notoriety for being one of the most resistant tumors to both chemo-therapy and radio-therapy, has increased the demand for the development of accurate diagnostic tests as well as targeted therapies [7, 25].

Starting from an *in vitro* disease system, we employed mass spectrometric based approaches to compare the cell surface proteome and the secretome of pancreatic cancer cells to the one of healthy control cell lines. The comparative proteomic approach made use of twelve primary patient derived pancreatic cancer cell lines, grown in serum-free conditions, developed by Dr. Christian Eisen (HI-STEM, DKFZ). These cell lines, termed PACOs, offer significant advantages over conventional serum dependent cell lines. First of all, for commercially available cell lines grown for multiple passages in FCS based medium, genetic alterations are often documented [140]. Second of all, FCS dependent cell lines have often been reported to diverge both from the original phenotype and the molecular characteristics of the original tumor, a phenomenon which aggravates with increasing number of passages [46, 140]. The transcriptomic profile of the original patient tumors matched that of tumor xenografts derived after injecting the corresponding PACO cell lines into immunodeficient mice (Pearson correlation coefficient > 0.8) [141]. Furthermore, the orthotopic mouse tumor generated by these cell lines showed histological features and molecular characteristics closely resembling those of the original malignant neoplasm (Noll, Eisen *et al*., manuscript in preparation). Thus, the serum-free PACO cell culture model is less likely to lead to the discovery of artifactual protein biomarkers, while at the same time offering the advantages of an easy to manipulate *in vitro* system. Additionally, Collisson *et al*. recently argued that PDAC tumors, far from reflecting a homogeneous disease, can be subclassified based on their transcriptomic profiling into three molecular subtypes: classical, exocrine-like and QM. The individual subtypes showed different responses to conventional therapies and could be correlated with distinct overall survivals [3, 141]. The PACO cell lines represent to our knowledge the only *in vitro* PDAC model to recapitulate all three molecular subtypes described by Collisson *et al*. (Noll, Eisen *et al*., manuscript in preparation).

The dynamic range of protein concentrations in complex biological raw samples exceeds the detection capabilities of current MS instruments. For this reason, extensive sample subfractionation and/or enrichment of subcellular compartments is a prerequisite for an efficient proteomic discovery experiment [72]. In this study, novel cell surface markers for future diagnostics and targeted therapies were identified by a proteomics approach utilizing *in vitro* biotinylation of cell surface proteins, followed by streptavidin pull down of the tagged subproteome. Using this technique, we identified more than 2500 proteins with a minimum of two proteotypic peptides. GO enrichment analysis revealed that the subcellular fractionation strategy has successfully lead to the detection of a significant number of

proteins annotated to the plasma membrane and the ECM (**Figure 11**). The list also included proteins annotated to intracellular membranous structures, such as ER proteins. Although such intracellular hits are usually discarded as contaminants, it is important to keep in mind that cellular re-localization is an often described event in human malignancies [29]. Thus, ER membrane proteins often end up being expressed on the surface of cancer cells, making them interesting targets for antibody based therapies, since they would normally be inaccessible in healthy tissues [29].

However, the list of proteins identified was not devoid of intracellular proteins, without membranous fraction annotations, such as ribosomal proteins, housekeeping proteins involved in glucose metabolism, components of the cytoskeleton and biotin carboxylases. Biotin carboxylases are enzymes known to require biotin as a cofactor [135]. Therefore, the identification of these proteins with a high number of peptides in non-biotinylated negative controls has to be expected. The remaining intracellular proteins have either unspecifically bound to the streptavidin column, or derive from intracellular proteins which have been released into the culture medium (e.g. by apoptotic cells). Additionally, despite the presence of an electrical charge on the sulfo-NHS moiety of the reactive biotin ester employed, some of the biotinylation reagent might still be internalized, especially by biotin transporters located on the plasma membrane. This can therefore lead to unspecific biotinylation of intracellular proteins [73, 142].

We interrogated the list of cell surface proteins for both universal PDAC biomarkers, as well as subtype specific markers, which could be used for the clinical subclassification of pancreatic cancer patients.

Out of the list of proteins expressed on PACO cell lines, but completely absent in the negative controls (**Table 6**), we selected two putative pan-PDAC markers for further validations: PCDH1 and LCN2. PCDH1 has never been associated with malignancies prior to our study and is typically expressed in the adult central nervous system and in the respiratory tract, at the bronchi level [118, 119]. LCN2 on the other hand, has been described as a tumor associated marker for many malignancies, including pancreatic cancer [124]. It has however captured our attention, since it is a secreted protein, which could be employed for the development of a blood based diagnostic test for PDAC, optionally in a panel with other pancreatic cancer biomarkers.

When searching for subtype specific biomarker candidates, we observed that the exocrine-like PDAC cell lines could clearly be separated from the ones representing the other two subtypes, expressing several unique proteins (**Table 7**). Two of them, namely – CDH17 and LGALS4 – were identified in every patient matched cell line belonging to the exocrine-like subtype, with a significant number of peptides. Therefore we selected the two proteins for further validation. The two proteins are expressed in healthy tissues of the digestive tract, but are largely absent from auxiliary glands, such as the liver and the pancreas [96, 110]. Their expression pattern is often altered in malignant development, being associated with both aggressive cancer behavior, as well as more differentiated neoplasms. Their role in cancer progression appears to be highly dependent on the tissue of origin [96, 110].

We could not identify any protein that could clearly distinguish the classical and QM molecular subtypes, as the proteins identified were often not present on all the patient-matched cell lines belonging to one subtype, or they were detected in more than one subtype (**Table 8**). Since the purpose of our investigation was the identification of true subtype specific biomarkers, and not protein markers reflecting an individual patient, we

excluded from validation experiments those proteins who could not be detected in all patient-matched cell lines belonging to one PDAC subtype. We have considered validating a marker defining two molecular subtypes, STEAP1, being present on the surface of cell lines belonging to the exocrine-like and the classical subtypes. The validation data for this marker are however not presented in this report, as they require additional optimizations.

The reason for the differences observed in the protein expression profiles between the exocrine-like and the other two molecular subtypes is unknown. One possible explanation could be that the different subtypes arise from different cells of origin. Both ductal and acinar cells (undergoing acinar to ductal metaplasia) have been hypothesized as the putative source of exocrine pancreatic neoplasms [143, 144]. However, no definite conclusion has been reached so far in the debate regarding the PDAC cell of origin [143].

We investigated as well if the putative protein biomarkers, identified in the course of the cell surface proteome analysis, could be employed as potential diagnostic markers in non-invasive screening protocols. In order for a marker to be eligible for such diagnostic purposes, it needs to be secreted or shed from the plasma membrane of cancer cell in amounts that would allow it's identification in body fluids of patients, preferably in blood [32]. One suggested way of detecting putative secreted biomarkers is the investigation of the supernatant of cultured cells, and its comparison to that of healthy counterparts [46]. For this purpose, we optimized the serum-free CSC medium in which the PACO cells were grown, by removing albumin and transferrin from the formulation, and incubated the cells with the new medium, for 48 hours. Albumin and transferrin subtraction from the CSC medium was dictated by the need to improve the dynamic range of our discovery experiments, as the high concentration of the two proteins would prohibit proper identification of less abundant secreted proteins.

The secretome analysis lead to the identification of more than 1700 proteins (excluding one hit wonders). Gene ontology analysis confirmed the enrichment of proteins with annotations containing the terms extracellular, ECM, plasma membrane, as well as vesicular systems in the list of secreted proteins (**Figure 15a**). We had previously reported that the GO enrichment analysis of the cell surface proteome revealed indistinguishable profiles between the three molecular subtypes. However, when enquiring the secretome profiles using GO annotations, we could observe some discrete, yet significant differences between the distinct PDAC subtypes (**Figure 15b**). Exocrine-like supernatants were enriched in proteins present in adherens junctions, whereas classical and QM secretomes were more likely to express proteins associated with the basement membrane. This observation could be complemented by the fact that the exocrine-like cell surface and secreted proteomes included multiple members of the cadherin family (**Table 7** and **Table 10**), known to be involved in the formation of adherens junctions [94]. Furthermore, we could distinguish several intracellular constituents among the proteins identified in the supernatant of cultured cells. Generally the cytoplasmic and organelle specific hits would be discarded from future validation experiments of putative secreted biomarkers. However, recent reports have indicated that unconventional secretion might be a common event in tumors, and the extracellular localization of typically intracellular proteins has been confirmed through IHC experiments [81]. Thus future validation experiments do not need to automatically exclude proteins with a GO annotation suggesting their intracellular localization.

It can be speculated if the cultured pancreatic cancer cells still express digestive enzymes specific to the exocrine pancreas (**Table 9**) and other proteases, potentially leading to the

release of non-tryptic peptides. Indeed, after investigating the non-tryptic peptides identified by the ProteinPilot software with higher than 95% confidence, we could determine that the percentage of non-tryptic peptides detected in all the secretome samples was almost double to the one detected for the *in vitro* biotinylated samples (7.6% versus 4.2%) (data not shown). Interestingly, the protein coverage of the secretome samples has been inferior to that observed in the cell surface proteome (data not shown). One possible explanation could be provided by the pre-digestion of the secreted proteins during the 48 hours incubation period, which is known to negatively affect the downstream proteomic analysis [84, 145].

Comparative analysis of PACO versus healthy control collected supernatants confirmed the presence of the previously selected putative pan-PDAC and exocrine-like protein biomarkers in the secretome of cultured cells (**Table 9** and **Table 10**). LCN2, which is constitutively released in the extracellular environment [129], was identified as expected with a large number of peptides in the supernatants of all PACO cell lines, while being completely absent in the secretome of healthy controls. PCDH1 release in the extracellular fraction was confirmed in more than 60% of the analyzed samples, despite being described in the literature as a membrane bound protein [119]. Its presence in the secretome could be the result of the protease based shaving of extracellular domains of membrane proteins, or unspecific release of plasma membrane proteins upon cell lysis. Indeed, PCDH1 peptides identified with a high confidence in the secretome samples mapped to the N-terminal extracellular domain, but not to the C-terminal cytoplasmic tail (data not shown). However, we could not exclude the contribution of potentially apoptotic cells to the secreted proteome. Future cell viability assays should therefore be performed, in order to evaluate the extent to which cell lysis products contribute to the overall secretome composition. Both CDH17 and LGALS4 were detected exclusively in the secretome of exocrine-like cells, thus confirming the initial cell surface discovery experiments (**Table 10**). LGALS4 was identified in more than half of the exocrine-like supernatant samples, whereas CDH17 signals were only detected in the secreted fraction of those cell lines with a high expression on the cell surface (PACO10 cells, **Supplementary Information**). These observation are in agreement with literature data, which confirm LGALS4 secretion in the extracellular environment through unconventional pathways [114]. The identification of CDH17 in the supernatant of exocrine-like cells could again be the result of protease mediated shaving, or unspecific lysis of cultured cells.

The chemically defined, serum free CSC medium used for the cultivation of PACO cells has the advantage that the necessary growth factors are provided at known concentration. By contrast, commercially available cell lines require serum supplements, with unknown chemical composition. Hence, we expect that removal of albumin from the PACO culture medium would not have an effect as severe as FCS withdrawal for commercially available cells (which is known to induce apoptosis and cell cycle arrest [81]) and would therefore not compromise the quality of the proteomic discovery. On the other hand, the two control cell lines were initially cultured in medium containing either added FCS or bovine pituitary extract. The medium composition optimal for the HPNE cell line does not vary significantly from the one of the CSC PACO medium. Consequently, we did not expect major metabolic and transcriptomic alterations when culturing the HPNE cells in the new CSC secretome medium. For the HPDE cell line however, the medium formulation varied substantially from the PACO standard culture medium. We are not sure if the morphological changes observed under the microscope (**Figure 12**) were induced by differences in the chemical composition of the advanced DMEM/F12 versus the keratinocyte based medium, or if they

were the result of bovine pituitary extract removal, which might have provided additional essential hormones and factors for this immortalized pancreatic cell line.

Following the completion of the discovery phase, we proceeded to validate the two pan-PDAC and exocrine-like protein biomarkers candidates, by using antibody based and antibody independent methods, *in vitro* as well as *in vivo*.

6.2 Validation of exocrine-like protein biomarker candidates

6.2.1 In vitro validations

We investigated the presence of CDH17 and LGALS4 on the cell surface and in the secretome of PACO cultured cells by means of both antibody based and antibody independent methods.

IF *in vitro* experiments confirmed the exclusive expression of the two proteins on the cell surface of all patient matched exocrine-like cell lines (**Figure 22** and **Figure 23**). As expected, cytoplasmic as well as limited nuclear staining could be reported for LGALS4 (**Figure 23**) [107]. The intensity of the staining profiles for the two markers varied between cell lines. CDH17 staining was most intense in the PACO10 cell line, while the weakest staining could be observed in PACO18 cells. LGALS4 staining intensity was the highest in PACO14 and PACO18 samples. Occasionally we could observe cytoplasmic fluorescence in some of the classical and QM cell lines. We estimated that the signal could be the result of autofluorescence, weak unspecific binding of the antibodies employed for the experiment, or might indicate a potential cytoplasmic localization for the protein of interest in some of the classical and QM PDAC cell lines. Isotype control stainings indicated that for some of the QM annotated cell lines, an unspecific cytoplasmic signal could indeed be detected (**Supplementary Figure 2**). However, not all classical and QM PACO cell lines revealed unspecific fluorescent signals in the respective isotype controls. Therefore, we could not exclude a potential cytoplasmic localization of the two biomarker candidates in some of the non-exocrine-like cell lines, prior to performing additional validation experiments.

In order to further validate the specificity of the two exocrine-like biomarker candidates, we performed antibody independent validation experiments, using a targeted mass spectrometric approach, termed SRM.

SRM experiments of total cell lysates and supernatants confirmed the exclusive presence of the two protein biomarker candidates in exocrine-like PDAC cells (**Figure 41**). Moreover, the targeted MS approach lead to the conclusion that the cytoplasmic staining detected in some of the classical/QM PACO cell lines was the result of unspecific background (**Figure 41** – absence of signals in whole cell lysates obtained from non exocrine-like cells). The presence of CDH17 and LGALS4 in the secreted fraction of exocrine-like cell lines (**Figure 41**, relative quantification achieved by SRM) makes the two proteins eligible for consideration as putative diagnostic biomarkers for patient stratification.

The upregulation of CDH17 and LGALS4 on the cell surface of exocrine-like PDAC cell lines appears to be (at least partially) dictated by transcriptional activation of the corresponding genes, according to the RT-qPCR quantification (**Figure 42**). The CDH17 gene expression profiles could be correlated with previous experimental data of Noll, Eisen *et al*., which have

identified the Hepatic Nuclear Factor 1-alpha (HNF1α) transcription factor as a marker for the exocrine-like PDAC subtype (Noll, Eisen *et al*., manuscript in preparation). HNF1α is one of the two transcription factors known to regulate the expression of CDH17 *in vivo* [99]. Moreover, the gene expression levels of the transcription factor correlated with the profiles observed for CDH17 mRNA levels in the exocrine-like cell lines (data not shown).

Tan *et al*. reported CDH17 and LGALS4 as markers describing the gastric carcinoma G-intestinal molecular subtype, correlated with a better response to 5-FU therapy and slightly improved overall survival compared to patients whose tumors did not express the two proteins [105]. Additionally, Takamura *et al*. identified galectin-3 as a putative interaction partner for CDH17 [95]. Galectin-3 appeared to bind to the extracellular domain of CDH17, more specifically to the carbohydrate residues of the N-glycosylated protein, and the interaction was inhibited by the presence of lactose [95]. In light of the two studies, we examined if the two putative exocrine-like biomarkers might interact at the plasma membrane level of PDAC cancer cells. Confocal IF costaining experiments revealed that CDH17 and LGALS4 indeed co-localize on the surface of exocrine-like PACO cells (**Figure 26**). Nonetheless, further investigations are required to confirm the interaction of the two proteins – such as co-IP (in the presence and absence of lactose solutions) or Fluorescence Resonance Energy Transfer (FRET) experiments. In FRET experiments, a donor chromophore (excited by a laser) can only transfer energy to an acceptor chromophore if the two light sensitive molecules are situated in close proximity. FRET is sensitive to even small changes in the distance between donor and acceptor molecule, thus making it a reliable technique for confirming direct interaction between two molecules [146, 147]. An additional method for detecting protein-protein interactions and pinpoint their subcellular localization is represented by the proximity ligation assay (PLA). The method combines antibody-based recognition of two protein targets with the amplification of a DNA reporter molecule, only possible if the two employed antibodies (containing attached oligonucleotides sequences) are located in close proximity. Ligated DNA molecules can be amplified via rolling circle amplification, and the corresponding products are visualized by hybridization [148, 149]. Nonetheless, to our knowledge, this is the first study ever to report a potential interaction between CDH17 and LGALS4.

6.2.2 In vivo validations

Next, we evaluated if the *in vitro* validated markers maintain their exclusive exocrine-like expression pattern in pancreatic tumors, when injected into the pancreas of immunodeficient NSG mice. Additionally, we also verified the expression levels of CDH17 and LGALS4 in the original patient tumors, when primary tissue material was available.

Tumor xenografts grown in NSG mice continued to express CDH17 and LGALS4 in exocrine-like malignancies, in patterns similar to those earlier described for the *in vitro* cultured cell lines. Thus, PACO10 CDH17 staining signal continued to be one of the most intense, whereas the PACO18 staining *in vivo* was once again very weak (**Figure 31**). In the case of LGALS4 staining, the signal was yet again highest in PACO18 samples. Moreover, the two exocrine-like marker candidates were also identified in novel exocrine-like tumor xenografts not included in the initial proteomics discovery experiments, thus further confirming their association with the exocrine-like molecular subtype. However, we could observe activation of CDH17 and LGALS4 expression in several tumors belonging to

the classical subtype, in particular PACO17 and PACO19 derived tumors. Additional tests are planned to evaluate if the increased expression of CDH17 can be associated with elevated levels of HNF-1α transcription factor, described by Noll, Eisen *et al.* as a biomarker for exocrine-like PDAC (Noll, Eisen *et al.,* manuscript in preparation). Tumors of the QM subtype were typically negative for the two markers. When comparing the staining intensities for the putative biomarkers, it became apparent that the signal detected in exocrine-like PDAC was typically higher than the one observed in classical tumors. Heterogeneity of PDAC tumors, including those derived in NSG mice from PACO cell lines, could present an explanation for the *in vivo* staining of some of the classical tumors. Thus, it is possible that one patient/cancerous growth harbors simultaneously two or more distinct PDAC molecular subtypes. Arguments supporting this theory can be partially provided by the PACO *in vitro* system. Based on the initial gene set enrichment analysis performed by Dr. C. Eisen (HI-STEM, DKFZ) using the PDA-assigner described by Collisson *et al.*, two PACO cell lines isolated from the same patient – PACO19 and PACO20 – were annotated to distinct subtypes: classical and respectively QM [141]. The re-classification of PACO20 as a classical cell line was based on more detailed transcriptomic and proteomic analysis, as mentioned in the *Results* chapter. Nonetheless, the exact classification of PACO20 will require additional investigations. Further analysis will therefore be required to determine if tumor heterogeneity represents a common event in PDAC.

The healthy pancreatic tissue was, to our surprise, positive for LGALS4, although literature data have never reported LGALS4 expression in the adult human pancreas [110]. Since the healthy tissue was collected from the healthy appearing area of a resected tumor (as evaluated at the time of the surgical procedure) and as we only had a limited number of healthy controls, we cannot exclude that the positive signals detected in the 'healthy' tissues are in fact associated with PDAC precancerous lesions. Further evaluation from a pathologist would be required prior to drawing a definite conclusion.

We thus turned our attention to the initial paper of Collisson *et al*., describing the gene expression patterns of the three molecular subtypes. We could observe that the original publication had listed both CDH17 and LGALS4 as genes whose transcription appears to be activated in tumors belonging to the classical subtype [3]. However, *in vivo* IF experiments indicated a strong expression of CDH17 and LGALS4 in the majority of our exocrine-like tumors, contradicting the data of Collisson *et al*. Furthermore, this upregulation could also be confirmed at mRNA level by RT-qPCR (**Figure 44**). The gene expression levels for the two targets were clearly higher in exocrine-like and some classical tumors (especially PACO19 derived xenografts), compared to healthy pancreas and QM tumors. Thus, it appears as if *in vivo*, both exocrine-like and classical tumors can selectively express the two protein markers.

The reason for this *in vivo* activation of the two putative exocrine-like markers in the classical subtype requires further studies. One possible explanation could be provided by the *in vitro* culture medium in which the PACO cells were isolated and maintained. It could be speculated that both the chemical composition and the selection of the growth factors lead to the enrichment of specific subpopulations *in vitro*, which may or may not harbor stem cell like properties. Some low expression levels of LGALS4 could, for example, already be detected *in vitro* (PACO19 cell line), albeit with only one peptide (**Table 7**). We can suspect that the tumor microenvironment leads to the transcriptional activation of the two genes, probably a means to better adapt to the new conditions. The nature of the PDAC stromal

compartment has baffled scientists working in the oncology field, since it is one of the only tumors described to be associated with a highly abundant desmoplastic stroma. Additionally, tumor blood vessels are only sparsely present, and many of them are nonfunctional [8]. The increased interstitial fluid pressure (IFP) recorded in PDAC tumors is suspected to be the main reason for the collapsing of tumor associated capillaries, leading to the formation of extended hypoxic, nutrient deficient areas [13]. LGALS4 overexpression has been associated with improved survival of cells cultured in FCS depleted medium compared to cells not expressing the protein [110]. Thus, LGALS4 *in vivo* activation could represent a strategy employed by PDAC cells to adapt to the new environment and to adjust to the scarce availability of nutrients, compared to the cultured system. In contrast, CDH17's role in cancer progression is more elusive, as very little is known about the signaling networks in which it participates. It has been linked to Wnt signaling and FAK activation after interacting with the β1 integrin receptor [102, 103]. We can therefore assume that CDH17 *in vivo* activation plays a role in cancer cell proliferation and migration. However, we cannot confirm whether the PDAC molecular subtypes described by Collisson *et al.* are stable or if they can transform into a new molecular subtype, based on environmental stimuli. As stated earlier, further investigations will also be required to evaluate if patients can harbor multiple molecular subtypes of PDAC in one malignant lesion.

As previously mentioned, Tan *et al*. acknowledged that patients with CDH17 and LGALS4 double positive gastric tumors have a better overall survival rate compared to patients whose tumors were double negative for the two proteins. Double positive tumors exhibited also a higher level of differentiation, consequently offering a possible explanation for the better prognosis associated with the disease [105]. Additionally, patients with PDAC tumors expressing high levels of CDH17 have been shown as well to have improved overall survival [95]. Collisson's original data, which reported high expression levels of both CDH17 and LGALS4 in classical PDAC tumors, suggested that patients diagnosed with classical tumors have a slightly better prognosis compared to patients harboring cancers of the exocrine-like and QM PDAC subtypes [3]. On the other hand, Noll, Eisen *et al*. identified exocrine-like PDAC tumors as being associated with the best overall survival (Noll, Eisen *et al*., manuscript in preparation). The reasons behind such contradicting data are most likely rooted either in the size of the patient cohort used for the comparison analysis, or in the applied technique (genomic analysis of microdissected tumors vs histological analyses). Either way, based on our experimental data, both classical and exocrine-like tumors express high levels of the two putative biomarkers compared to QM malignancies. We suggest that CDH17 and LGALS4 expression might correlate with a higher level of tumor differentiation, leading to better overall disease prognosis. In fact, classical and exocrine-like cancers typically develop ductal-like tumor structures, whereas QM tumors display a poorly differentiated histology (**Figure 30**).

We observed that metastatic lesions can also express the two protein markers, depending on the organ that has been colonized and the tumor of origin. Both PACO10 and PACO3 (originating from the same patient) exocrine-like liver metastases, expressed high levels of CDH17 (**Figure 32** and data not shown). However the CDH17 staining was negative in lung metastases. The presence of CDH17 in liver metastases has previously been documented for colorectal carcinoma, being in fact suggested that its expression in the primary cancer might be a prerequisite for future liver colonization of the cancer cells [102]. This could suggest a dual role for CDH17 in PDAC tumors: it could represent a marker of more differentiated primary tumors, while at the same time indicating a phenotype more likely to

produce liver metastasis. For PACO20 derived tumors on the other hand, metastatic lung lesions appear to upregulate CDH17, in contrast to the PACO10 derived lung tumors (**Figure 32**). We unfortunately were not able to test the expression levels in liver metastases, derived from the same cell line. Interestingly, LGALS4 expression in spontaneous PACO derived metastases correlated with the one observed for CDH17 (**Figure 34**). The observation that the two exocrine-like biomarker candidates are co-expressed by the same metastatic lesions might indicate a functional association of the two proteins, first suggested by the *in vitro* colocalization experiments. Further *in vivo* analysis will be required, prior to drawing a conclusion about the functional roles of CDH17 and LGALS4 in primary and metastatic PDAC tumors.

In addition, we investigated the presence of CDH17 and LGALS4 in the original patient tumors (**Figure 38** and **Figure 39**). CDH17 expression was clearly upregulated in those tumors from which the exocrine-like PACO cell lines were derived from. For LGALS4 however, the interpretation of the IF stainings was hindered by the high background detected in patient material, most likely originating from the tumor stroma. In the future, we would like to evaluate the IF stainings at a confocal microscope allowing spectral unmixing, which would therefore remove the autofluorescence signals from the high-stroma containing samples.

Important for the development of potential targeted therapies, the *in vivo* biotinylation experiments followed by mass spectrometric analysis, detected LGALS4 in PACO17 derived tumors with more than two peptides (**Table 13**). Albeit the *in vivo* experiments were to date performed on a small cohort, they do offer an initial confirmation that LGALS4 is accessible from the blood stream, depending on the size of the targeted therapeutical agent to be developed.

6.3 Validation of pan-PDAC protein biomarker candidates

6.3.1 In vitro validations

We could confirm the presence of PCDH1 on the cell surface of all *in vitro* grown PACO cell lines, the highest levels being detected in cell lines belonging to the QM subtype (**Figure 25**). Some cytoplasmic staining was identified in the healthy control cell lines, but no clear membrane staining was present in either the HPDE or the HPNE cells. However, no PCDH1 signal could be detected in the whole cell lysates of the healthy controls, as determined by WB (**Figure 29**), indicating that the cytoplasmic signals detected in IF experiments might have derived from unspecific background staining. The high protein expression levels of PCDH1 in PDAC cells were associated also with increased abundance of the corresponding mRNA, as determined by RT-qPCR (**Figure 43**).

IF experiments on paraffin embedded cells could confirm the presence of LCN2 associated with the plasma membrane or in the cytoplasm of PACO cell lines (**Figure 24**). The signal was highest in cells belonging to the exocrine-like and classical subtypes. We could detect some cytoplasmic staining in the HPNE control cell line. However, we could not detect LCN2 in any of the healthy control cells lines in our MS identification experiments. Additionally, LCN2 has not been reported in literature as being expressed by healthy pancreatic tissue [127]. Therefore, it is possible that the positive fluorescent signal detected in the HPNE cell line might represent an artifact introduced by the staining procedure. mRNA expression

levels detected by RT-qPCR, correlated with the protein levels previously evaluated in IF experiments (**Figure 43**), indicating that transcriptional activation was (at least partially) responsible for the overexpression of LCN2 in PDAC cultured PACO cells.

Isoform 2 of PCDH1 could be identified in several of the samples analyzed during the cell surface MS discovery experiment (**Figure 28**). Literature data report several splice isoforms for PCDH1, as well as multiple transcription initiating codons in the corresponding gene. The developmental status of the cell, as well as environmental signals from the niche, seem to influence the splicing pattern of the PCDH1 mRNA resulting in the expression of different PCDH1 protein isoforms [119]. Two main isoforms are most often reported for PCDH1, which are distinguishable by their molecular weight. Isoform 2 contains an additional cytoplain the asmic domain, which has been shown to mediate the interaction between PCDH1 and protein phosphatase 1-α [122]. WB analysis (**Figure 29**), revealed a single signal in the majority of cultured PACO cell lines, having a molecular weight of ~140 kDa. This molecular weight would correspond to isoform 2 of PCDH1 (molecular weight of polypeptide chain alone: ~120 kDa). However, we need to keep in mind that cadherins are often subject to N-glycosylation [94]. In the absence of data describing the extent of post-translational modifications of PCDH1, we cannot draw a definite conclusion about the main isoform present in our *in vitro* system. Alternative glycosylation patterns could also hinder the antibody based detection of different PCDH1 isoforms. WB detection following deglycosylation treatment would help elucidate which of the isoforms is predominantly expressed by PACO cells. Additionally, we are considering performing SRM experiments, with optimized transition selection for isoform specific precursor peptides, and resulting fragment ions. For PACO20 cell lysates, an additional band, with a higher molecular weight compared to the predominant isoform could also be detected, albeit with a weaker intensity (**Figure 29**). Despite the fact that IF experiments indicated a high protein expression level for PCDH1 in PACO16 cells, we could detect any signal corresponding to the protein in the total cell lysates in WB (**Figure 29**). One possible explanation could be that the protein concentration in PACO16 total lysates is below the detection limit of the WB method, whereas the intense IF signal might have originated mostly from the plasma membrane of cultured cells. An additional explanation could be provided by the presence of limited mouse fibroblast contamination in some of the PACO16 batches we analyzed, which might have led to the dilution of PACO16 specific proteins in the whole cell lysates.

6.3.2 In vivo validations

IF stainings performed on primary tumor xenografts confirmed the presence of both PCDH1 and LCN2 in pancreatic tumor mouse xenografts (**Figure 35** and **Figure 37**). Exocrine-like PDAC appeared to express the highest protein and mRNA LCN2 levels (**Figure 35** and **Figure 45**). In addition, LCN2 overexpression has often been correlated with improved disease prognosis [124], which could again be linked to the observations of Noll and Eisen, associating PDAC exocrine-like tumors with better overall survival (Noll, Eisen *et al*., manuscript in preparation). The PCDH1 signal *in vivo* was generally weak, with exocrine-like tumors and several QM derived cancers expressing higher levels, as evaluated by IF (**Figure 37**). mRNA PCDH1 levels were clearly upregulated in all analyzed PDAC samples compared to healthy pancreatic tissue (**Figure 45**).

Lipocalin 2 expression is typically down-regulated in metastatic lesions [124]. We investigated the expression levels of LCN2 in spontaneous metastases developed by tumor bearing mice and could report that the expression level of the protein lesion varied compared to the primary cancerous lesion, depending on the cell line of origin. For example, in PACO20 derived lung metastases, LCN2 protein expression was completely absent, correlating with literature reports. However, for the exocrine-like subtype derived liver and lung secondary tumors however, we could report a significant LCN2 overexpression compared to the primary cancer (**Figure 36**). The significance of this activation in some of the secondary lesions needs further investigation. PCDH1 expression in metastatic lesions has not yet been tested.

LCN2 could also be detected when performing the *in vivo* whole body perfusion of tumor bearing mice (**Table 13**). This again confirmed that LCN2 could represent a candidate for novel targeted PDAC therapies.

LCN2 expression could as well be detected in the patient original tumors, with the exception of the patient from which the PACO17 cell line was isolated (**Figure 40**). However, our observations were once again hindered by the high background present in the patient tumors. Due to insufficient patient tumor material available, we did not investigate the expression levels of PCDH1 in the original patient PDAC samples.

LCN2 activation in PDAC and its presence in the blood of pancreatic cancer patients has already been documented in literature [21]. The protein can be detected in the plasma of pancreatitis patients, as well as in asymptomatic individuals, harboring precancerous lesions [21]. Although this early activation might represent an impediment for developing LCN2 into a proper diagnostic tool, it is worth noting that both pancreatitis patients, and individuals bearing PanIN precursor lesions, have an increased risk of developing PDAC later in life [5] and could be better monitored if they are recognized as a high risk group.

In order for a putative cancer biomarker to successfully enter the clinic, extended validations, on large patient cohorts need to be performed prior to its recognition as a valuable clinical tool [32, 70, 150]. In future experiments, we plan to investigate the presence of the four protein biomarkers described in tissue microarrays and the corresponding blood of patients, harboring malignant and benign diseases. Thus, one can investigate which panel of markers can be used for diagnostic applications and/or targeted therapies.

7. Conclusions and outlook

We have validated PCDH1 as a novel marker for PDAC, and evaluated its applicability as a diagnostic tool, in combination with LCN2. Additionally we have identified molecular subtype specific protein markers – namely CDH17 and LGALS4, which can potentially be used for the stratification of patients and may help facilitate therapeutical decision making. CDH17 and LGALS4 were expressed exclusively on exocrine-like PDAC cells *in vitro*, and they maintained their expression *in vivo*. Additionally, we observed some ectopic activation of both CDH17 and LGALS4 in some of the classical mouse derived tumors, although generally the expression level was inferior to that documented for exocrine-like PDAC. Initial *in vitro* experiments revealed the colocalization of the two putative exocrine-like markers on the surface of cultured cells. So far, *in vivo* localization experiments have been hindered by the fact that the majority of our tissue samples have been fixed in formalin, and the commercially available antibodies we employed so far require distinct antigen retrieval protocols. For this reason, cryopreserved organ samples will be employed in the future, to evaluate if the putative interaction between CDH17 and LGALS4 can also be observed in tumor tissues.

The two panels of pan-PDAC and subtype specific putative markers will need to be additionally validated on larger cohorts of patients. In collaboration with Dr. M. Sprick (HI-STEM, DKFZ, Heidelberg) and Prof. Dr. W. Weichert (Pathology Department, Universitätsklinikum Heidelberg), we plan to investigate the four putative biomarkers on a PDAC TMA, originating from primary tumor material for which the molecular subclassification is already known.

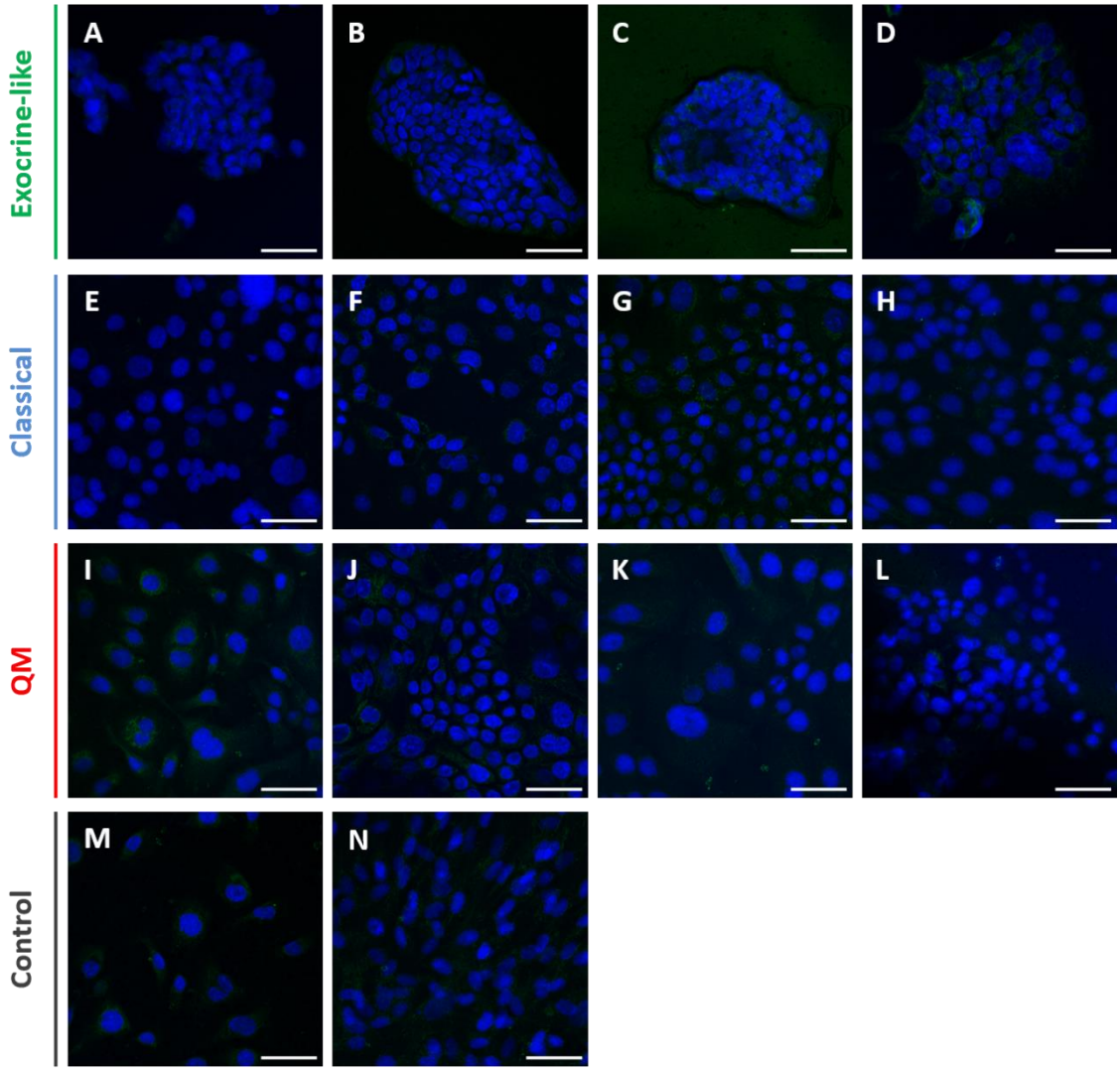
The two pan-PDAC and exocrine-like markers are also secreted by cancer cells *in vitro*, and could represent attractive diagnostic markers for clinical applications. We are currently investigating, using SRM and ELISA, if the presence of these markers can also be detected in the blood of tumor bearing mice.

The final goal of the current study is the development of specific monoclonal antibodies, which could be used for targeted therapies in pancreatic cancer. Our lab has recently selected specific antibodies against the two putative exocrine-like biomarkers CDH17 and LGALS4, using phage display technology (work of Lisa Maria Becker and Dr. Katharina Frey, Biomarker Discovery Group, HI-STEM). The two antibodies recognized the native proteins, evaluated by ELISA tests using recombinant proteins. Additionally, the antibodies produced as single chain Fv (scFv) fusion proteins could also recognize the proteins against which they were raised, when using methanol or formalin as fixation methods for the biological material (data not shown). The selected antibodies will need to first undergo affinity maturation before establishing their clinical applications as diagnostic and/or therapeutic tools. Should the antibodies be eligible for therapeutical delivery, biodistribution experiments will need to be performed, evaluating which antibody format would localize most efficiently within the tumor. Combinational therapies with agents targeting the tumor stroma, aimed to alleviate the augmented IFP reported in some tumors, can also be considered for an improved delivery. As we have mentioned, exocrine-like tumors appear to upregulate liver specific transcription factors, which might be associated with increased clearance of xenobiotics. A targeted antibody based therapy, potentially combined with immunostimulatory strategies, might be considered as an alternative for those patients

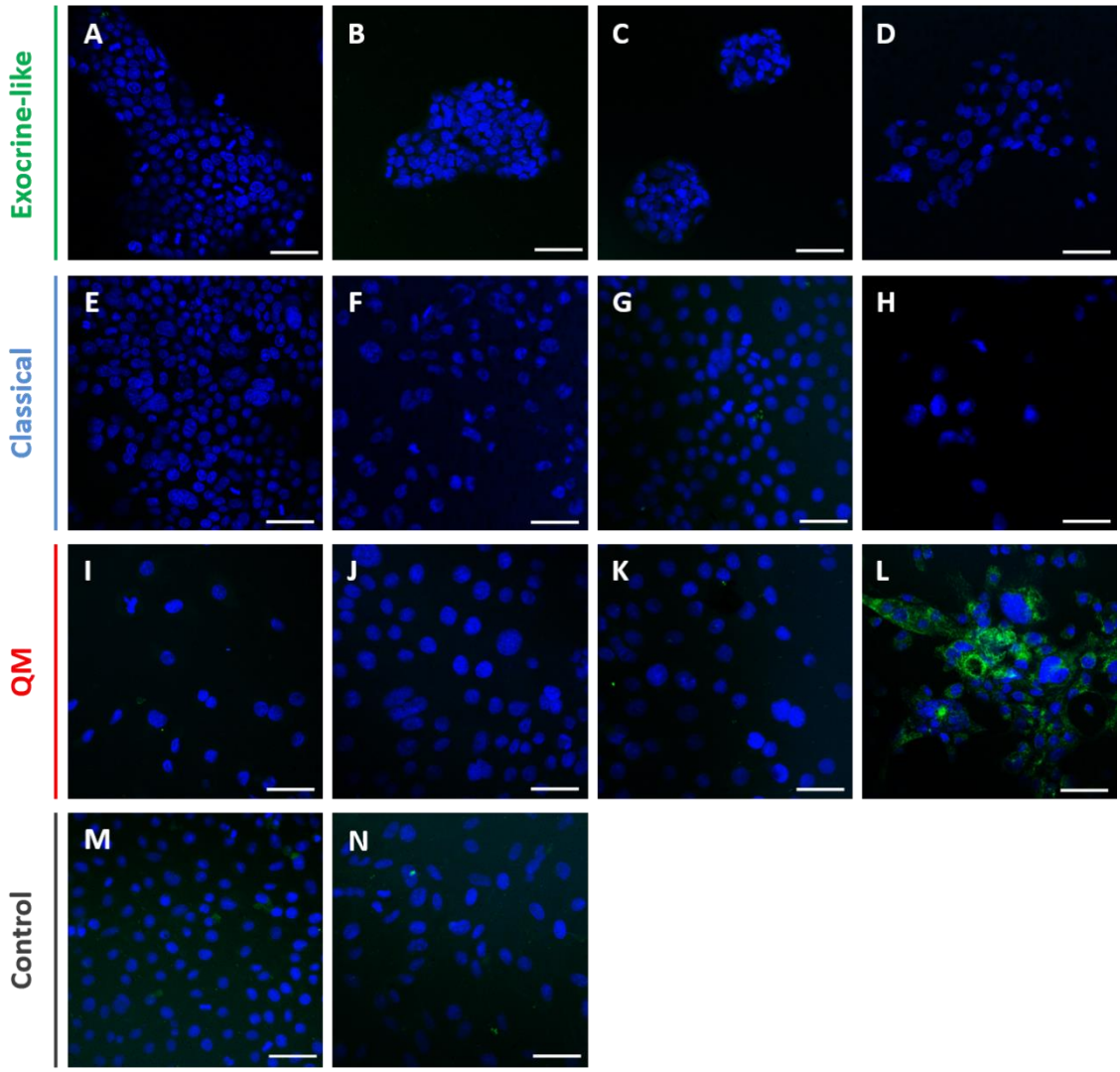
whose transcriptomic signature indicate the potential failure of conventional cancer therapies.

Due to time constraints, the current study did not investigate if the PDAC biomarker candidates discovered, are also associated with benign inflammatory pancreatic pathologies, such as pancreatitis. However, future experiments are planned interrogating the presence of the markers discussed in this work in healthy and diseased tissues, using both commercially available TMAs, as well as samples provided by the clinical hospital, Heidelberg. It is possible that some of the proteins for which we performed the early validation experiments are already present in precancerous lesions, which often progress to fully developed PDAC. If so, these biomarkers could be employed for the monitoring of high-risk patients.

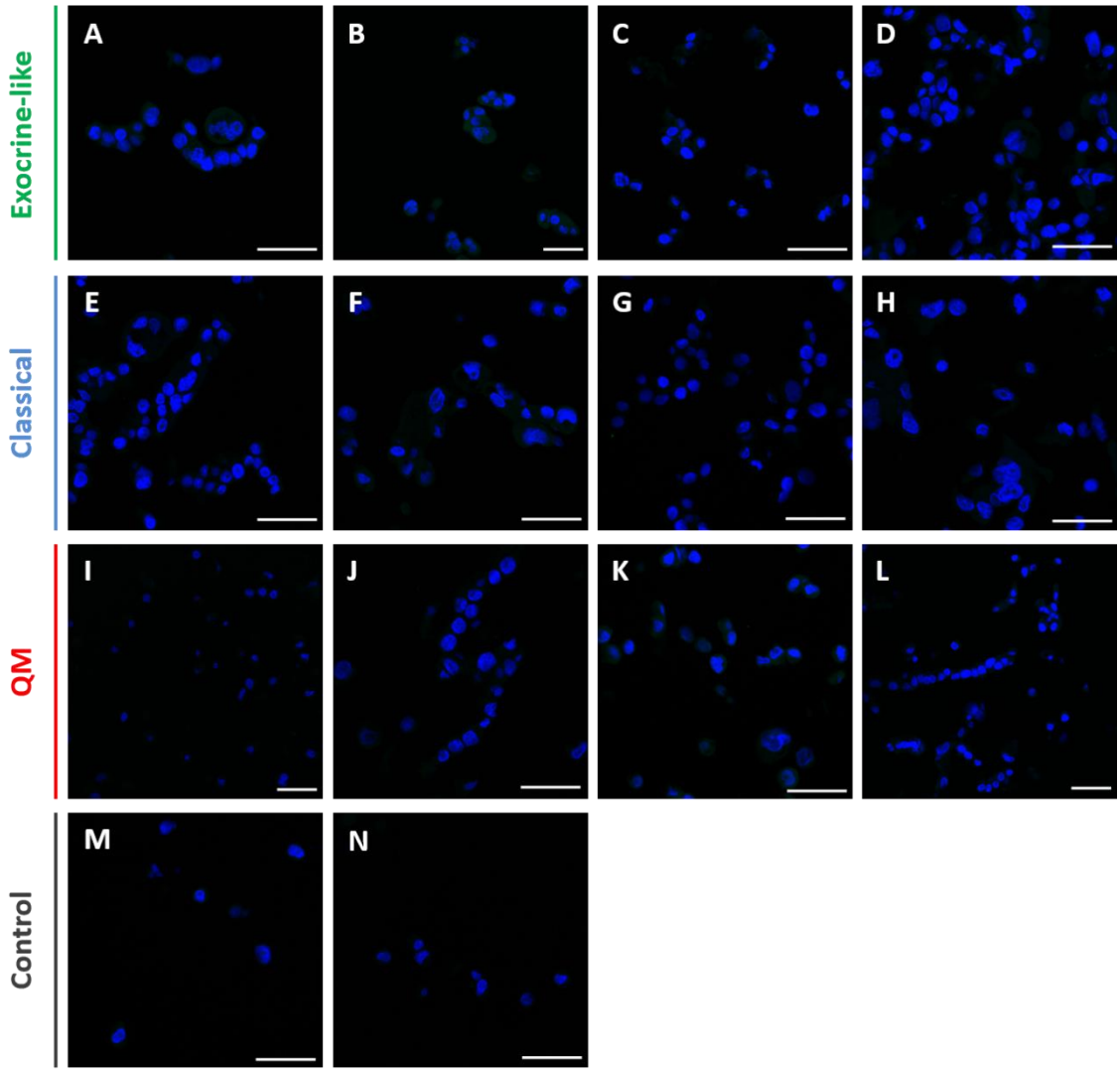
8. Supplementary information



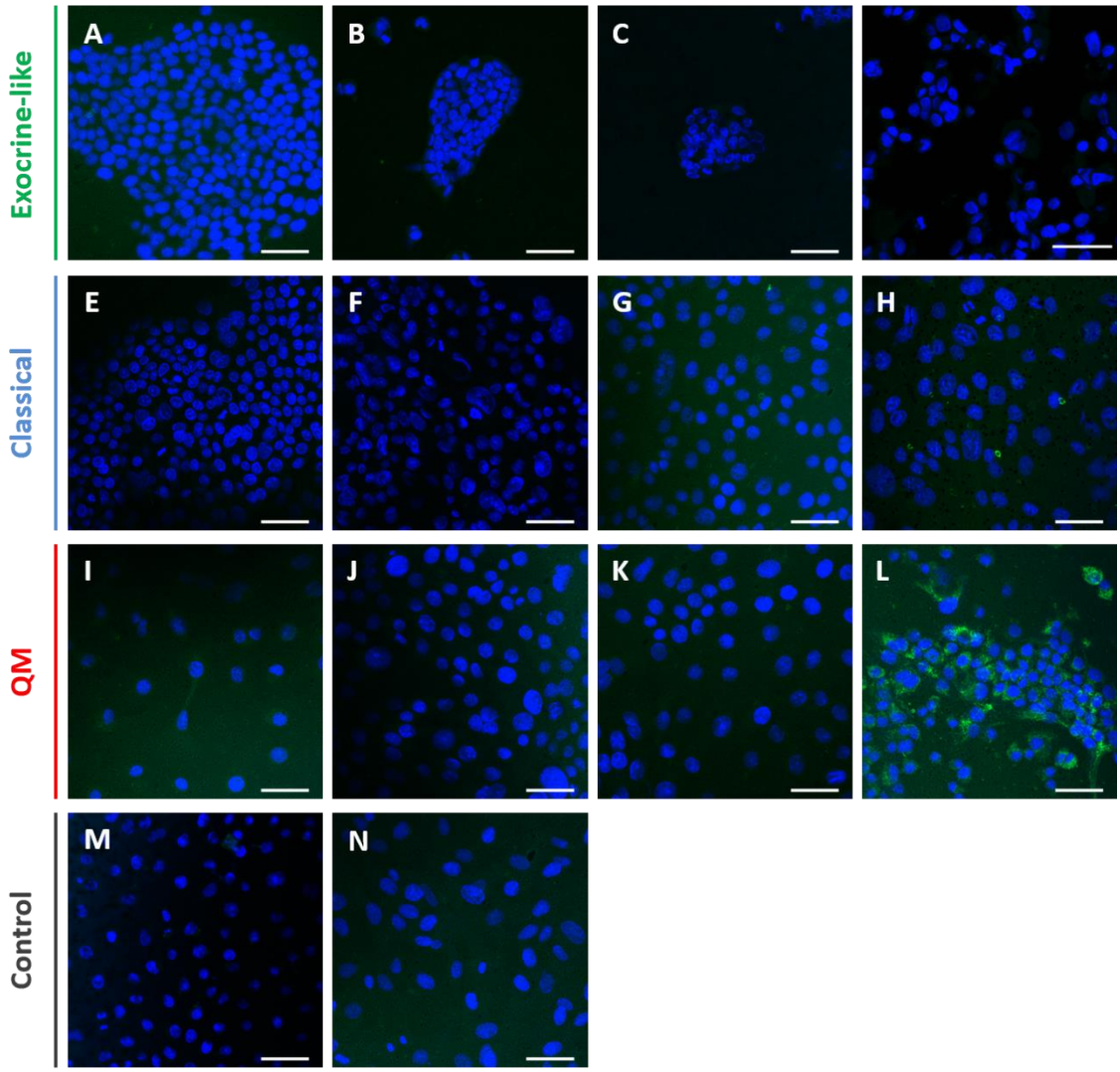
Supplementary Figure 1: *In vitro* IF validations: CDH17 – isotype controls. **A**: PACO3; **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO2; **F**: PACO17; **G**: PACO19; **H**: PACO20; **I**: PACO7; **J**: PACO8; **K**: PACO9; **L**: PACO16; **M**: HPDE; **N**: HPNE (Scale bars: 50µm)



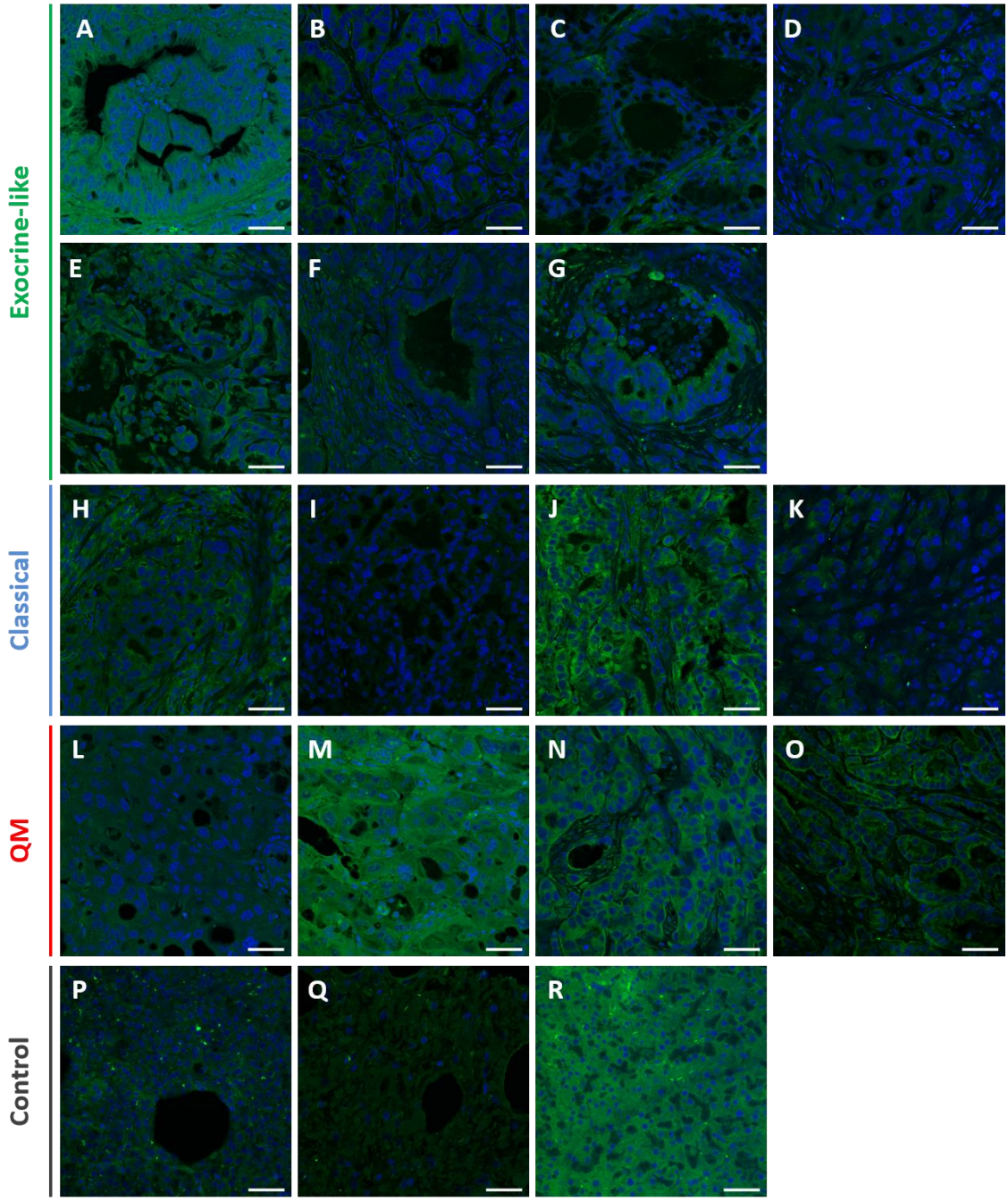
Supplementary Figure 2: *In vitro* IF validations: LGALS4 – isotype controls. **A**: PACO3; **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO2; **F**: PACO17; **G**: PACO19; **H**: PACO20; **I**: PACO7; **J**: PACO8; **K**: PACO9; **L**: PACO16; **M**: HPDE; **N**: HPNE (Scale bars: 50µm)



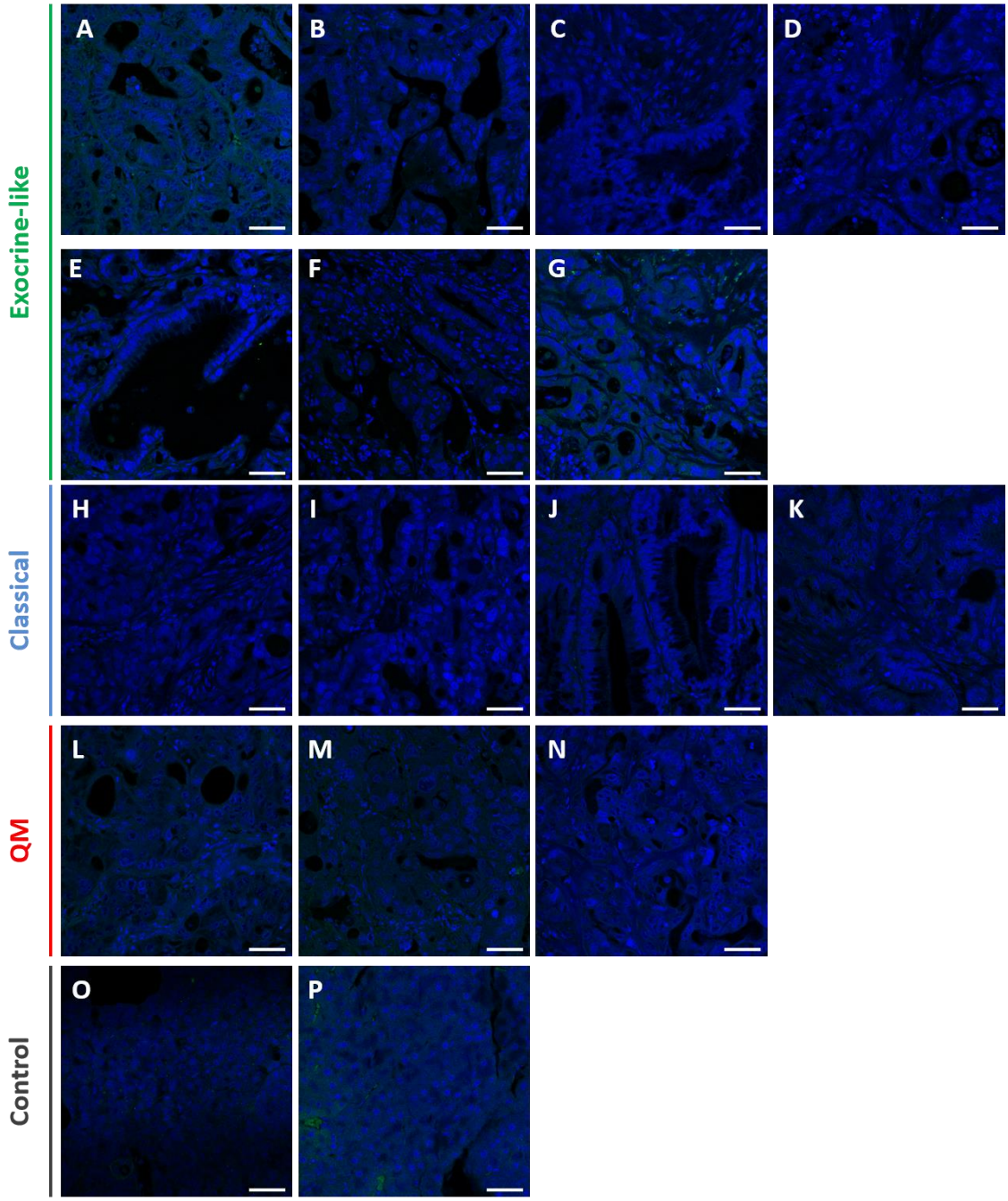
Supplementary Figure 3: *In vitro* IF validations: LCN2 – negative controls (without primary antibody). **A**: PACO3; **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO2; **F**: PACO17; **G**: PACO19; **H**: PACO20; **I**: PACO7; **J**: PACO8; **K**: PACO9; **L**: PACO16; **M**: HPDE; **N**: HPNE; (Scale bars: 50µm)



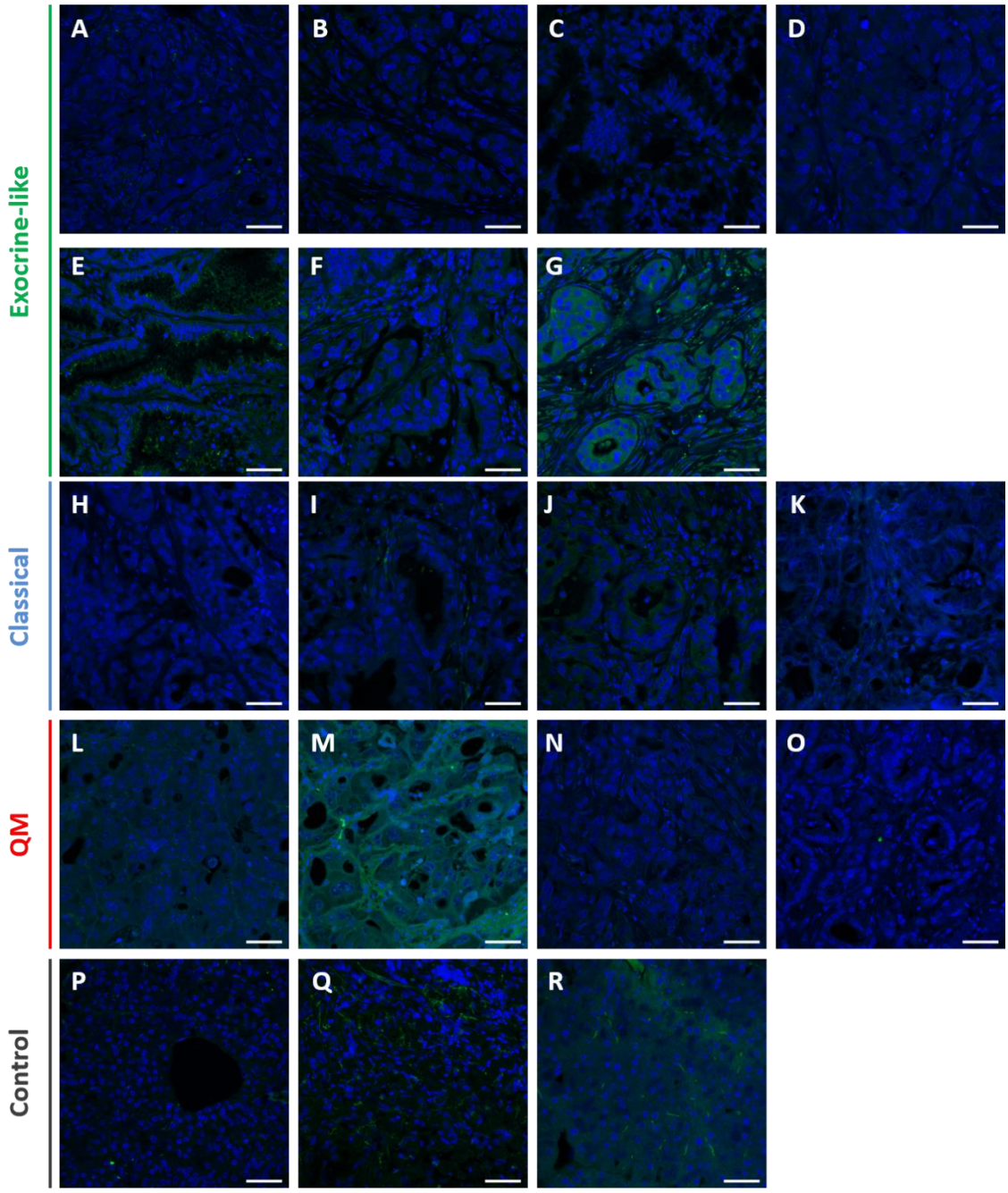
Supplementary Figure 4: *In vitro* IF validations: PCDH1 – isotype controls. **A**: PACO3; **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO2; **F**: PACO17; **G**: PACO19; **H**: PACO20; **I**: PACO7; **J**: PACO8; **K**: PACO9; **L**: PACO16; **M**: HPDE; **N**: HPNE; **O**: PACO10 isotype control; **P**: PACO14 isotype control (Scale bars: 50µm)



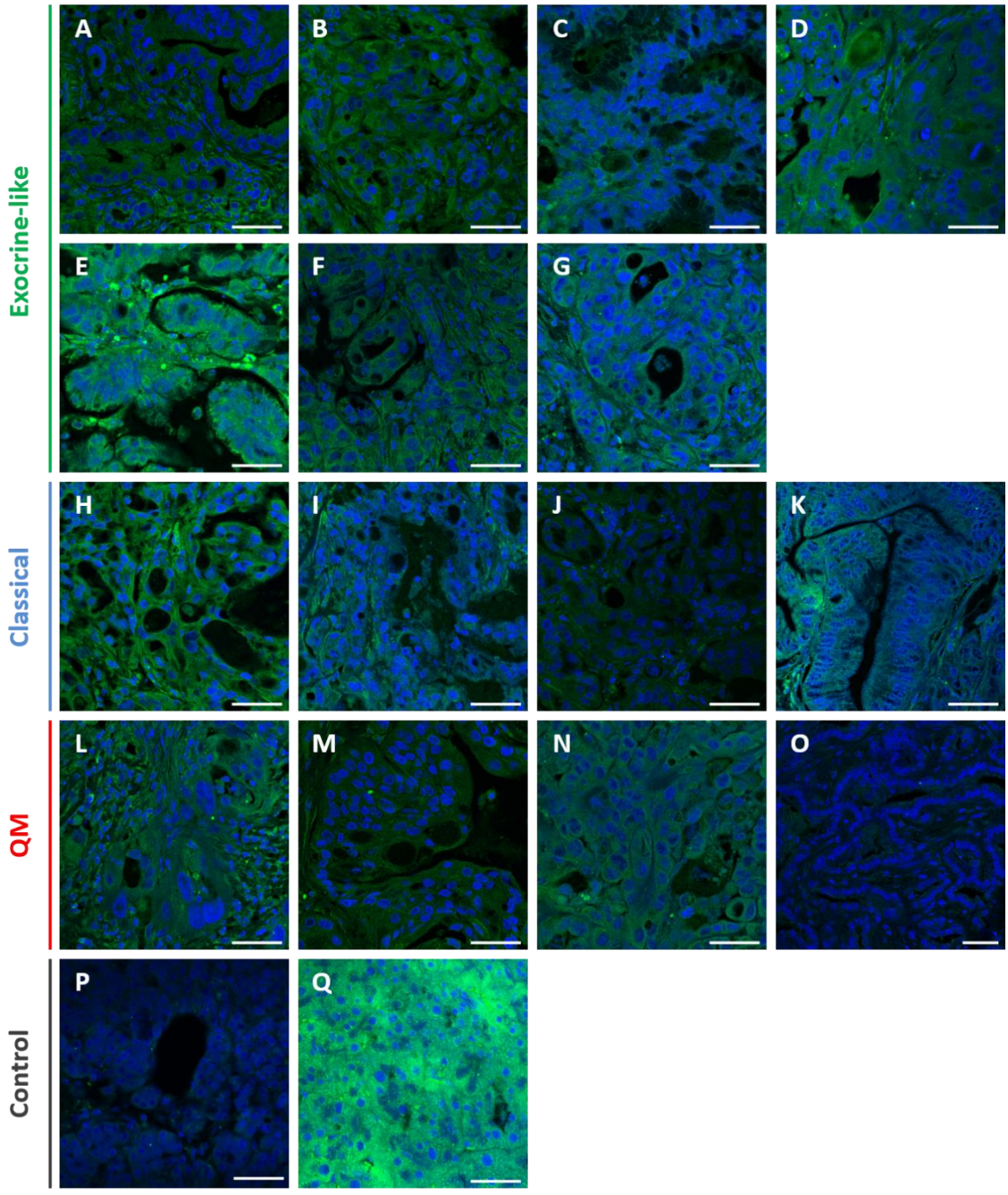
Supplementary Figure 5: *In vivo* validations - CDH17 isotype controls. **A**: PACO3, **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO25; **F** PACO26; **G**: PACO 28; **H**: PACO2; **I**: PACO17; **J**: PACO19; **K**: PACO20; **L**: PACO7; **M**: PACO8; **N**: PACO9; **O**: PACO16; **P**: human healthy pancreas patient # 4984; **Q**: human healthy pancreas patient # 4985; **R**: healthy mouse NSG pancreas. (Scale bars: 50 µm)



Supplementary Figure 6: *In vivo* validations – LGALS4 isotype controls. **A**: PACO3, **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO25; **F** PACO26; **G**: PACO 28; **H**: PACO2; **I**: PACO17; **J**: PACO19; **K**: PACO20; **L**: PACO7; **M**: PACO8; **N**: PACO9; **O**: human healthy pancreas patient # 4984; **P**: healthy mouse NSG pancreas. (Scale bars: 50 µm)



Supplementary Figure 7: *In vivo* validations – LCN2 isotype controls. **A**: PACO3, **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO25; **F** PACO26; **G**: PACO 28; **H**: PACO2; **I**: PACO17; **J**: PACO19; **K**: PACO20; **L**: PACO7; **M**: PACO8; **N**: PACO9; **O**: PACO16; **P**: human healthy pancreas patient # 4984; **Q**: human healthy pancreas patient # 4985; **R**: healthy mouse NSG pancreas. (Scale bars: 50 µm)



Supplementary Figure 8: *In vivo* validations – PCDH1 isotype controls. **A**: PACO3, **B**: PACO10; **C**: PACO14; **D**: PACO18; **E**: PACO25; **F** PACO26; **G**: PACO 28; **H**: PACO2; **I**: PACO17; **J**: PACO19; **K**: PACO20; **L**: PACO7; **M**: PACO8; **N**: PACO9; **O**: PACO16; **P**: human healthy pancreas patient # 4984; **Q**: healthy mouse NSG pancreas. (Scale bars: 50 µm)

Supplementary Material: Complete alignment of isoform 1 (upper sequence) and isoform 2 (lower sequence) of PCDH1, using ClustalW. In green: identified peptides common to both isoforms. In yellow: identified peptides specific to isoform 2 of PCDH1

```
sp|Q08174|PCDH1_HUMAN        MDSGAGGRRCPEAALLILGPPRMEHLRHSPGPGGQRLLLPSMLLALLLLL
sp|Q08174-2|PCDH1_HUMAN      MDSGAGGRRCPEAALLILGPPRMEHLRHSPGPGGQRLLLPSMLLALLLLL
                             **************************************************

sp|Q08174|PCDH1_HUMAN        APSPGHATRVVYKVPEEQPPNTLIGSLAADYGFPDVGHLYKLEVGAPYLR
sp|Q08174-2|PCDH1_HUMAN      APSPGHATRVVYKVPEEQPPNTLIGSLAADYGFPDVGHLYKLEVGAPYLR
                             **************************************************

sp|Q08174|PCDH1_HUMAN        VDGKTGDIFTTETSIDREGLRECQNQLPGDPCILEFEVSITDLVQNGSPR
sp|Q08174-2|PCDH1_HUMAN      VDGKTGDIFTTETSIDREGLRECQNQLPGDPCILEFEVSITDLVQNGSPR
                             **************************************************

sp|Q08174|PCDH1_HUMAN        LLEGQIEVQDINDNTPNFASPVITLAIPENTNIGSLFPIPLASDRDAGPN
sp|Q08174-2|PCDH1_HUMAN      LLEGQIEVQDINDNTPNFASPVITLAIPENTNIGSLFPIPLASDRDAGPN
                             **************************************************

sp|Q08174|PCDH1_HUMAN        GVASYELQAGPEAQELFGLQVAEDQEEKQPQLIVMGNLDRERWDSYDLTI
sp|Q08174-2|PCDH1_HUMAN      GVASYELQAGPEAQELFGLQVAEDQEEKQPQLIVMGNLDRERWDSYDLTI
                             **************************************************

sp|Q08174|PCDH1_HUMAN        KVQDGGSPPRASSALLRVTVLDTNDNAPKFERPSYEAELSENSPIGHSVI
sp|Q08174-2|PCDH1_HUMAN      KVQDGGSPPRASSALLRVTVLDTNDNAPKFERPSYEAELSENSPIGHSVI
                             **************************************************

sp|Q08174|PCDH1_HUMAN        QVKANDSDQGANAEIEYTFHQAPEVVRRLLRLDRNTGLITVQGPVDREDL
sp|Q08174-2|PCDH1_HUMAN      QVKANDSDQGANAEIEYTFHQAPEVVRRLLRLDRNTGLITVQGPVDREDL
                             **************************************************

sp|Q08174|PCDH1_HUMAN        STLRFSVLAKDRGTNPKSARAQVVVTVKDMNDNAPTIEIRGIGLVTHQDG
sp|Q08174-2|PCDH1_HUMAN      STLRFSVLAKDRGTNPKSARAQVVVTVKDMNDNAPTIEIRGIGLVTHQDG
                             **************************************************

sp|Q08174|PCDH1_HUMAN        MANISEDVAEETAVALVQVSDRDEGENAAVTCVVAGDVPFQLRQASETGS
sp|Q08174-2|PCDH1_HUMAN      MANISEDVAEETAVALVQVSDRDEGENAAVTCVVAGDVPFQLRQASETGS
                             **************************************************

sp|Q08174|PCDH1_HUMAN        DSKKKYFLQTTTPLDYEKVKDYTIEIVAVDSGNPPLSSTNSLKVQVVDVN
sp|Q08174-2|PCDH1_HUMAN      DSKKKYFLQTTTPLDYEKVKDYTIEIVAVDSGNPPLSSTNSLKVQVVDVN
                             **************************************************

sp|Q08174|PCDH1_HUMAN        DNAPVFTQSVTEVAFPENNKPGEVIAEITASDADSGSNAELVYSLEPEPA
sp|Q08174-2|PCDH1_HUMAN      DNAPVFTQSVTEVAFPENNKPGEVIAEITASDADSGSNAELVYSLEPEPA
                             **************************************************

sp|Q08174|PCDH1_HUMAN        AKGLFTISPETGEIQVKTSLDREQRESYELKVVAADRGSPSLQGTATVLV
sp|Q08174-2|PCDH1_HUMAN      AKGLFTISPETGEIQVKTSLDREQRESYELKVVAADRGSPSLQGTATVLV
                             **************************************************

sp|Q08174|PCDH1_HUMAN        NVLDCNDNDPKFMLSGYNFSVMENMPALSPVGMVTVIDGDKGENAQVQLS
sp|Q08174-2|PCDH1_HUMAN      NVLDCNDNDPKFMLSGYNFSVMENMPALSPVGMVTVIDGDKGENAQVQLS
                             **************************************************

sp|Q08174|PCDH1_HUMAN        VEQDNGDFVIQNGTGTILSSLSFDREQQSTYTFQLKAVDGGVPPRSAYVG
sp|Q08174-2|PCDH1_HUMAN      VEQDNGDFVIQNGTGTILSSLSFDREQQSTYTFQLKAVDGGVPPRSAYVG
                             **************************************************

sp|Q08174|PCDH1_HUMAN        VTINVLDENDNAPYITAPSNTSHKLLTPQTRLGETVSQVAAEDFDSGVNA
sp|Q08174-2|PCDH1_HUMAN      VTINVLDENDNAPYITAPSNTSHKLLTPQTRLGETVSQVAAEDFDSGVNA
                             **************************************************

sp|Q08174|PCDH1_HUMAN        ELIYSIAGGNPYGLFQIGSHSGAITLEKEIERRHHGLHRLVVKVSDRGKP
sp|Q08174-2|PCDH1_HUMAN      ELIYSIAGGNPYGLFQIGSHSGAITLEKEIERRHHGLHRLVVKVSDRGKP
                             **************************************************

sp|Q08174|PCDH1_HUMAN        PRYGTALVHLYVNETLANRTLLETLLGHSLDTPLDIDIAGDPEYERSKQR
sp|Q08174-2|PCDH1_HUMAN      PRYGTALVHLYVNETLANRTLLETLLGHSLDTPLDIDIAGDPEYERSKQR
                             **************************************************
```

```
sp|Q08174|PCDH1_HUMAN        GNILFGVVAGVVAVALLIALAVLVRYCRQREAKSGYQAGKKETKDLYAPK
sp|Q08174-2|PCDH1_HUMAN      GNILFGVVAGVVAVALLIALAVLVRYCRQREAKSGYQAGKKETKDLYAPK
                             **************************************************

sp|Q08174|PCDH1_HUMAN        PSGKASKGNKSKGKKSKSPKPVKPVEDEDEAGLQKSLKFNLMSDAPGDSP
sp|Q08174-2|PCDH1_HUMAN      PSGKASKGNKSKGKKSKSPKPVKPVEDEDEAGLQKSLKFNLMSDAPGDSP
                             **************************************************

sp|Q08174|PCDH1_HUMAN        RIHLPLNYPPGSPDLGRHYRSNSPLPSIQLQPQSPSASKKHQVVQDLPPA
sp|Q08174-2|PCDH1_HUMAN      RIHLPLNYPPGSPDLGRHYRSNSPLPSIQLQPQSPSASKKHQVVQDLPPA
                             **************************************************

sp|Q08174|PCDH1_HUMAN        NTFVGTGDTTSTGSEQYSDYSYRTNPPKYPSKQVGQPFQLSTPQPLPHPY
sp|Q08174-2|PCDH1_HUMAN      NTFVGTGDTTSTGSEQYSDYSYRTNPPKYPSKQLPHRRVTFSATSQAQEL
                             **********************************: :     :. . .:

sp|Q08174|PCDH1_HUMAN        HGAIWTEVWE----------------------------------------
sp|Q08174-2|PCDH1_HUMAN      QDPSQHSYYDSGLEESETPSSKSSSGPRLGPLALPEDHYERTTPDGSIGE
                             :..   . ::

sp|Q08174|PCDH1_HUMAN        --------------------------------------------------
sp|Q08174-2|PCDH1_HUMAN      MEHPENDLRPLPDVAMTGTCTRECSEFGHSDTCWMPGQSSPSRRTKSSAL


sp|Q08174|PCDH1_HUMAN        --------------------------------------------------
sp|Q08174-2|PCDH1_HUMAN      KLSTFVPYQDRGGQEPAGAGSPSPPEDRNTKTAPVRLLPSYSAFSHSSHD


sp|Q08174|PCDH1_HUMAN        -------------------------------------
sp|Q08174-2|PCDH1_HUMAN      SCKDSATLEEIPLTQTSDFPPAATPASAQTAKREIYL
```

Supplementary Material: Digital format content

- Complete list of proteins identified on the cell surface and secreted proteome of *in vitro* cultured PDAC and healthy pancreatic cell lines (Office Excel sheet)
- Complete list of cell surface and secreted proteins identified *in vitro* – combined data set
- Complete list of *in vivo* accessible proteins identified after performing whole body perfusion on NSG mice bearing orthotopic tumors (Office Excel sheet)

- Gene ontology complete pictures:
  - Gene ontology of the *in vitro* whole cell surface PDAC proteome
  - Gene ontology of the *in vitro* exocrine-like cell surface PDAC proteome
  - Gene ontology of the *in vitro* classical cell surface PDAC proteome
  - Gene ontology of the *in vitro* QM cell surface PDAC proteome
  - Gene ontology of the *in vitro* whole PDAC secretome
  - Gene ontology of the *in vitro* exocrine-like PDAC secretome
  - Gene ontology of the *in vitro* classical PDAC secretome
  - Gene ontology of the *in vitro* QM PDAC secretome
  - Gene ontology of the *in vivo* vascular accessible PDAC proteome – human proteome
  - Gene ontology of the *in vivo* vascular accessible PDAC proteome – mouse proteome

9. References

1. Siegel, R., et al., *Cancer statistics, 2014.* CA Cancer J Clin, 2014. **64**(1): p. 9-29.
2. Ferlay, J., et al., *Cancer incidence and mortality patterns in Europe: estimates for 40 countries in 2012.* Eur J Cancer, 2013. **49**(6): p. 1374-403.
3. Collisson, E.A., et al., *Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy.* Nat Med, 2011. **17**(4): p. 500-3.
4. Klein, A.P., *Identifying people at a high risk of developing pancreatic cancer.* Nat Rev Cancer, 2013. **13**(1): p. 66-74.
5. Distler, M., et al., *Precursor Lesions for Sporadic Pancreatic Cancer: PanIN, IPMN, and MCN.* Biomed Res Int, 2014. **2014**: p. 474905.
6. Niccolai, E., et al., *What is recent in pancreatic cancer immunotherapy?* Biomed Res Int, 2013. **2013**: p. 492372.
7. Cinar, P. and M.A. Tempero, *Monoclonal antibodies and other targeted therapies for pancreatic cancer.* Cancer J, 2012. **18**(6): p. 653-64.
8. Provenzano, P.P., et al., *Enzymatic targeting of the stroma ablates physical barriers to treatment of pancreatic ductal adenocarcinoma.* Cancer Cell, 2012. **21**(3): p. 418-29.
9. Hezel, A.F., et al., *Genetics and biology of pancreatic ductal adenocarcinoma.* Genes Dev, 2006. **20**(10): p. 1218-49.
10. Bryant, K.L., et al., *KRAS: feeding pancreatic cancer proliferation.* Trends Biochem Sci, 2014. **39**(2): p. 91-100.
11. Rhim, A.D., et al., *EMT and dissemination precede pancreatic tumor formation.* Cell, 2012. **148**(1-2): p. 349-61.
12. McAllister, F., et al., *Oncogenic Kras Activates a Hematopoietic-to-Epithelial IL-17 Signaling Axis in Preinvasive Pancreatic Neoplasia.* Cancer Cell, 2014. **25**(5): p. 621-37.
13. Provenzano, P.P. and S.R. Hingorani, *Hyaluronan, fluid pressure, and stromal resistance in pancreas cancer.* Br J Cancer, 2013. **108**(1): p. 1-8.
14. Lonardo, E., et al., *Pancreatic stellate cells form a niche for cancer stem cells and promote their self-renewal and invasiveness.* Cell Cycle, 2012. **11**(7): p. 1282-90.
15. Masamune, A. and T. Shimosegawa, *Signal transduction in pancreatic stellate cells.* J Gastroenterol, 2009. **44**(4): p. 249-60.
16. Gilkes, D.M., G.L. Semenza, and D. Wirtz, *Hypoxia and the extracellular matrix: drivers of tumour metastasis.* Nat Rev Cancer, 2014. **14**(6): p. 430-9.
17. Ozdemir, B.C., et al., *Depletion of Carcinoma-Associated Fibroblasts and Fibrosis Induces Immunosuppression and Accelerates Pancreas Cancer with Reduced Survival.* Cancer Cell, 2014.
18. Cruz-Monserrate, Z., et al., *Upregulation and redistribution of integrin alpha6beta4 expression occurs at an early stage in pancreatic adenocarcinoma progression.* Mod Pathol, 2007. **20**(6): p. 656-67.
19. Shen, R., et al., *The biological features of PanIN initiated from oncogenic Kras mutation in genetically engineered mouse models.* Cancer Lett, 2013. **339**(1): p. 135-43.
20. Besmer, D.M., et al., *Pancreatic ductal adenocarcinoma mice lacking mucin 1 have a profound defect in tumor growth and metastasis.* Cancer Res, 2011. **71**(13): p. 4432-42.
21. Moniaux, N., et al., *Early diagnosis of pancreatic cancer: neutrophil gelatinase-associated lipocalin as a marker of pancreatic intraepithelial neoplasia.* Br J Cancer, 2008. **98**(9): p. 1540-7.
22. Zimmermann, G., et al., *Small molecule inhibition of the KRAS-PDEdelta interaction impairs oncogenic KRAS signalling.* Nature, 2013. **497**(7451): p. 638-42.

23. Ghaneh, P., E. Costello, and J.P. Neoptolemos, *Biology and management of pancreatic cancer.* Postgrad Med J, 2008. **84**(995): p. 478-97.
24. Herreros-Villanueva, M., et al., *Adjuvant and neoadjuvant treatment in pancreatic cancer.* World J Gastroenterol, 2012. **18**(14): p. 1565-72.
25. Oberstein, P.E. and K.P. Olive, *Pancreatic cancer: why is it so hard to treat?* Therap Adv Gastroenterol, 2013. **6**(4): p. 321-37.
26. Wonganan, P., et al., *Just getting into cells is not enough: mechanisms underlying 4-(N)-stearoyl gemcitabine solid lipid nanoparticle's ability to overcome gemcitabine resistance caused by RRM1 overexpression.* J Control Release, 2013. **169**(1-2): p. 17-27.
27. Luedke, E., et al., *Monoclonal antibody therapy of pancreatic cancer with cetuximab: potential for immune modulation.* J Immunother, 2012. **35**(5): p. 367-73.
28. Ostrem, J.M., et al., *K-Ras(G12C) inhibitors allosterically control GTP affinity and effector interactions.* Nature, 2013. **503**(7477): p. 548-51.
29. Mellman, I., G. Coukos, and G. Dranoff, *Cancer immunotherapy comes of age.* Nature, 2011. **480**(7378): p. 480-9.
30. Sandin, L.C., et al., *Local CTLA4 blockade effectively restrains experimental pancreatic adenocarcinoma growth in vivo.* Oncoimmunology, 2014. **3**(1): p. e27614.
31. Solier, C. and H. Langen, *Antibody-based proteomics and biomarker research - current status and limitations.* Proteomics, 2014. **14**(6): p. 774-83.
32. Diamandis, E.P., *Cancer biomarkers: can we turn recent failures into success?* J Natl Cancer Inst, 2010. **102**(19): p. 1462-7.
33. Frank, R. and R. Hargreaves, *Clinical biomarkers in drug discovery and development.* Nat Rev Drug Discov, 2003. **2**(7): p. 566-80.
34. Cohn, J.N., *Introduction to surrogate markers.* Circulation, 2004. **109**(25 Suppl 1): p. IV20-1.
35. Ramskold, D., et al., *Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells.* Nat Biotechnol, 2012. **30**(8): p. 777-82.
36. Wheelock, C.E., et al., *Application of 'omics technologies to biomarker discovery in inflammatory lung diseases.* Eur Respir J, 2013. **42**(3): p. 802-25.
37. Velculescu, V.E., et al., *Serial analysis of gene expression.* Science, 1995. **270**(5235): p. 484-7.
38. Diamandis, E.P., *Next-generation sequencing: a new revolution in molecular diagnostics?* Clin Chem, 2009. **55**(12): p. 2088-92.
39. Picotti, P., B. Bodenmiller, and R. Aebersold, *Proteomics meets the scientific method.* Nat Methods, 2013. **10**(1): p. 24-7.
40. Geiger, T., J. Cox, and M. Mann, *Proteomic changes resulting from gene copy number variations in cancer cells.* PLoS Genet, 2010. **6**(9): p. e1001090.
41. Pan, S., et al., *Tissue proteomics in pancreatic cancer study: discovery, emerging technologies, and challenges.* Proteomics, 2013. **13**(3-4): p. 710-21.
42. Wilhelm, M., et al., *Mass-spectrometry-based draft of the human proteome.* Nature, 2014. **509**(7502): p. 582-7.
43. Gisel, A., et al., *miRNAs for the detection of multidrug resistance: overview and perspectives.* Molecules, 2014. **19**(5): p. 5611-23.
44. Zhang, Y., P. Yang, and X.F. Wang, *Microenvironmental regulation of cancer metastasis by miRNAs.* Trends Cell Biol, 2014. **24**(3): p. 153-60.
45. Benk, A.S. and C. Roesli, *Label-free quantification using MALDI mass spectrometry: considerations and perspectives.* Anal Bioanal Chem, 2012. **404**(4): p. 1039-56.
46. Pavlou, M.P. and E.P. Diamandis, *The cancer cell secretome: a good source for discovering biomarkers?* J Proteomics, 2010. **73**(10): p. 1896-906.
47. Vuckovic, D., et al., *Membrane proteomics by high performance liquid chromatography-tandem mass spectrometry: Analytical approaches and challenges.* Proteomics, 2013. **13**(3-4): p. 404-23.

48. Liu, Y., et al., *Mass spectrometric protein maps for biomarker discovery and clinical research.* Expert Rev Mol Diagn, 2013. **13**(8): p. 811-25.
49. Templin, M.F., et al., *Protein microarray technology.* Trends Biotechnol, 2002. **20**(4): p. 160-6.
50. Voshol, H., et al., *Antibody-based proteomics: analysis of signaling networks using reverse protein arrays.* FEBS J, 2009. **276**(23): p. 6871-9.
51. Zubarev, R.A., *The challenge of the proteome dynamic range and its implications for in-depth proteomics.* Proteomics, 2013. **13**(5): p. 723-6.
52. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, *Proteomics by mass spectrometry: approaches, advances, and applications.* Annu Rev Biomed Eng, 2009. **11**: p. 49-79.
53. Kim, M.S., et al., *A draft map of the human proteome.* Nature, 2014. **509**(7502): p. 575-81.
54. Gillette, M.A. and S.A. Carr, *Quantitative analysis of peptides and proteins in biomedicine by targeted mass spectrometry.* Nat Methods, 2013. **10**(1): p. 28-34.
55. Hebert, A.S., et al., *The one hour yeast proteome.* Mol Cell Proteomics, 2014. **13**(1): p. 339-47.
56. Bodenmiller, B., et al., *Multiplexed mass cytometry profiling of cellular states perturbed by small-molecule regulators.* Nat Biotechnol, 2012. **30**(9): p. 858-67.
57. Kratzer, A., H. Giral, and U. Landmesser, *High-density lipoproteins as modulators of endothelial cell functions: alterations in patients with coronary artery disease.* Cardiovasc Res, 2014.
58. Beckonert, O., et al., *Metabolic profiling, metabolomic and metabonomic procedures for NMR spectroscopy of urine, plasma, serum and tissue extracts.* Nat Protoc, 2007. **2**(11): p. 2692-703.
59. Zhou, W., L.A. Liotta, and E.F. Petricoin, *Cancer metabolism and mass spectrometry-based proteomics.* Cancer Lett, 2013.
60. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
61. Tang, N., P. Tornatore, and S.R. Weinberger, *Current developments in SELDI affinity technology.* Mass Spectrom Rev, 2004. **23**(1): p. 34-44.
62. Engholm-Keller, K. and M.R. Larsen, *Technologies and challenges in large-scale phosphoproteomics.* Proteomics, 2013. **13**(6): p. 910-31.
63. Tran, J.C., et al., *Mapping intact protein isoforms in discovery mode using top-down proteomics.* Nature, 2011. **480**(7376): p. 254-8.
64. Gillet, L.C., et al., *Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: a new concept for consistent and accurate proteome analysis.* Mol Cell Proteomics, 2012. **11**(6): p. O111 016717.
65. Deeb, S.J., et al., *Super-SILAC allows classification of diffuse large B-cell lymphoma subtypes by their protein expression profiles.* Mol Cell Proteomics, 2012. **11**(5): p. 77-89.
66. Yamaguchi, H. and M. Miyazaki, *Enzyme-immobilized reactors for rapid and efficient sample preparation in MS-based proteomic studies.* Proteomics, 2013. **13**(3-4): p. 457-66.
67. Elia, G., T. Fugmann, and D. Neri, *From target discovery to clinical trials with armed antibody products.* J Proteomics, 2014.
68. Hernandez, B., A. Parnell, and S.R. Pennington, *Why have so few proteomic biomarkers "survived" validation? (Sample size and independent validation considerations).* Proteomics, 2014.
69. Maryas, J., et al., *Proteomics in investigation of cancer metastasis: functional and clinical consequences and methodological challenges.* Proteomics, 2014. **14**(4-5): p. 426-40.
70. Diamandis, E.P., *The failure of protein cancer biomarkers to reach the clinic: why, and what can be done to address the problem?* BMC Med, 2012. **10**: p. 87.
71. Krall, N., J. Scheuermann, and D. Neri, *Small targeted cytotoxics: current state and promises from DNA-encoded chemical libraries.* Angew Chem Int Ed Engl, 2013. **52**(5): p. 1384-402.

72. Wu, C.C. and J.R. Yates, 3rd, *The application of mass spectrometry to membrane proteomics.* Nat Biotechnol, 2003. **21**(3): p. 262-7.
73. Strassberger, V., et al., *A novel reactive ester derivative of biotin with reduced membrane permeability for in vivo biotinylation experiments.* Proteomics, 2010. **10**(19): p. 3544-8.
74. Wu, C.C., et al., *A method for the comprehensive proteomic analysis of membrane proteins.* Nat Biotechnol, 2003. **21**(5): p. 532-8.
75. Strassberger, V., et al., *A comprehensive surface proteome analysis of myeloid leukemia cell lines for therapeutic antibody development.* J Proteomics, 2014. **99**: p. 138-51.
76. Kornhuber, J., et al., *The ceramide system as a novel antidepressant target.* Trends Pharmacol Sci, 2014. **35**(6): p. 293-304.
77. Nie, S., et al., *Glycoprotein biomarker panel for pancreatic cancer discovered by quantitative proteomics analysis.* J Proteome Res, 2014. **13**(4): p. 1873-84.
78. Yu, Y.Q., et al., *Enzyme-friendly, mass spectrometry-compatible surfactant for in-solution enzymatic digestion of proteins.* Anal Chem, 2003. **75**(21): p. 6023-8.
79. Eichelbaum, K., et al., *Selective enrichment of newly synthesized proteins for quantitative secretome analysis.* Nat Biotechnol, 2012. **30**(10): p. 984-90.
80. Stastna, M. and J.E. Van Eyk, *Secreted proteins as a fundamental source for biomarker discovery.* Proteomics, 2012. **12**(4-5): p. 722-35.
81. Villarreal, L., et al., *Unconventional secretion is a major contributor of cancer cell line secretomes.* Mol Cell Proteomics, 2013. **12**(5): p. 1046-60.
82. Hood, J.L., R.S. San, and S.A. Wickline, *Exosomes released by melanoma cells prepare sentinel lymph nodes for tumor metastasis.* Cancer Res, 2011. **71**(11): p. 3792-801.
83. Tiss, A., et al., *Serum peptide profiling using MALDI mass spectrometry: avoiding the pitfalls of coated magnetic beads using well-established ZipTip technology.* Proteomics, 2007. **7 Suppl 1**: p. 77-89.
84. Makawita, S., et al., *Integrated proteomic profiling of cell line conditioned media and pancreatic juice for the identification of pancreatic cancer biomarkers.* Mol Cell Proteomics, 2011. **10**(10): p. M111 008599.
85. Makridakis, M. and A. Vlahou, *Secretome proteomics for discovery of cancer biomarkers.* J Proteomics, 2010. **73**(12): p. 2291-305.
86. Marionneau, S., et al., *ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world.* Biochimie, 2001. **83**(7): p. 565-73.
87. Bustin, S.A., *Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays.* J Mol Endocrinol, 2000. **25**(2): p. 169-93.
88. Kidd, V. and T. Lion, *Debate round-table. Appropriate controls for RT-PCR.* Leukemia, 1997. **11**(6): p. 871-81.
89. Lion, T., *Current recommendations for positive controls in RT-PCR assays.* Leukemia, 2001. **15**(7): p. 1033-7.
90. Branca, R.M., et al., *HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics.* Nat Methods, 2014. **11**(1): p. 59-62.
91. Annesley, T.M., *Ion suppression in mass spectrometry.* Clin Chem, 2003. **49**(7): p. 1041-4.
92. Peterson, A.C., et al., *Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics.* Mol Cell Proteomics, 2012. **11**(11): p. 1475-88.
93. Kuzyk, M.A., et al., *Development of MRM-based assays for the absolute quantitation of plasma proteins.* Methods Mol Biol, 2013. **1023**: p. 53-82.
94. Angst, B.D., C. Marcozzi, and A.I. Magee, *The cadherin superfamily: diversity in form and function.* J Cell Sci, 2001. **114**(Pt 4): p. 629-41.
95. Takamura, M., et al., *Expression of liver-intestine cadherin and its possible interaction with galectin-3 in ductal adenocarcinoma of the pancreas.* Cancer Sci, 2003. **94**(5): p. 425-30.

96. Lee, N.P., et al., *Role of cadherin-17 in oncogenesis and potential therapeutic implications in hepatocellular carcinoma.* Biochim Biophys Acta, 2010. **1806**(2): p. 138-45.
97. Jung, R., et al., *Phylogenetic origin of LI-cadherin revealed by protein and gene structure analysis.* Cell Mol Life Sci, 2004. **61**(10): p. 1157-66.
98. Wendeler, M.W., et al., *Ksp-cadherin is a functional cell-cell adhesion molecule related to LI-cadherin.* Exp Cell Res, 2004. **294**(2): p. 345-55.
99. Zhu, R., et al., *HNF1alpha and CDX2 transcriptional factors bind to cadherin-17 (CDH17) gene promoter and modulate its expression in hepatocellular carcinoma.* J Cell Biochem, 2010. **111**(3): p. 618-26.
100. Gessner, R. and R. Tauber, *Intestinal cell adhesion molecules. Liver-intestine cadherin.* Ann N Y Acad Sci, 2000. **915**: p. 136-43.
101. Nollet, F., P. Kools, and F. van Roy, *Phylogenetic analysis of the cadherin superfamily allows identification of six major subfamilies besides several solitary members.* J Mol Biol, 2000. **299**(3): p. 551-72.
102. Bartolome, R.A., et al., *Cadherin-17 interacts with alpha2beta1 integrin to regulate cell proliferation and adhesion in colorectal cancer cells causing liver metastasis.* Oncogene, 2014. **33**(13): p. 1658-69.
103. Liu, L.X., et al., *Targeting cadherin-17 inactivates Wnt signaling and inhibits tumor growth in liver carcinoma.* Hepatology, 2009. **50**(5): p. 1453-63.
104. Panarelli, N.C., et al., *Tissue-specific cadherin CDH17 is a useful marker of gastrointestinal adenocarcinomas with higher sensitivity than CDX2.* Am J Clin Pathol, 2012. **138**(2): p. 211-22.
105. Tan, I.B., et al., *Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy.* Gastroenterology, 2011. **141**(2): p. 476-85, 485 e1-11.
106. Takamura, M., et al., *Loss of liver-intestine cadherin in human intrahepatic cholangiocarcinoma promotes angiogenesis by up-regulating metal-responsive transcription factor-1 and placental growth factor.* Int J Oncol, 2010. **36**(1): p. 245-54.
107. Yang, R.Y., G.A. Rabinovich, and F.T. Liu, *Galectins: structure, function and therapeutic potential.* Expert Rev Mol Med, 2008. **10**: p. e17.
108. Cummings, R.D. and F.T. Liu, *Galectins*, in *Essentials of Glycobiology*, A. Varki, et al., Editors. 2009: Cold Spring Harbor (NY).
109. Hokama, A., et al., *Induced reactivity of intestinal CD4(+) T cells with an epithelial cell lectin, galectin-4, contributes to exacerbation of intestinal inflammation.* Immunity, 2004. **20**(6): p. 681-93.
110. Huflejt, M.E. and H. Leffler, *Galectin-4 in normal tissues and cancer.* Glycoconj J, 2004. **20**(4): p. 247-55.
111. Barondes, S.H., et al., *Galectins. Structure and function of a large family of animal lectins.* J Biol Chem, 1994. **269**(33): p. 20807-10.
112. Danielsen, E.M. and B. van Deurs, *Galectin-4 and small intestinal brush border enzymes form clusters.* Mol Biol Cell, 1997. **8**(11): p. 2241-51.
113. Delacour, D., et al., *Galectin-4 and sulfatides in apical membrane trafficking in enterocyte-like cells.* J Cell Biol, 2005. **169**(3): p. 491-501.
114. Leffler, H., *Galectins structure and function--a synopsis.* Results Probl Cell Differ, 2001. **33**: p. 57-83.
115. Nagy, N., et al., *Refined prognostic evaluation in colon carcinoma using immunohistochemical galectin fingerprinting.* Cancer, 2003. **97**(8): p. 1849-58.
116. Huflejt, M.E., et al., *Strikingly different localization of galectin-3 and galectin-4 in human colon adenocarcinoma T84 cells. Galectin-4 is localized at sites of cell adhesion.* J Biol Chem, 1997. **272**(22): p. 14294-303.
117. Frank, M. and R. Kemler, *Protocadherins.* Curr Opin Cell Biol, 2002. **14**(5): p. 557-62.

118. Sano, K., et al., *Protocadherins: a large family of cadherin-related molecules in central nervous system.* EMBO J, 1993. **12**(6): p. 2249-56.
119. Koning, H., et al., *Characterization of protocadherin-1 expression in primary bronchial epithelial cells: association with epithelial cell differentiation.* FASEB J, 2012. **26**(1): p. 439-48.
120. Vanhalst, K., et al., *delta-Protocadherins: a gene family expressed differentially in the mouse brain.* Cell Mol Life Sci, 2005. **62**(11): p. 1247-59.
121. Koppelman, G.H., et al., *Identification of PCDH1 as a novel susceptibility gene for bronchial hyperresponsiveness.* Am J Respir Crit Care Med, 2009. **180**(10): p. 929-35.
122. Redies, C., K. Vanhalst, and F. Roy, *delta-Protocadherins: unique structures and functions.* Cell Mol Life Sci, 2005. **62**(23): p. 2840-52.
123. Flower, D.R., *The lipocalin protein family: structure and function.* Biochem J, 1996. **318 (Pt 1)**: p. 1-14.
124. Candido, S., et al., *Roles of neutrophil gelatinase-associated lipocalin (NGAL) in human cancer.* Oncotarget, 2014.
125. Tong, Z., et al., *Neutrophil gelatinase-associated lipocalin as a survival factor.* Biochem J, 2005. **391**(Pt 2): p. 441-8.
126. Tong, Z., et al., *Neutrophil gelatinase-associated lipocalin: a novel suppressor of invasion and angiogenesis in pancreatic cancer.* Cancer Res, 2008. **68**(15): p. 6100-8.
127. Cowland, J.B. and N. Borregaard, *Molecular characterization and pattern of tissue expression of the gene for neutrophil gelatinase-associated lipocalin from humans.* Genomics, 1997. **45**(1): p. 17-23.
128. Leung, L., et al., *Lipocalin2 promotes invasion, tumorigenicity and gemcitabine resistance in pancreatic ductal adenocarcinoma.* PLoS One, 2012. **7**(10): p. e46677.
129. Kaur, S., et al., *Potentials of plasma NGAL and MIC-1 as biomarker(s) in the diagnosis of lethal pancreatic cancer.* PLoS One, 2013. **8**(2): p. e55171.
130. Slater, E.P., et al., *LCN2 and TIMP1 as Potential Serum Markers for the Early Detection of Familial Pancreatic Cancer.* Transl Oncol, 2013. **6**(2): p. 99-103.
131. Xu, B., et al., *Lipocalin-2 is associated with a good prognosis and reversing epithelial-to-mesenchymal transition in pancreatic cancer.* World J Surg, 2013. **37**(8): p. 1892-900.
132. Xu, B., et al., *Treatment of pancreatic cancer using an oncolytic virus harboring the lipocalin-2 gene.* Cancer, 2012. **118**(21): p. 5217-26.
133. Furukawa, T., et al., *Long-term culture and immortalization of epithelial cells from normal adult human pancreatic ducts transfected by the E6E7 gene of human papilloma virus 16.* Am J Pathol, 1996. **148**(6): p. 1763-70.
134. Roesli, C., D. Neri, and J.N. Rybak, *In vivo protein biotinylation and sample preparation for the proteomic identification of organ- and disease-specific antigens accessible from the vasculature.* Nat Protoc, 2006. **1**(1): p. 192-9.
135. Baez-Saldana, A., et al., *Effects of biotin on pyruvate carboxylase, acetyl-CoA carboxylase, propionyl-CoA carboxylase, and markers for glucose and lipid homeostasis in type 2 diabetic patients and nondiabetic subjects.* Am J Clin Nutr, 2004. **79**(2): p. 238-43.
136. Fritz, S., et al., *Role of serum carbohydrate antigen 19-9 and carcinoembryonic antigen in distinguishing between benign and invasive intraductal papillary mucinous neoplasm of the pancreas.* Br J Surg, 2011. **98**(1): p. 104-10.
137. Lardon, J., et al., *Stem cell marker prominin-1/AC133 is expressed in duct cells of the adult human pancreas.* Pancreas, 2008. **36**(1): p. e1-6.
138. Drexler, S.K., et al., *SIGIRR/TIR-8 is an inhibitor of Toll-like receptor signaling in primary human cells and regulates inflammation in models of rheumatoid arthritis.* Arthritis Rheum, 2010. **62**(8): p. 2249-61.
139. Andoh, A., et al., *Local secretion of complement C3 in the exocrine pancreas: ductal epithelial cells as a possible biosynthetic site.* Gastroenterology, 1996. **110**(6): p. 1919-25.

140. McQueen, H.A., et al., *Stability of critical genetic lesions in human colorectal carcinoma xenografts.* Br J Cancer, 1991. **63**(1): p. 94-6.
141. Eisen, C., *Development and investigation of a novel model system representing all three subtypes of pancreatic ductal adenocarcinoma reveals novel biomarkers and distinct drug sensitivities*. 2012, Heidelberg University.
142. Zempleni, J., et al., *Sodium-dependent multivitamin transporter gene is regulated at the chromatin level by histone biotinylation in human Jurkat lymphoblastoma cells.* J Nutr, 2009. **139**(1): p. 163-6.
143. Stanger, B.Z. and Y. Dor, *Dissecting the cellular origins of pancreatic cancer.* Cell Cycle, 2006. **5**(1): p. 43-6.
144. Gu, G., J.R. Brown, and D.A. Melton, *Direct lineage tracing reveals the ontogeny of pancreatic cell fates during mouse embryogenesis.* Mech Dev, 2003. **120**(1): p. 35-43.
145. Tonack, S., et al., *iTRAQ reveals candidate pancreatic cancer serum biomarkers: influence of obstructive jaundice on their performance.* Br J Cancer, 2013. **108**(9): p. 1846-53.
146. Schmid, J.A. and A. Birbach, *Fluorescent proteins and fluorescence resonance energy transfer (FRET) as tools in signaling research.* Thromb Haemost, 2007. **97**(3): p. 378-84.
147. Periasamy, A., *Fluorescence resonance energy transfer microscopy: a mini review.* J Biomed Opt, 2001. **6**(3): p. 287-91.
148. Thymiakou, E. and V. Episkopou, *Detection of signaling effector-complexes downstream of bmp4 using PLA, a proximity ligation assay.* J Vis Exp, 2011(49).
149. Baan, B., et al., *In situ proximity ligation detection of c-Jun/AP-1 dimers reveals increased levels of c-Jun/Fra1 complexes in aggressive breast cancer cell lines in vitro and in vivo.* Mol Cell Proteomics, 2010. **9**(9): p. 1982-90.
150. Surinova, S., et al., *On the development of plasma protein biomarkers.* J Proteome Res, 2011. **10**(1): p. 5-16.