Mass Spectrometry-Based Protein Biomarker Discovery in Pancreatic Cancer

Zhou, Qimin

2020

Document Version:
Publisher's PDF, also known as Version of record

Mass Spectrometry-Based Protein Biomarker Discovery in Pancreatic Cancer

QIMIN ZHOU
DEPARTMENT OF SURGERY, CLINICAL SCIENCES LUND | LUND UNIVERSITY

Mass Spectrometry-Based Protein Biomarker Discovery in Pancreatic Cancer

Qimin Zhou, M.D.

DOCTORAL DISSERTATION

by due permission of the Faculty of Medicine, Lund University, Sweden.
To be defended at Lecture Room 3, Main building, Skåne University Hospital,
Lund, April 28, 2020, at 1:00 pm.

Faculty opponent
Professor Malin Sund, Department of Surgical and Perioperative Sciences,
Umeå University, Umeå, Sweden

Main Supervisor: Associate Professor Daniel Ansari
Co-supervisor: Professor Roland Andersson
Co-supervisor: Professor György Marko-Varga

Organisation
LUND UNIVERSITY
Department of Surgery, Clinical Sciences Lund, Lund University, Lund, Sweden
Author: Qimin Zhou, M.D.

Document name
Doctoral dissertation

Date of issue
April 28, 2020

Sponsoring organisation

Title and subtitle: Mass Spectrometry-Based Protein Biomarker Discovery in Pancreatic Cancer

Abstract

Background: Pancreatic cancer has the lowest survival rate among all the major cancer types. Although recent decades have seen advances in diagnostic imaging, surgical techniques, perioperative care and oncological treatment, this has not been translated into major improvements in clinical outcome. The 5-year survival rate remains less than 10% for all stages. One important unmet clinical need is biomarkers of clinical utility that can be used for early detection, prognostication and guidance of treatment.

Aim: The aim of this thesis was to develop and validate protein biomarkers for diagnosis, prognosis and prediction of treatment response in pancreatic cancer.

Methods: Mass spectrometry (MS)-based proteomic profiling of fresh frozen tissue specimens from pancreatic cancer patients and control subjects was conducted to identify potential protein biomarkers. These were subsequently verified by targeted proteomics (parallel reaction monitoring (PRM)) and bioinformatic analysis. Selected biomarker candidates were further validated in larger patient cohorts by tissue microarray-based immunohistochemistry studies, serum immunoassay measurements and in vitro experiments.

Results/conclusions:

(I) A proteolytic digestion protocol was optimised for MS-based proteomics studies. Urea in-solution digestion at room temperature (24 ± 2 °C) was found to be superior to traditional proteolysis at 37 °C, presenting several advantages such as fewer experimentally-induced post-translational modifications (carbamylation and pyroglutamic acid modifications), increased identification of peptides and proteins, and improved protein quantification by reducing coefficients of variations.

(II) Some 165 potential protein biomarkers were identified in pancreatic cancer tissues and a panel of 45 biomarker candidates was verified by targeted MS. The novel protein BASP1 was significantly associated with favourable survival and positive response to adjuvant chemotherapy in pancreatic cancer patients. Bioinformatic analysis indicated that BASP1 interacts with Wilms tumour protein WT1. Patients with negative BASP1 and high WT1 expression had the poorest outcomes.

(III) Prognostic analysis of YAP1 demonstrated a significant correlation with lower survival, at both mRNA expression levels (TCGA cohort) and protein expression levels (Lund cohort). Inhibiting the YAP1/TEAD interaction interfered with the expression of AREG, CTGF, CYR61 and MSLN in pancreatic cancer cells, which suggests that YAP1 transcriptional activity may affect the evolution and persistence of a fibrotic tumour microenvironment.

(IV) Expression of AGP1 in pancreatic cancer tissues is significantly correlated with poor survival. Circulating levels of AGP1 and CA 19-9 yielded a high diagnostic accuracy (AUC 0.963) for discrimination of resectable pancreatic cancer patients against healthy controls.

Keywords pancreatic cancer, proteomics, mass spectrometry, biomarkers, diagnosis, prognosis, prediction, BASP1, WT1, YAP1, AGP1

Classification system and/or index terms (if any)

Supplementary bibliographical information		Language: English
ISSN and key title 1652-8220		ISBN 978-91-7619-908-4
Recipient's notes	Number of pages 84	Price
	Security classification	

Mass Spectrometry-Based Protein Biomarker Discovery in Pancreatic Cancer

Qimin Zhou, M.D.

LUND
UNIVERSITY

Cover photo (3D protein structure of BASP1) by courtesy of Dr Daniel Ansari.
The image was created using Protean 3D software from DNASTAR Inc.

Department of Surgery, Clinical Sciences Lund,
Lund University, Lund, Sweden

ISBN 978-91-7619-908-4
ISSN 1652-8220

Printed in Sweden by Media-Tryck, Lund University
Lund 2020

The man who has made up his mind to win will never say ‘impossible’.

- Napoleon Bonaparte

有志者，事竟成，破釜沉舟，百二秦关终属楚；

苦心人，天不负，卧薪尝胆，三千越甲可吞吴。

——蒲松龄

Table of Contents

Original Articles

The thesis is based on the following papers, which are referred to in the text by their Roman numerals. The papers are appended at the end of the thesis:

I. Betancourt LH, Sanchez A, Pla I, Kuras M, **Zhou Q**, Andersson R, Marko-Varga G. Quantitative assessment of urea in-solution Lys-C/trypsin digestions reveals superior performance at room Temperature over traditional proteolysis at 37 °C. Journal of Proteome Research 2018;17:2556-61.

II. **Zhou Q**, Andersson R, Hu D, Bauden M, Kristl T, Sasor A, Pawlowski K, Pla I, Said Hilmersson K, Zhou M, Lu F, Marko-Varga G, Ansari D. Quantitative proteomics identifies brain acid-soluble protein 1 (BASP1) as a prognostic biomarker candidate in pancreatic cancer. EBioMedicine 2019;43:282-94.

III. **Zhou Q**, Bauden M, Andersson R, Hu D, Marko-Varga G, Xu J, Sasor A, Dai H, Pawlowski K, Shen X, Said Hilmersson K, Chen X, Ansari D. YAP1 is an independent prognostic marker in pancreatic cancer and associated with extracellular matrix remodelling. Journal of Translational Medicine 2020;18:77.

IV. **Zhou Q**, Andersson R, Hu D, Bauden M, Sasor A, Bygott T, Pawlowski K, Pla I, Marko-Varga G, Ansari D. Alpha-1-acid glycoprotein 1 is upregulated in pancreatic ductal adenocarcinoma and confers a poor prognosis. Translational Research 2019;212:67-79.

Abstract

Background: Pancreatic cancer has the lowest survival rate among all the major cancer types. Although recent decades have seen advances in diagnostic imaging, surgical techniques, perioperative care and oncological treatment, this has not been translated into major improvements in clinical outcome. The 5-year survival rate remains less than 10% for all stages. One important unmet clinical need is biomarkers of clinical utility that can be used for early detection, prognostication and guidance of treatment.

Aim: The aim of this thesis was to develop and validate protein biomarkers for diagnosis, prognosis and prediction of treatment response in pancreatic cancer.

Methods: Mass spectrometry (MS)-based proteomic profiling of fresh frozen tissue specimens from pancreatic cancer patients and control subjects was conducted to identify potential protein biomarkers. These were subsequently verified by targeted proteomics (parallel reaction monitoring (PRM)) and bioinformatic analysis. Selected biomarker candidates were further validated in larger patient cohorts by tissue microarray-based immunohistochemistry studies, serum immunoassay measurements and in vitro experiments.

Results/conclusions: (I) A proteolytic digestion protocol was optimised for MS-based proteomics studies. Urea in-solution digestion at room temperature (24 ± 2 °C) was found to be superior to traditional proteolysis at 37 °C, presenting several advantages such as fewer experimentally-induced post-translational modifications (carbamylation and pyroglutamic acid modifications), increased identification of peptides and proteins, and improved protein quantification by reducing coefficients of variations.

(II) Some 165 potential protein biomarkers were identified in pancreatic cancer tissues and a panel of 45 biomarker candidates was verified by targeted MS. The novel protein BASP1 was significantly associated with favourable survival and positive response to adjuvant chemotherapy in pancreatic cancer patients. Bioinformatic analysis indicated that BASP1 interacts with Wilms tumour protein WT1. Patients with negative BASP1 and high WT1 expression had the poorest outcomes.

(III) Prognostic analysis of YAP1 demonstrated a significant correlation with lower survival, at both mRNA expression levels (TCGA cohort) and protein expression levels (Lund cohort). Inhibiting the YAP1/TEAD interaction interfered with the expression of AREG, CTGF, CYR61 and MSLN in pancreatic cancer cells, which suggests that YAP1 transcriptional activity may affect the evolution and persistence of a fibrotic tumour microenvironment.

(IV) Expression of AGP1 in pancreatic cancer tissues is significantly correlated with poor survival. Circulating levels of AGP1 and CA 19-9 yielded a high diagnostic accuracy (AUC 0.963) for discrimination of resectable pancreatic cancer patients against healthy controls.

Selected abbreviations

ACN	acetonitrile
ADEX	aberrantly differentiated endocrine exocrine
AGC	automatic gain control
AGP1	alpha-1-acid glycoprotein 1
AJCC	American Joint Committee on Cancer
Ambic	ammonium bicarbonate
AUC	area under the curve
BASP1	brain acid soluble protein 1
CA 19-9	carbohydrate antigen 19-9
CI	confidence interval
CT	computed tomography
ctDNA	circulating tumour DNA
CV	coefficient variation
DDA	data-dependent acquisition
DFS	disease-free survival
DTT	dithiothreitol
ECLIA	electro-chemiluminescence immunoassay
ELISA	enzyme-linked immunosorbent assay
ESI	electrospray ionisation
EUS	endoscopic ultrasonography
FA	formic acid
FDR	false discovery rate
FFPE	formalin-fixed paraffin-embedded
FPKM	Fragments Per Kilobase of exon per Million reads
FTICR	Fourier-transform ion cyclotron resonance
GC	gas chromatography
HPLC	high-performance liquid chromatography
HR	hazard ratio
IAA	iodoacetamide
IF	immunofluorescence
IHC	immunohistochemistry
IPA	Ingenuity Pathway Analysis
IPMN	intraductal papillary mucinous neoplasm
LC	liquid chromatography

LC-MS/MS	liquid chromatography tandem mass spectrometry
LIT	linear ion trap
LTQ-Orbitrap	linear trap quadrupole-Orbitrap
MALDI	matrix-assisted laser desorption/ionisation
MCN	mucinous cystic neoplasm
MS	mass spectrometry
MSI-H/dMMR	microsatellite instability high/deficient mismatch repair
NCE	normalised collision energy
OS	overall survival
PanIN	pancreatic intraepithelial neoplasia
PCA	Principal Component Analysis
PRM	parallel reaction monitoring
PRTC	peptide retention time mixture
PSMs	peptide-to-spectrum matches
PTMs	post-translational modifications
RCS	retrospective cohort study
ROC	Receiver Operator Characteristic
RT	room temperature
RT qPCR	real-time quantitative PCR
TCGA	The Cancer Genome Atlas
TMA	tissue microarray
TNM	Tumour, Node, Metastasis.
TOF	time-of-flight
TQ	triple quadrupole
WT1	wilms tumour protein
YAP1	yes-associated protein 1

1. Introduction

1.1 The pancreas: history, anatomy and physiology

The pancreas was first described as a distinct organ by the Greeks about 300BCE (1). However, its function remained unknown for many centuries. In 1642, Johann Wirsüng discovered the main pancreatic duct (2), and in 1724, Giovanni Santorini described the accessory pancreatic duct, but the priority of the discovery is still a matter of controversy (3, 4). In 1846, Claude Bernard demonstrated the role of the pancreas in digestion (5), and in 1869, Paul Langerhans described the pancreatic islets (6). Frederick Banting and Charles Best discovered insulin in 1921 (7).

The pancreas is a retroperitoneal organ that lies transversely in the upper abdomen with a length of 15-20 cm and an average weight of about 75-100 grams (see Figure 1). Anatomically, it can be divided into head, neck, body and tail. The head lies in the C-loop of the duodenum. The neck of the pancreas is the part of the organ located anterior to the confluence of the superior mesenteric-portal vein, and the body extends to the left from the pancreatic neck. The tail of the pancreas is the narrowest part and continues towards the hilum of the spleen. The main pancreatic duct runs from the tail to the head where it merges with the common bile duct to open at the major duodenal papilla. The accessory pancreatic duct begins within the head of the pancreas and opens at the minor duodenal papilla.

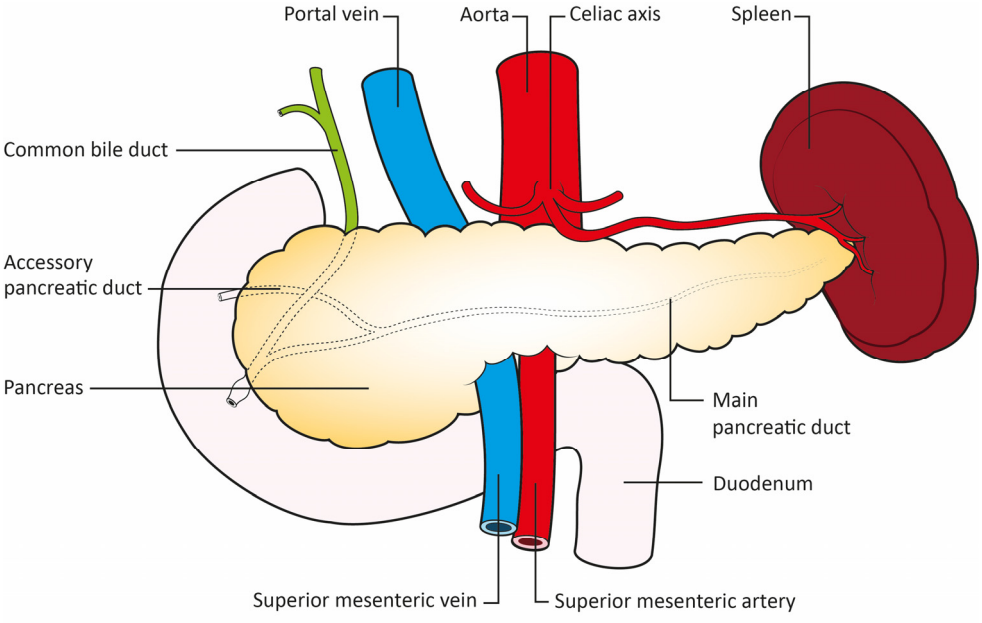


Figure 1. Anatomy of the pancreas. Image courtesy of Dr Daniel Ansari and medical illustrator Jan Funke.

The pancreas is a unique gland that has both exocrine and endocrine functions (Figure 2). The exocrine tissue consists of acini which secrete pancreatic juice into small pancreatic ducts which converge to form larger ducts and ultimately the main pancreatic duct that carries the pancreatic juice to the duodenum. Approximately 1.5-2 L of pancreatic juice is produced every day. Pancreatic juice contains water, electrolytes and enzymes. Some pancreatic enzymes (amylase, lipase, deoxyribonuclease and ribonuclease) are secreted by the acinar cells in their active forms, while others (trypsinogen, chymotrypsinogen, procarboxypeptidase and proelastase) are secreted as inactive proenzymes which are activated in the lumen of the proximal intestine. The activated enzymes aid in food digestion. The pancreatic exocrine section is regulated by hormonal factors including secretin and cholecystokinin and by neural factors.

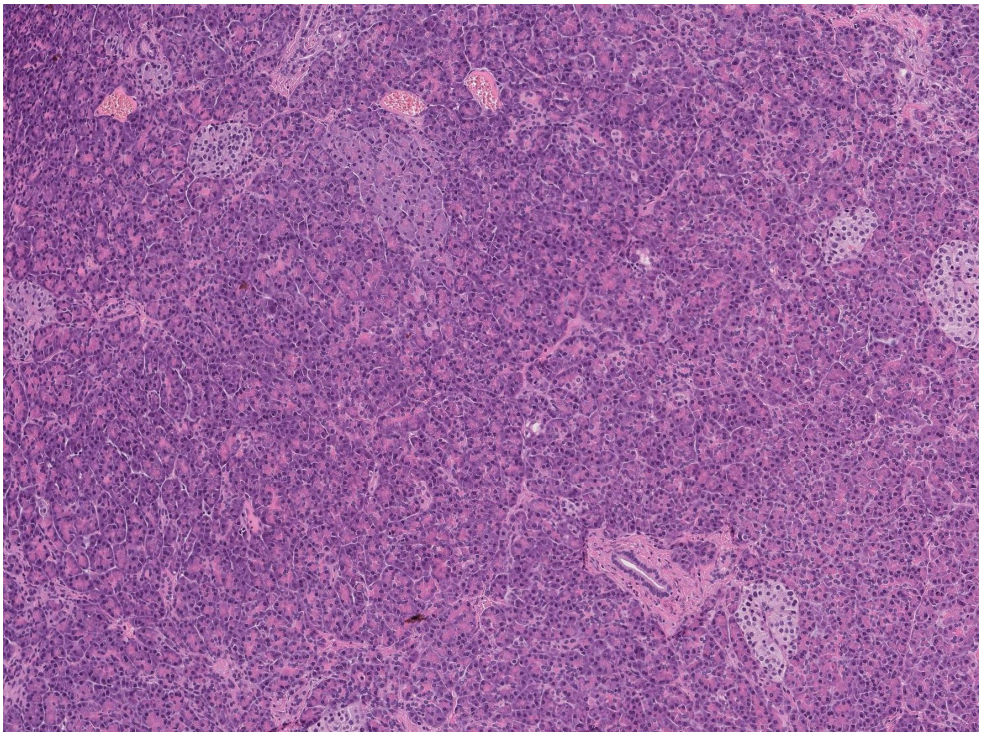


Figure 2. Normal pancreas. Acini occupy most of the microscopic field but small pancreatic ducts and islets of Langerhans are also present. Image courtesy of Dr Agata Sasor.

The endocrine pancreas is composed of the islets of Langerhans which make up less than 2% of the total pancreatic mass. The islets contain 5 major hormone-producing cells. These include alpha cells involved in the secretion of glucagon, beta cells (insulin), delta cells (somatostatin), PP cells (pancreatic polypeptide) and epsilon cells (ghrelin). Glucagon increases blood glucose concentrations and insulin lowers them. Somatostatin inhibits the release of glucagon and insulin. Pancreatic polypeptide can regulate both exocrine and endocrine pancreatic secretions. Ghrelin has several functions including the stimulation of appetite and growth hormone release.

1.2 Pancreatic cancer

The first known description of pancreatic cancer is credited to Giovanni Battista Morgagni in 1761 (8). However, the lack of a microscopic evaluation makes the true diagnosis uncertain, and it was not until 1858 when Jacob Mendez Da Costa revisited Morgagni's original work and provided a histological diagnosis that pancreatic cancer could be established as a genuine disease entity (9). Histologically, most pancreatic cancers (about 90%) are ductal adenocarcinomas arising within the

exocrine component of the pancreas (Figure 3), which is the focus of this study. Over recent decades, considerable progress has been made in the clinical, pathological and molecular understanding of pancreatic ductal adenocarcinoma, but the prognosis of those with the disease remains very poor (10).

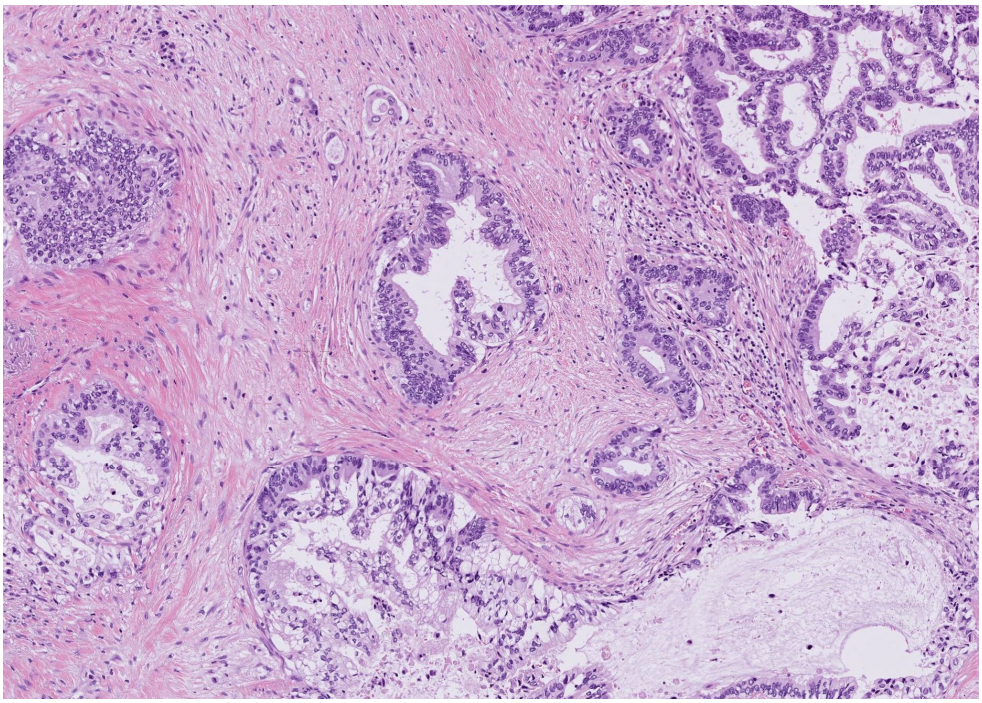


Figure 3. Pancreatic ductal adenocarcinoma. Poorly formed glands are present in a relatively abundant stroma. Notice the cribriform pattern in the upper right corner. Image courtesy of Dr Agata Sasor.

1.3 Epidemiology and risk factors

The incidence of pancreatic cancer is 13.3 cases per 100,000 individuals annually in Sweden (11). In China, the incidence is lower, but it has markedly increased in recent years (12). Although the mortality rates for most cancer types have declined over recent decades, those for pancreatic cancer have risen (13). Pancreatic cancer has now surpassed breast cancer to become the third leading cause of cancer-related mortality in Western countries (14). Projections for 2030 suggest that pancreatic cancer will become the second leading cause of cancer death if no major advances in the diagnosis and treatment of the disease are made (15). The current 5-year survival rate for pancreatic cancer is 9%, according to the American Cancer Society (14). This is the lowest survival rate among major organ cancers (see Figure 4).

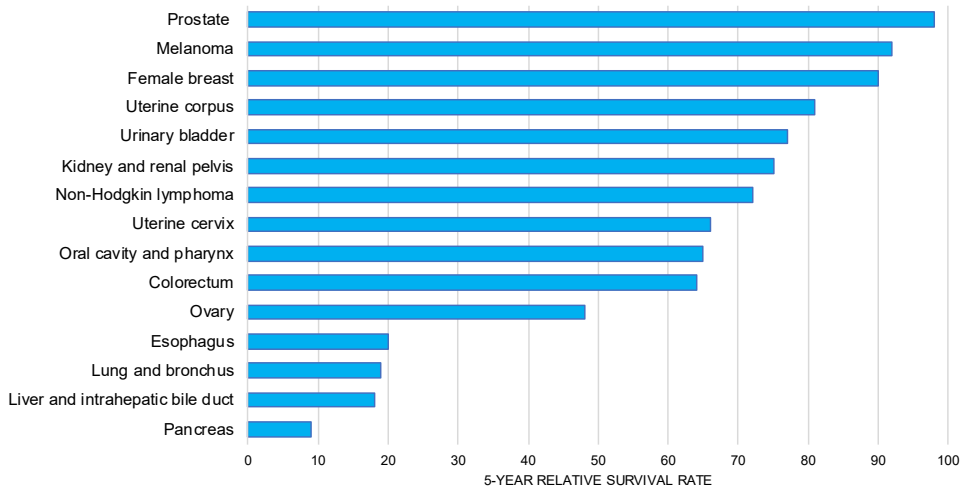


Figure 4. Five-year relative survival rates for major organ cancers in the United States (2009 to 2015). Pancreatic cancer is associated with the lowest survival rate. Adapted from reference (14).

Pancreatic cancer is an age-related disease with a peak incidence between 60 and 80 years of age. It is uncommon in people below 50, and patients in this age group make up only 6% of cases (16). The causes remain unknown, but the best-established risk factors are smoking, diabetes mellitus, obesity, chronic pancreatitis and pancreatic cystic lesions (10, 17). Approximately 5-10% of patients have an inherited predisposition, such as familial pancreatic cancer (BRCA2, PALB2), hereditary pancreatitis (PRSS1, SPINK1), familial atypical multiple mole melanoma syndrome (CDKN2A), hereditary breast and ovarian cancer syndrome (BRCA1, BRCA2, PALB2), Peutz-Jeghers syndrome (STK11), Lynch syndrome (MLH1, MSH2, MSH6), Li-Fraumeni syndrome (TP53) and ataxia-telangiectasia (ATM) (18, 19).

1.4 Tumour development and molecular pathology

Pancreatic ductal adenocarcinoma is known to arise from non-invasive precursor lesions, the most important of which is pancreatic intraepithelial neoplasia (PanIN) (20). Others include intraductal papillary mucinous neoplasm (IPMN) and mucinous cystic neoplasm (MCN) (21).

PanINs represent a series of proliferative intraepithelial lesions within the pancreatic ducts, which can be subcategorised into several stages from PanIN-1A to PanIN-3 (see Figure 5). PanIN-1A is composed of a flat mucinous epithelium, while the PanIN-1B lesion shows papillary hyperplasia. In PanIN-2 lesions, the cells begin to display moderate nuclear changes, such as loss of nuclear polarity, nuclear crowding,

pleomorphism, pseudostratification and hyperchromasia. PanIN-3 lesions (carcinoma *in situ*) are characterised by extensive nuclear atypia, loss of polarity and increased mitosis.

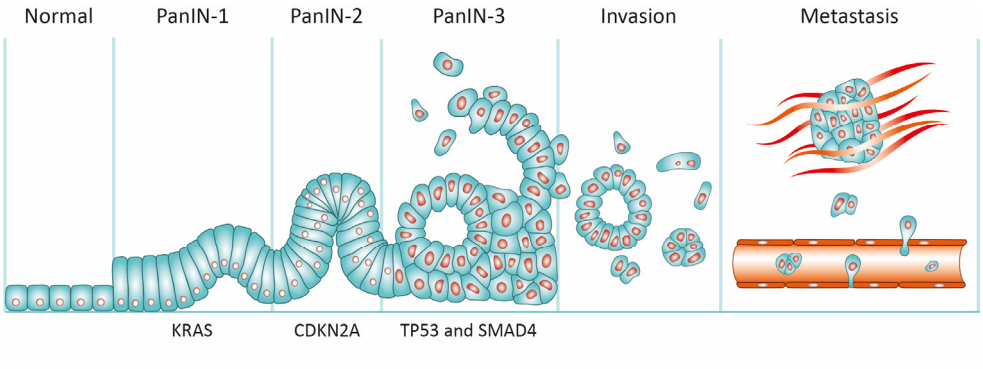


Figure 5. Development of pancreatic cancer from precursor lesions (PanINs) to metastatic disease. Image courtesy of Dr Daniel Ansari and medical illustrator Jan Funke.

More than 90% of PanIN lesions harbour KRAS mutations, which is considered to be an early genetic event. Subsequent mutations in CDKN2A, TP53, and SMAD4 tumour suppressors are observed with increasing frequency in PanIN-2 and PanIN-3 lesions. The stepwise PanIN progression model is well-established (20, 22, 23), but has been challenged by those suggesting a more rapid tumour progression (24).

Pancreatic cancer is a heterogeneous disease. Recent genomic and transcriptomic analyses have defined potential subgroups with distinct biology and therapeutic vulnerabilities. In 2011, Collisson et al. (25) proposed three epithelial subtypes: classical, quasimesenchymal and exocrine-like. In 2015, Moffitt et al. (26) proposed epithelial subtypes (classical and basal-like) and stromal subtypes (normal and activated). In 2016, Bailey et al. (27) defined four molecular subtypes: pancreatic progenitor, immunogenic, squamous and aberrantly differentiated endocrine exocrine (ADEX). In 2019, Maurer et al. (28) revisited the previous molecular classifications to provide a unified molecular taxonomy. Evidence was found of certain epithelial subtypes (classical and basal-like) and stromal subtypes (immune-rich and extracellular matrix-rich), but the exocrine/ADEX subtype and immunogenic subtype could not be verified and were likely due to non-transformed microenvironmental cells.

1.5 Clinical presentation

Pancreatic cancer is often referred to as the ‘silent killer’ because most patients do not experience symptoms until the tumour is well-advanced. Pain is a typical presenting symptom and tends to correlate with perineural invasion by tumour cells

and advanced disease (29). Tumours located in the head of the pancreas may present with obstructive jaundice. Approximately 80% of pancreatic cancer patients have a pathological glucose tolerance test or frank diabetes at the time of diagnosis (30). The occurrence of new-onset diabetes in patients above 50 may be a harbinger of pancreatic cancer and provide a clue to early detection (31). Additional signs and symptoms of pancreatic cancer include weight loss, lethargy, panniculitis, depression and gastric outlet obstruction (32). Pancreatic cancer has been linked to venous thrombosis and some patients may develop migratory thrombophlebitis (Trousseau's syndrome) (33).

1.6 Imaging

Multidetector computed tomography (CT) is the preferred method for initial diagnosis and staging of pancreatic cancer. The standard triphasic protocol for pancreatic assessment involves arterial, late arterial and venous phases, where the tumour is visualised in relation to major vessels (the celiac axis, superior mesenteric artery and vein, and the portal vein) and to distant organs. Additional diagnostic investigations may be required in some patients. Endoscopic ultrasonography (EUS) is valuable when pancreatic cancer is suspected but there is no detectable lesion on CT. EUS is also the recommended method for obtaining tissue for diagnostic purposes. A tissue biopsy is usually not needed in patients who are planned for upfront surgery, but is required before the initiation of chemotherapy.

1.7 Staging

Pancreatic cancer is staged using the American Joint Committee on Cancer (AJCC) TNM classification system (Tables 1 and 2) (34). The tumours are further categorised into resectable, borderline resectable and unresectable groups. Resectable pancreatic cancer includes most stage I or II tumours. Borderline resectable pancreatic cancer indicates that the tumour may have spread to nearby blood vessels, but is still considered to be surgically removable. Unresectable tumours can be either locally advanced (stage III) or metastatic (stage IV).

Table 1. AJCC TNM classification, 8th edition (34)

T = primary tumour	N = regional lymph node	M = distant metastasis
TX: primary tumour cannot be assessed **T0:** No evidence of primary tumour **Tis:** carcinoma *in situ* (includes PanIN-3, IPMN with high-grade dysplasia, ITPN with high-grade dysplasia, and MCN with high-grade dysplasia) **T1:** tumour ≤2 cm in greatest dimension **T1a:** tumour ≤0.5 cm in greatest dimension **T1b:** tumour >0.5 and <1 cm in greatest dimension **T1c:** tumour 1-2 cm in greatest dimension **T2:** tumour >2 cm and ≤4 cm in greatest dimension **T3:** tumour >4 cm in greatest dimension **T4:** tumour involves celiac axis, superior mesenteric artery and/or common hepatic artery	**NX:** Regional lymph nodes cannot be assessed **N0:** no regional lymph node metastasis **N1:** Metastasis in 1-3 regional lymph nodes **N2:** Metastasis in ≥4 regional lymph nodes	**M0:** no distant metastasis **M1:** distant metastasis

AJCC, American Joint Committee on Cancer; IPMN, intraductal papillary mucinous neoplasm; ITPN, intraductal tubulopapillary neoplasm; MCN, mucinous cystic neoplasm; PanIN, pancreatic intraepithelial neoplasia; TNM, Tumour, Node, Metastasis.

Table 2. AJCC stage groups (34)

T	N	M	Stage
Tis	N0	M0	0
T1	N0	M0	IA
T1	N1	M0	IIB
T1	N2	M0	III
T2	N0	M0	IB
T2	N1	M0	IIB
T2	N2	M0	III
T3	N0	M0	IIA
T3	N1	M0	IIB
T3	N2	M0	III
T4	Any N	M0	III
Any T	Any N	M1	IV

AJCC, American Joint Committee on Cancer.

1.8 Treatment

Clinical management of pancreatic cancer requires a multidisciplinary approach. Standard treatment options include surgery, chemotherapy, chemoradiotherapy and supportive care (Figure 6). The choice of treatment depends on the stage and the general condition of the patient. Targeted therapy and immunotherapy are emerging treatment options in selected patients with specific molecular alterations.

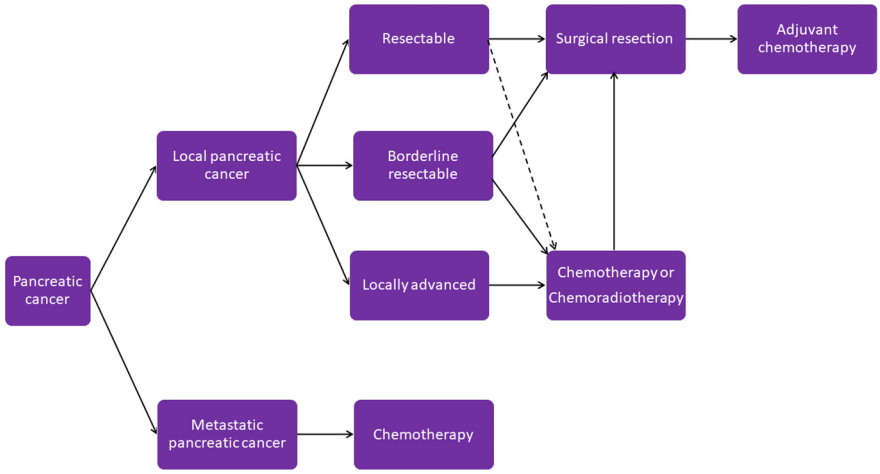


Figure 6. Standard treatment options in pancreatic cancer.

1.8.1 Surgery

Surgical resection is the only potentially curative therapy for pancreatic cancer. The choice of operative procedure is based on the location of the tumour and may involve pancreatoduodenectomy (the Whipple procedure), distal pancreatectomy, or total pancreatectomy. For selected patients with vascular involvement, it is appropriate to perform vascular resection and reconstruction if R0 resection can be achieved (35).

1.8.2 Chemotherapy

Chemotherapy is an integral part of pancreatic cancer treatment and may be used at any stage. For patients with resectable disease, chemotherapy has traditionally been used after surgery (adjuvant chemotherapy). The use of chemotherapy before surgery (neoadjuvant chemotherapy) in upfront resectable disease is currently being evaluated in clinical trials (36). For patients with advanced pancreatic cancer, chemotherapy can be used to downstage the tumour for surgery or as palliative treatment. Gemcitabine remains the most widely used agent (37, 38), but many other chemotherapeutic options are available with proven clinical benefit, including FOLFIRINOX (39, 40), capecitabine (41), albumin-bound paclitaxel (42) and liposomal irinotecan (43).

1.8.3 Radiotherapy

Radiotherapy is usually administered together with chemotherapy. Chemoradiotherapy is typically reserved for patients with borderline or locally

advanced pancreatic cancer. However, the clinical benefits of chemoradiotherapy remain unclear and there is little evidence of clear survival improvement (10).

1.8.4 Targeted therapy and immunotherapy

Molecularly targeted therapy is still in its infancy. The EGFR inhibitor erlotinib has shown modest survival benefits when used in combination with gemcitabine in patients with advanced pancreatic cancer (44). A small subset of patients with mutations in the BRCA1 or BRCA2 genes may benefit from the PARP inhibitor olaparib, which can be used for the treatment of advanced pancreatic cancer that has not progressed after at least 4 months of first-line platinum-based chemotherapy (45, 46). A small subset of patients with an NTRK gene fusion may benefit from NTRK inhibitors such as larotrectinib and entrectinib (47). The immune checkpoint inhibitor pembrolisumab may also be used in a small subset of patients that display microsatellite instability high/deficient mismatch repair (MSI-H/dMMR) (48).

1.8.5 Supportive care

Most patients with pancreatic cancer will eventually require palliative care. Biliary and duodenal obstruction is treated with surgical, endoscopic or radiological interventions. Management of pain is important and, apart from conventional pharmacotherapy, it may be possible to use a neurolytic celiac plexus block.

1.9 Biomarkers for pancreatic cancer

A biomarker is defined as a ‘characteristic that is measured as an indicator of normal biological processes, pathogenic processes or responses to an exposure or intervention’ (49). Biomarkers for pancreatic cancer may be genes, RNA, proteins or metabolites and they can be measured in tumour tissue and body fluids such as blood, urine, saliva and pancreatic cyst fluid. Tumour tissue has the highest concentration of tumour biomarkers and one strategy is to first search for biomarkers in cancer tissue and then look for these tumour-derived biomarkers in non-invasive biofluids such as blood.

1.9.1 Diagnostic biomarkers

The purpose of a diagnostic biomarker is to detect or confirm the presence of a disease or condition, or to identify a specific disease subtype. The STARD guidelines should be followed when developing potential diagnostic biomarkers (50).

There is currently no approved biomarker for the early detection of pancreatic cancer. Carbohydrate antigen 19-9 (CA 19-9) is the only biomarker routinely used in the clinical management of pancreatic cancer and can be used for disease monitoring. However, CA 19-9 is not recommended for diagnosing pancreatic cancer due to limited sensitivity and specificity, lack of expression in patients with a Lewis-negative genotype and increased expression in several benign conditions (51). The diagnostic value of CA 19-9 can be improved when used in combination with additional markers. Table 3 shows a list of investigational diagnostic markers in pancreatic cancer.

Table 3. Selected blood-based markers for diagnosis of pancreatic cancer

Biomarker	Category	Technique	AUC	Reference
CEMIP	Protein	ELISA	0.94	(52)
MIC-1	Protein	ELISA, ECLIA	0.89	(53)
TSP-1 + CA 19-9	Protein	MS	0.86	(54)
TSP-2 + CA 19-9	Protein	ELISA, MS	0.96	(55)
29-protein biomarker panel	Protein	Antibody microarray	0.96	(56)
15 microRNA panel	microRNA	RT qPCR	0.96 at diagnosis; 0.60 <5 yrs before diagnosis	(57)
4-nucleosome panel + CA 19-9	Nucleosome	ELISA	0.98	(58)
GPC1(+) circulating exosomes	Exosome	MS, flow cytometry	1.0	(59)
CancerSEEK	16 ctDNA mutations + 8 proteins	Multianalyte test	Sensitivity 70% at 99.5% specificity	(60)

AUC, area under the curve; ctDNA, circulating tumour DNA; ECLIA, electro-chemiluminescence immunoassay; ELISA, enzyme-linked immunosorbent assay; MS, mass spectrometry; RT qPCR, real-time quantitative PCR.

1.9.2 Prognostic biomarkers

The purpose of a prognostic biomarker is to provide information on the likelihood of a clinical event such as disease recurrence or death, irrespective of the treatment. The REMARK guidelines should be followed when developing prognostic markers (61).

Several hundred prognostic biomarker candidates have been reported in pancreatic cancer tissue, of which the most consistent and potentially informative are shown in Table 4.

Table 4. Selected tissue-based prognostic markers in pancreatic cancer

Marker	HR for OS	P-value	Reference
Bax	0.31 (0.17-0.56)	<0.001	(62)
Bcl-2	0.41 (0.27-0.63)	<0.001	(62)
COX-2	1.39 (1.13-1.71)	0.002	(62)
E-cadherin	1.80 (1.33-2.42)	<0.001	(62)
Ki-67	2.42 (1.87-3.14)	0.005	(62)
MUC4	2.01 (1.42-2.86)	<0.001	(63)
p16	0.63 (0.43-0.92)	0.02	(64)
S100A2	3.23 (1.58-6.62)	0.001	(62)
Stromal SPARC	1.53 (1.05-2.24)	0.03	(65)
Survivin	0.46 (0.29-0.73)	0.001	(62)
VEGF	1.51 (1.18-1.92)	0.001	(64)

HR, hazard ratio; OS, overall survival.

1.9.3 Predictive biomarkers

The purpose of a predictive biomarker is to provide information on the likely clinical benefit of a specific treatment. Recent years have seen a rapid introduction of predictive biomarkers in the management of several different malignancies, including BRAF V600 mutations in melanoma and EGFR, ALK, ROS1, KRAS and BRAF mutations in lung cancer.

A list of potential predictive biomarkers is shown in Table 5.

Table 5. Potential predictive biomarkers in pancreatic cancer

Drug	Predictive biomarker	References
Gemcitabine	hENT1	(66-69)
FOLFIRINOX	TS (5-FU), CES2 (irinotecan), BRCA1/2 (oxaliplatin), PALB2 (oxaliplatin)	(70-74)
Erlotinib	No established marker	(75, 76)
Nab-paclitaxel	No established marker	(77)
PARP inhibitor	BRCA1/2	(45, 46)
NTRK inhibitor	NTRK	(47)
Immune checkpoint inhibitor	PD-L1, MSI-H/dMMR	(48, 78-81)

1.10 Proteomics

Proteins are the major functional units of all living organisms. They have important roles in the maintenance of life and their dysfunction contributes to the development of numerous diseases. Proteomics is defined as the large-scale study of all proteins in a given organism or biological system. Progress within the field of proteomics is dependent on advances in proteomics technology and instrumentation.

1.11 Mass spectrometry

Mass spectrometry is a fundamental technological tool in proteomics research. Modern mass spectrometry provides great depth of proteome analysis which cannot be achieved with other methods such as two-dimensional gel electrophoresis or protein microarrays.

Mass spectrometry identifies and quantifies molecules based on their mass-to-charge (m/z) ratio. A mass spectrometer consists of three parts: 1) an ion source that converts analyte molecules into gas-phase ions; 2) a mass analyser that separates ionised analytes based on m/z ratio, and 3) a detector. The development of electrospray ionisation (ESI) and matrix-assisted laser desorption/ionisation (MALDI) have transformed modern MS-based protein analysis.

The mass analyser is a key feature of MS methodology. There are several types of mass analysers including the linear ion trap (LIT), linear trap quadrupole-Orbitrap (LTQ-Orbitrap), triple quadrupole (TQ), time-of-flight (TOF), and Fourier-transform ion cyclotron resonance (FTICR) mass analyser. Their analytical performances are summarised in Table 6.

Table 6. Performance characteristics of commonly used mass spectrometers (82)

Instrument	Sensitivity	Mass resolution	Mass accuracy (ppm)
LIT	Femtomole	2000	100
LTQ-Orbitrap	Femtomole	100 000	2
TQ	Attomole	2000	100
TOF	Attomole	10 000	2-5
FTICR	Femtomole	500 000	<2

Separation technologies are very important for protein MS analysis of complex biological samples. The two most common methods of separation are gas chromatography (GC) and liquid chromatography (LC).

Figure 7 describes a typical proteomics workflow using clinical samples.

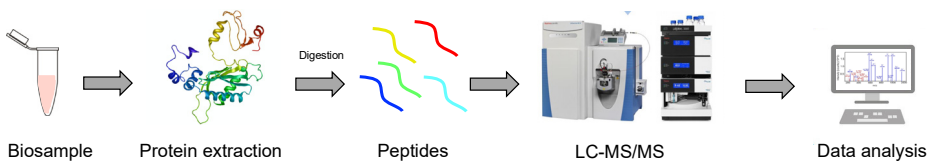


Figure 7. MS-based proteomics workflow.

1.12 MS-based proteomic biomarker discovery for pancreatic cancer

Sample preparation is the first step in proteomics analysis. During this step, the proteins are either analysed intact (top-down analysis) or enzymatically digested into peptides (bottom-up analysis, also known as shotgun proteomics). Bottom-up proteomics is the most established method for large scale MS-based analysis of complex biosamples and often uses tandem data acquisition in which peptides are subjected to collision-activated dissociation. Quantification of protein concentrations can be done using either stable isotope-labelling methods (e.g. iTRAQ, ICAT and SILAC) or label-free methods.

There are two general approaches to MS-based protein analysis: discovery proteomics and targeted proteomics. In the former, protein identification is prioritised and each sample is analysed over a longer period and in more depth, thus reducing the number of samples that can be processed. In targeted proteomics, a limited number of features is monitored to provide the highest sensitivity and allow for increased throughput of samples. Parallel reaction monitoring (PRM) is a novel method for targeted MS-based proteomics based on the parallel monitoring of all fragments from the precursor ion.

Targeted proteomics often follows discovery proteomics to quantify specific proteins found during the discovery phase.

2. Aim of the thesis

The general aim of this thesis was to employ an MS-based proteomics approach to the discovery and orthogonal validation of protein biomarkers for pancreatic cancer. The specific aims of the individual studies were:

I. to optimise a traditional proteolytic digestion protocol for MS-based quantitative proteomics studies;

II. to identify disease-related protein markers for pancreatic cancer via MS-based quantitative proteomic profiling of fresh-frozen tumour and normal specimens;

III. to evaluate the prognostic utility and biological significance of YAP1 in pancreatic cancer tissue; and

IV. to explore the role of AGP1 as a potential biomarker for improved prognostication and non-invasive diagnosis of pancreatic cancer.

3. Material and methods

3.1 Study design

Table 7 provides an overview of the study designs used in this thesis. The workflow of study II is illustrated in Figure 8.

Table 7. General description of the studies in this thesis

Article	I	II	III	IV
Design	Experimental	RCS + experimental	RCS + experimental	RCS
Study material	Mice; cell line	Human tissue; cell line	Human tissue; cell line	Human tissue + serum
Number	6+10; 1	10+10+143; 1	176+140; 1	140+110
Methods	LC-MS/MS	LC-MS/MS, IHC, IF	IHC, IF, drug inhibition	IHC, serum immunoassay

RCS, retrospective cohort study; IHC, immunohistochemistry; LC-MS/MS, liquid chromatography tandem mass spectrometry; IF, immunofluorescence.

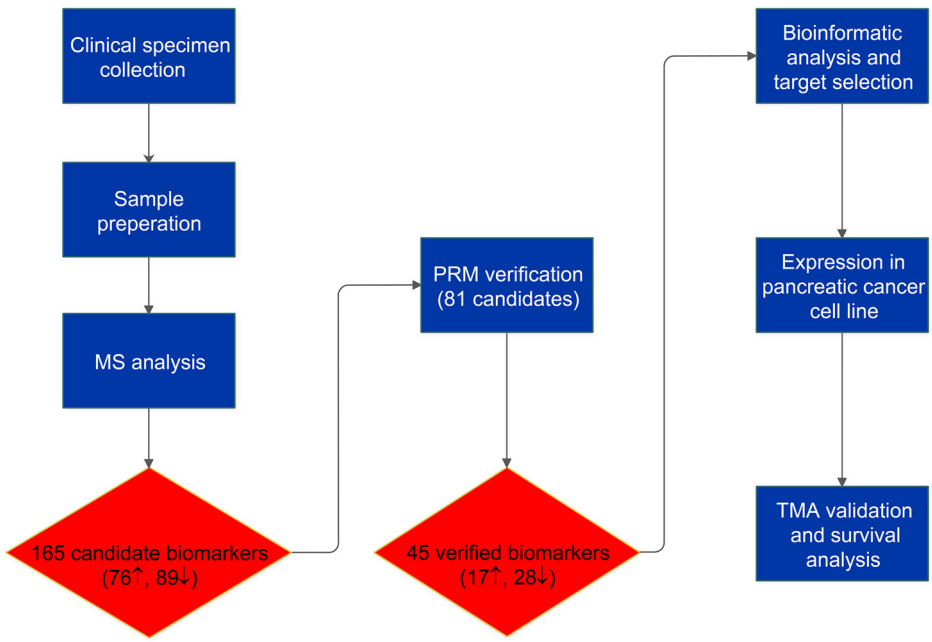


Figure 8. Methodological workflow of study II. PRM, parallel reaction monitoring; TMA, tissue microarray.

3.2 Study populations

Patient samples

For MS analysis, fresh frozen pancreatic cancer tissue specimens (n = 10) were prospectively collected from patients undergoing pancreaticoduodenectomy due to tumours located in the head of the pancreas between July 2013 and April 2015 at the Department of Surgery, Skåne University Hospital, Lund, Sweden. Age-gender-matched fresh frozen pancreatic tissues were biopsies (Articles II, III, IV, n = 10) from organ donors free of any pancreatic disease and were obtained from Lund University Diabetes Centre (LUDC) and analysed as healthy controls (HC). Written informed consent was obtained from the participating patients.

Immunohistochemical (IHC) target verification was performed using tissue microarrays (TMAs) from archival formalin-fixed paraffin-embedded (FFPE) resection specimens from 144 patients with pancreatic ductal adenocarcinoma who underwent curative-intent pancreatic surgery between 1995 and 2017 at Skåne University Hospital and Malmö, Sweden. As additional controls, histologically normal pancreas tissue was obtained from 10 patients who underwent surgical resection for benign pancreatic lesions made up of 7 serous cystadenomas, 1 mucinous cystadenoma, 1 pancreatic pseudotumour and 1 thrombosed splenic artery aneurysm.

Serum samples were prospectively collected from 110 individuals, comprised of 52 patients with resectable pancreatic cancer, 24 with benign pancreatic disease, and 34 healthy controls between 2012 and 2017 at the same institution. Patient sera were obtained at diagnosis and before any surgical intervention. Healthy control serum was obtained from donors at the blood donation centre in Lund.

All tumour samples were re-evaluated by a senior pancreatic pathologist (A.S.) and staged according to the 8th version of AJCC TNM staging system (34).

Publicly available transcriptomics data was retrieved from 176 pancreatic cancer patients from The Cancer Genome Atlas (TCGA) (83-85) to perform complementary mRNA expression analysis.

Cell lines and animals

Proteomics experiments were performed using pancreatic cancer cell line Panc-1 (ATCC-LGC Standards, Manassas, VA, USA), SK-MEL-28 cultured cells (RRID: CVCL_0526), spleen tissue from an adult male Sprague Dawley rat and human pancreatic cancer xenografts. The xenografts were generated in genetically-identical NMRI-nu mice (Janvier Labs, Saint-Berthevin Cedex, France) by

inoculation of human pancreatic cancer cell line Capan-1 (ATCC, Manassas, VA, USA).

Ethics approval

Ethics approval for all human studies was granted by the local Ethics Committee for clinical research at Lund University, Sweden (Ref 2010/684, 2012/661, 2015/266, 2017/320). All animal experiments were performed in a dedicated animal operating room following the guidelines set by the Swedish government and Lund University, and were approved by the regional ethics committee.

3.3 Sample preparation for the MS study

In study I, 125 µg protein from cell lysate of SK-MEL was dissolved in 50 mM ammonium bicarbonate (Ambic) buffer and 6 M urea, further reduced with 10 mM dithiothreitol (DTT) for 1 h at 37 °C and alkylated by 20 mM iodoacetamide (IAA) for 30 min in the dark at room temperature (RT). After reduction and alkylation, each sample was 2-fold diluted with 50 mM Ambic and aliquoted into five replicates with 25 µg of each as experimental replicates of the protein digests. Three of those aliquots were digested by Lys-C with an enzyme/protein ratio of 1:50 w/w for 6 h at 37 °C. The mixture samples were then 3-fold diluted with 50 mM Ambic and further digested by trypsin with a trypsin/protein ratio of 1:50 w/w overnight at 37 °C. The digestion of the remaining two aliquots of the reduced and alkylated cell lysate was the same, but they were incubated at RT (24 ± 2 °C).

Tissue lysates from pancreatic cancer xenografts and rat spleen in 50 mM Ambic and 4 M urea were reduced with 10 mM DTT for 1 h at 37 °C and further alkylated by applying 20 mM IAA for 30 min at RT in the dark. After 4-fold dilution with 50 mM Ambic, each sample was split into six aliquots of 25 µg each as experimental replicates. Three of these aliquots were randomly chosen for Lys-C digestion at an enzyme/protein ratio of 1:50 w/w for 7 h at 37 °C. These samples were then 2-fold diluted with 50 mM Ambic and digested at 37 °C overnight by trypsin at a trypsin/protein ratio of 1:50 w/w. The digestion procedure for the remaining three aliquots was the same, but they were incubated at RT.

All digestions were quenched by adding formic acid (FA) to a final concentration of 1%, and then peptide desalting was performed with ultramicrospin C18 columns following the manufacturer's instructions. Desalted and dried peptides were resuspended in 50 µL 0.1% FA and assayed using the Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific, Rockford, IL) to determine peptide concentration.

In studies II, III, and IV, fresh frozen tissue specimens were pulverised in liquid N_2 with a dry-ice cooled mortar and pestle. The fine tissue powder was then

homogenised in an extraction buffer containing 500 mM Tris-Cl (pH 8), 6 M guanidine-HCl in 50 mM Ambic, together with a protease and phosphatase inhibitor cocktail. The acquired homogenised mixture was subsequently subjected to four freeze-thaw cycles followed by 20 min ultrasonication in a 0 °C bath and a short centrifugation to precipitate the debris. Proteins dissolved in the supernatant were reduced with 15 mM DTT for 1 h at 60 °C and then alkylated using 50 mM IAA for 30 min at RT. Absolute ethanol at -20 °C was used for the protein precipitation with a sample to ethanol ratio of 1:9. From this, precipitated proteins were resolved in 50 mM Ambic, further determined by BCA assay for concentration. Following this, 130 µg proteins from each tissue sample were digested at 37 °C overnight by applying Mass Spec Grade Trypsin/Lys-C Mix with a final enzyme/protein ratio of 1:100 w/w. Finally, digested samples were dried in a speed vacuum concentrator and resuspended in 50 µl 0.1% FA (mobile phase A). The final peptide concentration was confirmed employing a peptide determination assay. Later, 25 fmol peptide retention time mixture (PRTC) (Thermo-Fisher Scientific, Rockford, IL, USA) consisting of 15 peptides was added to each sample, which enabled normalisation and acted as a control for chromatographic performance.

3.4 LC-MS/MS analysis

In study I, a ThermoEasy nLC 1000 system coupled with a Q Exactive Plus mass spectrometer was used for LC−MS/MS analysis. Initial loading amounts of peptides was 1 µg onto the trap column (Acclaim PepMap 100 precolumn, 75 µm ID × 2 cm, C18, 3 mm, 100 Å), which was subsequently separated via an analytical column (EASY-Spray column, 75 µm ID × 25 cm, PepMap RSLC C18, 2 mm, 100 Å) by using an acetonitrile gradient in 0.1% FA with a flow rate of 300 nL/min at a column temperature of 35 °C for 80 mins. Every sample was analysed by triplicate injections. All instruments were supplied by Thermo Fisher Scientific (San José, CA).

In studies II, III, and IV, 1 µg peptides from resuspended peptide mixture aliquots of corresponding samples were separated using a high-performance liquid chromatography (HPLC) system (EASY- nLCTM 1000) and analysed by conjunct Q Exactive quadrupole-Orbitrap mass spectrometer using a nanospray ion source. Peptide lysates were injected to the analytical system at a flow rate of 300 nL/min, and separated using a 132 min gradient programme of 5-22% acetonitrile (ACN) in 0.1% FA followed by an 18 min gradient of 22-38% ACN in 0.1% FA. A central two-column system consisting of an EASY-Spray analytical column (25 cm x 75 µm ID, particle size of 2µm, pore size at 100 Å, C18) and the acclaim pre-column (2cm x 75 µm ID, particle size of 3 µm, pore size at 100 Å, C18) was used for peptide separation. Each replicate of sample was measured twice in a randomised procedure. The raw data files from the duplicates were combined and processed by Proteome Discoverer software (Thermo-Fisher) Version 1.4 focusing on all high confidence peptides.

The Q Exactive system was operated in a positive data-dependent acquisition (DDA) mode with automatic shifting between the full scan MS and subsequent MS/MS acquisition. Based on the highest intensity precursors, 15 data-dependent higher energy collision dissociations of MS/MS scans were enforced, and a full MS scan was performed in the Orbitrap detector for peptide searching. A resolution of 70,000 at 200 m/z was employed to detect the MS1 survey scans of the eluted peptides, together with a recording window of m/z of between 400. and 1,600. The automatic gain control (AGC) target was established as 1×10^6 with an injection time of 100 ms, and the normalised collision energy (NCE) was fixed at 27.0% for all scans. Finally, the resolution of the data-dependent MS2 scans was set to 17,500 at 200 m/z with the value for the AGC target and injection time of 5×10^5 and 80 ms, respectively.

3.5 Targeted MS verification

To verify the presence of potential biomarker proteins, targeted MS analysis using PRM was performed. Depending on the MS spectra obtained from the 40 discovery measurements, 1 or 2 unique peptides of each targeted protein were selected based on detection frequencies (> 50%, missed cleavage = 0, p-value < 0.05) and rankings of peptide intensities and peptide spectrum matches. Eventually, a spectral library of 81 selected proteins containing 150 peptides was developed.

Those selected proteins were reanalysed in the same cohort as the MS discovery phase. Briefly, proteins extracted from 18 fresh frozen samples were reduced, alkylated and digested using the same protocols described above. Due to inadequate tissue sample volume, we had to exclude 2 pancreatic cancer subjects. For the evaluation, 1 μg of the sample peptide was subjected to the MS system, operating the PRM assay programme in a time-scheduled acquisition method with a retention time of +/-5 min. The resolution setting was 35,000 and AGC targeted 5×10^5, with a maximum injection time of 50 ms. The chromatographic peak width was set to 30s, and the normalised collision energy to 26% with an isolation window of 2 m/z. For relative quantification calculations, Skyline software was used for the PRM study (86).

3.6 MS data processing

Raw MS data files from study I were analysed using MaxQuant version 1.6.0.1 with the Andromeda Search engine. Protein isoforms were precluded when searching the UniProtKB human database (released 9 July 2016) and the UniProtKB Rattus norvegicus database (released 18 Jan 2017). Matches from the contaminant protein database and the decoy database were excluded as a default setting, and search implements were fixed at 20 ppm and 0.02 Da for precursor and fragment ion tolerances, respectively. The carbamidomethylation of cysteine residues was set as

the fixed modification, while other protein modifications were analysed as dynamic modifications. These were acetylation, carbamylation of lysine residues, oxidation of methionine residues, carbamylation of protein N-termini and pyroglutamic acid modification. Two missed cleavage sites and the 'match-between-runs' option were allowed, and filtering of high confidence at the peptide and protein level was applied by applying a false discovery rate (FDR) of 0.01.

In studies II, III, and IV, the MS/MS raw data output files obtained from the combined randomised measurements of the replicates were processed using Proteome Discoverer software (Thermo Fisher) version 1.4. When selecting the MS spectra, a minimum precursor mass of 350 Da, a maximum precursor mass of 5,000 Da and a signal-to-noise (s/n) threshold of 1.5 should be set. The settings for the search engine SEQUEST HT(87) were as follows: precursor mass tolerance and fragment mass tolerance at 10 ppm and 0.02 Da, respectively, with trypsin as the enzyme in use and 1 missed cleavage site accepted. To increase the number of identified peptides, carbamidomethyl (+57.021 Da; C) was set as a fixed modification, and multiple variable modifications were included when searching through the UniProtKB human database(88). These were: such oxidation (+15.995 Da; M, P); methyl (+14.016 Da; K, R); dimethyl (+28.031 Da; K, R); acetyl (+42.011 Da; K); trimethyl (+42.047 Da; K, R); and glygly (+114.043 Da; K). The percolator was used for the high confidence peptides and proteins, and the FDR cut-off limit value was fixed at 0.01, along with applying a precursor ions area detector in the search engine for peptide quantification.

3.7 Tissue microarray

To validate the expression of candidate protein biomarkers in pancreatic cancer tissues and investigate the relationship between expression levels of those markers and clinicopathological parameters (Articles II, III, IV), a total of 144 resected and archived FFPE pancreatic cancer specimens were used in constructing the TMA. Briefly, the tumour core of each tissue sample was located and marked by the pathologist (A.S.) and sampled suing an automated tissue array vehicle, Alphelys (Minicore, Plaisir, France). A total of 4 cores (diameter 2 mm) of cancer tissue from each specimen were obtained and fixed into paraffin blocks, which were then stored at -4 °C. When needed, 3 µm thick slides were sectioned from these TMA blocks for further IHC analysis.

3.8 Immunohistochemistry analysis

The TMA-based IHC analysis was performed to illustrate the expression pattern of candidate biomarkers in pancreatic tissue, and to evaluate the correlation between biomarker expression level and the patient's clinicopathological parameters (Article II, III, IV). Briefly, TMA blocks were sectioned into 3 µm slides and attached to

individual glass which were re-treated using an automated PT Link system (Dako, Agilent Technologies, Glostrup, Denmark) for 20 min at 97 °C in low pH EnVision Flex retrieval solution (Dako, Agilent Technologies, Glostrup, Denmark). After deparaffinisation, rehydration and antigen-retrieval, the TMA slides were incubated with corresponding primary antibodies (see Table 8) in an optimised dilution at 4 °C overnight. Thereafter, the slides were incubated with respective secondary antibodies (see Table 8), followed by staining with avidin-biotin-peroxidase complex (Vectastain Elite ABC-HRP Kit, Vector Laboratories, Cat No. PK-6100, Burlingame, CA) to amplify the staining signal. Chromogen diaminobenzidine (DAB) (Vector Laboratories) was used for antibody-antigen complex visualisation, and Mayer's haematoxylin (Histolab, Gothenburg, Sweden) was used for cell nuclei counterstaining. Finally, the slides were dehydrated and cleared in graded alcohol and xylene, respectively, and mounted using Pertex (Histolab). The immune-stained TMA slides were examined using an Aperio scanscope scanner (Leica Biosystems, Wetzlar, Germany) and the images sent to a group of independent observers for further evaluation (A. Sasor, M. Bauden, H. Dai, J. Xu, and X. Chen). The observers were blinded to clinical and outcome data.

Table 8. Primary and secondary antibodies for IHC analysis

	Primary antibodies	Secondary antibodies
Study II	rabbit anti-human BASP1 (dilution 1: 100; Cat No. HPA045218, Atlas Antibodies)	biotinylated goat anti-rabbit (dilution 1:200; Cat No. BA-1000, Vector Laboratories)
	mouse anti-human WT1 (clone 6F-H2, Ready-to-Use, Cat No. IS05530-2, DAKO)	biotinylated horse anti-mouse (dilution 1:200; Cat No. BA-2000, Vector Laboratories)
Study III	rabbit anti-human YAP1 (dilution 1: 200; Cat No. #14074, Cell Signalling)	biotinylated goat anti-rabbit (dilution 1:200; Cat No. BA-1000, Vector Laboratories)
Study IV	rabbit anti-human AGP1 (dilution 1: 50; Cat No. HPA046438, Atlas Antibodies)	biotinylated goat anti-rabbit (dilution 1:200; Cat No. BA-1000, Vector Laboratories)

In studies II and IV, scoring of the BASP1, WT1 and AGP1 staining was based on the proportion of positive reacted tumour cells and their reaction intensity. IHC results were marked as follows: negative (0), weak (1), moderate (2), and strong (3). Tumours showing > 10% of stained cancer cells were considered positive, and when tumours presented heterogeneous staining, the potent pattern was used for final scoring. In Article III, a semiquantitative approach using the H-score was employed (89, 90). Briefly, YAP1 protein staining intensities were marked as either negative (0), weak (1+), moderate (2+), or strong (3+), which was multiplied by the percentage of reacted cells at each staining intensity level. The final H-scores were calculated using the following formula:

H-score = 0 × (% cells staining negative [0] × 100) + 1 × (% cells staining weakly [1+] × 100) + 2 × (% cells staining moderately [2+] × 100) + 3 × (% cells staining strongly [3+] × 100).

3.9 Immunofluorescence analysis

To elucidate the intracellular localisation of candidate biomarkers (Article II), an eight-well chamber slide (Lab-Tek II Chamber Slide System, Nunc) was employed for Panc-1 cell culture (approximately 8×10^3 cells/well). After 48h of culturing, the cancer cells were then fixed with 4% formaldehyde and permeabilised with 1% Triton X-100, blocked with 5% goat serum, and eventually incubated with mouse anti-human WT1 (DAKO, clone 6F-H2, Cat No. IS05530-2) for 2 hours at room temperature. Thereafter, the cells were washed and incubated with goat-anti-mouse Alexa Fluor 594 by dilution of 1:500 (Cat No. A11032, Invitrogen) for 1 hour at room temperature in the dark. Then the cells were further blocked by 5% donkey serum, along with incubation of rabbit anti-human BASP1 (Atlas Antibodies, dilution 1:50, Cat No. HPA045218) for 2 hours at room temperature. After washing, donkey-anti-rabbit Alexa Fluor 488 with a dilution of 1:500 (Cat No. A21206, Invitrogen) was introduced at room temperature for 1 hour. Finally, the nuclei of the cells were stained with DAPI. For immunofluorescence evaluation, an imaging system consisting of a Nikon Eclipse 80i microscope and Nikon DS-Qi1 camera was employed, together with analyzing software (NIS-Elements, Nikon Instruments Inc., Melville, NY, USA).

In study III, to assess the YAP1 expression features, Panc-1 cells were cultured in 6-well plates with a density of 50,000 cells per well. After 48 h, the cells were fixed by 4% paraformaldehyde (Histolab, Västra Frölunda, Sweden) and then stained using primary rabbit anti-human YAP1 (dilution 1: 250, Cell Signalling) and incubated with Alexa Fluor 488 conjugated donkey-anti-rabbit secondary antibody at a dilution of 1:200 (Invitrogen, USA). DAPI (NucBlue, Molecular probes, Life Technologies, USA) was applied to mark the cell nucleus. For image processing, the Cellomics ArrayScan platform VTI HCS (ThermoScientific, Rockford, IL) and Bioapplication software were used. When the cell population reached 2,000 per well, high content images were taken by a multiparameter fluorescent microscopic imaging system and quantified with the fluorescence intensities of the selected channel (Alexa 488).

3.10 mRNA expression data analysis

In study III, transcriptomics data of YAP1 was obtained from 176 pancreatic cancer patients sequenced by TCGA (83-85). The RNA-seq data were analysed depending on the number of Fragments Per Kilobase of exon per Million reads (FPKM).

3.11 Serum analysis

In study IV, levels of the potential diagnostic biomarker AGP1 in serum samples were measured using an immunoturbidimetric method at the Department of Clinical Chemistry and Pharmacology, University and Regional Laboratories Region, Skåne, Sweden (an authorised clinical laboratory for routine patient blood sample tests). The evaluation platform was the Cobas 6000 analyser (Roche Diagnostics, Mannheim, Germany). Which c501 module was used depended on the IFCC-IUPAC-coding system (NPU19873). Channels at 340 and 660 nm were used for bi-chromatic measurement of formed AGP1-antibody complex in the serum samples. The CFAS protein calibrator (Roche Diagnostics) was used for calibration according to the standard method using the international protein calibrator CRM 470. The established detection range was 0.1 – 6.0 g/L and the total coefficient variation (CV) was 4.9% and 3.6% at 0.37 g/L and 0.95 g/L, respectively.

Immunometric sandwich analysis was performed for CA 19-9 concentrations by using an electrochemiluminescence immunoassay (ECLIA) detection technique depending on the Ruthenium (Ru) derivate. A Cobas immunoanalyser was employed for the measurements, which were accomplished in line with the IFCC-IUPAC-coding system NPU01450 at the same institution as the AGP1 evaluations. The standard method was used (11776193 122, 2016-11, V23, Roche Diagnostics) as stated in the CA 19-9 immunoassay references. The detection range was fixed between 0.6 – 10,000 kU/L with a CV of 5% and 4% at 20 kU/L and 80 kU/L, respectively.

3.12 Statistics and bioinformatics

For statistical analyses of MS data, Perseus software (91) version 1.5.6.0 and 1.6.0.7 were applied. Proteins detected in less than 50% of samples in each study group were excluded. After $\log_2$ transformation, the normalisation of protein intensities was performed by subtracting the median intensity of all the proteins from the same sample. Subsequent data imputation of missing values was generated by restoring vanished intensities from a normal distribution with a width of 0.3 and downshift of 0. A Two-Sample Student's T-test (two-tailed) and permutation-based FDR correction were carried out to compare protein levels between the groups. Proteins with q-value < 0.01 were considered as significantly differentially expressed with the established settings such as S0 = 2 on both sides and FDR = 0.01. These were the parameters that defined the artificial variance within groups and controlled the relative variations of the resulted p-values.

For bioinformatic analysis, several tools were used to explore the networks involved in the biological relationship between identified protein biomarker candidates. Three publicly-available analytical search engines PANTHER (92), STRING (93),

and KOBAS (94) were used for protein classification and potential pathway network identification of the significantly altered proteins in Article II. To unravel the involvements of selected biomarker candidates in possible protein-protein interactions and disease-linked functional pathways, the Ingenuity Pathway Analysis (IPA, Qiagen, Inc. Redwood City, CA, USA) (95) software was used in Articles II, III, and IV. This toolset builds on a literature-derived knowledge base, generating biological networks containing more than 40,000 nodes representing mammalian genes and their products (transcripts, proteins and miRNA), along with up to 1,480,000 interactions between them. IPA enables the identification of direct or putative interactors of targeted proteins and enrichment of related functional pathways. Subcellular localisation of potential protein biomarkers (Article II) was manually assessed using the UniProt database (96) and their gene ontology terms were identified using PANTHER (92). IceLogos were created using a web application (97).

For IHC analysis (Articles II, III, IV), the expression levels of biomarker candidates were dichotomised into negative/positive, high/low or strong/weak groups. To assess the potential correlation of biomarker expression levels with patients clinicopathological characteristics, the Wilcoxon-Mann-Whitney test, Fisher's exact test or χ^2 were used for continuous and categorical variables, respectively. Kaplan-Meier analysis and the log-rank test were applied to calculate the cumulative probability of OS and evaluate the statistical difference between groups. The hazard ratios (HRs) of death for different prognostic factors were calculated using univariate and multivariate analysis (Cox regression proportional hazards models) and a value of $p < 0.05$ was considered statistically significant.

For serum immunoassay (Article IV), the Wilcoxon rank-sum (Mann-Whitney) test was used to assess AGP1 and CA 19-9 serum expression differences between groups, and the Spearman rank-order correlation was performed to evaluate potential correlations between AGP1 and CA 19-9 levels. The R function lrm, from the package rms (98), was used for Logistic Regression (LR). Diagnostic functional parameters were calculated to separate pancreatic cancer, benign pancreatic disease and healthy controls by using the R package pROC (99), including the Receiver Operator Characteristic (ROC) curves, the Areas Under the Curve (AUCs) and the detection sensitivities and specificities. To find the optimal cut-off serum concentration values for AGP1 in discriminating pancreatic cancer from controls, Youden's index (J) was used (100), and calculated by the formula: J = sensitivity + specificity – 1.

Statistical evaluation was conducted with GraphPad Prism version 7 (La Jolla, CA), SPSS version 23.0 (SPSS Inc., Chicago, IL), Stata MP statistical package version 13.1 and R(101) programming language version 3.5.1 (R Foundation for Statistical Computing).

4. Results

4.1 Optimisation of a traditional MS sample preparation protocol

In conventional proteomic studies, urea-containing buffer solutions are generally used for protein denaturation and solubilisation of cells and tissues (102). However, the involvement of isocyanic acid (a degradation product of urea) in these buffers causes carbamylation of proteins and peptides, leading to incomplete digestion of, particularly at higher temperatures and for longer incubations (103, 104). The ionisation efficiency of carbamylated peptides is also lower than that of normal peptides, which can significantly reduce MS performance in identifying and quantifying proteins (105). In Article I, the aim was to optimise the traditional urea-containing digestion procedure, offering practical solutions for minimising these disadvantages.

It was shown that digestion with low urea concentration contributes to low peptide miscleavage in all samples and at all conditions. When the digestions took place in 1 M urea, the miscleavage frequency was 11% at 37 °C and 14% at RT (24 ± 2 °C). In 0.5 M urea, the frequency of miscleavages was < 7% and < 9% at 37 °C and RT, respectively. The number of identified peptides and proteins was also consistently increased when performing digestion at RT (see Figure 9), and identified peptides and proteins were increased from 5% to 23% and from 8% to 39%, respectively, particularly in the urea concentration of 0.5 M. There was also a trend to achieve higher peptide abundance when experiments performed at RT than at 37 °C (Figure 10).

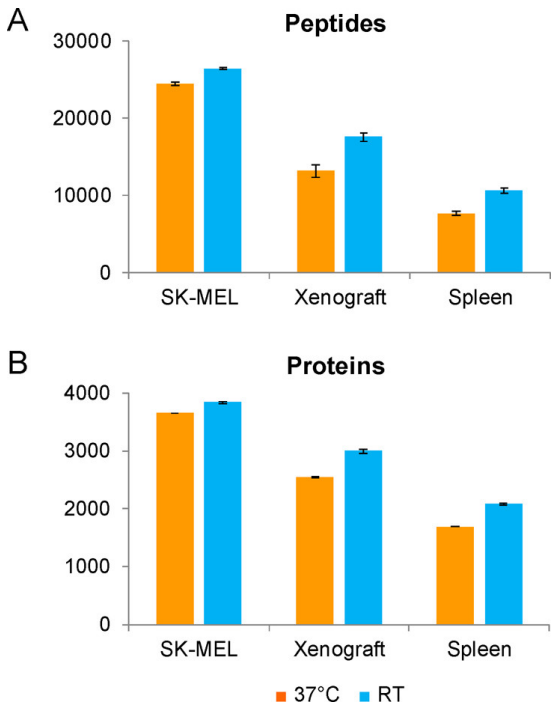


Figure 9. Comprehensive comparison of (A) peptides and (B) proteins identified in urea-containing buffer digestions at 37 °C and RT. Samples from SK-MEL cells, pancreatic cancer xenograft tissues, and spleen tissues (Data are reported as mean ± SD).

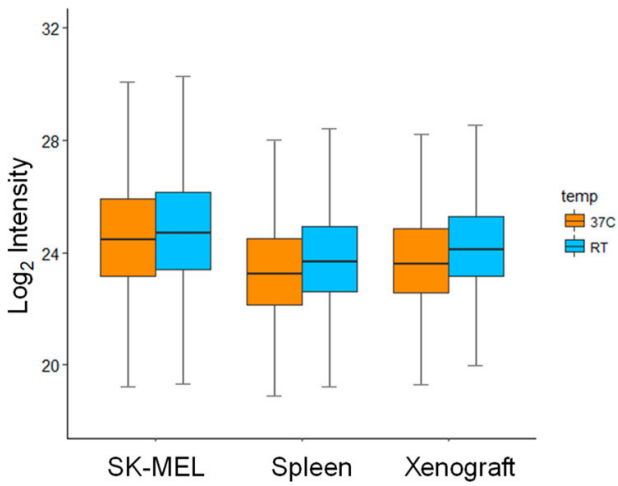


Figure 10. The intensities of common peptides for the digestions at 37 °C and RT visualized using box plots. Samples from SK-MEL cells, pancreatic cancer xenograft tissues, and spleen tissues.

To evaluate the differences in terms of protein abundance between those two enzymatic digestion conditions, t-tests were performed and volcano plots were generated filtering by fold change ≥ 2 and FDR < 0.05 (see Figure 11). Notably, when proteolysis was performed at RT, there were 213 (from SK-MEL cells), 517 (pancreatic cancer xenograft) and 525 (from rat spleen) proteins presented at significantly higher abundance. Conversely, when proteolysis was conducted at 37 °C, there only 85, 57, and 123 proteins were identified at higher abundance from SK-MEL cells, pancreatic cancer xenograft, and rat spleen, respectively. Similar or lower CV distributions were achieved when digestion underwent at RT compared to performed at 37 °C.

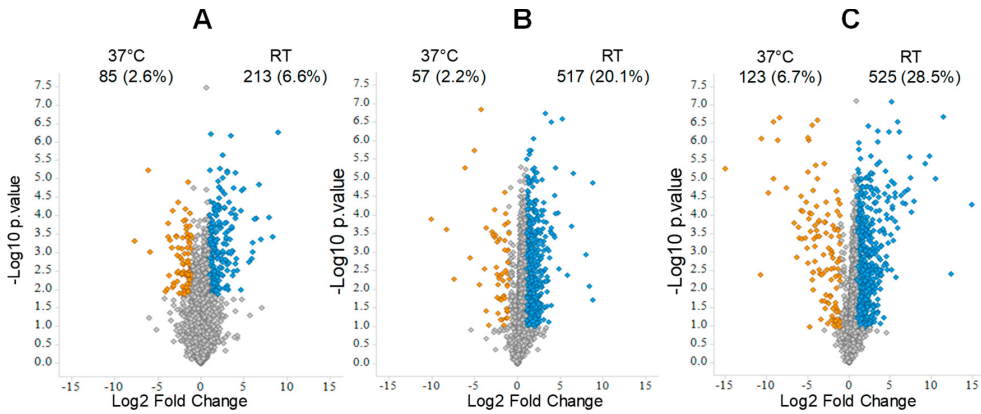


Figure 11. Volcano plots showing different protein abundances for the digestions at 37 °C and RT for SK-MEL cells (A), pancreatic cancer xenograft tissues (B), and spleen tissues (C). Label-free quantification was used. The numbers and percentages of proteins that were significantly affected by the digestion conditions are highlighted.

To investigate the potentially temperature-affected post-translational modifications (PTMs), the incidences of carbamylation and pyroglutamic acid formation were studied. Interestingly, a decrease in N-terminal carbamylated peptides of over 40% was observed when the digestion took place at RT rather than at 37 °C (Figure 12), and the highest reduction was up to 15-fold, which was seen in the SKMEL lysate.

As in previous studies, carbamylation was found more often in the α-amino groups of peptides than the ε-amino groups of lysine residues (104), leading to a reduction of 0.25% for both digestion environment temperatures (Figure 12). A 25-60% reduction in peptide-to-spectrum matches (PSMs) corresponding to lysine carbamylation modification was found for digestions at RT. Finally, there was a significant decrease (30-50%) of pyroglutamic acid-containing peptides in the digestions performed at RT, leading to less than 0.6% of PSMs being assigned this PTM (Figure 12).

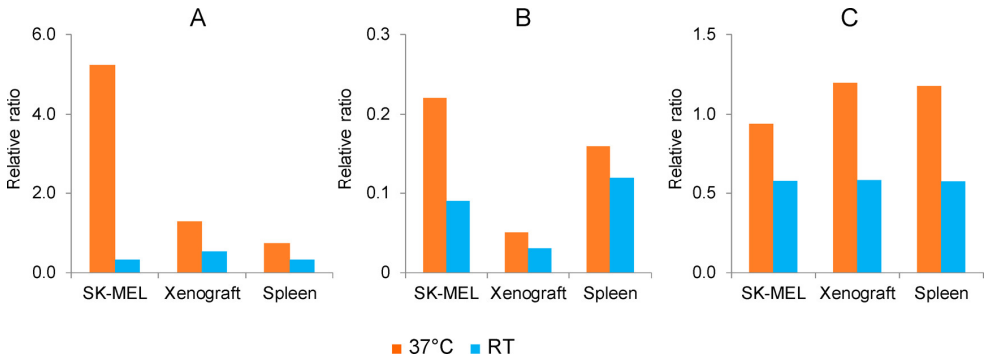


Figure 12. Ratios of PSMs in terms of (A) N-terminal carbamylation, (B) lysine carbamylation, and (C) pyroglutamic acid-containing peptides for the digestions conducted at 37 °C and RT. The ratios were calculated by dividing the number of PSMs assigned to a specific modification by the total number of PSMs, including all triplicate injections on the LC−MS/MS.

In summary, digesting the protein mixtures with a urea-containing solution in RT have multiple advantages than in traditional 37 °C environment, including identification and quantification of peptides and proteins, a minimised impact on miscleavage rates, as well as a greater reduction of PTMs such as carbamylation and pyroglutamic acid formation.

4.2 Overview of proteomic discovery and verification

Label-free MS quantitative analysis

The aim of study II was to discover novel disease-specific protein biomarkers for pancreatic cancer by MS-based proteomics profiling fresh frozen tissue specimens of pancreatic cancer and normal pancreas.

A total of 4,138 and 2,950 proteins from cancerous and normal samples were identified and quantified, respectively. The quantification was based on one or more unique peptides from the corresponding protein. The quantified proteins included 2,264 from the pancreatic cancer group and 2,354 from the healthy control group (Figure 13). After filtering of the missing values, 1,959 proteins were statistically confirmed.

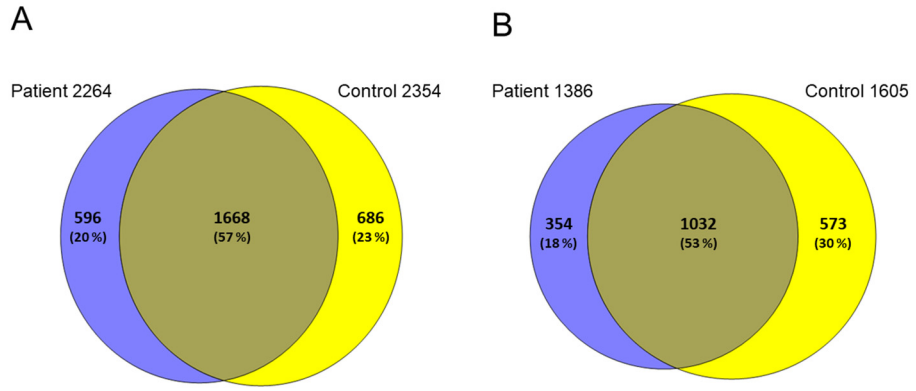


Figure 13. Overview of protein identification and quantification. (A) Venn diagram of 2950 proteins quantified with one or more unique peptides. (B) Venn diagram of 1959 quantified proteins after filtering of missing values.

To determine and visualise a generic pattern of protein clustering and the abundance variations between and within sampling groups, the quantified protein-based two-dimensional Principal Component Analysis (PCA) was projected. A clear separation of proteins from the cancer and control samples was observed, using the value from the $\log_2$-ratio of individual samples over the average value from the gross samples (Figure 14A). This indicated a robust difference in the protein expression between the two groups. Following the statistical criteria described above, some 165 proteins were significantly differentially expressed between the two experimental groups, and these are potential protein biomarkers for pancreatic cancer (Figure 14B).

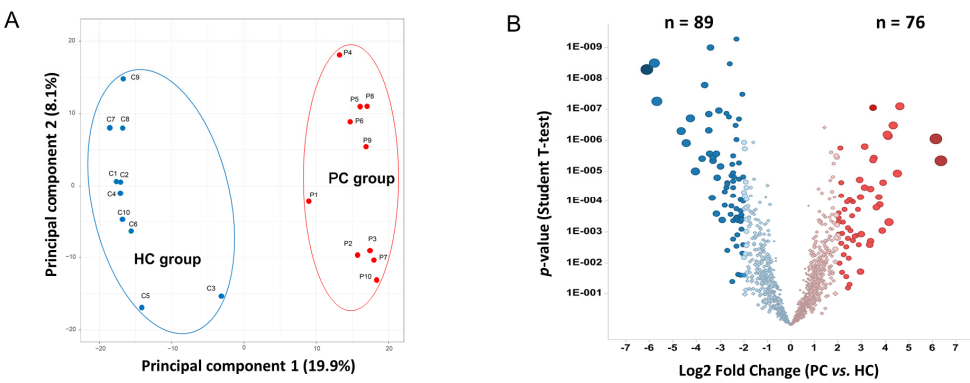


Figure 14. The generic proteome patterns of pancreatic cancer (PC) patients and healthy controls (HC). (A) two-dimensional PCA analysis of the two study groups. Data points are marked with the specimen identifiers (sample number, P-patient/C-control); blue data points denote healthy control pancreas tissues, and red data points denote pancreatic cancer tissues. (B) Volcano plots of all identified proteins in this study; the dark red and dark blue dots indicate the significantly up- and down-regulated proteins in pancreatic cancer compared to controls, respectively (the size of dots represent fold changes).

To gain an overview of expression classification and involved pathways of the 165 potential pancreatic cancer biomarkers, the web-based open-access tool boxes PANTHER and STRING were applied. In the PANTHER system, these proteins were classified according to molecular function, cellular component, biological process and protein class. In terms of molecular function, the proteins were divided into seven groups: binding (41.2%), catalytic activity (31.1%), structural molecule activity (21.0%), transporter activity (4.2%), receptor activity, antioxidant activity (0.8%) and signal transducer activity (0.8%) (Figure 15A). When cellular components were examined, more than one-third of the proteins originated from the cell part (36.6%), followed by organelles (27.4%), macromolecular complexes (21.5%) and other cellular components including the extracellular region (9.7%), membrane (2.7%) and cell junction (1.1%) (Figure 15B). Regarding biological processes, 29.5% participated in the cellular process, followed by the metabolic process (25.5%) and biological regulation (10.4%). The remaining 34.6% were involved in growth, reproduction, biological adhesion, developmental process and immune system processes (Figure 15C). The candidates were further classified into 19 subgroups, including nucleic acid binding (31.5%), hydrolase (12.0%) and signalling molecule (8.7%) (Figure 15D).

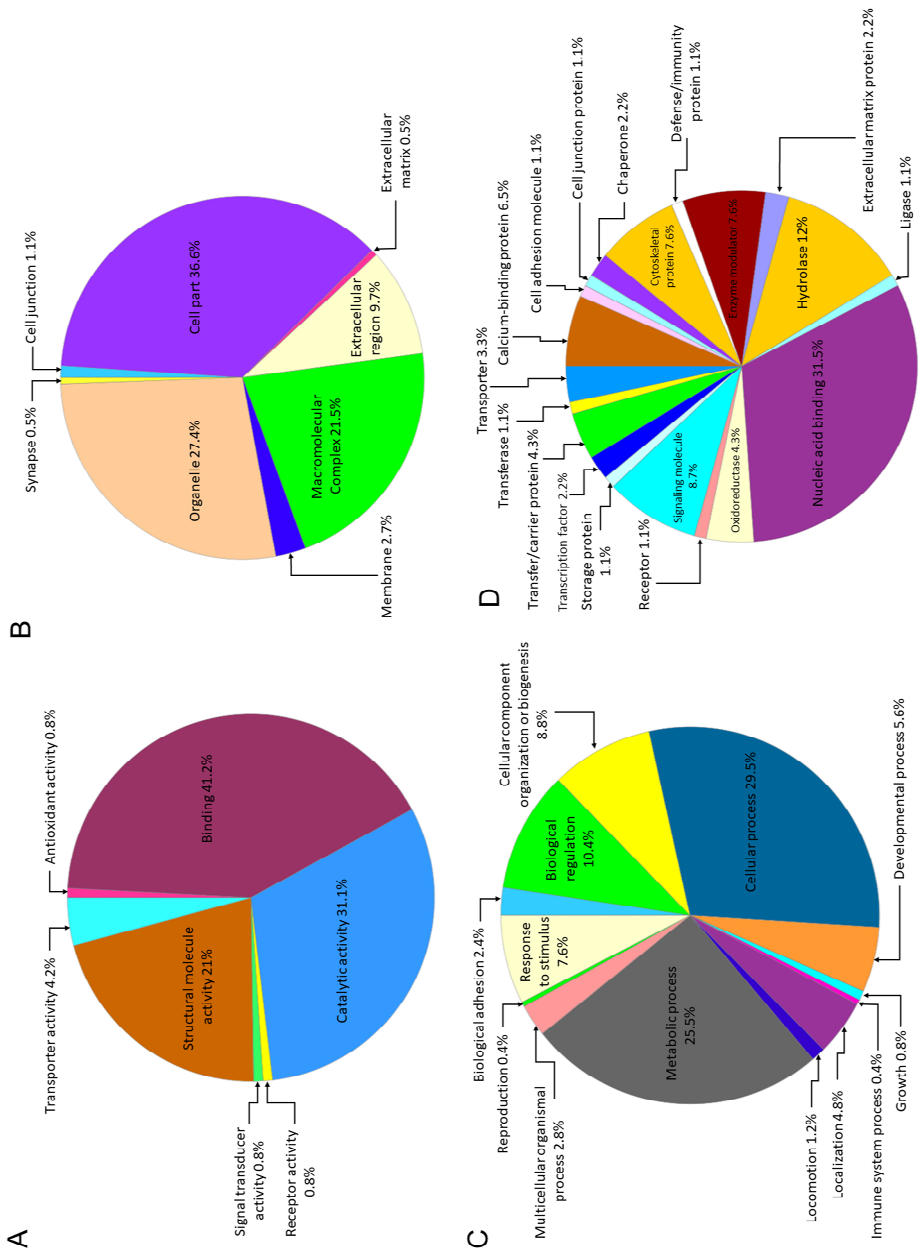


Figure 15. Classifications of the 165 potential protein biomarkers using the PANTHER system. (A) Classification according to molecular function. (B) Classification according to the cellular component. (C) Classification according to the biological process. (D) Classification according to protein class.

The STRING database was used to investigate the functional protein-protein interactions among the 165 differentially expressed proteins. As TP53 and KRAS mutations are the pivotal factors in pancreatic carcinogenesis, both were added to the interaction analysis manually. With the setting of high confidence (minimum required interaction score of 0.7), 559 interactions were identified and significantly enriched ($p < 1.0E-16$). Notably, ten proteins showed a close connection with TP53, of which ALB, CTNNB1, THBS1, VCAN and YWHAB were upregulated and HSPA9, PHGDH, PPIF, RPS27A and TPT1 were downregulated in pancreatic cancer patients compared to the controls. Another group of seven proteins formed a tight network with KRAS. Of the KRAS associated proteins, FGA, FGB, FGG, FN1 and YWHAB were upregulated while PEBP1 and RPS27A were downregulated in pancreatic cancer versus controls (Figure 16).

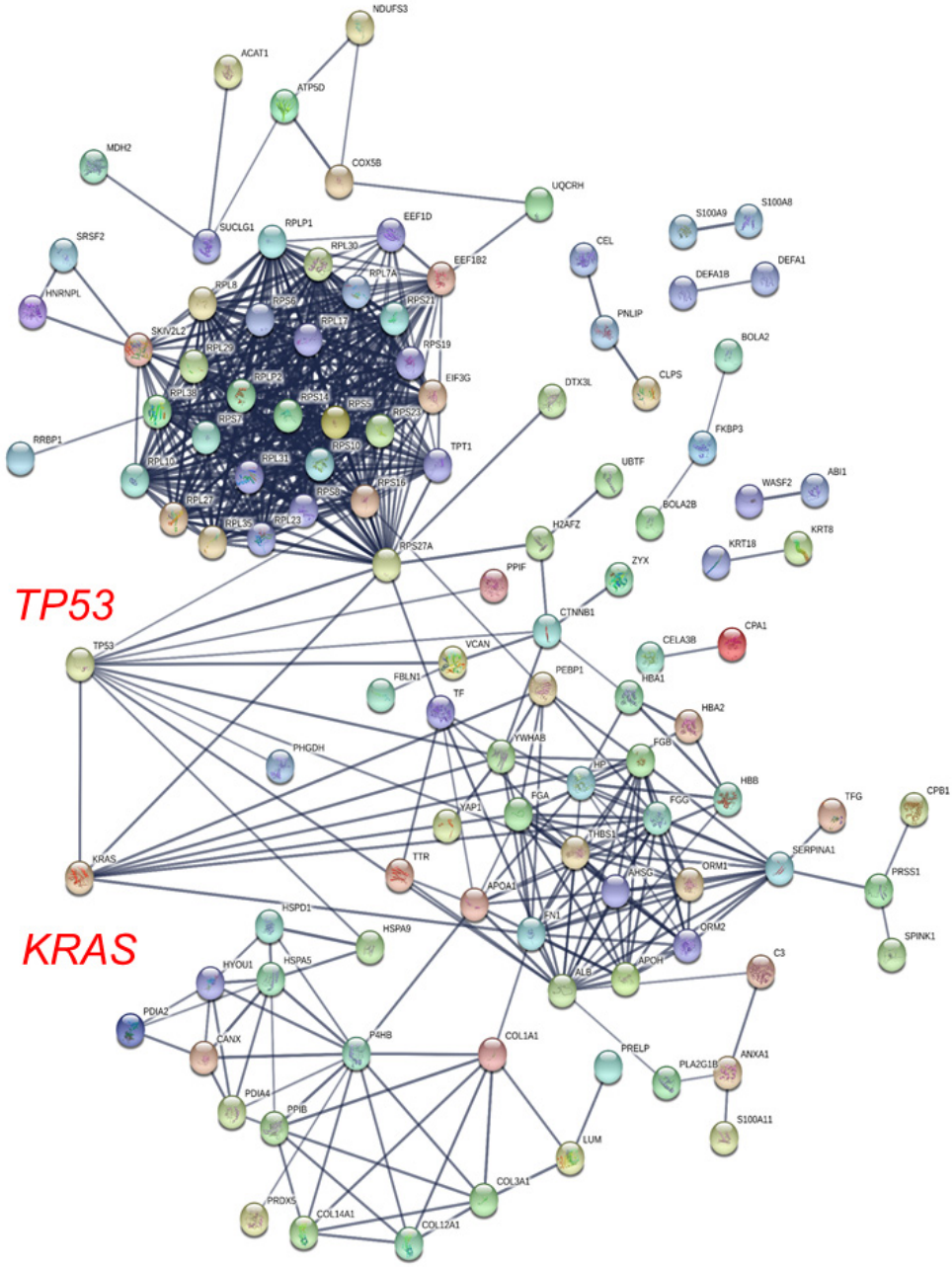


Figure 16. Protein-protein interactions of the 165 potential protein biomarkers. Interaction networks were generated from STRING database. TP53 and KRAS were added manually to probe potentially related pathways.

Finally, a consensus heatmap corresponding to the 165 significantly altered proteins was created. The overexpressed and down-regulated proteins were ranked in order of $\log_2$ fold changes, and the subcellular location of each protein was annotated. Gene ontology (GO) terms relating to biological processes were obtained from the PANTHER toolset for these potential biomarkers (Figure 17).

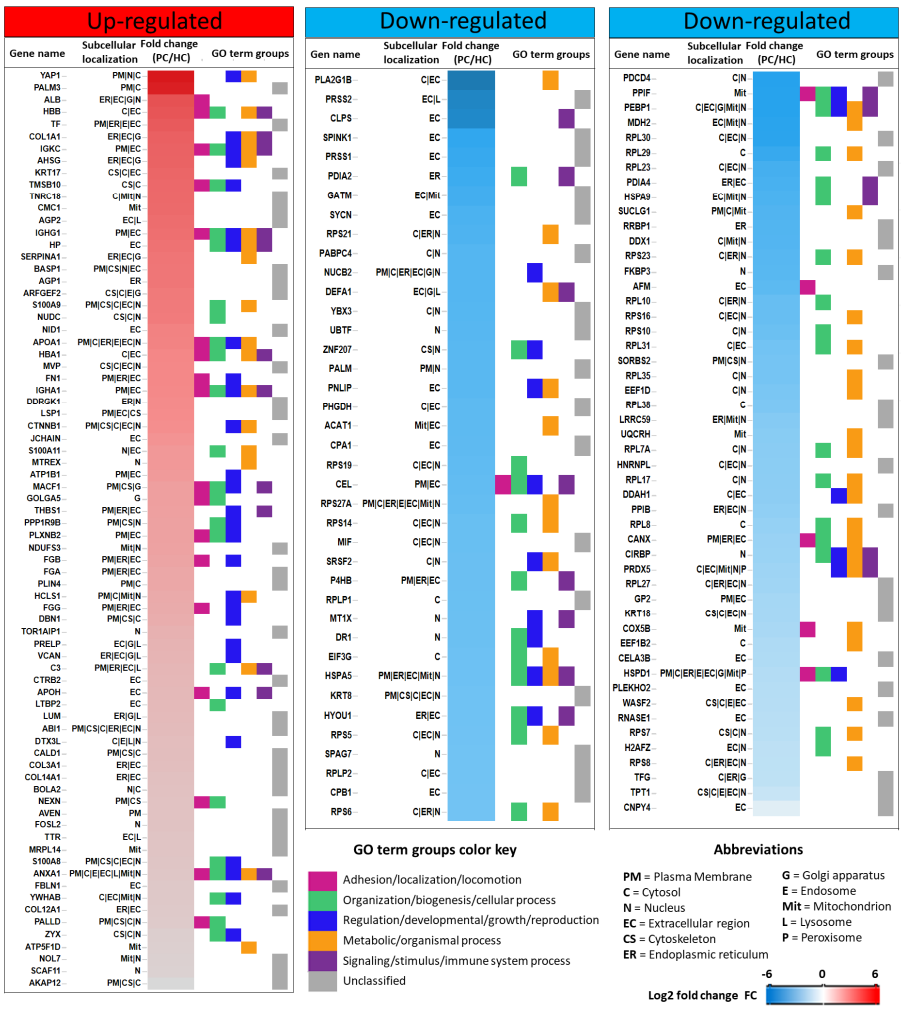


Figure 17. Overview of the potential protein biomarkers for pancreatic cancer. The up- and down-regulated proteins were ranked by $\log_2$ fold change and the annotated subcellular location of each protein is also shown. The PANTHER gene ontology (GO) analysis illustrated that 165 potential protein biomarkers were related to multiple biological processes (e.g. localization, biogenesis and signaling).

Targeted MS verification of selected biomarker candidates

To verify the altered expressions of the potential markers, a PRM assay was performed of the same cohort. Briefly, 81 proteins with unique peptides were selected for this targeted study, and a panel of 45 candidate proteins was eventually identified and quantified, including 17 up-regulated and 28 down-regulated proteins ($p < 0.01$, Tables 9 and 10). Based on the 45 verified protein biomarker candidates, a consensus clustering heat map was generated showing clear segregation between pancreatic cancer and controls (Figure 18).

Table 9. PRM verified upregulated proteins in PC compared to HC (ranked according to fold change)

Sr. no.	UniProt accession	Gene	Protein name	Unique peptide	P-value	Fold change (PC/HC)
1	P02647	APOA1	Apolipoprotein A-I	K.LLDNWDSVTSTFSK.L	1.8E-08	39.12
2	B9A064	IGLL5	Immunoglobulin lambda-like polypeptide 5	K.VTVLGQPK.A	3.5E-09	35.02
3	P02765	AHSG	Alpha-2-HS-glycoprotein	K.FSVVYAK.C	2.1E-09	27.47
4	P0DOY2	IGLC2	Immunoglobulin lambda constant 2	K.AAPSVTLFPPSSEELQANK.A	1.6E-09	24.42
5	P02763	AGP1	Alpha-1-acid glycoprotein 1	R.YVGGQEHFAHLLILR.D	4.6E-06	24.25
6	P01857	IGHG1	Immunoglobulin heavy constant gamma 1	K.GPSVFPLAPSSK.S	2.2E-10	23.59
7	P01834	IGKC	Immunoglobulin kappa constant	K.VDNALQSGNSQESVTEQDSK.D	5.1E-11	22.78
8	P01876	IGHA1	Immunoglobulin heavy constant alpha 1	K.TPLTATLSK.S	1.3E-09	20.82
9	P02787	TF	Serotransferrin	K.EGYYGYTGAFR.C	1.6E-10	19.84
10	P02768	ALB	Serum albumin	K.DDNPNLPR.L	4.5E-09	19.70
11	P01009	SERPINA1	Alpha-1-antitrypsin	K.AVLTIDEK.G	6.2E-10	17.03
12	P80723	BASP1	Brain acid soluble protein 1	K.ETPAATEAPSSTPK.A	1.7E-05	12.91
13	P06703	S100A6	Protein S100-A6	K.LQDAEIAR.L	1.5E-05	12.13
14	Q05707	COL14A1	Collagen alpha-1(XIV) chain	R.YTAILNQIPSHSSSIR.T	6.5E-12	10.70
15	P16401	HIST1H1B	Histone H1.5	K.ATGPPVSELITK.A	1.6E-08	9.85
16	P23142	FBLN1	Fibulin-1	K.IIEVEEEQEDPYLNDR.C	4.8E-08	8.34
17	P52566	ARHGDIB	Rho GDP-dissociation inhibitor 2	K.TLLGDGPVVTDPK.A	1.9E-08	5.46

HC, healthy controls; PC, pancreatic cancer; PRM, parallel reaction monitoring.

Table 10. PRM verified downregulated proteins in PC compared to HC (ranked according to fold change)

Sr. no.	UniProt accession	Gene	Protein name	Unique peptide	P-value	Fold change (HC/PC)
1	P04054	PLA2G1B	Phospholipase A2	R.AVWQFR.K	1.1E-06	56.89
2	P16233	PNLIP	Pancreatic triacylglycerol lipase	R.TGYTQASQNIR.I	8.7E-06	51.98
3	P09093	CELA3A	Chymotrypsin-like elastase family member 3A	R.WNWWGSTVK.K	1.7E-06	47.18
4	P04118	CLPS	Colipase	K.TLYGIYYK.C	1.2E-05	44.63
5	P19835	CEL	Bile salt-activated lipase	K.LGLLGDSVDIFK.G	4.5E-06	39.95
6	P07478	PRSS2	Trypsin-2	R.TLDNDILLIK.L	3.5E-06	36
7	P15085	CPA1	Carboxypeptidase A1	K.TEPVPDQDELDQLSK.A	2.4E-06	35.02
8	Q13087	PDIA2	Protein disulfide-isomerase A2	K.NFEQVAFDETK.N	3.4E-05	30.06
9	P07477	PRSS1	Trypsin-1	K.TLNNDIMLIK.L	3.3E-05	28.44
10	P09210	GSTA2	Glutathione S-transferase A2	K.LALIQEK.T	9.0E-05	10.78
11	O43175	PHGDH	D-3-phosphoglycerate dehydrogenase	K.TLGILGLGR.I	1.6E-06	10.27
12	Q13310	PABPC4	Polyadenylate-binding protein 4	K.SGVGNVFIK.N	8.4E-08	9
13	P07237	P4HB	Protein disulfide-isomerase	K.VDATEESDLAQQYGVR.G	1.1E-07	7.62
14	P43307	SSR1	Translocon-associated protein subunit alpha	K.GEDFPANNIVK.F	7.4E-07	7.21
15	P16989	YBX3	Y-box-binding protein 3	K.GAEAANVTGPDGVPVEGSR.Y	9.0E-07	6.36
16	Q9P2E9	RRBP1	Ribosome-binding protein 1	K.LLATEQEDAAVAK.S	1.8E-06	5.5
17	Q96AG4	LRRC59	Leucine-rich repeat-containing protein 59	K.LQQLPADFGR.L	3.9E-05	5.21
18	P13667	PDIA4	Protein disulfide-isomerase A4	K.VEGFPTIYFAPSGDK.K	1.9E-06	4.82
19	P11021	HSPA5	78 kDa glucose-regulated protein	K.NQLTSNPENTVFDAK.R	5.6E-07	4.5
20	Q9Y4L1	HYOU1	Hypoxia up-regulated protein 1	K.AANSLEAFIFETQDK.L	9.3E-05	4.23
21	O94760	DDAH1	N(G),N(G)-dimethylarginine dimethylaminohydrolase 1	R.ALPESLGQHALR.S	4.4E-07	3.66
22	P24534	EEF1B2	Elongation factor 1-beta	K.YGPADVEDTTGSGATDSK.D	2.0E-04	3.41
23	P30086	PEBP1	Phosphatidylethanolamine-binding protein 1	K.LYEQLSGK.-	1.4E-05	3.36
24	P63220	RPS21	40S ribosomal protein S21	K.DHASIQMNVAEVDK.V	4.7E-05	3.34
25	P61247	RPS3A	40S ribosomal protein S3a	K.TTDGYLLR.L	9.9E-06	3.34
26	P62263	RPS14	40S ribosomal protein S14	K.TPGPGAQSALR.A	2.9E-06	3.12
27	P30050	RPL12	60S ribosomal protein L12	K.IGPLGLSPK.K	1.8E-04	2.97
28	Q92734	TFG	Protein TFG	K.LLSNDEVTIK.Y	1.9E-05	2.43

HC, healthy controls; PC, pancreatic cancer; PRM, parallel reaction monitoring.

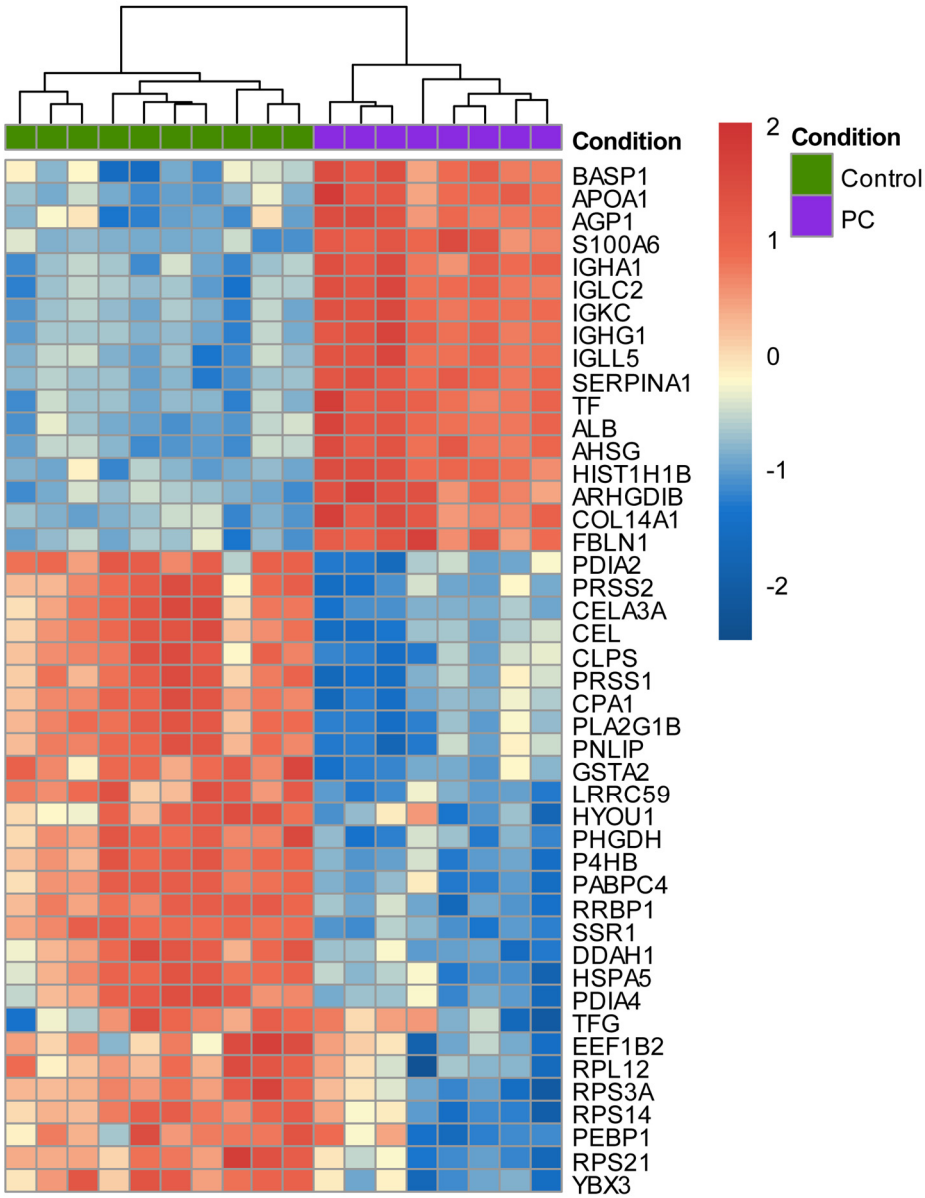


Figure 18. Heat map of the 45 PRM verified biomarker candidates. A clear separation between pancreatic cancer and controls can be observed. The names of the protein biomarker candidates are listed to the right.

To query the underlying interactions of those biomarker candidates from the verified panel, the open-source toolset FUNRICH (106) version 3.0 was used. The significantly upregulated and downregulated proteins were analysed separately, and each network presented tight interactions among the corresponding proteins (Figure 19).

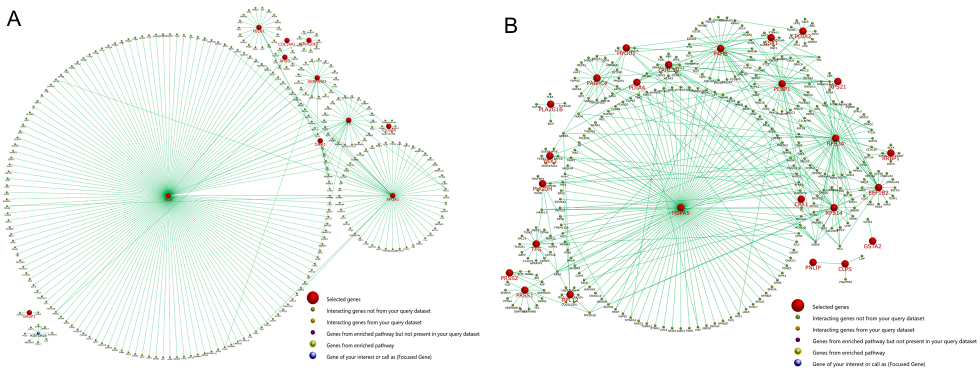
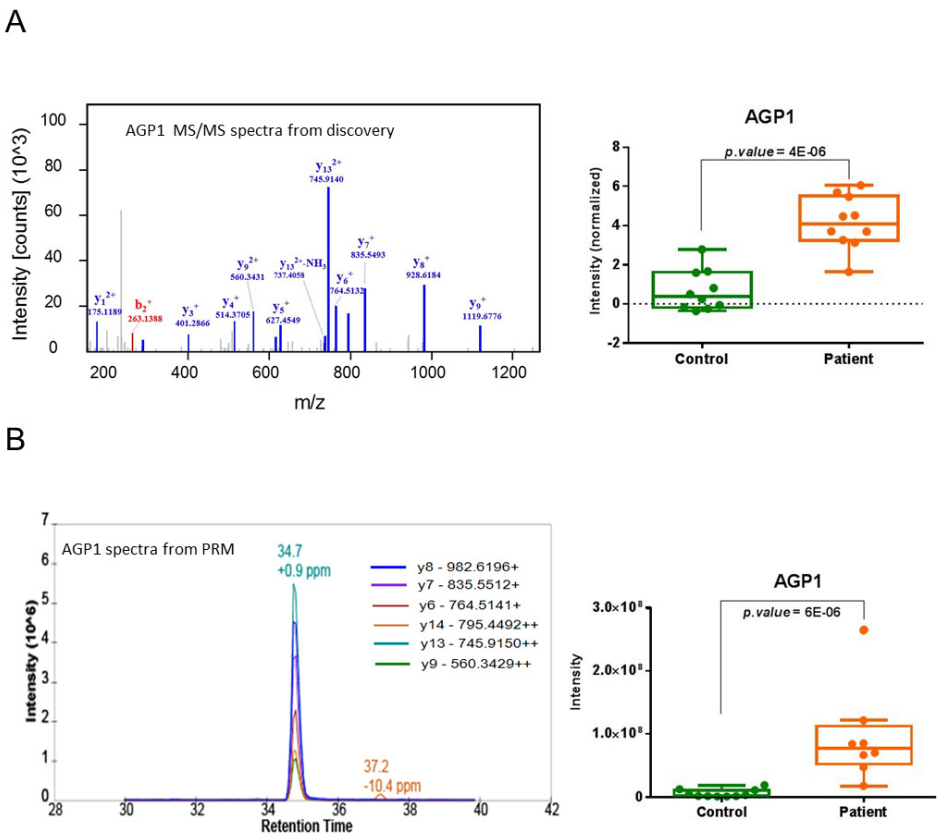


Figure 19. Protein-protein interactions of the 45 PRM-verified protein biomarker candidates. (A) Interactions of the 17 PRM-verified upregulated biomarker candidates in the pancreatic cancer group. (B) Interactions of the 28 PRM-verified downregulated biomarker candidates in the pancreatic cancer group.

4.3 Validation of diagnostic biomarker candidates

Initially, we were interested in extracellular proteins from the MS-discovered potential protein biomarkers which could be detected in serum and potentially be developed for in non-invasive diagnosis or prognosis of pancreatic cancer (Article IV). Therefore, we went thought the 165-protein list and cross-referenced it to database records and current literature, and eventually 52 extracellular proteins stood out. These candidate markers were further subjected to targeted MS verification, and 16 were successfully re-confirmed by the targeted label-free PRM: AGP1, AHSG, ALB, APOA1, CLPS, COL14A1, FBLN1, HYOU1, PLA2G1B, PNLIP, PRSS1, PRSS2, P4HB, SERPINA1, S100A6 and TF. By quantification of the unique peptides SDVVYTDWK and YVGGQEHFAHLLILR, AGP1 was found to be a top-ranked protein ($p = 5E\text{-}06$) both in the MS discovery study and the targeted MS verification study (Figure 20). Thus, AGP1 was selected as a biomarker candidate for the next stage, which was large cohort patient sample validations.



Figures 20. Proteomic-based MS discovery and targeted PRM verification of AGP1. (A) MS spectra of AGP1 (based on unique peptide YVGGQEHFAHLLILR) in the label-free quantitative MS discovery phase (left); box-plot showing the different expression levels of AGP1 in pancreatic cancer patients and matched normal controls (right). (B) The PRM transitions (based on unique peptide YVGGQEHFAHLLILR) for targeted MS verification of AGP1 (left); box-plot demonstrating different expression levels of AGP1 in pancreatic cancer patients and matched normal controls (right).

To find the protein-protein interactions and disease-linked pathways in which AGP1 is involved, we performed IPA analysis to construct a network of potential protein-protein interactions and possible functional relationships by extracting information from the literature and the databases. A group of extracellular proteins including several interleukins, interferon gamma, and tumour necrosis factor were found to physically or functionally interact with AGP1 in the resulting protein networks (Figure 21).

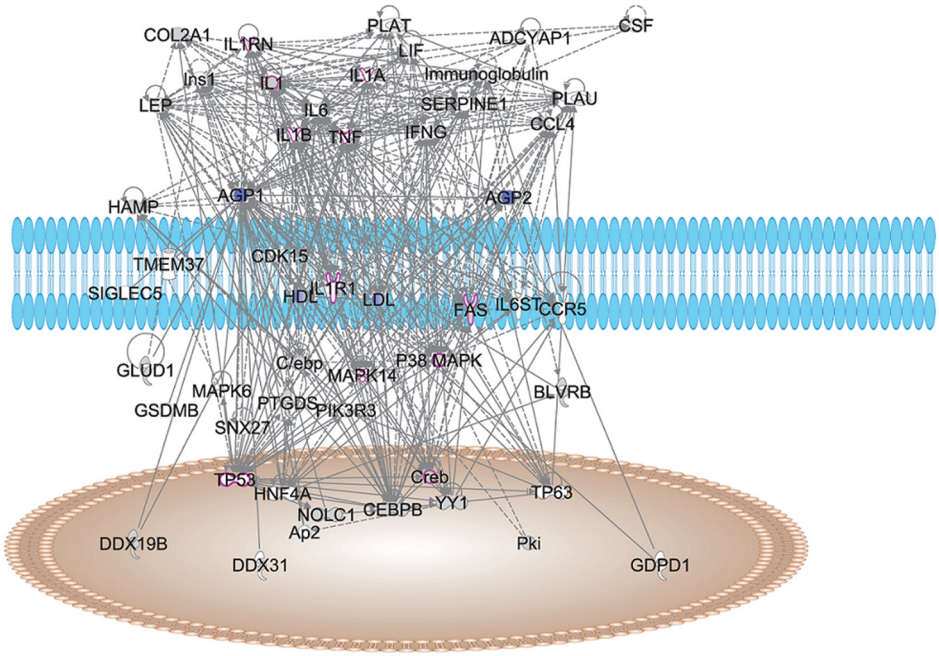


Figure 21. IPA-based bioinformatic analysis of AGP1. A network was built of proteins that physically and/or functionally interact with AGP1. Cell nucleus illustrated as brown oval and cell membrane depicted in blue. Direct relationships (e.g. protein-protein interactions) are shown using solid lines, while indirect relationships (e.g. regulation of expression) are shown using dashed lines. Proteins from the canonical MAPK signalling pathway are marked by magenta contours.

AGP1 was found to be involved in multiple well-established cancer-associated signalling pathways such as p38 and MAPK14 (enrichment p = 1E-10) and IL-10 and IL-6 pathways. Some cancer-related transcriptional regulators were found to interplay with the AGP1-centred protein networks, including Creb, HNF4A, TP53 and YY1. Among many other significantly enriched canonical pathways, the Acute Phase Response signalling pathway (enrichment p = 1E-21) was top-ranked, followed by associated proteins previously proposed as diagnostic markers for many gastrointestinal cancers. Many proteins involved in AGP1-centred functional network are not well characterised. For example, the recently discovered transmembrane structure molecule TMEM37 has only been reported to be a prognostic marker in colon cancer (107), and there is no further associated study, to the best of our knowledge, elucidating its biological activity in malignancy.

To determine if the circulating serum levels of AGP1 were consistent with the MS results and could enable non-invasive detection of pancreatic cancer, levels of AGP1 and the traditional pancreatic cancer diagnostic marker CA 19-9 were measured in 110 serum samples, including 52 patients with resectable pancreatic cancer, 24 with benign pancreatic disease and 34 healthy controls. The results

indicated that the levels of AGP1 in serum were significantly higher in patients with pancreatic cancer compared with the healthy controls ($p < 0.001$; Figure 22A). There were no significant differences in AGP1 levels between pancreatic cancer patients and those with benign pancreatic disease. CA 19-9 levels were significantly higher in pancreatic cancer patients compared to the benign pancreatic disease group ($p < 0.001$) and to the healthy control group ($p < 0.001$; Figure 22B).

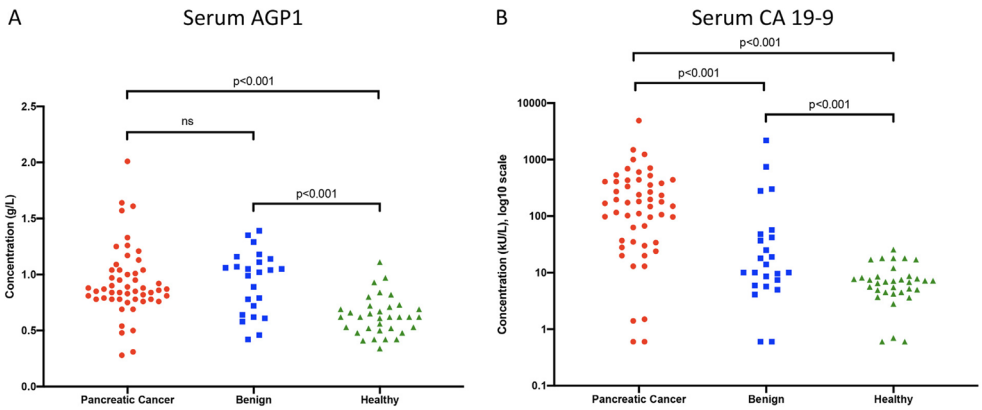


Figure 22. Scatter dot plots of serum AGP1 and CA 19-9 levels in pancreatic cancert and control groups. (A) Serum AGP1 levels in pancreatic cancer, benign pancreatic disease and healthy controls. (B) Serum CA 19-9 levels in pancreatic cancer, benign pancreatic disease and healthy controls. Wilcoxon rank-sum (Mann-Whitney) test. AGP1, alpha-1-acid glycoprotein 1.

The Spearman correlation between AGP1 and CA 19-9 was 25.3%, indicating that the two markers complement one another well. AGP1 displayed an AUC of 0.837 for the discrimination of pancreatic cancer from healthy controls, with a sensitivity of 86.5% at 82.4% specificity. CA 19-9 provided a lower sensitivity at 75% but with a specificity of 100% (AUC 0.919), when the standard cut-off of 37 kU/L was used. The maximum Youden's Index was applied to determine the optimal cut-off concentration for AGP1, which was found to be 0.74 g/L. Combining AGP1 with CA 19-9 increased the AUC to 0.963 (Figure 23). For discrimination of pancreatic cancer from healthy and benign groups, AGP1 alone provided an AUC of 0.678, which was increased to 0.798 when AGP1 was combined with CA 19-9.

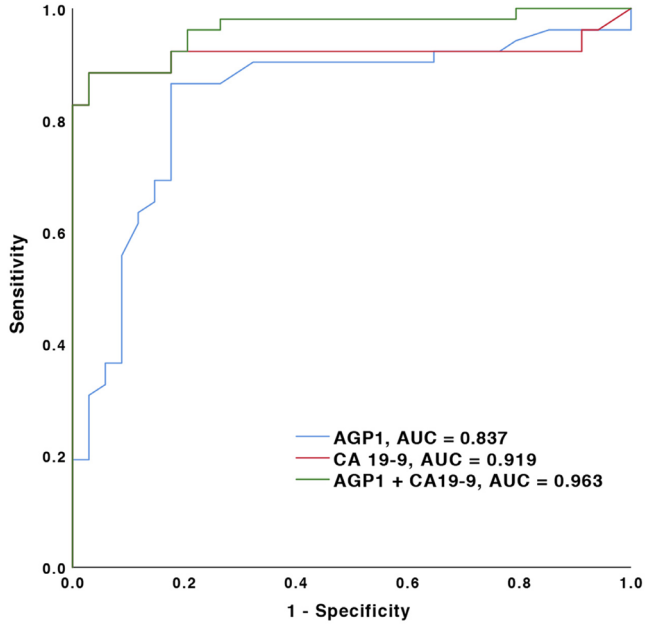
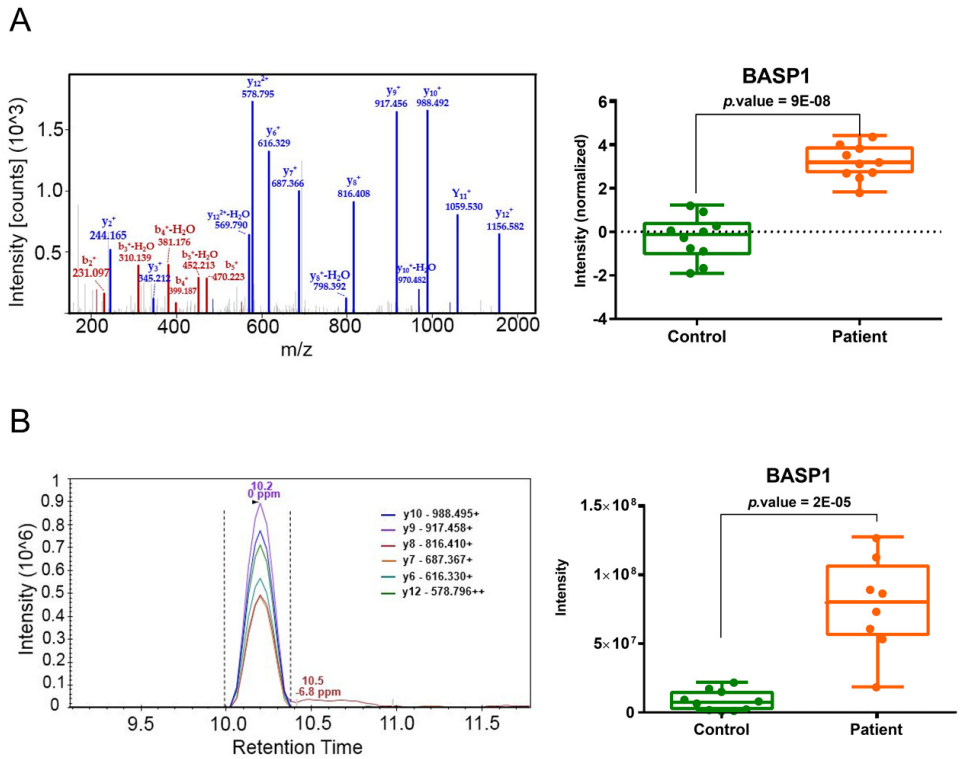


Figure 23. Diagnostic performance of serum AGP1 and CA 19-9 levels for detection of pancreatic cancer against healthy individuals.

4.4 Validation of prognostic biomarker candidates

In study II, we focused on BASP1, which was discovered by our MS analysis and found to be a novel protein in the context of pancreatic cancer (Figure 24).



Figures 24. MS-based discovery and targeted PRM verification of BASP1. (A) MS spectra of BASP1 (based on unique peptide ETPAATEAPSSTPK) in the label-free quantitative MS discovery phase (left), box-plot illustrating the expression levels of BASP1 in pancreatic cancer patients and matched normal controls (right). (B) The PRM transitions (based on unique peptide ETPAATEAPSSTPK) for targeted MS verification of BASP1 (left), box-plot demonstrating expression levels of BASP1 in pancreatic cancer patients and matched normal controls (right).

The IPA-based bioinformatic analysis suggests that BASP1 interacts with WT1 and several other proteins from the Pancreatic Adenocarcinoma Signalling pathway (Figure 25).

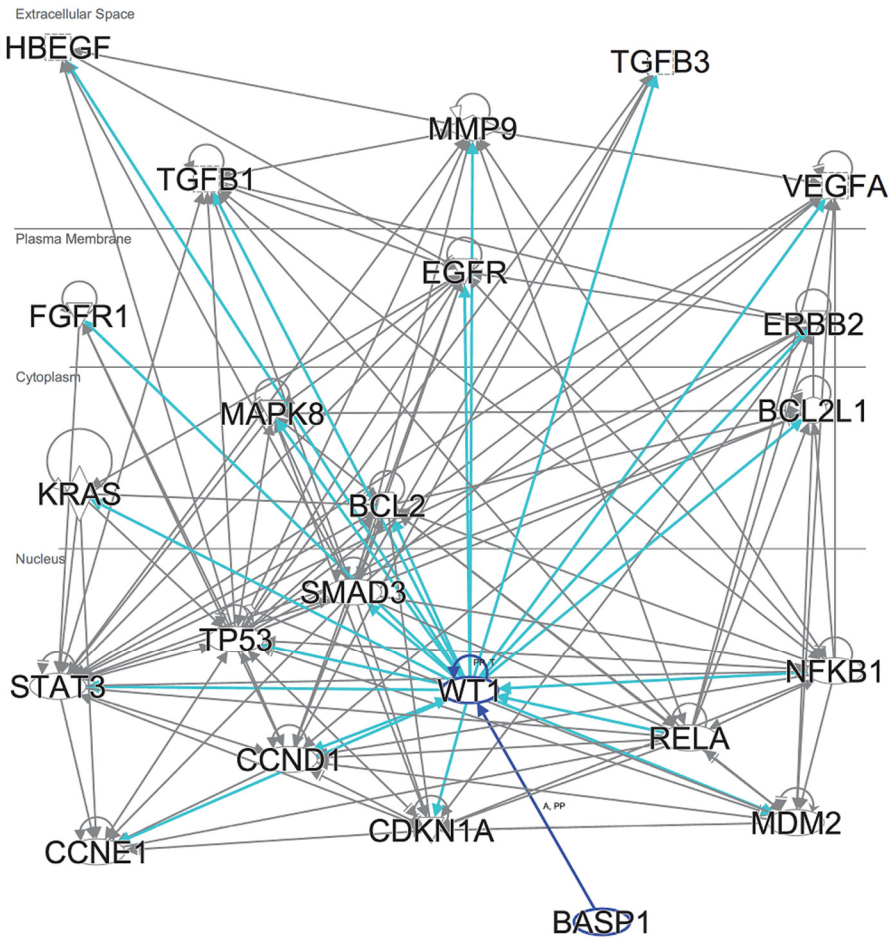


Figure 25. IPA-based bioinformatic analysis of BASP1. The network was built with all proteins in the pancreatic adenocarcinoma signaling pathway that have direct biological interactions with BASP1 or WT1.

The expression pattern of BASP1 and WT1 in both patient-derived tumour cells and tissue specimens is shown in Figure 26.

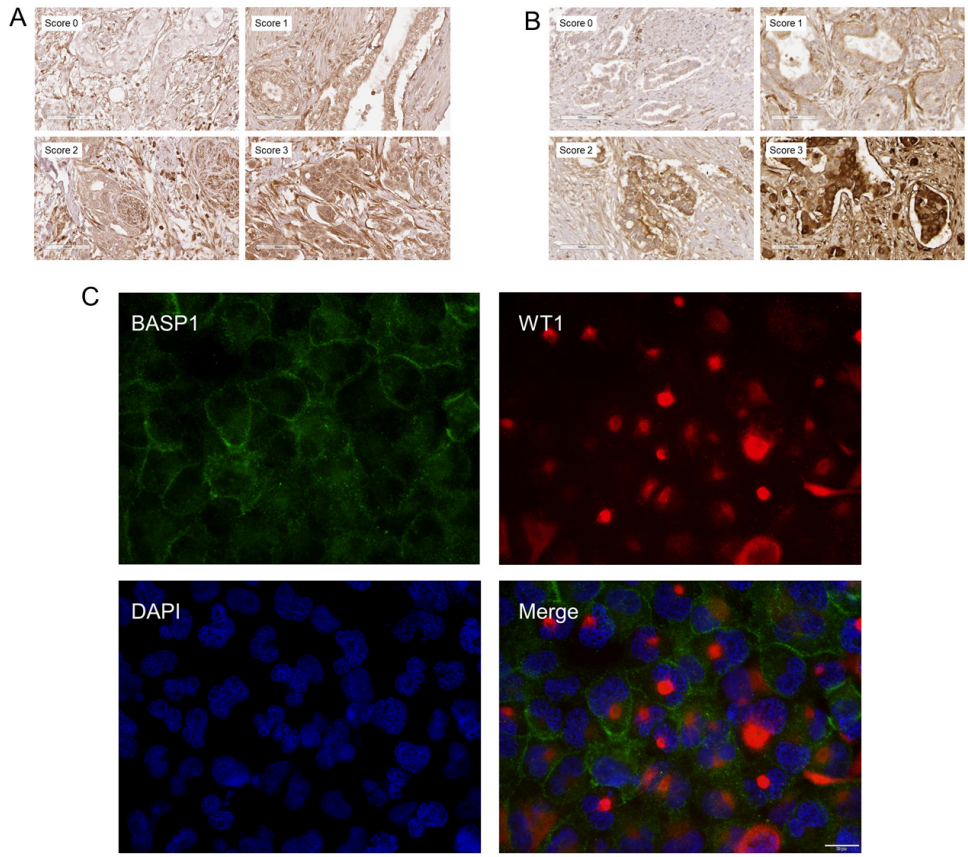


Figure 26. Immunohistochemistry and immunofluorescence analysis of BASP1 and WT1. (A) Representative photomicrographs exemplifying different staining patterns of BASP1 in pancreatic cancer tissue samples. (B) Representative photomicrographs illustrating different staining patterns of WT1 in pancreatic cancer tissue samples. (C) Panc-1 cancer cells were labeled with antibodies for BASP1, WT1, and DAPI; green represents BASP1 (mostly expressed in cytoplasm and plasma membrane), red represents WT1 (detected in cytoplasm and mainly perinuclearly localised), and blue represents nuclear DNA staining by DAPI.

Kaplan-Meier analysis showed that BASP1 expression in the pancreatic tissue was associated with a significantly prolonged OS compared to patients without BASP1 expression ($p = 0.022$, Figure 27A). Multivariate Cox regression analysis confirmed BASP1 expression as an independent prognostic factor for favourable survival (HR = 0.468, 95% CI 0.257-0.852, $p = 0.013$, Table 11). In contrast, pancreatic cancer patients with high WT1 expression in pancreatic tissue had significantly shorter OS compared to those with low WT1 expression ($p = 0.028$, Figure 27B). High WT1

expression was found to be an independent factor for worse OS (HR of 1.636, 95% CI 1.083-2.473, p = 0.019, Table 11).

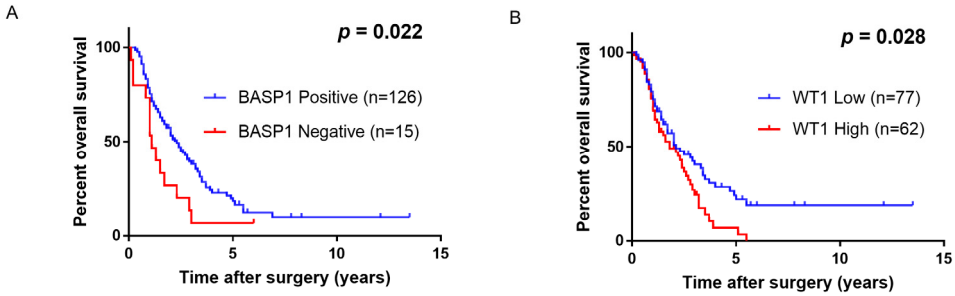


Figure 27. Prognostic implications of (A) BASP1 and (B) WT1 expression in pancreatic cancer.

Subgroup analysis showed that the best survival was observed in patients with positive expression of BASP1 and low WT1 expression, whereas patients with negative BASP1 expression and high WT1 expression had the poorest outcomes (p = 0.0001, Figure 28). The multivariable Cox regression analysis highlighted negative BASP1 expression and high WT1 expression as an independent factor associated with significantly shortened OS (HR 3.536, 95% CI 1.336–9.362, p = 0.011). This data suggests that BASP1 may act as a tumour suppressor rescuing the oncogenic effect of overexpressed WT1.

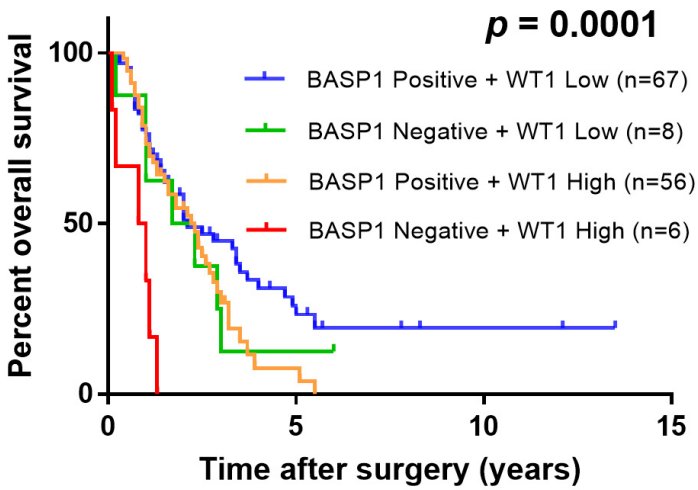


Figure 28. Prognostic subgroup analysis of BASP1 and WT1. Pancreatic cancer patients with positive BASP1 and low WT1 expression presented the best prognosis, while pancreatic cancer patients with negative BASP1 and high WT1 demonstrated the worst survival.

In study III, we explored the prognostic value of YAP1, which was the top upregulated protein in pancreatic cancer tissue as identified by MS analysis (Figure 29).

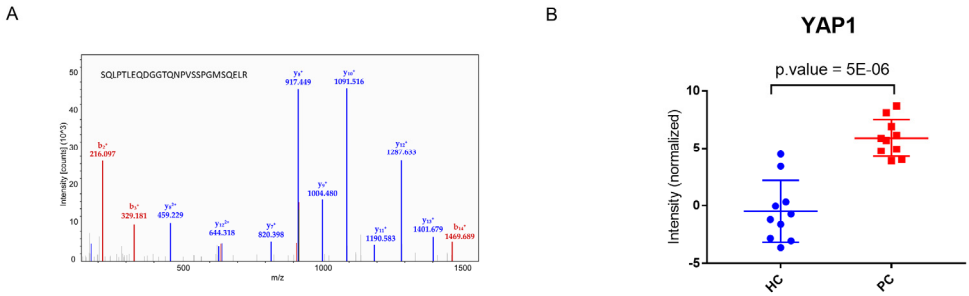


Figure 29. YAP1 is the top 1 upregulated protein in pancreatic cancer tissue. (A) MS spectra for YAP1 (based on the unique peptide SQLPTLEQDGGTQNPVSSPGMSQELR). (B) Box plot showing YAP1 expression levels in pancreatic cancer (PC) tissues compared to healthy controls (HC).

The prognostic utility of YAP1 was first evaluated at the mRNA expression level. Open-access mRNA data from TCGA was analysed in 176 patients with pancreatic cancer (83-85). The Kaplan-Meier plots revealed that the high expression of YAP1 was significantly correlated with poor outcome, as illustrated in Figure 30 (p = 0.0002). No patient reached 5-year survival in the YAP1 high expression group, while 32% of YAP1 low expression patients survived beyond 5-years.

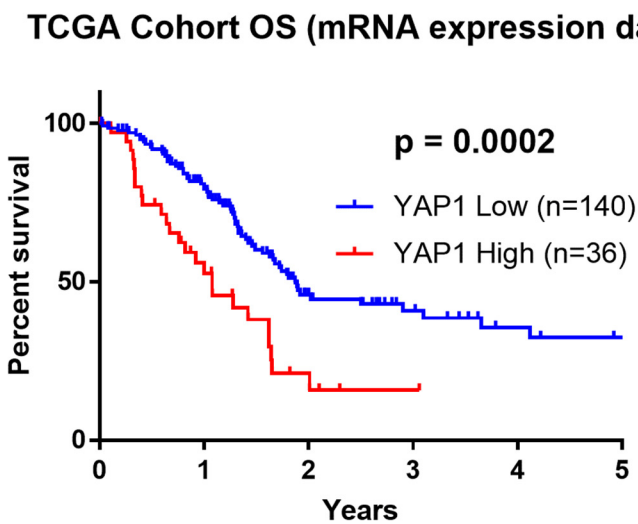


Figure 30. Prognostic role of YAP1 mRNA expression within the TCGA pancreatic cancer dataset.

YAP1 protein expression was evaluated by tissue microarray and IHC analysis in a local cohort of 140 resected pancreatic cancer patients. Protein level survival analysis revealed that pancreatic cancer patients with high expression of YAP1 had significantly reduced overall survival compared with low YAP1 expression patients (p = 0.001, Figure 31A). High YAP1 expression also correlated significantly with decreased disease-free survival (p = 0.005, Figure 31B).

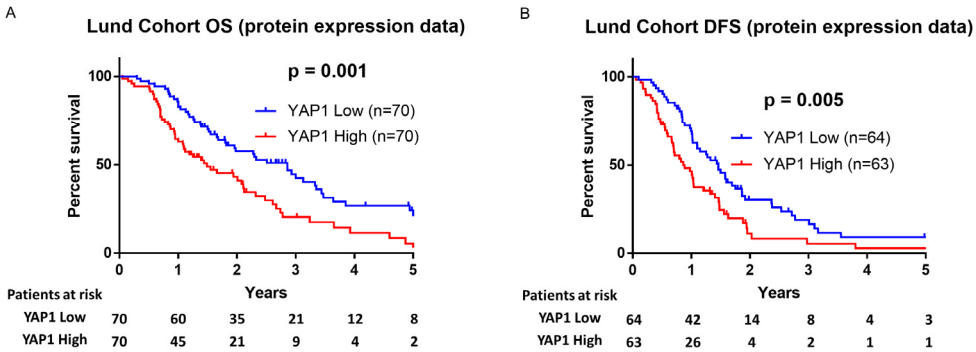


Figure 31. Prognostic role of YAP1 protein expression in the Lund pancreatic cancer dataset. (A) Overall survival (OS). (B) Disease-free survival (DFS).

Multivariant Cox regression analysis confirmed that high YAP1 protein expression was as an independent prognostic factor for poor OS (HR of 1.87, 95% CI 1.224-2.855, p = 0.004) and reduced DFS (HR of 1.95, 95% CI 1.299-2.927, p = 0.001) (Table 11).

To further delineate the functional role of YAP1 in pancreatic cancer, IPA-based bioinformatic analysis was performed. These analyses showed that YAP1 is involved in mechanotransduction and interacts with cell membrane proteins PATJ and PIEZO1. YAP1 is also an indirect regulator of both PIEZO1 and PIEZO2. The cytokine EDN1 is directly related to YAP1 and is also a regulator of the degenerin/epithelial sodium channels (DEG/ENaC, marked in Figure 32 as SCNN1A, SCNN1B, SCNN1G and SCNN1D). Tight junction signalling proteins related to YAP1 include CTNNA1, MPDZMPP5, OCLN, PATJ and TJP2. Epithelial adherens junction signalling proteins related to YAP1 include CDH1, CTNNA1, CTNNA2, EGFR, FGF1, PARD3 and ZYX. Examples of secreted proteins involved in creating a profibrotic microenvironment include AREG, CTGF, CYR61, FGF1 and MSLN; these YAP1 target genes are highlighted in Figure 32 and were chosen for further in vitro confirmation.

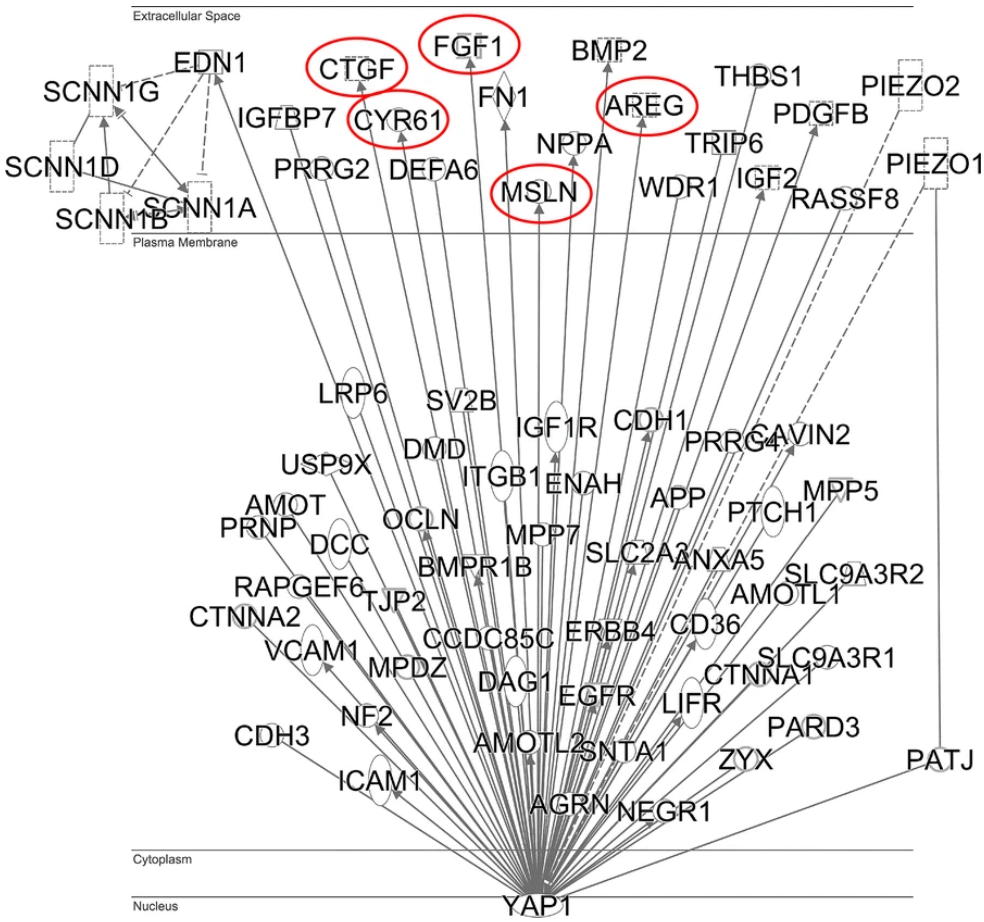


Figure 32. Ingenuity Pathway Analysis showing the plasma membrane and extracellular proteins directly related to YAP1.

The expression of selected YAP1 target gene products was evaluated after treatment of the pancreatic cancer cell line Panc-1 with three substances interrupting YAP-TEAD interaction: Super-TDU, Verteporfin and CA3 (Figure 33). We found that inhibition of YAP1/TEAD interaction interferes with the expression of AREG, CTGF, CYR61 and MSLN, suggesting that YAP1 transcriptional activity may affect the development and persistence of a fibrotic tumour microenvironment.

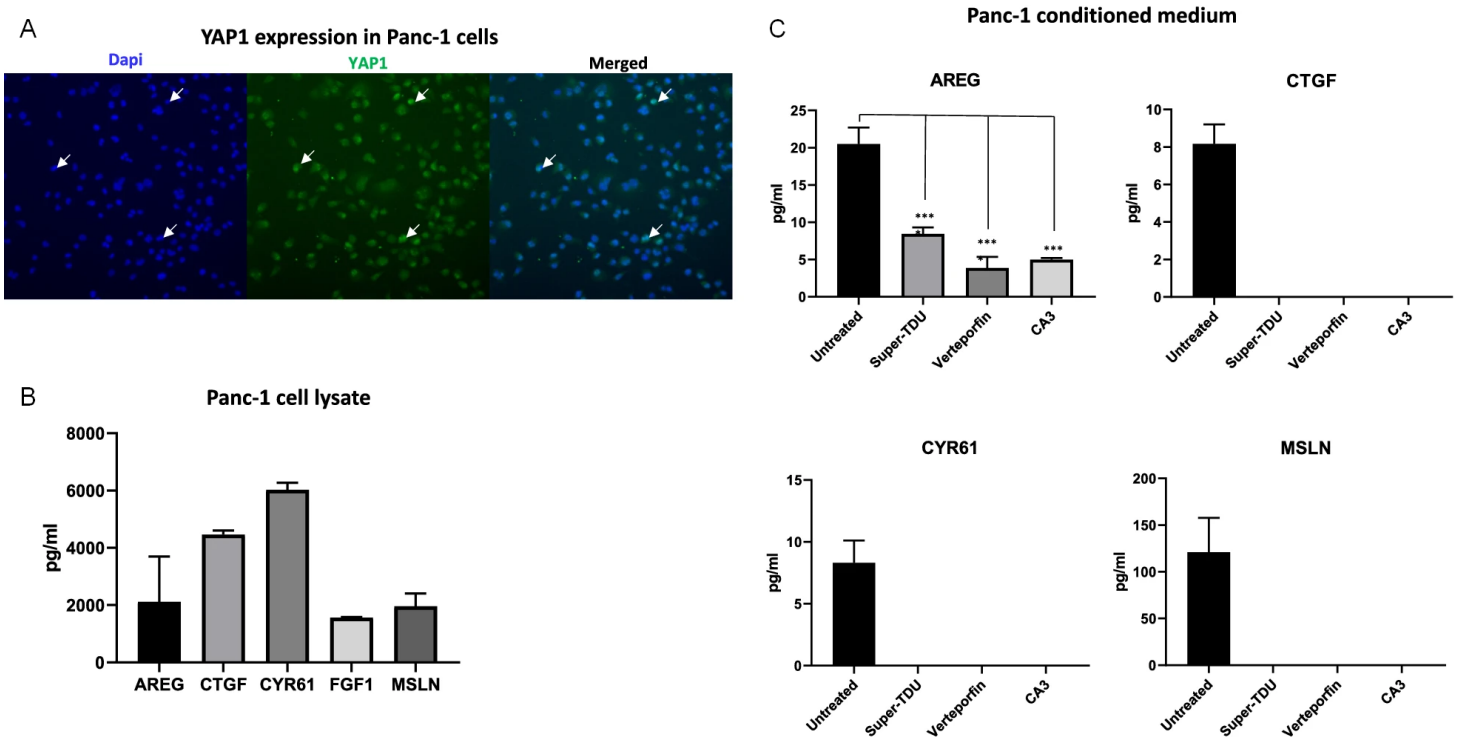


Figure 33. In vitro experiments of YAP1 and selected target genes in pancreatic cancer cells. (A) Expression pattern of the YAP1 protein in Panc-1 cells. Representative immunofluorescence staining of endogenous YAP1 in Panc-1 cells. The arrows indicate YAP1 nuclear accumulation. (B) Concentrations of YAP1 target genes in Panc-1 cell lysates. C) Concentrations of YAP1 target genes in conditioned medium from Panc-1 cells in, which were subjected to maximal tolerable doses of drugs blocking the YAP1/TEAD interaction.

In study IV, the prognostic value of the MS-identified AGP1 protein was investigated. Kaplan-Meier survival analysis revealed that pancreatic cancer patients with positive expression of AGP1 in pancreatic cancer tissue have significantly reduced OS compared with negative expression patients (p = 0.003, Figure 34).

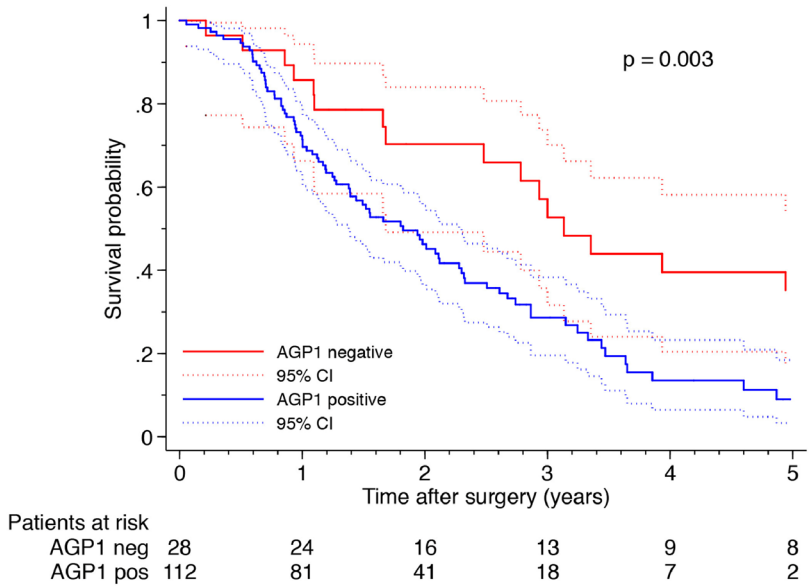


Figure 34. AGP1 expression is associated with poor survival in pancreatic cancer patients after surgical resection.

The univariable analysis highlighted that high expression levels of AGP1 as factors associated with shorter OS (HR of 2.22; 95% CI 1.30-3.79, P = 0.004), and multivariate analysis revealed high expression levels of AGP1 as an independent risk factor for poor OS (HR of 1.87; 95% CI 1.08-3.24, P = 0.026, Table 11).

Table 11. Univariable and multivariable analyses results of prognostic biomarker candidates

Variables	Survival	HR	95% CI	p-value
AGP1 positive, univariable	OS	2.22	1.30-3.79	0.004
AGP1 positive, multivariable	OS	1.87	1.08-3.24	0.026
BASP1 positive, univariable	OS	0.52	0.30-0.92	0.025
BASP1 positive, multivariable	OS	0.47	0.26-0.85	0.013
WT1 High, univariable	OS	1.56	1.05-2.33	0.029
WT1 High, multivariable	OS	1.64	1.08-2.47	0.019
YAP1 High, univariable	OS	1.92	1.29-2.85	0.001
YAP1 High, multivariable	OS	1.87	1.22-2.86	0.004
YAP1 High, univariable	DFS	1.75	1.18-2.61	0.006
YAP1 High, multivariable	DFS	1.95	1.30-2.93	0.001

CI, confidence interval; DFS, disease-free survival; HR, hazard ratio; OS, overall survival.

4.5 Validation of predictive biomarker candidates

In study II, we showed that pancreatic cancer patients with high levels of BASP1 expression (Score 3) had improved OS when administered adjuvant chemotherapy compared to those without adjuvant chemotherapy (p = 0.020, Figure 35A). In contrast, in pancreatic cancer patients with low BASP1 expression levels (score 0, 1 and 2) the positive effect of adjuvant chemotherapy was not observed (p = 0.603, Figure 35B). Our collected results suggest that BASP1 may function both as a marker for favourable prognosis and as a predictive biomarker for positive adjuvant chemotherapy response.

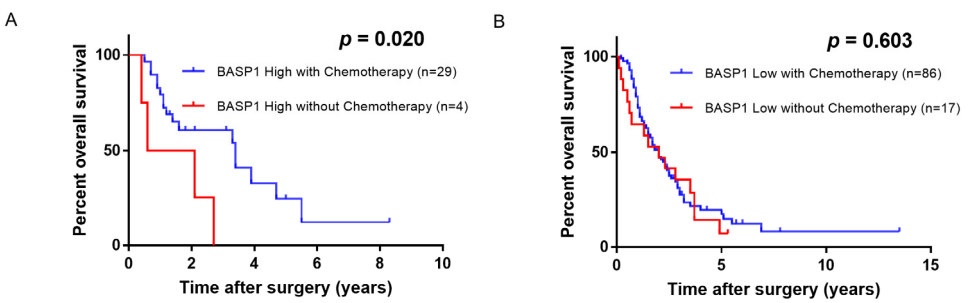


Figure 35. BASP1 as a predictive biomarker in pancreatic cancer tissue. (A) Pancreatic cancer patients with high BASP1 expression showed prolonged OS with adjuvant chemotherapy. (B) In pancreatic cancer patients with low BASP1 expression, adjuvant chemotherapy did not impact OS.

In pancreatic cancer patients with strong expression of WT1, adjuvant chemotherapy had no effect on OS (p = 0.335, Figure 36A). However, in pancreatic cancer patients with weak-to-moderate WT1 expression, adjuvant chemotherapy could significantly enhance OS (p = 0.006, Figure 36B). These findings indicate that WT1 expression is correlated with chemoresistance in pancreatic cancer.

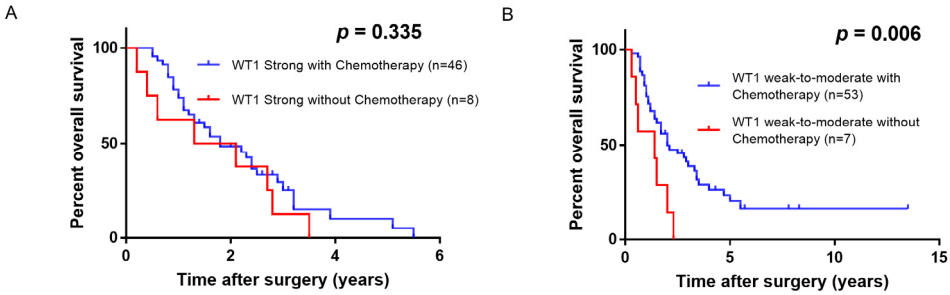


Figure 36. WT1 as a predictive biomarker in pancreatic cancer tissue. (A) In pancreatic cancer patients with strong expression of WT1, adjuvant chemotherapy showed no effect on OS. (B) In pancreatic cancer patients with weak-to-moderate WT1 expression, adjuvant chemotherapy could significantly enhance OS.

5. Discussion

Pancreatic cancer remains an intractable disease that often evades early diagnosis and defies treatment. It has the highest mortality rate among all cancer types. To overcome these dismal statistics, pancreatic cancer needs new, cross-disciplinary solutions integrating discoveries in molecular biology with clinical research in detection, risk stratification and treatment. This thesis represents just such a cross-disciplinary effort to improve the management of pancreatic cancer by the development and validation of new types of protein biomarkers which are of clinical utility.

Mass spectrometry (MS) has become the method of choice in clinical protein science. The sensitivity of MS instruments has enabled the accurate identification and quantitation of proteins in complex biological samples. In study I, we modified a sample preparation protocol to evaluate its effect on results from MS analysis. In studies II to IV we identified novel protein biomarkers for pancreatic cancer by discovery and targeted MS analysis, pathway analysis and antibody-based validation in tissue and serum. The general workflow presented in this thesis may also apply to the identification of biomarkers for other complex and multifaceted diseases.

Methodological considerations (study I)

Whether the benefits observed for urea digests of proteomic samples at room temperature apply to other commonly used denaturants such as guanidinium chloride, detergents or organic solvents is currently unclear and will be the subject of future investigations. However, although incubation of the enzymatic digest at room temperature is feasible and showed some promise, further validation is necessary before this protocol can become routine.

Methodological considerations (studies II-IV)

A particular strength of our study was that healthy pancreas biopsies were used as control systems in the biomarker discovery phase. These unique and rare specimens were acquired from organ donors. Previous proteomic studies have generally used histologically normal tissue adjacent to the tumour as a control. However, the regions adjacent to tumours have been found to have many aberrant morphologic and phenotypic alterations as predicted by the ‘field cancerisation theory’ (108, 109). The choice of healthy tissue as a comparative material for identification and further development of a discriminative biomarker is therefore preferable.

There are some potential limitations in our study that must be acknowledged. The fresh-frozen samples used in the discovery phase were limited in number. The tissue microarray samples were accrued over a long period and could have been vulnerable to changes in histopathological characterisation, treatment and follow-up. However, all tissue specimens were re-evaluated by a dedicated pancreas pathologist to confirm the diagnosis and uniformity of histopathological evaluation. Chemotherapy regimens varied during the study period, but most patients received gemcitabine-based chemotherapy.

6. Conclusions

The major conclusions reached in the studies in this thesis were:

I. Urea in-solution digestion at room temperature (24 ± 2 °C) is superior to traditional proteolysis at 37 °C, presenting several advantages such as reducing peptides with carbamylation and pyroglutamic acid modifications, increasing the identification of peptides and proteins, and obtaining lower coefficients of variations for protein quantitation.

II. MS-based proteomic profiling of tissue specimens from pancreatic cancer patients and healthy controls identified 165 potential protein biomarkers and verified a novel panel of 45 biomarker candidates for pancreatic cancer. The novel protein BASP1 was significantly associated with favourable survival and positive response to adjuvant chemotherapy in pancreatic cancer patients, while its putative interaction partner WT1 exhibited the opposite effect on patient outcomes. We propose that BASP1 and WT1 could potentially be used as biomarkers for monitoring the prognosis and predicting the outcomes of patients with pancreatic cancer.

III. YAP1 was found to be a candidate prognostic biomarker associated with recurrence and unfavourable survival of pancreatic cancer patients. Inhibition of YAP1/TEAD interaction interferes with the expression of AREG, CTGF, CYR61, and MSLN in pancreatic cancer cells, which suggests that interrupting YAP1 transcriptional activity may affect the development and persistence of fibrotic tumour microenvironment of pancreatic cancer.

IV. Expression of AGP1 in pancreatic cancer tissue served as an independent prognostic factor of early recurrence and poor survival of pancreatic cancer patients. Circulating levels of AGP1 is a candidate biomarker for non-invasive pancreatic cancer diagnosis.

7. Future study

Research in biomarkers presents great opportunities but also challenges to improve outcomes in pancreatic cancer.

We have demonstrated the feasibility of MS-based proteomic profiling of patient-derived tissue specimens for biomarker development in pancreatic cancer. The clinical utility of the selected candidates has been validated in large, clinically well-annotated tissue and serum samples. However, the current interpretation was based on specimens from a single biobank. Thus, to confirm and develop the clinical utility of those biomarker candidates, external cohorts of clinical tissue and biofluid samples collected and stored by a standard procedure in different institutes, and participants from different ethnicities are needed for future validation.

The proposed biomarker candidates must undergo further validation to confirm, for example, their efficacy with the early detection function. It is also important to consider which patients to test. As pancreatic cancer is relatively uncommon in the general population, screening needs to be focused on high-risk groups. Such groups include individuals with an inherited predisposition, patients with cystic lesions, individuals with symptoms that suggest pancreatic cancer and individuals with new-onset diabetes at over 50 years of age. We identified several extracellular and secreted proteins in our MS studies. These may also be included in a multiplex serum panel for early-stage diagnosis of pancreatic cancer.

Likewise, for evaluating the function of predictive biomarker candidates, clear results could only be obtained by prospective cohort-based studies in a randomised setting. It would be valuable to develop a predictive tissue panel which is evaluated at the time of diagnosis by sampling tissue via EUS before initiation of treatment.

8. Acknowledgements

I would like to express my sincere gratitude to everyone who supported me during the last few years. Without your guidance and encouragement, this thesis would not have been possible.

I would like specifically to thank:

Associate professor Daniel Ansari, my main supervisor; you are one of the most talented and hard-working people that I have ever met. I was very lucky to have you as my tutor during my PhD studies. You are an excellent role model as a surgeon-scientist, being able to combine clinical work with innovative, translational research, a path that I also want to pursue.

Professor Roland Andersson, my co-supervisor, for bringing me to Sweden and the Pancreas Lab. I cannot imagine how much effort you have put into making the double-PhD project come true. Apart from tutoring in science, your generosity and optimistic attitude towards life has inspired me.

Professor György Marko-Varga, my other co-supervisor, for giving me the opportunity to be part of the Proteomics Lab. You have taught me many things, and not only about clinical protein science.

Lazaro Hiram Betancourt, my tutor in the proteomics world, for teaching me all the experimental skills necessary to making me confident in the laboratory.

Agata Sasor for the huge contributions to the TMA constructions and, particularly, the time consuming immunohistochemical scoring of our protein markers.

Theresa Kristl for the outstanding work with the MS analysis, which is one of the corner stones of my thesis.

Monika Bauden for all your support with the experiments. We would have never achieved our research milestones without your help.

Katarzyna Said Hilmersson for help with the immunohistochemistry slides and the in vitro studies. Your kindness made me feel at home.

Krzysztof Pawłowski for all the generous help with the comprehensive bioinformatic analysis and the complicated investigation of protein networks and signal cascades.

Indira Pla for all the input and statistical support in the MS studies.

Aniel Sanchez for sharing your vast knowledge and providing an optimistic attitude towards science.

Jeovanis Gil Valdes for kind help with the MS data interpretation.

Roger Appelqvist for friendly and supportive help in the Lab.

Thomas Bygott for kind help with the statistics.

Linus Aronsson for double checking clinicopathological information of the patients.

Jonatan Eriksson and Yutaka Sugihara for the happy times we have spent together.

Magdalena Kuras, Melinda Rezeli, Henriett Kovacs-Oskolas, Karin Behrens, Charlotte Welinder, Henrik Lindberg, and all members from the CPS&I team for your kind help and sharing good times with me.

Dingyuan Hu, an amazing office mate and a good friend.

Monica Keidser, for kind help with every step of my PhD journey, from the administration to the PhD defence. I want to thank you for your great contributions behind these papers.

Jianfeng Xu, Hua Dai, Xi Chen, and Mengtao Zhou for helping with the immunohistochemistry evaluations and for your valuable comments on my work.

Yuanguang Li, Guangqiang Xu, Xunzi Xie, Longfei Huang and professor David Bryder for great support when the PhD project seemed insurmountable.

Yu Zhang, Mei Shen, Hongzhe Li, Haixia Wan, Wen Li, Haoran Yu, Chengjun Wu, Andy McCourt, Carlos Urey, Talía Velasco-Hernández, Isabel Hidalgo, Antonio Boza Serrano, Hooi Ching Lim, Maciej Ciesla, Anna Konturek-Ciesla, Magdalena Madej, Parashar Dhapola, Yuxiang Zhu, Hong Yan, Susan Lan, Kun Luo, and Ping Li, and many others, for the memorable moments we spent together.

Professor Lu Fan, my Chinese supervisor, without whom it would not have been possible for me to get this PhD.

Finally, I would like to thank my family and especially Yuan Ouyang, I love you all, and could not have done this without you.

9. References

1. Busnardo AC, DiDio LJ, Tidrick RT, Thomford NR. History of the pancreas. Am J Surg 1983;146:539-50.
2. Howard JM, Hess W, Traverso W. Johann Georg Wirsung (1589-1643) and the pancreatic duct: the prosector of Padua, Italy. J Am Coll Surg 1998;187:201-11.
3. Kleinerman R, John A, Etienne D, Turner B, Shoja MM, Tubbs RS, et al. Giovanni Domenico Santorini (1681-1737): a prominent physician and meticulous anatomist. Clin Anat 2014;27:545-7.
4. Stern CD. A historical perspective on the discovery of the accessory duct of the pancreas, the ampulla 'of Vater' and pancreas divisum. Gut 1986;27:203-12.
5. McClusky DA, 3rd, Skandalakis LJ, Colborn GL, Skandalakis JE. Harbinger or hermit? Pancreatic anatomy and surgery through the ages--part 1. World J Surg 2002;26:1175-85.
6. Baskin DG. A Historical Perspective on the Identification of Cell Types in Pancreatic Islets of Langerhans by Staining and Histochemical Techniques. J Histochem Cytochem 2015;63:543-58.
7. Karamitsos DT. The story of insulin discovery. Diabetes Res Clin Pract 2011;93 Suppl 1:S2-8.
8. Morgagni GB. De Sedibus et Causis Morborum per Anatomen Indagatis 1761.
9. Da Costa J. On the morbid anatomy and symptoms of cancer of the pancreas. JB Lippincott & Company 1858.
10. Kamisawa T, Wood LD, Itoi T, Takaori K. Pancreatic cancer. Lancet 2016;388:73-85.
11. Cancer i siffror 2018. Cancerfonden och Socialstyrelsen i samarbete. Socialstyrelsens artikelnummer: 2018--6--10.
12. Chen W, Zheng R, Baade PD, Zhang S, Zeng H, Bray F, et al. Cancer statistics in China, 2015. CA Cancer J Clin 2016;66:115-32.
13. Wu W, He X, Yang L, Wang Q, Bian X, Ye J, et al. Rising trends in pancreatic cancer incidence and mortality in 2000-2014. Clin Epidemiol 2018;10:789-97.
14. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin 2020;70:7-30.
15. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 2014;74:2913-21.

16. Ansari D, Althini C, Ohlsson H, Andersson R. Early-onset pancreatic cancer: a population-based study using the SEER registry. Langenbecks Arch Surg 2019;404:565-71.

17. Decker GA, Batheja MJ, Collins JM, Silva AC, Mekeel KL, Moss AA, et al. Risk factors for pancreatic adenocarcinoma and prospects for screening. Gastroenterol Hepatol (N Y) 2010;6:246-54.

18. Ryan DP, Hong TS, Bardeesy N. Pancreatic adenocarcinoma. N Engl J Med 2014;371:1039-49.

19. Jones S, Hruban RH, Kamiyama M, Borges M, Zhang X, Parsons DW, et al. Exomic sequencing identifies PALB2 as a pancreatic cancer susceptibility gene. Science 2009;324:217.

20. Hruban RH, Maitra A, Goggins M. Update on pancreatic intraepithelial neoplasia. Int J Clin Exp Pathol 2008;1:306-16.

21. Distler M, Aust D, Weitz J, Pilarsky C, Grutzmann R. Precursor lesions for sporadic pancreatic cancer: PanIN, IPMN, and MCN. Biomed Res Int 2014;2014:474905.

22. Yachida S, Iacobuzio-Donahue CA. Evolution and dynamics of pancreatic cancer progression. Oncogene 2013;32:5253-60.

23. Yachida S, Jones S, Bozic I, Antal T, Leary R, Fu B, et al. Distant metastasis occurs late during the genetic evolution of pancreatic cancer. Nature 2010;467:1114-7.

24. Notta F, Chan-Seng-Yue M, Lemire M, Li Y, Wilson GW, Connor AA, et al. A renewed model of pancreatic cancer evolution based on genomic rearrangement patterns. Nature 2016;538:378-82.

25. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med 2011;17:500-3.

26. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet 2015;47:1168-78.

27. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016;531:47-52.

28. Maurer C, Holmstrom SR, He J, Laise P, Su T, Ahmed A, et al. Experimental microdissection enables functional harmonisation of pancreatic cancer subtypes. Gut 2019;68:1034-43.

29. Bapat AA, Hostetter G, Von Hoff DD, Han H. Perineural invasion and associated pain in pancreatic cancer. Nat Rev Cancer 2011;11:695-707.

30. Wang F, Herrington M, Larsson J, Permert J. The relationship between diabetes and pancreatic cancer. Mol Cancer 2003;2:4.

31. Pannala R, Basu A, Petersen GM, Chari ST. New-onset diabetes: a potential clue to the early diagnosis of pancreatic cancer. Lancet Oncol 2009;10:88-95.

32. Vincent A, Herman J, Schulick R, Hruban RH, Goggins M. Pancreatic cancer. Lancet 2011;378:607-20.

33. Ansari D, Ansari D, Andersson R, Andren-Sandberg A. Pancreatic cancer and thromboembolic disease, 150 years after Trousseau. Hepatobiliary Surg Nutr 2015;4:325-35.

34. Amin MB, Edge S, Greene F, et al, eds. AJCC Cancer Staging Manual. 8th ed. New York: Springer; 2017.

35. Delpero JR, Sauvanet A. Vascular Resection for Pancreatic Cancer: 2019 French Recommendations Based on a Literature Review From 2008 to 6-2019. Front Oncol 2020;10:40.

36. Labori KJ, Lassen K, Hoem D, Gronbech JE, Soreide JA, Mortensen K, et al. Neoadjuvant chemotherapy versus surgery first for resectable pancreatic cancer (Norwegian Pancreatic Cancer Trial - 1 (NorPACT-1)) - study protocol for a national multicentre randomized controlled trial. BMC Surg 2017;17:94.

37. Burris HA, Moore MJ, Andersen J, Green MR, Rothenberg ML, Madiano MR, et al. Improvements in survival and clinical benefit with gemcitabine as first-line therapy for patients with advanced pancreas cancer: A randomized trial. J Clin Oncol 1997;15:2403-13.

38. Neoptolemos JP, Stocken DD, Friess H, Bassi C, Dunn JA, Hickey H, et al. A randomized trial of chemoradiotherapy and chemotherapy after resection of pancreatic cancer. N Engl J Med 2004;350:1200-10.

39. Conroy T, Desseigne F, Ychou M, Bouche O, Guimbaud R, Becouarn Y, et al. FOLFIRINOX versus gemcitabine for metastatic pancreatic cancer. N Engl J Med 2011;364:1817-25.

40. Conroy T, Hammel P, Hebbar M, Ben Abdelghani M, Wei AC, Raoul JL, et al. FOLFIRINOX or Gemcitabine as Adjuvant Therapy for Pancreatic Cancer. N Engl J Med 2018;379:2395-406.

41. Neoptolemos JP, Palmer DH, Ghaneh P, Psarelli EE, Valle JW, Halloran CM, et al. Comparison of adjuvant gemcitabine and capecitabine with gemcitabine monotherapy in patients with resected pancreatic cancer (ESPAC-4): a multicentre, open-label, randomised, phase 3 trial. Lancet 2017;389:1011-24.

42. Von Hoff DD, Ervin T, Arena FP, Chiorean EG, Infante J, Moore M, et al. Increased survival in pancreatic cancer with nab-paclitaxel plus gemcitabine. N Engl J Med 2013;369:1691-703.

43. Wang-Gillam A, Li CP, Bodoky G, Dean A, Shan YS, Jameson G, et al. Nanoliposomal irinotecan with fluorouracil and folinic acid in metastatic pancreatic

cancer after previous gemcitabine-based therapy (NAPOLI-1): a global, randomised, open-label, phase 3 trial. Lancet 2016;387:545-57.

44. Moore MJ, Goldstein D, Hamm J, Figer A, Hecht JR, Gallinger S, et al. Erlotinib plus gemcitabine compared with gemcitabine alone in patients with advanced pancreatic cancer: a phase III trial of the National Cancer Institute of Canada Clinical Trials Group. J Clin Oncol 2007;25:1960-6.

45. Golan T, Hammel P, Reni M, Van Cutsem E, Macarulla T, Hall MJ, et al. Maintenance Olaparib for Germline BRCA-Mutated Metastatic Pancreatic Cancer. N Engl J Med 2019;381:317-27.

46. Califf RM. Biomarker definitions and their applications. Exp Biol Med (Maywood) 2018;243:213-21.

47. Drilon A, Laetsch TW, Kummar S, DuBois SG, Lassen UN, Demetri GD, et al. Efficacy of Larotrectinib in TRK Fusion-Positive Cancers in Adults and Children. N Engl J Med 2018;378:731-9.

48. Lemery S, Keegan P, Pazdur R. First FDA Approval Agnostic of Cancer Site - When a Biomarker Defines the Indication. N Engl J Med 2017;377:1409-12.

49. FDA-NIH Biomarker Working Group. BEST (Biomarkers, EndpointS, and other Tools) Resource. Silver Spring (MD): Food and Drug Administration (US); Bethesda (MD): National Institutes of Health (US), www.ncbi.nlm.nih.gov/books/NBK326791/ (2016, accessed 9 March 2020).

50. Cohen JF, Korevaar DA, Altman DG, Bruns DE, Gatsonis CA, Hooft L, et al. STARD 2015 guidelines for reporting diagnostic accuracy studies: explanation and elaboration. BMJ Open 2016;6:e012799.

51. Ballehaninna UK, Chamberlain RS. The clinical utility of serum CA 19-9 in the diagnosis, prognosis and management of pancreatic adenocarcinoma: An evidence based appraisal. J Gastrointest Oncol 2012;3:105-19.

52. Lee HS, Jang CY, Kim SA, Park SB, Jung DE, Kim BO, et al. Combined use of CEMIP and CA 19-9 enhances diagnostic accuracy for pancreatic cancer. Sci Rep 2018;8:3383.

53. Yang Y, Yan S, Tian H, Bao Y. Macrophage inhibitory cytokine-1 versus carbohydrate antigen 19-9 as a biomarker for diagnosis of pancreatic cancer: A PRISMA-compliant meta-analysis of diagnostic accuracy studies. Medicine (Baltimore) 2018;97:e9994.

54. Jenkinson C, Elliott VL, Evans A, Oldfield L, Jenkins RE, O'Brien DP, et al. Decreased Serum Thrombospondin-1 Levels in Pancreatic Cancer Patients Up to 24 Months Prior to Clinical Diagnosis: Association with Diabetes Mellitus. Clin Cancer Res 2016;22:1734-43.

55. Kim J, Bamlet WR, Oberg AL, Chaffee KG, Donahue G, Cao XJ, et al. Detection of early pancreatic ductal adenocarcinoma with thrombospondin-2 and CA19-9 blood markers. Sci Transl Med 2017;9.

56. Mellby LD, Nyberg AP, Johansen JS, Wingren C, Nordestgaard BG, Bojesen SE, et al. Serum Biomarker Signature-Based Liquid Biopsy for Diagnosis of Early-Stage Pancreatic Cancer. J Clin Oncol 2018;36:2887-94.

57. Franklin O, Jonsson P, Billing O, Lundberg E, Ohlund D, Nystrom H, et al. Plasma Micro-RNA Alterations Appear Late in Pancreatic Cancer. Ann Surg 2018;267:775-81.

58. Bauden M, Pamart D, Ansari D, Herzog M, Eccleston M, Micallef J, et al. Circulating nucleosomes as epigenetic biomarkers in pancreatic cancer. Clin Epigenetics 2015;7:106.

59. Melo SA, Luecke LB, Kahlert C, Fernandez AF, Gammon ST, Kaye J, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. Nature 2015;523:177-82.

60. Cohen JD, Li L, Wang Y, Thoburn C, Afsari B, Danilova L, et al. Detection and localization of surgically resectable cancers with a multi-analyte blood test. Science 2018;359:926-30.

61. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer 2005;93:387-91.

62. Jamieson NB, Carter CR, McKay CJ, Oien KA. Tissue biomarkers for prognosis in pancreatic ductal adenocarcinoma: a systematic review and meta-analysis. Clin Cancer Res 2011;17:3316-31.

63. Huang X, Wang X, Lu SM, Chen C, Wang J, Zheng YY, et al. Clinicopathological and prognostic significance of MUC4 expression in cancers: evidence from meta-analysis. Int J Clin Exp Med 2015;8:10274-83.

64. Smith RA, Tang J, Tudur-Smith C, Neoptolemos JP, Ghaneh P. Meta-analysis of immunohistochemical prognostic markers in resected pancreatic cancer. Br J Cancer 2011;104:1440-51.

65. Han W, Cao F, Chen MB, Lu RZ, Wang HB, Yu M, et al. Prognostic Value of SPARC in Patients with Pancreatic Cancer: A Systematic Review and Meta-Analysis. PLoS One 2016;11:e0145803.

66. Bird NT, Elmasry M, Jones R, Psarelli E, Dodd J, Malik H, et al. Immunohistochemical hENT1 expression as a prognostic biomarker in patients with resected pancreatic ductal adenocarcinoma undergoing adjuvant gemcitabine-based chemotherapy. Br J Surg 2017;104:328-36.

67. Farrell JJ, Elsaleh H, Garcia M, Lai R, Ammar A, Regine WF, et al. Human equilibrative nucleoside transporter 1 levels predict response to gemcitabine in patients with pancreatic cancer. Gastroenterology 2009;136:187-95.

68. Greenhalf W, Ghaneh P, Neoptolemos JP, Palmer DH, Cox TF, Lamb RF, et al. Pancreatic cancer hENT1 expression and survival from gemcitabine in patients from the ESPAC-3 trial. J Natl Cancer Inst 2014;106:djt347.

69. Spratlin J, Sangha R, Glubrecht D, Dabbagh L, Young JD, Dumontet C, et al. The absence of human equilibrative nucleoside transporter 1 is associated with reduced survival in patients with gemcitabine-treated pancreas adenocarcinoma. Clin Cancer Res 2004;10:6956-61.

70. Capello M, Lee M, Wang H, Babel I, Katz MH, Fleming JB, et al. Carboxylesterase 2 as a Determinant of Response to Irinotecan and Neoadjuvant FOLFIRINOX Therapy in Pancreatic Ductal Adenocarcinoma. J Natl Cancer Inst 2015;107.

71. Khanna R, Morton CL, Danks MK, Potter PM. Proficient metabolism of irinotecan by a human intestinal carboxylesterase. Cancer Res 2000;60:4725-8.

72. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature 2015;518:495-501.

73. Villarroel MC, Rajeshkumar NV, Garrido-Laguna I, De Jesus-Acosta A, Jones S, Maitra A, et al. Personalizing cancer treatment in the age of global genomic analyses: PALB2 gene mutations and the response to DNA damaging agents in pancreatic cancer. Mol Cancer Ther 2011;10:3-8.

74. Hu YC, Komorowski RA, Graewin S, Hostetter G, Kallioniemi OP, Pitt HA, et al. Thymidylate synthase expression predicts the response to 5-fluorouracil-based adjuvant therapy in pancreatic cancer. Clin Cancer Res 2003;9:4165-71.

75. Boeck S, Jung A, Laubender RP, Neumann J, Egg R, Goritschan C, et al. KRAS mutation status is not predictive for objective response to anti-EGFR treatment with erlotinib in patients with advanced pancreatic cancer. J Gastroenterol 2013;48:544-8.

76. Propper D, Davidenko I, Bridgewater J, Kupcinskas L, Fittipaldo A, Hillenbach C, et al. Phase II, randomized, biomarker identification trial (MARK) for erlotinib in patients with advanced pancreatic carcinoma. Ann Oncol 2014;25:1384-90.

77. Le N, Sund M, Vinci A, Pancreas Gcgo. Prognostic and predictive markers in pancreatic adenocarcinoma. Dig Liver Dis 2016;48:223-30.

78. Luchini C, Bibeau F, Ligtenberg MJL, Singh N, Nottegar A, Bosse T, et al. ESMO recommendations on microsatellite instability testing for immunotherapy in cancer, and its relationship with PD-1/PD-L1 expression and tumour mutational burden: a systematic review-based approach. Ann Oncol 2019;30:1232-43.

79. Arora S, Velichinskii R, Lesh RW, Ali U, Kubiak M, Bansal P, et al. Existing and Emerging Biomarkers for Immune Checkpoint Immunotherapy in Solid Tumors. Adv Ther 2019.

80. Le DT, Durham JN, Smith KN, Wang H, Bartlett BR, Aulakh LK, et al. Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. Science 2017;357:409-13.

81. Strimbu K, Tavel JA. What are biomarkers? Curr Opin HIV AIDS 2010;5:463.

82. Yates JR, Ruse CI, Nakorchevsky A. Proteomics by mass spectrometry: approaches, advances, and applications. Annu Rev Biomed Eng 2009;11:49-79.

83. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, et al. A pathology atlas of the human cancer transcriptome. Science 2017;357.

84. Smith EB. Carcinoma of the pancreas--early diagnostic presumptive signs. J Natl Med Assoc 1974;66:496-8.

85. Khorana AA, Fine RL. Pancreatic cancer and thromboembolic disease. Lancet Oncol 2004;5:655-63.

86. Henderson CM, Shulman NJ, MacLean B, MacCoss MJ, Hoofnagle AN. Skyline Performs as Well as Vendor Software in the Quantitative Analysis of Serum 25-Hydroxy Vitamin D and Vitamin D Binding Globulin. Clin Chem 2018;64:408-10.

87. Tabb DL. The SEQUEST family tree. J Am Soc Mass Spectrom 2015;26:1814-9.

88. Chen C, Huang H, Wu CH. Protein bioinformatics databases and resources. Protein Bioinformatics: Springer; 2017. p. 3-39.

89. Ishibashi H, Suzuki T, Suzuki S, Moriya T, Kaneko C, Takizawa T, et al. Sex steroid hormone receptors in human thymoma. J Clin Endocrinol Metab 2003;88:2309-17.

90. John T, Liu G, Tsao MS. Overview of molecular testing in non-small-cell lung cancer: mutational analysis, gene copy number, protein expression and other biomarkers of EGFR for the prediction of response to tyrosine kinase inhibitors. Oncogene 2009;28 Suppl 1:S14-23.

91. Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 2016;13:731-40.

92. Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res 2010;38:D204-D10.

93. Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, et al. STRING 8—a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res 2008;37:D412-D6.

94. Xie C, Mao X, Huang J, Ding Y, Wu J, Dong S, et al. KOBAS 2.0: a web server for annotation and identification of enriched pathways and diseases. Nucleic Acids Res 2011;39:W316-22.

95. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 2014;30:523-30.

96. Chen C, Huang H, Wu CH. Protein Bioinformatics Databases and Resources. Methods Mol Biol 2017;1558:3-39.

97. Colaert N, Helsens K, Martens L, Vandekerckhove J, Gevaert K. Improved visualization of protein consensus sequences by iceLogo. Nature methods 2009;6:786-7.

98. Harrell Jr, FE (2018). rms: Regression Modeling Strategies. R package version 5.1-2. https://CRAN.R-project.org/package=rms

99. Robin X, Turck N, Hainard A, Tiberti N, Lisacek F, Sanchez JC, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinformatics 2011;12:77.

100. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32-5.

101. Team RC. R: A language and environment for statistical computing. 2013.

102. Herbert B. Advances in protein solubilisation for two-dimensional electrophoresis. Electrophoresis 1999;20:660-3.

103. McCarthy J, Hopwood F, Oxley D, Laver M, Castagna A, Righetti PG, et al. Carbamylation of proteins in 2-D electrophoresis--myth or reality? J Proteome Res 2003;2:239-42.

104. Sun S, Zhou JY, Yang W, Zhang H. Inhibition of protein carbamylation in urea solution using ammonium-containing buffers. Anal Biochem 2014;446:76-81.

105. Tenga MJ, Lazar IM. Impact of peptide modifications on the isobaric tags for relative and absolute quantitation method accuracy. Anal Chem 2011;83:701-7.

106. Pathan M, Keerthikumar S, Ang CS, Gangoda L, Quek CY, Williamson NA, et al. FunRich: An open access standalone functional enrichment and interaction network analysis tool. Proteomics 2015;15:2597-601.

107. Li C, Shen Z, Zhou Y, Yu W. Independent prognostic genes and mechanism investigation for colon cancer. Biol Res 2018;51:10.

108. Aran D, Camarda R, Odegaard J, Paik H, Oskotsky B, Krings G, et al. Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat Commun 2017;8:1077.

109. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer 1953;6:963-8.

Paper I

Journal of proteome research
Cite This: J. Proteome Res. 2018, 17, 2556−2561
Technical Note
pubs.acs.org/jpr

Quantitative Assessment of Urea In-Solution Lys-C/Trypsin Digestions Reveals Superior Performance at Room Temperature over Traditional Proteolysis at 37 °C

Lazaro Hiram Betancourt,*[,†,∥] Aniel Sanchez,[†,∥] Indira Pla,[†] Magdalena Kuras,[†] Qimin Zhou,[‡] Roland Andersson,[‡] and Gyorgy Marko-Varga[†,§]

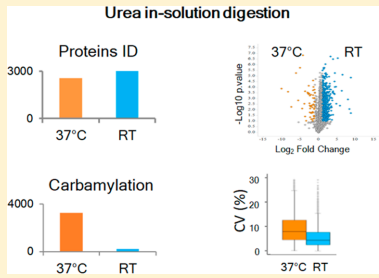
[†]Division of Clinical Protein Science and Imaging, Department of Clinical Sciences (Lund) and Department of Biomedical Engineering, Lund University, SE-221 84 Lund, Sweden

[‡]Department of Clinical Sciences Lund (Surgery), Lund University, and Skåne University Hospital, SE-221 84 Lund, Sweden

[§]First Department of Surgery, Tokyo Medical University, Shinjuku-ku, Tokyo 160-8402, Japan

Supporting Information

ABSTRACT: Urea-containing buffer solutions are generally used in proteomic studies to aid protein denaturation and solubilization during cell and tissue lysis. It is well-known, however, that urea can lead to carbamylation of peptides and proteins and, subsequently, incomplete digestion of proteins. By the use of cells and tissues that had been lysed with urea, different solution digestion strategies were quantitatively assessed. In comparison with traditional proteolysis at 37 °C, urea in-solution digestion performed at room temperature improved peptide and protein identification and quantitation and had a minimum impact on miscleavage rates. Furthermore, the signal intensities and the number of carbamylated and pyroglutamic acid-modified peptides decreased. Overall, this led to a reduction in the negative effects often observed for such modifications. Data are available via ProteomeXchange with identifier PXD009426.



KEYWORDS: *urea, in-solution digestion, carbamylation, temperature, mass spectrometry, proteomics, label-free quantification*

INTRODUCTION

Urea-containing buffer solutions are generally used in proteomic studies during the lysis of cells and tissues, thereby contributing to protein denaturation and solubilization.[1] It is well-known, however, that such buffers can lead to carbamylation of peptides and proteins via the reaction with isocyanic acid, a degradation product of urea.[2−5] This reaction is exacerbated at higher temperatures and long incubation periods.[2,6]

Carbamylation has several detrimental effects on the analysis of peptides and proteins, which have been thoroughly described in the literature.[5,7] Modification of both the α- and ε-amino groups of lysine residues renders peptides and proteins unsuitable for N-terminal sequencing. In addition, the blocked groups are unable to react with isobaric tag reagents such as iTRAQ and TMT that are commonly used in quantitative proteomics. Enzymatic digestion by trypsin and endopeptidase Lys-C are also inhibited, as carbamylation of the side chains of lysine residues prevents hydrolysis of the peptide bond by the enzymes. Carbamylation creates a subpopulation of peptides with different masses and retention times that contributes to the complexity of the proteomic sample.[7] There is also a negative impact on peptide and protein identification and quantitation, as the ionization efficiency of carbamylated peptides is reduced.[8] Carbamylated products that are artificially introduced during sample preparation also affect studies investigating the in vivo occurrence of this modification. It has been reported that immunopurification of lysine-acetylated peptides is hampered by the copurification of lysine-carbamylated peptides.[9]

Furthermore, the analysis of samples enzymatically digested in the presence of urea often reveals a high percentage of missed cleavage sites. This fact is attributed to the denaturation of the protease resulting in a reduction in digestion efficiency.[10]

Solutions to face these challenges have ranged from avoiding or completely removing urea from the sample preparation protocol to procedures that include urea but minimize the aforementioned disadvantages. For the latter approach, inhibition of protein carbamylation in solutions containing urea has been demonstrated by using ammonium-based buffers.[7] To reduce the level of cyanate, it has also been recommended to freshly prepare the urea solutions and further deionize prior to use. In addition, the sample should be maintained at a low temperature to reduce the urea decomposition rate.[11−13]

Received: April 6, 2018
Published: May 29, 2018

DOI: 10.1021/acs.jproteome.8b00228
J. Proteome Res. 2018, 17, 2556−2561

A recent study provided some directives concerning the occurrence of carbamylation in current proteomic literature.[9] The authors found that the highest levels of carbamylation in published data sets occurred when protein reduction was performed at 56 °C. Consequently, these researchers reduced protein disulfide bonds at room temperature. A noticeable decrease in the levels of peptide carbamylation from 5.8% to 1.1% was observed. This confirmed that temperature is a key factor in maintaining a low level of the modification.

A significant amount of effort has also been devoted to increasing the efficiency of enzymatic digestion of proteomic samples by reducing the level of missed cleavage peptides and thus improving the quantitation accuracy, sensitivity, and reproducibility of mass spectrometric measurements.[14,15] For most proteomic studies, trypsin is still the main protease of choice, although the combination of endopeptidase Lys-C and trypsin has steadily gained widespread acceptance.[16,17] Endopeptidase Lys-C can cleave peptide bonds at higher concentrations of urea. Consequently, partial digestion of proteins by Lys-C aids the action of trypsin following dilution of the denaturant.

In this study, total lysates from cell culture and tissues were used to quantitatively evaluate urea-based digestion methods on a proteome-wide scale. Our key findings were that incubation of the enzymatic digest at room temperature reduced the level of undesirable amino acid modifications, such as carbamylation and pyroglutamic acid formation. In addition, the overall number of identified peptides and proteins increased, as did the relative abundance of a significant number of proteins. Other factors such as diluting the urea concentration positively contributed to the overall performance of the LC−MS/MS analyses.

MATERIALS AND METHODS

Urea In-Solution Digestion

A lysate of SK-MEL cells containing 125 μg of protein in 50 mM Ambic buffer and 6 M urea was reduced with 10 mM DTT for 1 h at 37 °C and alkylated using 20 mM iodoacetamide for 30 min in the dark at RT. The sample was diluted 2-fold with 50 mM Ambic buffer and divided into five aliquots of 25 μg each for experimental replicates of the protein digests. Three aliquots were digested with Lys-C at an enzyme:protein ratio of 1:50 w/w for 6 h at 37 °C. The samples were then diluted 3-fold with 50 mM Ambic buffer and further digested at 37 °C overnight with trypsin at a trypsin:protein ratio of 1:50 w/w. The remaining two aliquots of the reduced and alkylated cell lysate were digested as above but incubated at RT. Standard room temperature in the laboratory was 24 ± 2 °C.

Tissue lysates in 50 mM Ambic buffer and 4 M urea were reduced with 10 mM DTT for 1 h at 37 °C and alkylated using 20 mM iodoacetamide for 30 min at RT in the dark. Each sample was diluted 4-fold with 50 mM Ambic buffer and divided into six aliquots of 25 μg each for experimental replicates of the protein digests. Three aliquots were digested with Lys-C at an enzyme:protein ratio of 1:50 w/w for 7 h at 37 °C. The samples were then diluted 2-fold with 50 mM Ambic buffer and further digested at 37 °C overnight with trypsin at a trypsin:protein ratio of 1:50 w/w. The remaining three aliquots of the tissue lysates were digested as above and incubated at RT.

Digestion was quenched by addition of formic acid to a final concentration of 1%. Peptides were desalted with ultra-microspin C18 columns according to the instructions supplied by the manufacturer. Desalted and dried peptides were resuspended in 50 μL of 0.1% formic acid, and the peptide concentration was measured using the Pierce Quantitative Colorimetric Peptide Assay (Thermo Fisher Scientific, Rockford, IL).

LC−MS/MS Analysis

LC−MS/MS was performed using a ThermoEasy nLC 1000 system (Thermo Fisher Scientific) coupled online to a Q-Exactive Plus mass spectrometer (Thermo Fisher Scientific, San José, CA). The peptides (∼1 μg) were initially loaded onto a trap column (Acclaim PepMap 100 precolumn, 75 μm i.d. × 2 cm, C18, 3 mm, 100 Å; ThermoFisher Scientific, San José, CA) and then separated on an analytical column (EASY-Spray column, 75 μm i.d. × 25 cm, PepMap RSLC C18, 2 mm, 100 Å; ThermoFisher Scientific, San José, CA) using an 80 min acetonitrile gradient in 0.1% formic acid at a flow rate of 300 nL/min and a column temperature of 35 °C. Each sample was analyzed in triplicate injections.

Data Analysis

Raw files were analyzed with MaxQuant version 1.6.0.1 using the Andromeda Search engine. The raw files were searched against the UniProtKB human database (released July 9, 2016) or UniProtKB *Rattus norvegicus* database (released Jan 18, 2017) both excluding isoforms. The default contaminant protein database and the decoy database were used, and matches from these databases were excluded. The search implemented 20 ppm and 0.02 Da precursor and fragment ion tolerances, respectively. Carbamidomethylation of cysteine residues was a fixed modification. Oxidation of methionine residues, carbamylation at protein N-termini, lysine carbamylation, pyroglutamic acid modification, and acetylation were dynamic modifications. Two missed cleavage sites were allowed. The "match-between-runs" option was enabled. Filters: high confidence at the peptide and protein level were applied (FDR 0.01). The statistical analyses were performed in Perseus version 1.5.6.0 and Graphpad Prism version 7. Data were normally distributed by $\log_2$ transformation (normalization) of protein intensities, and standardization was done by subtracting the median of the $\log_2$-transformed intensities per sample. It was assumed that most proteins do not change their abundance. IceLogos were generated using the web application.[18,19]

Further details regarding chemicals and reagents, cell and tissue lysis, protein extraction, and method of LC−MS/MS analysis are presented in the Supporting Information.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[20] partner repository with the data set identifier PXD009426.

RESULTS AND DISCUSSION

Effect of Temperature on Peptide and Protein Identification and Quantitation

The effects of the urea concentration and the temperature at which the enzymatic digestion was performed were evaluated using SK-MEL cell, pancreatic tumor xenograft, and rat spleen lysates. An overview of the experimental workflow is given in Figure 1. Protein extracts were digested in tandem with Lys-C followed by trypsin at 37 °C or RT. Depending on the sample, the final urea concentration was 0.5 or 1 M.

DOI: 10.1021/acs.jproteome.8b00228
J. Proteome Res. 2018, 17, 2556−2561

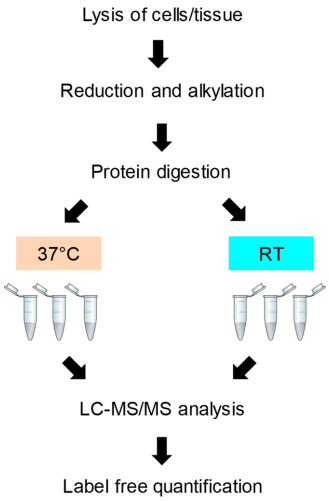


Figure 1. Schematic workflow overview. SK-MEL cells and tissue (pancreatic tumor xenograft and rat spleen) samples were lysed in 6 and 4 M urea, respectively. After reduction and alkylation, the samples were subjected to enzymatic digestion at 37 °C or RT. For the cell-derived proteins, the digestion was performed with Lys-C in 2 M urea and with trypsin in 1 M urea. Tissue-derived proteins were digested with Lys-C in 1 M urea and with trypsin in 1 M urea. Two to three technical replicates were carried out for all enzymatic digestions (see Materials and Methods). The digested peptide samples were then analyzed in triplicate injections by LC−MS/MS and subjected to label-free quantification.

Relatively low peptide miscleavage rates were observed for all of the samples and conditions. When the digestions were performed in 1 M urea, the frequency of miscleavages was 11% and 14% for the 37 °C and RT experiments, respectively. For the digests performed in 0.5 M urea, the frequency of miscleavages was <7% and <9% at 37 °C and RT, respectively.

For all of the samples, a consistent increase in the number of identified peptides and proteins was observed when the digestion was performed at RT (Figure 2). Increase rates of 5−23% and 8−39% were observed for the proteins and peptides, respectively (Table S-2 in the Supporting Information). The greatest increases were seen at a urea concentration of 0.5 M. These results were accompanied by a slight improvement in the average sequence coverage of proteins of up to 3% (Table S-2). It is noteworthy that for all of the samples, 89−95% of the proteins identified in the digests performed at 37 °C were also present in the digests performed at RT (Figure S-1). These results are similar to the overlap in proteins obtained within experimental replicates and within individual LC−MS/MS analyses of the same digestion conditions (data not shown).

Figure S-2 shows the spanned dynamic range of abundance of the peptides identified from both digestion conditions. The graph confirms the increase in the number of identified peptides when the digestions were performed at RT. More importantly, however, the peptides generated at RT had a trend toward higher abundance than those generated at 37 °C. This fact is also reflected in the higher median intensity of all quantitated peptides (Figure 3).

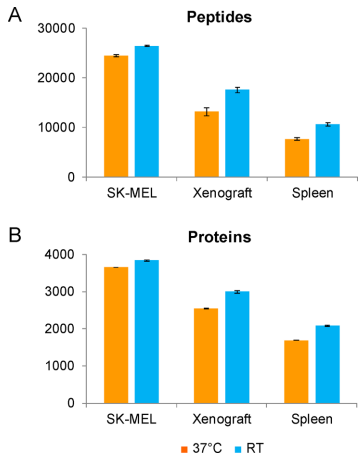


Figure 2. Global comparison of (A) peptides and (B) proteins identified in Lys-C/trypsin digestions at 37 °C and RT for SK-MEL, xenograft, and spleen samples. Data are reported as mean ± SD for all experimental replicates.

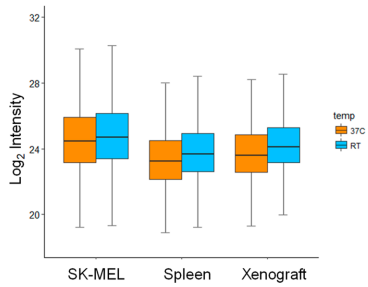


Figure 3. Box plot representation of the intensities of common peptides for the digestions at 37 °C and RT.

t tests were performed to determine significant differences in protein abundance between the two enzymatic digestion conditions. Figure 4 presents volcano plots of $-\log_{10}(p \text{ value})$ versus $\log_2$(fold change of RT/37 °C). When applying stringent filtering criteria (fold change ≥ 2, FDR < 0.05), 213, 517, and 525 proteins from the SK-MEL cells, pancreatic tumor xenograft, and spleen, respectively, were identified with significantly higher abundance when proteolysis was performed at RT. Conversely, only 85, 57, and 123 proteins with higher abundance were identified when proteolysis was performed at 37 °C.

Figure 5 shows the distribution of the coefficient of variation (CV) obtained for the quantitated proteins in all of the samples and under all of the digestion conditions. Compared with the digestions performed at 37 °C, similar or lower CV distributions were obtained at RT. For both temperatures, 86−93% of the proteins quantitated had CV values below 20% (Figure S-3). Particularly for the digestions of SK-MEL and the pancreatic tumor xenograft samples at RT, higher numbers of proteins displayed CV values below 10%. Furthermore, the correlations for all of the quantitated proteins within the

DOI: 10.1021/acs.jproteome.8b00228
J. Proteome Res. 2018, 17, 2556−2561

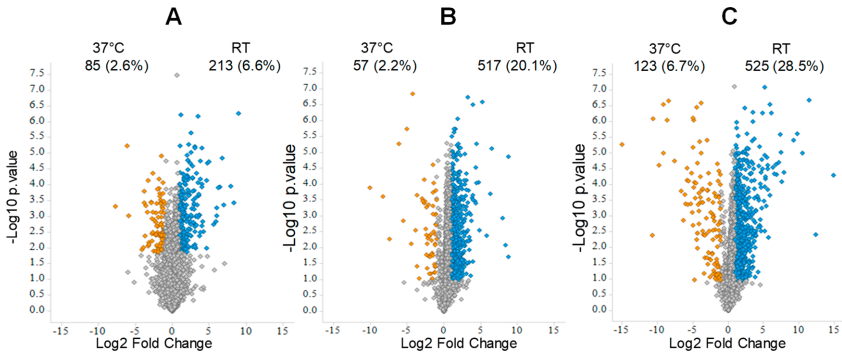


Figure 4. Volcano plot showing different protein abundances for the digestions at 37 °C and RT for (A) SK-MEL cells, (B) pancreatic tumor xenograft tissue, and (C) rat spleen tissue. Following LC−MS analysis and label-free quantification, *t*-test-based significance values ($\log_{10}(p$ value)) were plotted versus $\log_2$(intensity ratio for all proteins between different digestion conditions). The numbers and percentages of proteins significantly affected by the digestion conditions are indicated.

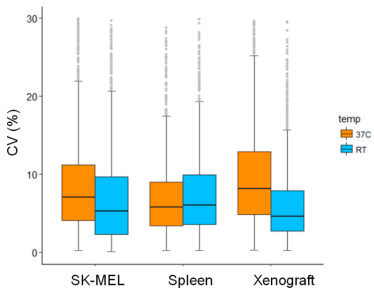


Figure 5. Box plot representation of the coefficient of variation (CV) for proteins commonly quantitated in the digestions performed at 37 °C and RT. Outliers are represented by discontinuous gray dots.

technical replicates were similar (r > 0.99) (Figure S-4). This indicated that both experiments were reproducible and that the differences in performance between the digestions at 37 °C and RT were exclusively related to the change in the digestion temperature.

Analysis of Peptide Carbamylation and Pyroglutamic Acid Formation

The incidence of post-translational modifications (PTMs) that may be affected by temperature changes, i.e., carbamylation and pyroglutamic acid formation, were investigated. Figure 6 shows the average ratio of peptide-to-spectrum matches (PSMs) of post-translationally modified peptides at 37 °C and RT. Regardless of temperature, the spleen and pancreatic tumor xenograft tissue samples had reduced numbers (<2%) of N-terminal carbamylated peptides (Figure 6A). These results suggest that dilution of the urea to 0.5 M is a key factor in drastically reducing peptide carbamylation. Nevertheless, more than a 40% decrease was observed in the PSMs assigned to N-terminal carbamylated peptides in the digestions performed at RT. The highest reduction (15-fold) corresponded to the SK-MEL lysate, i.e., the digestion carried out in 1 M urea, but for all of the samples digested at RT, carbamylation was detected in less than 0.5% of PSMs (Figure 6A).

Carbamylation occurred more frequently at α-amino groups of peptides compared with ε-amino groups of lysine residues, which was observed to a lower extent (<0.25%) for both digestion temperatures (Figure 6B). This result was consistent with previous reports.[2,5,7] Nevertheless, a 25−60% decrease in

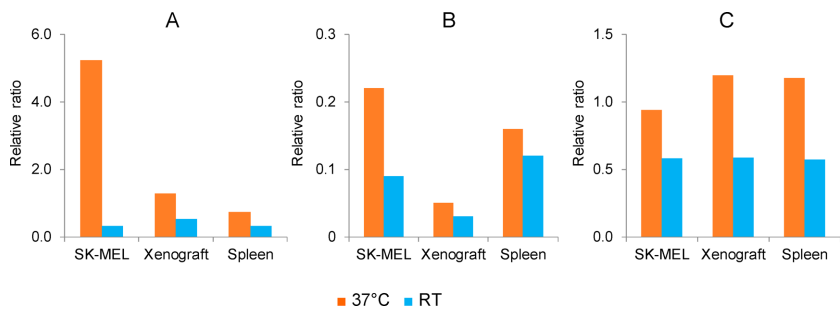


Figure 6. Ratios of PSMs assigned to (A) N-terminal carbamylation, (B) lysine carbamylation, and (C) pyroglutamic acid-containing peptides for the digestions performed at 37 °C and RT. The ratios were calculated by dividing the number of PSMs assigned to a modification by the total number of PSMs, considering triplicate injection on the LC−MS/MS.

DOI: 10.1021/acs.jproteome.8b00228
J. Proteome Res. 2018, 17, 2556−2561

the number of PSMs corresponding to this modification was detected for the digestions performed at RT. The majority of carbamylated lysines were found at an internal position of the peptide, i.e., miscleaved lysines, suggesting that the modification mainly took place during sample storage or most probably during reduction of disulfide bonds, which was performed at 37 °C. Notably, compared with the enzymatic digestion process, this stage of the sample preparation had little influence on the overall carbamylation of the final peptide set. The results also indicated that regardless of the temperature of the enzymatic digestion, carbamylation of lysine had little impact on the overall miscleavage rate, since it represented less than 6% of all PSMs assigned to miscleaved peptides.

Furthermore, less than 0.6% of PSMs were assigned to pyroglutamic acid-containing peptides in the digestions performed at RT (Figure 6C). This corresponded to a reduction of 30−50% compared with pyroglutamic acid-containing peptides detected in the digestions at 37 °C. Thus, the advantages of digesting the protein mixtures at a lower temperature were confirmed with the concomitant result of a greater reduction in these modifications.

The effect of the digestion temperature on the oxidation of methionine residues was also investigated. A slight increase of 5% in the number of PSMs at RT for this PTM was evident (data not shown). We hypothesized that the relatively higher solubility of oxygen in water at RT was responsible for such an outcome.

Figure S-5 shows plots of intensity versus retention time of identified peptides for one experimental replicate per sample condition. Comparison of the intensities of the modified peptides at the two temperatures (Figure S-5A−C) and the peptides uniquely identified in the digestion performed at RT (Figure S-5D−F) showed that these peptide groups share similar intensities and are detected across the entire gradient. Such observations point to signal competition and an interfering effect of the modified peptides in the digestion performed at 37 °C. Also, a search for the number of acquired MS/MS per segment of retention time showed no significance differences between the two digestion conditions (data not shown). At 37 °C, PTM peptides are generated at a higher rate, which reduces the number of nonmodified peptides selected for MS/MS analysis. Thus, both the number and intensity of PTM peptides prevent the detection and sequencing of additional peptides in the digestion performed at 37 °C.

As indicated earlier, peptides generated from the digests performed at RT displayed higher signal intensities after LC−MS/MS analysis. Nevertheless, carbamylated and pyroglutamic acid-containing peptides displayed lower intensities across most of the samples (excluding the spleen). This result was confirmed by analyzing the mean intensities and standard deviations of the carbamylated and pyroglutamic acid-containing peptides generated at 37 °C that were also present in the digestion performed at RT (Figure S-6).

To some extent, the overall increases in the number and abundance of peptides and proteins identified in the digests performed at RT could be explained by a signal suppression effect caused by the higher number and intensity of the PTM peptides generated at 37 °C. Factors affecting the solubility of peptides and proteins at these temperatures, e.g., molecular mass and GRAVY index were also analyzed. However, a direct association could not be established between these physicochemical properties and the proteins and peptides that differed in abundance in the digestions performed at 37 °C and RT (data not shown). Performing the enzymatic digestions at these temperatures but in the complete absence of urea and with or without the use of other denaturants such as guanidinium chloride, detergents, or organic solvents will rule out the peptide carbamylation and could aid in measuring the impact of this PTM on the digestion outcome. Furthermore, other factors that are outside the scope of this study should be taken into consideration. These include, e.g., the relationship of the enzymatic activities of Lys-C and trypsin with temperature in the presence of urea.

Finally, to obtain a global overview of frequently carbamylated N-terminal amino acid residues, iceLogo plots were generated (Figure S-7). Glycine and alanine were shown to be the most frequently carbamylated residues at the peptide N-terminus. The results suggested that under our experimental conditions that afford low peptide carbamylation rates, N-terminal amino acids with the lowest steric impediment have the greatest chance of reacting rapidly with urea degradation products.

CONCLUSIONS

Urea in-solution digestion at room temperature offers several advantages compared with proteolysis at 37 °C. The number and mass spectral signal intensity of peptides with carbamylation and pyroglutamic acid modifications are reduced, consequently leading to a reduction in the negative effects of these modifications. The number of identified peptides and proteins increased, and a significant number of proteins displayed higher abundance. The slight increase in the number of miscleaved peptides in the digestions performed at RT had no detrimental effect on the overall efficiency of the proteolytic digestion. Furthermore, similar or lower coefficients of variation were obtained for protein quantitation when the digestion was performed at room temperature.

Whether the benefits observed in this study for urea in-solution digests of proteomic samples at room temperature are applicable to other commonly used denaturants (e.g., guanidinium chloride, detergents, or organic solvents) is a matter of debate and will be the subject of future investigations.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jproteome.8b00228.

Chemicals and reagents and methods for protein extraction and LC−MS/MS analysis; number of proteins and peptides identified for lysates digested at 37 °C and RT (Table S-1); average protein sequence coverages for the digestions at 37 °C and RT (Table S-2); overlaps of identified proteins for enzymatic digestions at 37 °C and RT (Figure S-1); distribution of label-free quantified peptides (Figure S-2); frequency distribution of the CV for quantified proteins (Figure S-3); protein intensity correlations between experimental replicates (Figure S-4); plots of intensities vs retention time of PTM peptides and peptides uniquely detected in the digestion at RT (Figure S-5); mean intensities and standard deviations of PTM peptides (Figure S-6); analysis of frequently carbamylated N-terminal amino acid residues (Figure S-7) (PDF)

DOI: 10.1021/acs.jproteome.8b00228
J. Proteome Res. 2018, 17, 2556−2561

AUTHOR INFORMATION

Corresponding Author

*E-mail: lazaro.betancourt@med.lu.se. Telephone: 46-46-222 37 21.

ORCID

Lazaro Hiram Betancourt: 0000-0001-8207-7041

Author Contributions

‖L.H.B. and A.S. contributed equally to this work.

Notes

The authors declare no competing financial interest.

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE[20] partner repository with the data set identifier PXD009426.

ACKNOWLEDGMENTS

This study was supported by a joint grant from the Swedish Research Council, Vinnova, SSF, under the program Biomedical Engineering for Better Health and the Berta Kamprad Foundation. This work was also supported by grants from the National Research Foundation of Korea, funded by the Government of the Republic of Korea (MSIP; 2012M3A9D1054520, 2015K1A1A2028365, 2015M3A9C4076321), and the Brain Korea 21 Plus Project, Republic of Korea.

REFERENCES

(1) Herbert, B. Advances in Protein Solubilisation for Two-Dimensional Electrophoresis. *Electrophoresis* **1999**, *20*, 660−663.

(2) Stark, G. R. Reactions of Cyanate with Functional Groups of Proteins. III. Reactions with Amino and Carboxyl Groups. *Biochemistry* **1965**, *4* (6), 1030−1036.

(3) Marier, J. R.; Rose, D. Determination of Cyanate, and a Study of Its Accumulation in Aqueous Solutions of Urea. *Anal. Biochem.* **1964**, *7* (3), 304−314.

(4) McCarthy, J.; Hopwood, F.; Oxley, D.; Laver, M.; Castagna, A.; Righetti, P. G.; Williams, K.; Herbert, B. Carbamylation of Proteins in 2-D Electrophoresis–Myth or Reality? *J. Proteome Res.* **2003**, *2* (3), 239−242.

(5) Kollipara, L.; Zahedi, R. P. Protein Carbamylation: In Vivo Modification or in Vitro Artefact? *Proteomics* **2013**, *13* (6), 941−944.

(6) Gerding, J. J. T.; Koppers, A.; Hagel, P.; Bloemendal, H. Cyanate Formation in Solutions of Urea II. Effect of Urea on the Eye Lens Protein α-Crystallin. *Biochim. Biophys. Acta, Protein Struct.* **1971**, *243* (3), 374−379.

(7) Sun, S.; Zhou, J. Y.; Yang, W.; Zhang, H. Inhibition of Protein Carbamylation in Urea Solution Using Ammonium-Containing Buffers. *Anal. Biochem.* **2014**, *446* (1), 76−81.

(8) Tenga, M. J.; Lazar, I. M. Impact of Peptide Modifications on the Isobaric Tags for Relative and Absolute Quantitation Method Accuracy. *Anal. Chem.* **2011**, *83* (3), 701−707.

(9) Martinez-Val, A.; Garcia, F.; Ximénez-Embún, P.; Martínez Teresa-Calleja, A.; Ibarz, N.; Ruppen, I.; Munoz, J. Urea Artifacts Interfere with Immuno-Purification of Lysine Acetylation. *J. Proteome Res.* **2017**, *16* (2), 1061−1068.

(10) Chiva, C.; Ortega, M.; Sabidó, E. Influence of the Digestion Technique, Protease, and Missed Cleavage Peptides in Protein Quantitation. *J. Proteome Res.* **2014**, *13* (9), 3979−3986.

(11) Sigma. Urea solution. https://www.sigmaaldrich.com/content/dam/sigma-aldrich/docs/Sigma/Datasheet/u4883dat.pdf.

(12) Wingfield, P. T. Use of Protein Folding Reagents. *Curr. Protoc. Protein Sci.* **1995**, DOI: 10.1002/0471140864.psa03as00.

(13) Wiśniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M. Universal Sample Preparation Method for Proteome Analysis. *Nat. Methods* **2009**, *6* (5), 359−362.

(14) Glatter, T.; Ludwig, C.; Ahrné, E.; Aebersold, R.; Heck, A. J. R.; Schmidt, A. Large-Scale Quantitative Assessment of Different in-Solution Protein Digestion Protocols Reveals Superior Cleavage Efficiency of Tandem Lys-C/trypsin Proteolysis over Trypsin Digestion. *J. Proteome Res.* **2012**, *11* (11), 5145−5156.

(15) Wiśniewski, J. R.; Mann, M. Consecutive Proteolytic Digestion in an Enzyme Reactor Increases Depth of Proteomic and Phosphoproteomic Analysis. *Anal. Chem.* **2012**, *84* (6), 2631−2637.

(16) Guo, Z.; Cheng, J.; Sun, H.; Sun, W. A Qualitative and Quantitative Evaluation of the Peptide Characteristics of Microwave- and Ultrasound-Assisted Digestion in Discovery and Targeted Proteomic Analyses. *Rapid Commun. Mass Spectrom.* **2017**, *31* (16), 1353−1362.

(17) Schniers, A.; Anderssen, E.; Fenton, C. G.; Goll, R.; Pasing, Y.; Paulssen, R. H.; Florholmen, J.; Hansen, T. The Proteome of Ulcerative Colitis in Colon Biopsies from Adults - Optimized Sample Preparation and Comparison with Healthy Controls. *Proteomics: Clin. Appl.* **2017**, *11* (11−12), 1700053.

(18) Ghent University. iceLogo. https://iomics.ugent.be/icelogoserver/create.

(19) Colaert, N.; Helsens, K.; Martens, L.; Vandekerckhove, J.; Gevaert, K. Improved Visualization of Protein Consensus Sequences by iceLogo. *Nat. Methods* **2009**, *6* (11), 786−787.

(20) Vizcaíno, J. A.; Csordas, A.; Del-Toro, N.; Dianes, J. A.; Griss, J.; Lavidas, I.; Mayer, G.; Perez-Riverol, Y.; Reisinger, F.; Ternent, T.; et al. 2016 Update of the PRIDE Database and Its Related Tools. *Nucleic Acids Res.* **2016**, *44* (D1), D447−D456.

DOI: 10.1021/acs.jproteome.8b00228
J. Proteome Res. 2018, 17, 2556−2561

SUPPORTING INFORMATION

Quantitative assessment of urea in-solution Lys-C/trypsin digestions reveals superior performance at room temperature over traditional proteolysis at 37°C.

Lazaro Hiram Betancourt[1¤*], Aniel Sanchez[1¤], Indira Pla[1], Magdalena Kuras[1], Qimin Zhou[2], Roland Andersson[2] and Gyorgy Marko-Varga[1,3]

1) Div. Clinical Protein Science & Imaging, Dept. of Clinical Sciences (Lund) and Dept. of Biomedical Engineering, Lund University, Lund, Sweden

2) Department of Clinical Sciences Lund (Surgery), Lund University, Lund, Sweden

Skane University Hospital, Lund, Sweden

3) First Dept. of Surgery, Tokyo Medical University, Tokyo, Japan

¤ Contributed equally to this work

*Corresponding author

Table of contents

Materials and methods

Chemicals and reagents

Dithiothreitol (DTT), iodoacetamide, ammonium bicarbonate (Ambic), α-cyano-4-hydroxycinnamic acid (CHCA) and urea were purchased from Sigma–Aldrich (St. Louis, MO, USA). Water and organic solvents were all LC–MS grade and supplied by Merck (Darmstadt, Germany). Endoproteinase Lys-C was obtained from Wako (Osaka, Japan) sequencing-grade modified trypsin and trypsin/Lys-C mix were purchased from Promega (Madison, WI, USA). Ultra-microspin C18 columns and C18 tips were purchased from The Nest Group (Southborough, MA, USA) and ThermoFisher Scientific (Walthman, MA, USA), respectively. Proteomics experiments were performed using SK-MEL-28 cultured cells (RRID:CVCL_0526), spleen tissue from an adult male Sprague Dawley rat and human pancreatic cancer xenografts. These were generated in genetically identical NMRI-nu mice (Janvier Labs, Saint-Berthevin Cedex, France) by inoculation of human pancreatic cancer cell line Capan-1 (ATCC, Manassas, VA, USA) originating from liver metastasis derived from pancreas adenocarcinoma in the head of the human pancreas. All animal handling was performed in a dedicated room and received proper animal care in accordance with the guidelines of the Swedish Government and the Lund University (Lund, Sweden). This study was approved by the local ethical committee at Lund University.

Protein extraction using automated ultrasound technology

Protein extraction was performed on a SK-MEL-28 cell pellet and on sectioned (10 µm) frozen tissues from pancreatic xenografts and rat spleen, using the Bioruptor plus, model UCD-300 (Diagenode). The lysis buffer contained 6 M (for SK-MEL cell pellet) or 4 M Urea (for tissue samples) in 100 mM ammonium bicarbonate. Tissue samples (10 sections) were

extracted with 100 µl of lysis buffer. After brief vortexing and agitation for 5 min, the samples were installed into the tube holder in the Bioruptor. The sonication was performed as follows: 40 cycles, each consisting of 15 sec at high power (on) with 15 sec rest (off) in between, at 4°C. Following centrifugation at 10,000 g for 10 minutes at 4°C, the solubilized protein content of each sample was determined with a colorimetric micro BCA Protein Assay Kit according to the manufactures' instructions (Thermo Fisher Scientific, Rockford, IL).

Cell and tissue lysis and protein extraction

Protein extraction was performed on SK-MEL cells and frozen, sectioned tissues (pancreatic tumor xenograft and rat spleen) using the Bioruptor plus (model UCD-300, Diagenode). Cells were lysed with 200 µL 6 M urea and 50 mM ammonium bicarbonate (Ambic) while tissue sections were lysed with 100 µL 4 M urea and 100 mM ammonium bicarbonate (Ambic). After briefly vortexing and agitating for 5 min, the samples were placed in the tube holder of the Bioruptor. Sonication was performed as follows: 40 cycles at 4°C. Each cycle consisted of 15 s pulse at high power (on) with 15 s rest (off). Following centrifugation at 10,000×g for 10 min at 4°C, the solubilised protein content of each sample was determined with a colorimetric Micro BCA protein assay kit according to the instructions supplied by the manufacturer (ThermoFisher Scientific, Rockford, IL).

LC-MS/MS analysis

The mass spectrometer was operated using a data-dependent top-15 method. Full MS scans were acquired from m/z 400–1600 at a resolution of 70,000 (at m/z 200), a target AGC value of 3×106 and a maximum injection time of 60 ms. Selected ions were fragmented in the HCD collision cell with a normalised collision energy of 30%. Tandem mass spectra were acquired in the Orbitrap mass analyser at a resolution of 17,500 (at m/z 200), a target AGC

value of 5×104 and maximum injection time of 50 ms. The ion selection threshold was 5×104 counts and dynamic exclusion was 20 s.

Table S-1. Summary table for SK-MEL, pancreatic tumor xenograft and rat spleen lysates digested in tandem with Lys-C and Trypsin at 37°C or at room temperature (RT). Shown are the number of proteins and peptides identified in triplicate experiments, each measured three times by LC-MS/MS. na: not analyzed

	SK-MEL				**Xenograf**				**Spleen**			
	37°C		**RT**		**37°C**		**RT**		**37°C**		**RT**	
Replicate	peptides	proteins	peptides	proteins	peptides	proteins	peptides	proteins	peptides	proteins	peptides	proteins
1	24326	3651	26478	3848	13599	2542	17979	3032	7943	1818	10234	2058
2	24353	3647	26276	3826	13651	2547	16966	2973	7416	1818	10657	2083
3	24721	3660	na	na	12197	2553	17702	2985	7492	1812	10931	2101
Average	24467	3653	26377	3837	13149	2547	17549	2996	7617	1694	10607	2080
STD	221	7	143	16	825	6	524	31	285	3	351	22

Table S-2. Average protein sequence coverages obtained for SK-MEL, pancreatic tumor xenograft and rat spleen lysates digested in tandem with Lys-C and Trypsin at 37°C or at room temperature (RT). These values considered the average of the sequence coverages obtained for all the proteins identified in the three experimetal replicates.

Temperature	SK-MEL	Xenograft	Spleen
37°C	23.4	20.6	19.9
RT	23.6	22.8	22.5

Figure S-1. Overlaps of all identified proteins in triplicates experiments of the enzymatic digestions at 37°C or at room temperature (RT).

A) SK-MEL, B) pancreatic tumor xenograft and C) rat spleen

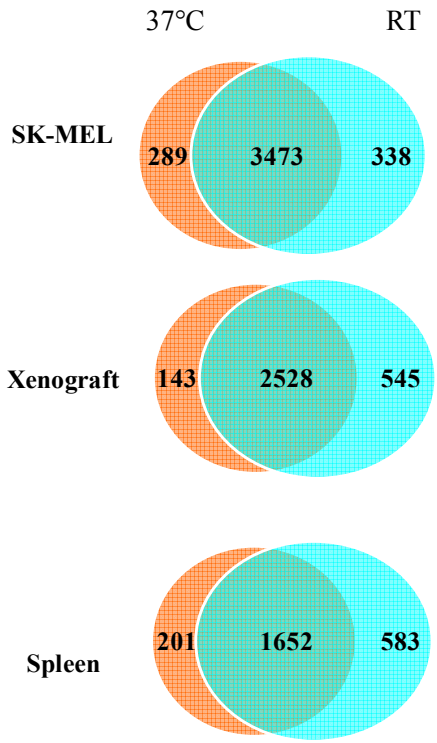


Figure S-2. Distribution of label free peptides quantified in the digestions at 37°C and at RT.

A) SK-MEL, B) pancreatic tumor xenograft and C) rat spleen tissue.

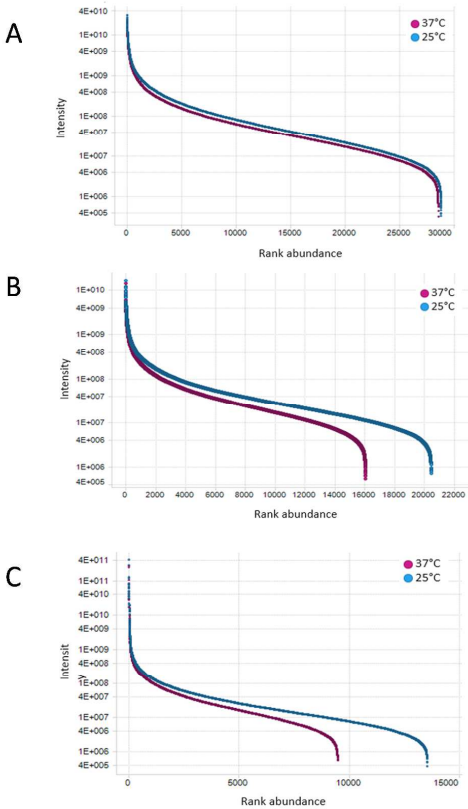


Figure S-3. Frequency distribution in percentage of the coefficient of variation (CV), for the proteins commonly quantified in the digestions at 37°C and at RT. The percentage of quantified proteins with CV below 10% and 20% are shown.

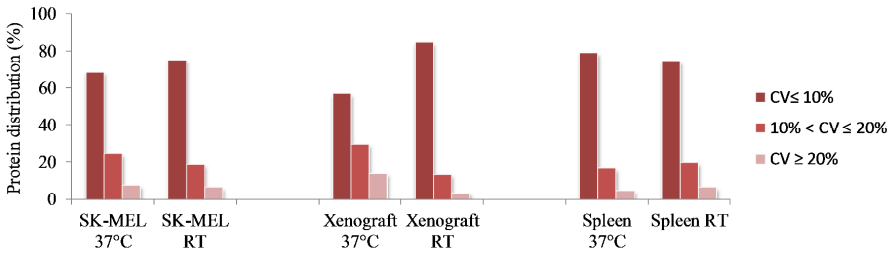


Figure S-4. Pairwise scatter plots showing (log2 transformed) protein intensity correlations between the experimental replicates (Pearson correlation, r >0.98) A) SK-MEL , B) pancreatic tumor xenograft and C) spleen.

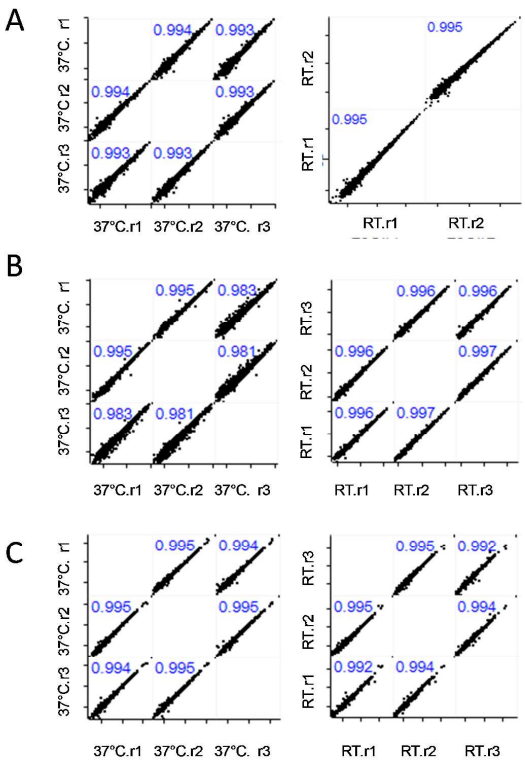


Figure S-5. Plots of intensities vs. retention time of PTM peptides and peptides uniquely detected in the digestion at room temperature (RT) for one experimental replicate. A and D) SK-MEL-28, B and E) Xenograft and C and F) spleen. ● PTMs at 37°C , ● PTMs at RT and ● Peptide uniquely detected at RT

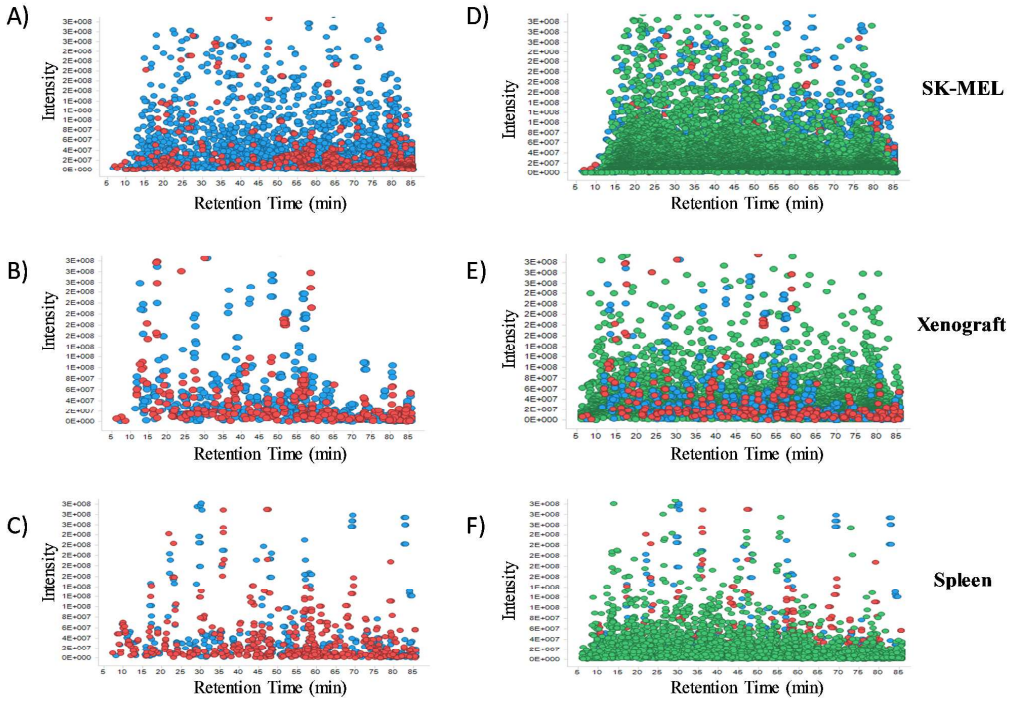


Figure S-6. Mean intensities and standard deviations of carbamylated peptides and pyroglutamic acid containing peptides commonly detected in the digestion generated at 37°Ca nd at RT

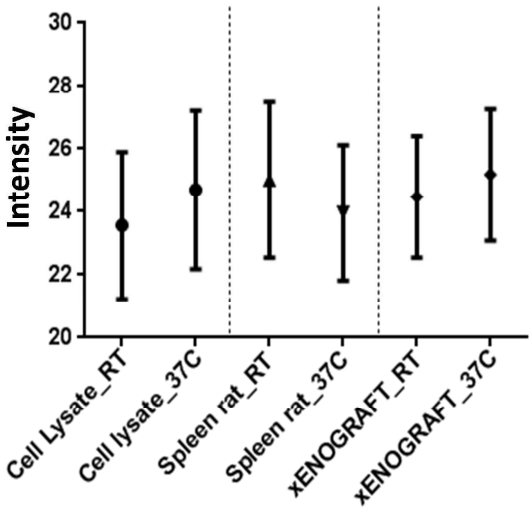
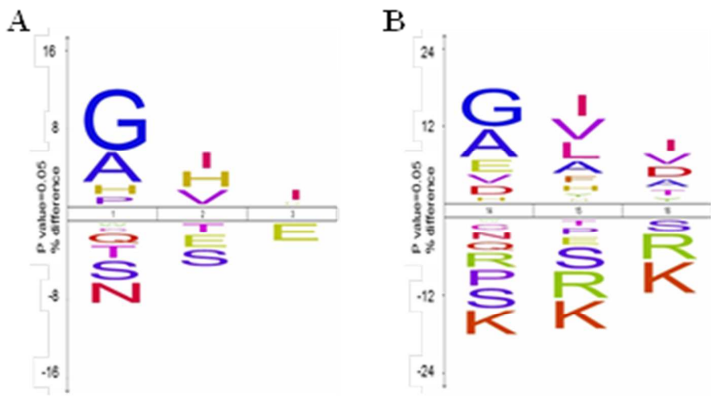


Figure S-7. Analysis of frequently carbamylated N-terminal amino acid residues under the experimental conditions of this study. IceLogo analysis were performed using all carbamylated peptides irrespective of the sample and digest condition. (A) Unmodified peptides from all experiments were used as a reference sequence set and (B) tryptic peptides from the human and rat SwissProt databases were used as reference sequence.



Paper II

EBioMedicine 43 (2019) 282–294

Contents lists available at ScienceDirect

EBioMedicine

journal homepage: www.ebiomedicine.com

EBioMedicine
Published by THE LANCET

Quantitative proteomics identifies brain acid soluble protein 1 (BASP1) as a prognostic biomarker candidate in pancreatic cancer tissue

Qimin Zhou [a,b,c], Roland Andersson [a], Dingyuan Hu [a], Monika Bauden [a], Theresa Kristl [a], Agata Sasor [d], Krzysztof Pawłowski [e,f], Indira Pla [g], Katarzyna Said Hilmersson [a], Mengtao Zhou [c], Fan Lu [b], György Marko-Varga [g], Daniel Ansari [a,*]

[a] Department of Surgery, Clinical Sciences Lund, Lund University and Skåne University Hospital, Lund, Sweden
[b] The Eye Hospital, School of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou, Zhejiang, China
[c] Key Laboratory of Diagnosis and Treatment of Severe Hepato-Pancreatic Diseases of Zhejiang Province, The First Affiliated Hospital of Wenzhou Medical University, Wenzhou, China
[d] Department of Pathology, Skåne University Hospital, Lund, Sweden
[e] Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences, Warsaw, Poland
[f] Department of Translational Medicine, Lund University, Malmö, Sweden
[g] Department of Biomedical Engineering, Clinical Protein Science and Imaging, Lund University, Lund, Sweden

ARTICLE INFO

Article history:
Received 5 March 2019
Received in revised form 27 March 2019
Accepted 3 April 2019
Available online 11 April 2019

Keywords:
Pancreatic cancer
Mass spectrometry
Biomarkers
BASP1
WT1
Prognosis
Chemotherapy response

ABSTRACT

Background: Pancreatic cancer is a heterogenous disease with a poor prognosis. This study aimed to discover and validate prognostic tissue biomarkers in pancreatic cancer using a mass spectrometry (MS) based proteomics approach.

Methods: Global protein sequencing of fresh frozen pancreatic cancer and healthy pancreas tissue samples was conducted by MS to discover potential protein biomarkers. Selected candidate proteins were further verified by targeted proteomics using parallel reaction monitoring (PRM). The expression of biomarker candidates was validated by immunohistochemistry in a large tissue microarray (TMA) cohort of 141 patients with resectable pancreatic cancer. Kaplan-Meier and Cox proportional hazard modelling was used to investigate the prognostic utility of candidate protein markers.

Findings: In the initial MS-discovery phase, 165 proteins were identified as potential biomarkers. In the subsequent MS-verification phase, a panel of 45 candidate proteins was verified by the development of a PRM assay. Brain acid soluble protein 1 (BASP1) was identified as a new biomarker candidate for pancreatic cancer possessing largely unknown biological and clinical functions and was selected for further analysis. Importantly, bioinformatic analysis indicated that BASP1 interacts with Wilms tumour protein (WT1) in pancreatic cancer. TMA-based immunohistochemistry analysis showed that BASP1 was an independent predictor of prolonged survival (HR 0.468, 95% CI 0.257–0.852, $p = .013$) and predicted favourable response to adjuvant chemotherapy, whereas WT1 indicated a worsened survival (HR 1.636, 95% CI 1.083–2.473, $p = .019$) and resistance to chemotherapy. Interaction analysis showed that patients with negative BASP1 and high WT1 expression had the poorest outcome (HR 3.536, 95% CI 1.336–9.362, $p = .011$).

Interpretation: We here describe an MS-based proteomics platform for developing biomarkers for pancreatic cancer. Bioinformatic analysis and clinical data from our study suggest that BASP1 and its putative interaction partner WT1 can be used as biomarkers for predicting outcomes in pancreatic cancer patients.

Abbreviations: ACN, acetonitrile; AGC, automatic gain control; AJCC, American joint committee on cancer; AMBIC, ammonium bicarbonate; BASP1, brain acid soluble protein 1; CI, confidence interval; DAB, diaminobenzidine; DDA, data-dependent acquisition; DTT, dithiothreitol; FDR, false discovery rate; FFPE, formalin-fixed paraffin-embedded; GO, gene ontology; HC, healthy controls; HPLC, high-performance nanoflow liquid chromatography; HR, hazard ratio; IAA, iodoacetamide; IHC, immunohistochemistry; IPA, ingenuity pathway analysis software; LC-MS/MS, chromatography-tandem mass spectrometry; MS, mass spectrometry; OS, overall survival; p.p.m, parts per million; PC, pancreatic cancer; PCA, principal component analysis; PM, plasma membrane; PRM, parallel reaction monitoring; PRTC, peptide retention time mixture; s/n, signal-to-noise; TMA, tissue microarray; WT1, Wilms tumour protein..

* Corresponding author at: Department of Surgery, Clinical Sciences Lund, Lund University and Skåne University Hospital, SE-221 85 Lund, Sweden
E-mail address: daniel.ansari@med.lu.se (D. Ansari).

https://doi.org/10.1016/j.ebiom.2019.04.008

Research in context

Evidence before this study

Pancreatic cancer is a heterogenous disease. There is a lack of molecular markers that can accurately predict the course of the disease and response to therapy. New prognostic and predictive biomarkers are urgently needed in order to characterise individual tumour biology and select optimal treatment.

Added value of this study

We conducted global and targeted mass spectrometry (MS)-based protein profiling of fresh frozen pancreatic cancer tissue specimens and healthy pancreas. Brain acid soluble protein 1 (BASP1) was found to be significantly upregulated in pancreatic cancer. External validation by tissue microarray (TMA) and immunohistochemistry in a large cohort showed that BASP1 overexpression significantly correlated to survival and response to chemotherapy in patients with pancreatic cancer. Pathway analysis linked to clinical data suggested that BASP1 interacts with Wilms tumour protein (WT1) in pancreatic cancer.

Implications of all the available evidence

Our study depicts how an MS-based proteomics platform can aid in biomarker development for pancreatic cancer. The results indicate that BASP1 and its putative interaction partner WT1 are useful biomarkers for predicting the outcomes of pancreatic cancer patients, although further validation in prospective clinical cohorts are necessary.

1. Background

Pancreatic cancer is an almost uniformly fatal disease. Tremendous efforts have been made to elucidate the mechanisms underlying pancreatic cancer in order to develop effective treatments. Although there have been significant scientific advancements, pancreatic cancer survival rates remain stagnant with a 5-year survival rate of 9%. In the United States, 56,770 patients are predicted to be diagnosed with pancreatic cancer and 45,750 individuals will die from the disease in 2019 [1]. Despite the continuous overall decline in the death rates from most cancer forms, both incidence and mortality rates for pancreatic cancer have increased during the past decade [2]. It is projected that pancreatic cancer will become the second leading cause of cancer-related death by the year 2030 [3].

Surgical resection is the only curative treatment option, yet only about 15–20% of patients are eligible for up-front radical surgery [4]. Furthermore, despite complete surgical resection and adjuvant chemotherapy, >60% of patients develop recurrences within 2 years post-operatively [5]. No molecular marker has yet been able to accurately predict the course of the disease or response to therapy [6]. Therefore, molecular markers are not used in routine clinical management of pancreatic cancer. To improve patient outcomes, novel prognostic and predictive biomarkers are needed in order to characterise individual tumour biology and select optimal treatment.

Proteomic profiling of biological samples has been shown to be a valuable approach for biomarker discovery in many cancers [7–10]. Analyses of patient serum samples and formalin-fixed paraffin-embedded (FFPE) tissue specimens using proteomic-based technologies have greatly increased the pool of potential biomarkers for pancreatic cancer detection and monitoring [11,12]. However, high abundance proteins in serum samples and chemical modifications acquired during the sample preparation of FFPE specimens hinder the accurate detection of low abundant and disease-specific proteins [13,14]. When examining disease-specific molecular information, including altered protein expression and post-translational modifications, fresh frozen tissues are considered superior for MS-based proteomics analysis [15].

In the present study, we utilised a quantitative proteomics approach using fresh frozen pancreatic cancer tissue specimens and healthy pancreas. Brain acid soluble protein 1 (BASP1) was found to be significantly upregulated in pancreatic cancer. Overexpression of BASP1 was closely correlated to survival and response to chemotherapy when examined in a large cohort by clinicopathological analysis. Based on further bioinformatic data mining coupled with clinical data analysis, we suggest that BASP1 interacts with Wilms tumour protein (WT1) in pancreatic cancer.

2. Materials and methods

2.1. Study design

The methodological workflow of the present study is illustrated in Additional file 1: Fig. S1. A Nano-liquid chromatography-tandem mass spectrometry (LC-MS/MS) platform was used for identification of candidate protein biomarkers for pancreatic cancer. Parallel Reaction Monitoring (PRM) was used for verification of protein biomarker candidates. Comprehensive bioinformatics analyses of candidate proteins and biological interaction partners were conducted to characterise functional relevance. Antibody-based validation was performed in a pancreatic cancer cell line and resected pancreatic cancer tissues from a larger cohort (Table 3). Protein expression levels were then integrated with clinicopathological information for survival analyses.

2.2. Patients and tissue samples

For MS analysis, fresh frozen pancreatic cancer tissue samples ($n = 10$ for MS discovery, $n = 8$ for targeted MS) were prospectively collected from patients undergoing pancreaticoduodenectomy due to tumours located in the head of the pancreas between July 2013 and April 2015 at the Department of Surgery, Skåne University Hospital, Lund, Sweden. Age and gender-matched fresh frozen normal pancreas ($n = 10$) from organ donors free of any pancreatic disease were obtained from Lund University Diabetes Center and used as healthy controls (HC). Written informed consent was obtained from participating patients. For tissue microarray (TMA) and immunohistochemistry (IHC) analysis, FFPE tissue samples ($n = 143$) were included from a retrospective cohort of pancreatic cancer patients who underwent surgery with curative intent from 1995 to 2017 at Skåne University Hospital in Lund and Malmö, Sweden. Following antibody optimisation and staining, biomarker expression could be evaluated in 141 of the 143 (98.6%) of tumour samples included in the TMA. All samples were re-evaluated by a pancreatic pathologist to confirm the diagnosis and uniformity of staging. The REMARK guidelines were followed where applicable [16].

2.3. MS studies

2.3.1. Tissue sample preparation

Individual fresh frozen tissue samples were pulverised in liquid N_2 and thoroughly homogenised in an extraction buffer consisting of 500 mM Tris-Cl [pH 8], 6 M guanidine-HCl in 50 mM ammonium bicarbonate (AMBIC) along with protease and phosphatase inhibitor cocktail. The obtained extracts were then subjected to 4 freeze-thaw cycles, followed by ultrasonic bath for 20 min at 0 °C. The soluble proteins were then reduced with 15 mM dithiothreitol (DTT) for 60 min at 60 °C, alkylated using 50 mM iodoacetamide (IAA) for 30 min at room temperature in the dark, precipitated with a sample to ethanol

(99.5%) ratio of 1:9 at −20 °C. The protein precipitates were dissolved in 50 mM AMBIC and digested at 37 °C overnight using Mass Spec Grade Trypsin/Lys-C Mix (Promega, Madison, WI, USA), with an enzyme to protein ratio of 1:100. The digested samples were dried and dissolved in 50 μl 0.1% Formic Acid (mobile phase A), and the concentration was specified using Pierce quantitative colorimetric peptide assay from Thermo Scientific (Rockford, IL, USA). Finally, to enable normalisation and as a control of the chromatographic performance, 25 fmol peptide retention time mixture (PRTC) (Thermo Fisher) consisting of 15 peptides was added to each sample.

2.3.2. LC-MS/MS analysis

The analytical platform, including a high-performance nanoflow liquid chromatography (HPLC) system (EASY-nLCTM™ 1000) and a Plus Hybrid Quadrupole-Orbitrap mass spectrometer (Q Exactive™) equipped with a nanospray ion source (EASY-Spray™), were manufactured by Thermo Fisher Scientific (Bremen, Germany). Individual samples containing 1 μg of peptide mixture in mobile phase A were injected at a flow rate of 300 nl min^{-1}, separated by a 132 min gradient of 5–22% acetonitrile (ACN) in mobile phase A, followed by an 18 min gradient of 22–38% ACN in mobile phase A. Subsequent separation was conducted by a two-column system including the EASY-Spray analytical column (25 cm × 75 μm ID, particle size 2 μm, pore size 100 Å, PepMap C18) tandem with the Acclaim pre-column (2 cm × 75 μm ID, particle size 3 μm, pore size 100 Å, PepMap C18). The Orbitrap system was operated in the positive data-dependent acquisition (DDA) mode with an automatic switch between the full scan MS and MS/MS acquisition. On the precursors with the highest intensity, 15 data-dependent higher energy collision dissociation MS/MS scans were implemented. For the peptide detection, a full MS survey scan was performed in the Orbitrap detector. The MS scans with a resolution of 70,000 at 200 *m/z*, recording window between 400.0 and 1600.0 *m/z*, and automatic gain control (AGC) target value of 1 × 10^6 with a maximum injection time of 100 ms. The resolution of the data dependent MS/MS scans was fixed of 17,500 at 200 *m/z*, values for the AGC target of 5 × 10^5 and maximum injection time was 80 ms. The normalised collision energy was set on 27.0% for all scans.

2.3.3. Targeted proteomics analysis

PRM analysis was performed to verify differentially expressed proteins. One or 2 unique peptides of each targeted protein were selected from the discovery measurements, depending on detection frequencies >50%, missed cleavage = 0 and *p*-value <.05, along with peptide intensities and ranking of peptide spectrum matches. Finally, a spectral library of 81 selected proteins (from the 165 differentially expressed proteins as well as the proteins only detectable in one condition) including 150 peptides was created. Owing to inadequate tissue sample volume, we had to exclude 2 pancreatic cancer subjects from the PRM phase. The proteins extracted from 18 fresh frozen samples (8 pancreatic cancer samples vs. 10 healthy controls) were reduced, alkylated, and digested as described previously in sample preparation. One microgram of the sample was injected into the LC-MS/MS system, and the PRM assay was set in a time-scheduled acquisition mode with a retention time +/− 5 min and resolution at 35000 (AGC target to 5 × 10^5, maximum injection time of 50 ms). The chromatographic peak width was 30s, normalised collision energy on 26.0%, and the isolation window of 2 *m/z*. Skyline software was used for relative quantification in the PRM study [17].

2.3.4. MS data analysis

Each sample was measured in duplicate by LC-MS/MS in a randomised order. The raw files generated from the duplicates were combined and evaluated using Proteome Discoverer software (Thermo Fisher) Version 1.4 focusing on high confidence peptides only. The spectra selection settings: minimum and maximum precursor mass at 350 Da and 5000 Da, respectively; signal-to-noise (*s/n*) threshold 1.5. Parameters for SEQUEST HT [18] were set as follows: precursor mass tolerance of 10 ppm (p.p.m); fragment mass tolerance of 0.02 Da; trypsin as the enzyme; one missed cleavage site was accepted. Based on the UniProtKB human database [19], dynamic modifications were included, such as: methyl (+14.016 Da; K, R), dimethyl (+28.031 Da; K, R), acetyl (+42.011 Da; K), trimethyl (+42.047 Da; K, R), glygly (+114.043 Da; K), oxidation (+15.995 Da; M), and the fixed modification carbamidomethyl (+57.021 Da; C). The percolator was applied for the processing node, and the false discovery rate (FDR) value was set to 0.01. To quantify the peptides, the precursor ions area detector was used in the search engine (Proteome discoverer; Thermo Scientific), protein groups identified ≥2 peptides from all samples were considered for further analysis and only unique peptides were used for protein quantification.

2.4. Tissue microarray construction and immunohistochemistry

Archival FFPE pancreatic cancer specimens from the larger validation cohort were subjected to TMA. Employing an automated tissue array instrument (Minicore® 3, Alphelys, Plaisir, France), 4 cores of cancer tissue from each specimen (diameter at 2 mm, selected by a pathologist) were extracted and fixed into paraffin blocks. After quality control, the TMA blocks were sectioned into 3 μm thick slides for IHC analysis.

IHC was performed as described previously [20]. Briefly, after deparaffinisation, rehydration and antigen-retrieval, TMA-slides were incubated with primary antibodies (rabbit anti-human BASP1 (dilution 1: 100; Cat No. HPA045218, Atlas Antibodies); mouse anti-human WT1 (clone 6F-H2, Ready-to-Use, Cat No. IS05530-2, DAKO)) overnight at 4 °C. Next, slides were incubated with second antibody (for BASP1, biotinylated goat anti-rabbit (dilution 1:200; Cat No. BA-1000, Vector Laboratories, Burlingame, CA); for WT1, biotinylated horse anti-mouse (dilution 1:200, Vector Laboratories, Cat No. BA-2000)) followed by staining with avidin-biotin-peroxidase complex (Vectastain Elite ABC-HRP Kit, Cat No. PK-6100, Vector Laboratories, Burlingame, CA). The sections were then incubated with chromogen diaminobenzidine (DAB) (Cat No. SK-4100, Vector Laboratories, Burlingame, CA) and counter stained with haematoxylin and mounted with xylene based medium. The IHC scoring was performed by an experienced pancreas pathologist (A.S.) who was blinded to the clinical information. Scoring was based on the percentage of positive tumour cells and the staining intensity. IHC results were scored as follows: 0 = negative; 1 = weak; 2 = moderate; and 3 = strong. For tumours that showed heterogeneous staining, the predominant pattern was taken into account for scoring.

2.5. Cell culture and immunofluorescence

The human pancreatic cancer cell line, PANC-1, was purchased from ATCC-LGC Standards (Manassas, VA, USA). The cells were maintained in Dulbecco's modified Eagle's medium (DMEM; Life Technologies, CA, USA) supplemented with 10% fetal bovine serum and antibiotics (100 U/ml penicillin and 100 μg/ml streptomycin) in a humified 5% CO_2 atmosphere at 37 °C.

For investigating intracellular localisation, PANC-1 cells were cultured (8 × 10^3 cells/well) in eight-well chamber slides (Lab-Tek II Chamber Slide System, Nunc). After 48 h, the cells were fixed with 4% formaldehyde, then permeabilised with 1% Triton X-100, blocked with 5% goat serum and incubated with mouse anti-human WT1 (clone 6F-H2, Ready-to-Use; Cat No. IS05530–2, DAKO) at room temperature for 2 h. After washing, cells were moved into dark environment, Goat-anti-Mouse Alexa Fluor 594 (dilution 1: 500; Cat No. A11032, Invitrogen) was added at room temperature for 1 h. Subsequently, the cells were blocked with 5% donkey serum and incubated with rabbit anti-human BASP1 (dilution 1: 50; Cat No. HPA045218, Atlas Antibodies) at room temperature for 2 h. Following washing, Donkey-anti-Rabbit Alexa Fluor 488 (dilution 1: 500; Cat No. A21206, Invitrogen)

was added at room temperature for 1 h. Finally, the cells were incubated with DAPI to stain the nuclei. Positive staining was visualised using a Nikon Eclipse 80i microscope with a Nikon DS-Qi1 camera and analysed using NIS-Elements software (Nikon Instruments Inc.; Melville, NY, USA).

2.6. Statistics and bioinformatics

Perseus software [21] version 1.6.0.7 was used for the statistical analysis of the MS results. The protein intensities were log2 transformed and normalised by subtracting the median intensity of all the proteins per sample. Replacing the missing values from a normal distribution was performed though data imputation by using the following settings: width 0.3 and downshift 0. A Two-Sample Student's *t*-test (two-tailed) followed by permutation-based FDR correction was performed to compare protein levels between the groups. The settings included S0 = 2, which is a parameter used to calculate the relative difference (ratio of change in protein expression to standard deviation) between group means. It defines the within groups variance, the relative importance of the resulted *p*-values, and the difference between means of log2 intensities [22]. Finally, the proteins with FDR adjusted *p*-value (or *q*-value) of 0.01 were considered as differentially expressed.

For bioinformatic analysis of networks involving the biological relationship between BASP1 and WT1, the Ingenuity Pathway Analysis software (IPA, Qiagen, Inc. Redwood City, CA, USA) was used. This toolset builds upon a literature-derived relationship knowledge base. A network involving all direct interactors of these proteins was built and analysed for pathway enrichment and functional annotations. Additionally, differentially expressed proteins between pancreatic cancer and healthy controls samples from MS discovery were mapped onto the BASP1/WT1 network. Subcellular localisation of significantly up- and down-regulated proteins in pancreatic cancer versus healthy control samples was manually assessed using UniProt [19] (https://www.uniprot.org/). PANTHER [23] (http://www.pantherdb.org/) was employed to identify gene ontology terms of the significantly differentially expressed proteins.

For IHC analysis, the correlation between the expression levels of protein biomarkers and clinicopathological parameters was estimated using the Mann-Whitney *U* test for continuous variables and Fisher's exact test or χ^2 for categorical variables. Kaplan–Meier analysis was used to calculate the cumulative probability of overall survival (OS), log-rank tests were used to evaluate the differences. Prognostic factors were calculated using univariable and multivariable analysis (Cox proportional hazards regression model). A value of $p < .05$ was considered statistically significant.

Statistical evaluation was conducted with Perseus software version 1.6.0.7, SPSS version 23.0 (SPSS Inc., Chicago, IL, USA), GraphPad Prism v.7 (La Jolla, CA, USA), and R [24] programming language version 3.5.1 (R Foundation for Statistical Computing, https://www.r-project.org/).

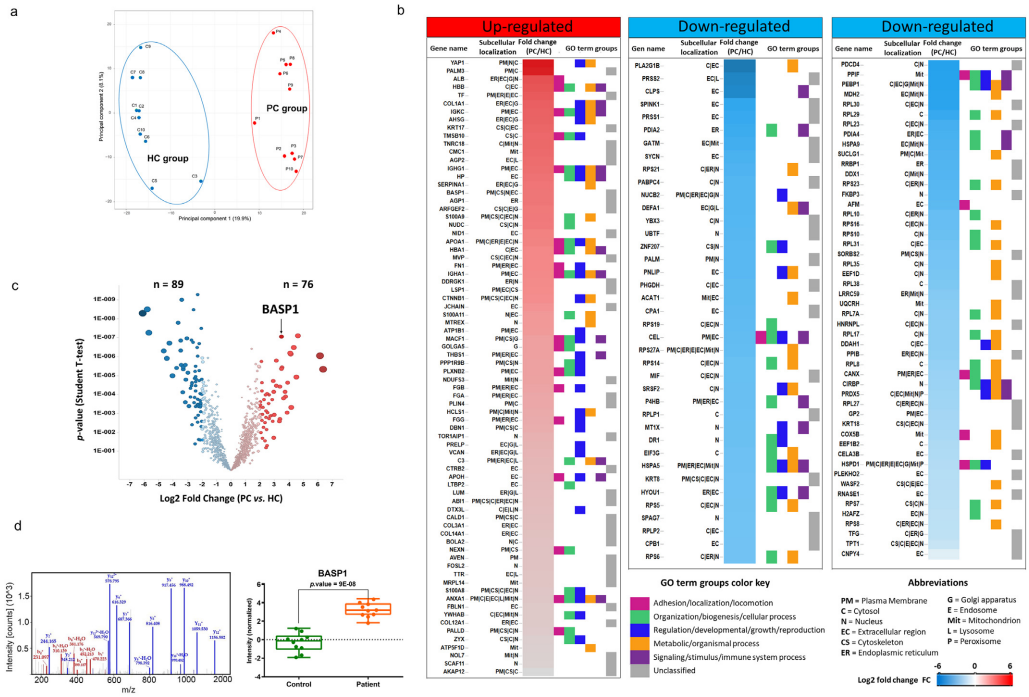


Fig. 1. Mass spectrometry (MS) discovery study. **a.** Principal Component Analysis of quantified proteins, a complete separation of pancreatic cancer (PC) and healthy control (HC) groups was observed. **b.** Heatmap of the 165 significantly altered proteins. The up-regulated and down-regulated proteins are ranked by log2 fold change and the subcellular location of each protein is also presented. PANTHER gene ontology (GO) analysis showed that GO terms of the 165 proteins related to biological processes such as localisation, biogenesis and signalling etc. **c.** Volcano plot of all proteins that were identified in this study, the dark red and dark blue dots denote the significantly up- and down-regulated proteins in PC compared to HC, respectively (*q*- value < .01 and log2 fold change over 2 in Student's *t*-test, the size of dots represent fold changes). BASP1 was one of the top-ranked candidate protein biomarkers. **d.** The label-free quantitative MS discovery, MS spectra of BASP1 (left), box-plot showing relative expression levels of BASP1 in PC and HC (right).

Table 1
PRM verified upregulated proteins in pancreatic cancer compared to healthy controls (ranked according to fold change).

Sr. no.	UniProt accession	Gene	Protein name	Unique peptide	*P* value	Fold change (PC/NC)
1	P02647	APOA1	Apolipoprotein A-I	K.LLDNWDSVTSTFSK.L	1.8E-08	39.12
2	B9A064	IGLL5	Immunoglobulin lambda-like polypeptide 5	K.VTVLGQPK.A	3.5E-09	35.02
3	P02765	AHSG	Alpha-2-HS-glycoprotein	K.FSVVYAK·C	2.1E-09	27.47
4	P0DOY2	IGLC2	Immunoglobulin lambda constant 2	K.AAPSVTLFPPSSEELQANK.A	1.6E-09	24.42
5	P02763	AGP1	Alpha-1-acid glycoprotein 1	R.YVGGQEHFAHLLILR.D	4.6E-06	24.25
6	P01857	IGHG1	Immunoglobulin heavy constant gamma 1	K.GPSVFPLAPSSK·S	2.2E-10	23.59
7	P01834	IGKC	Immunoglobulin kappa constant	K.VDNALQSGNSQESVTEQDSK.D	5.1E-11	22.78
8	P01876	IGHA1	Immunoglobulin heavy constant alpha 1	K.TPLTATLSK·S	1.3E-09	20.82
9	P02787	TF	Serotransferrin	K.EGYYGYTGAFR.C	1.6E-10	19.84
10	P02768	ALB	Serum albumin	K.DDNPNLPR.L	4.5E-09	19.70
11	P01009	SERPINA1	Alpha-1-antitrypsin	K.AVLTIDEK.G	6.2E-10	17.03
12	P80723	BASP1	Brain acid soluble protein 1	K.ETPAATEAPSSTPK.A	1.7E-05	12.91
13	P06703	S100A6	Protein S100-A6	K.LQDAEIAR.L	1.5E-05	12.13
14	Q05707	COL14A1	Collagen alpha-1(XIV) chain	R.YTAILNQIPSHSSSIR.T	6.5E-12	10.70
15	P16401	HIST1H1B	Histone H1.5	K.ATGPPVSELITK.A	1.6E-08	9.85
16	P23142	FBLN1	Fibulin-1	K.IIEVEEEQEDPYLNDR.C	4.8E-08	8.34
17	P52566	ARHGDIB	Rho GDP-dissociation inhibitor 2	K.TLLGDGPVVTDPK.A	1.9E-08	5.46

Abbreviations: NC, normal controls; PC, pancreatic cancer; PRM, parallel reaction monitoring.

3. Results

3.1. Identification of candidate biomarkers for pancreatic cancer

Representative fresh frozen pancreatic cancer ($n = 10$) and healthy control (n = 10) tissue samples were analysed using a LC-MS/MS platform. A total of 4138 proteins were identified (Additional file 2: Table S1) and 2950 proteins were quantified with one or more unique peptides (Additional file 3: Table S2). Among the quantified proteins, 2264 proteins were present in the pancreatic cancer group and 2354 proteins in the healthy control group, respectively. To demonstrate the general pattern of protein abundance variation within and between different groups, a two-dimensional Principal Component Analysis (PCA) was performed based on all quantified proteins by an online tool ClustVis [25]. Using the log2-ratio of each sample over the mean of all samples, a complete separation of the pancreatic cancer and healthy control groups was observed (Fig. 1a).

By employing the criteria of FDR adjusted *p*-value (or *q*-value) of 0.01 , S0 = 2, the number of peptides >1 and the fold change >2 as a cut-off, a total of 165 proteins with two or more unique peptides were significantly differentially expressed between the two experimental groups (Fig. 1b). A volcano plot of significantly upregulated and down-regulated proteins is presented in Fig. 1c.

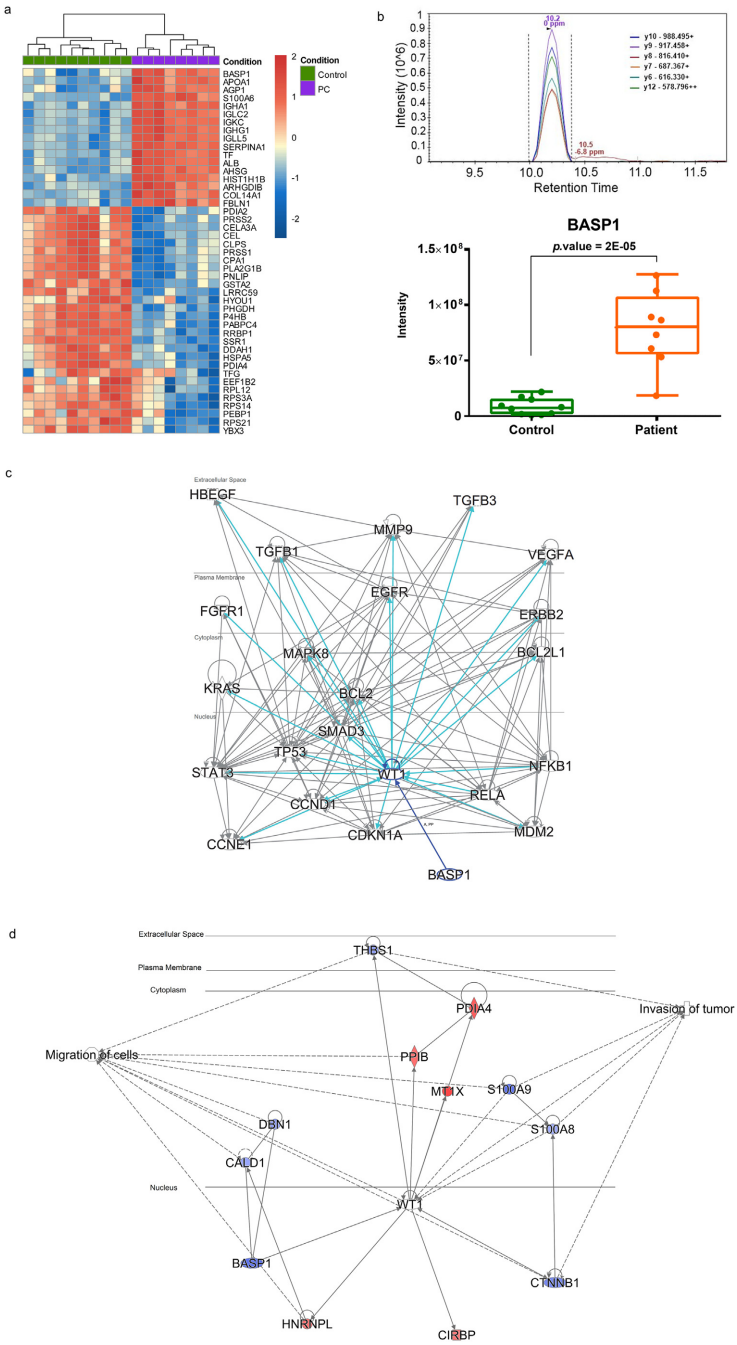
3.2. Development of targeted protein assays using PRM

To verify the differential expression changes of potential protein biomarkers from MS discovery, PRM was employed based on the same samples from the MS discovery phase ($n = 8$ in the pancreatic cancer group and $n = 10$ in the healthy control group). Eighty-one proteins with one or two unique peptides for each protein were selected and a panel of 45 proteins were successfully detected and quantified. Among these proteins, 17 proteins were significantly up-regulated ($p < .01$), while 28 proteins were down-regulated in pancreatic cancer

Table 2
PRM verified downregulated proteins in pancreatic cancer compared to healthy controls (ranked according to fold change).

Sr. no.	UniProt accession	Gene	Protein name	Unique peptide	*P* value	Fold change (NC/PC)
1	P04054	PLA2G1B	Phospholipase A2	R.AVWQFR.K	1.1E-06	56.89
2	P16233	PNLIP	Pancreatic triacylglycerol lipase	R.TGYTQASQNIR.I	8.7E-06	51.98
3	P09093	CELA3A	Chymotrypsin-like elastase family member 3A	R.WNWWGSTVK·K	1.7E-06	47.18
4	P04118	CLPS	Colipase	K.TLYGIYYK·C	1.2E-05	44.63
5	P19835	CEL	Bile salt-activated lipase	K.LGLLGDSVDIFK.G	4.5E-06	39.95
6	P07478	PRSS2	Trypsin-2	R.TLDNDILLIK.L	3.5E-06	36
7	P15085	CPA1	Carboxypeptidase A1	K.TEPVPDQDELDQLSK.A	2.4E-06	35.02
8	Q13087	PDIA2	Protein disulfide-isomerase A2	K.NFEQVAFDETK·N	3.4E-05	30.06
9	P07477	PRSS1	Trypsin-1	K.TLNNDIMLIK.L	3.3E-05	28.44
10	P09210	GSTA2	Glutathione S-transferase A2	K.LALIQEK.T	9.0E-05	10.78
11	O43175	PHGDH	D-3-phosphoglycerate dehydrogenase	K.TLGILGLGR.I	1.6E-06	10.27
12	Q13310	PABPC4	Polyadenylate-binding protein 4	K.SGVGNVFIK·N	8.4E-08	9
13	P07237	P4HB	Protein disulfide-isomerase	K.VDATEESDLAQQYGVR.G	1.1E-07	7.62
14	P43307	SSR1	Translocon-associated protein subunit alpha	K.GEDFPANNIVK·F	7.4E-07	7.21
15	P16989	YBX3	Y-box-binding protein 3	K.GAEAANVTGPDGVPVEGSR.Y	9.0E-07	6.36
16	Q9P2E9	RRBP1	Ribosome-binding protein 1	K.LLATEQEDAAVAK·S	1.8E-06	5.5
17	Q96AG4	LRRC59	Leucine-rich repeat-containing protein 59	K.LQQLPADFGR.L	3.9E-05	5.21
18	P13667	PDIA4	Protein disulfide-isomerase A4	K.VEGFPTIYFAPSGDK·K	1.9E-06	4.82
19	P11021	HSPA5	78 kDa glucose-regulated protein	K.NQLTSNPENTVFDAK.R	5.6E-07	4.5
20	Q9Y4L1	HYOU1	Hypoxia up-regulated protein 1	K.AANSLEAFIFETQDK.L	9.3E-05	4.23
21	O94760	DDAH1	N(G), N(G)-dimethylarginine dimethylaminohydrolase 1	R.ALPESLGQHALR.S	4.4E-07	3.66
22	P24534	EEF1B2	Elongation factor 1-beta	K.YGPADVEDTTGSGATDSK.D	2.0E-04	3.41
23	P30086	PEBP1	Phosphatidylethanolamine-binding protein 1	K.LYEQLSGK.-	1.4E-05	3.36
24	P63220	RPS21	40S ribosomal protein S21	K.DHASIQMNVAEVDK·V	4.7E-05	3.34
25	P61247	RPS3A	40S ribosomal protein S3a	K.TTDGYLLR.L	9.9E-06	3.34
26	P62263	RPS14	40S ribosomal protein S14	K.TPGPGAQSALR.A	2.9E-06	3.12
27	P30050	RPL12	60S ribosomal protein L12	K.IGPLGLSPK·K	1.8E-04	2.97
28	Q92734	TFG	Protein TFG	K.LLSNDEVTIK·Y	1.9E-05	2.43

Abbreviations: NC, normal controls; PC, pancreatic cancer; PRM, parallel reaction monitoring.


a
Condition
Control
PC
BASP1
APOA1
AGP1
S100A6
IGHA1
IGLC2
IGKC
IGHG1
IGLL5
SERPINA1
TF
ALB
AHSG
HIST1H1B
ARHGDIB
COL14A1
FBLN1
PDIA2
PRSS2
CELA3A
CEL
CLPS
PRSS1
CPA1
PLA2G1B
PNLIP
OSTA2
LRRC59
HYOU1
PHGDH
P4HB
PABPC4
RRBP1
SSR1
DDAH1
HSPA5
PDIA4
TFG
EEF1B2
RPL12
RPS3A
RPS14
PEBP1
RPS21
YBX3
b
Intensity (10^6)
Retention Time
y10 - 988.495+
y9 - 917.458+
y8 - 816.410+
y7 - 687.367+
y6 - 616.330+
y12 - 578.796++
BASP1
p.value = 2E-05
Intensity
Control
Patient
c
Extracellular Space
Plasma Membrane
Cytoplasm
Nucleus
HBEGF
TGFB3
MMP9
TGFB1
VEGFA
EGFR
FGFR1
ERBB2
MAPK8
BCL2L1
KRAS
BCL2
SMAD3
TP53
STAT3
WT1
NFKB1
RELA
CCND1
CDKN1A
MDM2
CCNE1
BASP1
d
Extracellular Space
Plasma Membrane
Cytoplasm
Nucleus
THBS1
PDIA4
Invasion of tumor
Migration of cells
PPIB
MYLX
S100A9
S100A8
DBN1
CALD1
WT1
BASP1
CTNNB1
HNRNPL
CIRBP

versus healthy controls, respectively (Tables 1 and 2). From the panel of 45 verified candidates, 16 extracellular proteins emerged that could theoretically be detected in serum and potentially be applied in non-invasive diagnosis and/or prognosis prediction, including S100A6, TF, FBLN1, HYOU1, PNLIP, P4HB, AHSG, PLA2G1B, AGP1, PRSS1, PRSS2, APOA1, ALB, SERPINA1, CLPS, and COL14A1 as previously reported by our group [26]. Subsequently, a consensus clustering heat map was created based on the 45 verified proteins and a clear discrimination between pancreatic cancer and healthy controls was observed (Fig. 2a).

3.3. Selection of BASP1 for further validation

BASP1 is a neuron enriched Ca(2+)-dependent calmodulin-binding protein with unknown function in pancreatic cancer. BASP1 was established as a top-ranked protein, being significantly up-regulated in the pancreatic cancer group by a fold change of 11.24, $p = 9E\text{-}08$ (Fig. 1d). Notably, based on quantification of the following unique peptides: SDGAPASDSKPGSSEAAPSSK and ETPAATEAPSSTPK, BASP1 presented as one of the most reproducible candidates, being significantly up-regulated in the pancreatic cancer group with a fold change of 12.91 and $p = 2E\text{-}05$ (Fig. 2b). As a potential novel biomarker, BASP1 was selected for further validation by bioinformatic and clinical association studies.

3.4. BASP1 is functionally related to WT1

In order to obtain an unbiased overview of the BASP1 functional relationships in a biological context, Ingenuity Pathway Analysis (IPA) was used to create a network involving all proteins with direct relationships (e.g. physical interaction or direct activation) to BASP1. This analysis, building upon a literature-derived relationship knowledge base, yielded a network including 412 proteins that were significantly enriched and involved in several canonical pathways (e.g. pancreatic adenocarcinoma signalling, regulation of the epithelial-mesenchymal transition pathway, ILK signalling, Additional file 4: Table S3) as well as tumorigenic conditions (e.g. apoptosis, cell migration, angiogenesis). Furthermore, among the top upstream regulators automatically identified by the IPA algorithm for the BASP1 interactor set, several well-known tumour-related signalling proteins emerged (e.g. TP53, TNF, TGFB1, EGF, HRAS).

Interestingly, the pathway analysis suggests that the link between BASP1 and pancreatic cancer is via WT1, and there are 21 proteins from the pancreatic adenocarcinoma signalling pathway that interact with WT1 (enrichment p-value 3E-16, Fig. 2c). Among these, extracellular signalling molecules TGFB1, TGFB3, VEGFA, HBEGF, receptor tyrosine kinases EGFR1, ERBB2 and FGFR1, apoptosis regulators BCL2, BCL2L1 and the recognised pancreatic cancer-related transcription regulator TP53, KRAS, and MAPK8 were annotated. Mapping of the differentially expressed proteins into the BASP1/WT1 network provided 11 hits out of 165 (Fig. 2d). Markedly, according to IPA analysis, most of these proteins are involved in cellular migration and tumour invasion processes.

3.5. BASP1 and WT1 expression in tumour samples and cancer cell line

The expression levels of BASP1 and WT1 were assessed in a larger cohort of pancreatic cancer patients by TMA-IHC. The clinical characteristics of the pancreatic cancer patients are shown in Table 3. Based on the validation cohort, 141 patients were successfully scored for BASP1 and 139 patients for WT1, respectively. Both markers were evaluable in 137 patients. In the BASP1 cohort ($n = 141$), 15 (10.6%) tissue samples from pancreatic cancer patients showed negative staining (Score 0) and 126 (89.4%) samples displayed positive staining, where 25 (17.7%) samples were scored as weak (Score 1), 66 (46.8%) as moderate (Score 2), and 35 (24.8%) as strong (Score 3). The majority of the staining was observed accentuated in the cytoplasm/plasma membrane (PM), accompanied by weak nuclear staining (Fig. 3a). Interestingly, 135 (97.1%) pancreatic cancer tissue samples had positive staining of WT1 protein in the WT1 cohort ($n = 139$), and only 4 (2.9%) were observed as loss of positivity (Score 0). Moreover, the WT1 staining was predominantly presented in the cytoplasm of pancreatic tumour cells, while nuclear immunostaining was weak. Furthermore, the positively stained tissue samples were subdivided into weak 22 (15.8%, Score 1), moderate 51 (36.7%, Score 2), and strong 62 (44.6%, Score 3) staining (Fig. 3b).

In order to study the dual expression patterns of BASP1 and WT1 in human pancreatic cancer cell line, we performed immunofluorescence staining of BASP1 and WT1 in PANC-1 cell line. In accordance with our IHC results, BASP1 was mostly expressed in cytoplasm and PM, while WT1 was detected in the cytoplasm and mostly with perinuclear localisation (Fig. 3c).

3.6. BASP1 expression is an independent predictor of favourable survival

Kaplan-Meier analysis showed that pancreatic cancer patients with positive BASP1 expression had significantly prolonged overall survival (OS) compared to patients with negative BASP1 expression (median survival, 27.7 vs. 13.3 months, respectively, $p = .022$, Fig. 4a). The univariable Cox regression analysis indicated that apart from BASP1 positive expression ($p = .025$), three other variables, including smoking history ($p = .015$), presenting symptoms at diagnosis ($p = .044$), and histological grade ($p = .041$), were correlated with OS. In multivariable Cox regression analysis, positive BASP1 expression remained an independent prognostic factor with a hazard ratio (HR) of 0.468, 95% confidence interval (CI) 0.257–0.852, and $p = .013$ (Table 4).

3.7. High BASP1 expression predicts beneficial response to adjuvant chemotherapy

In the BASP1 cohort, patients with high expression of BASP1 (Score 3) exhibited significantly improved OS when they received adjuvant chemotherapy compared to those without adjuvant chemotherapy (median survival, 40.5 vs. 7.2 months, respectively, $p = .020$, Fig. 4b). No correlation to adjuvant chemotherapy ($p = .603$) was observed in patients with low expression of BASP1 (score 0, 1, and 2, Fig. 4c). These results suggest that BASP1 may function both as a marker for favourable prognosis and as a predictive biomarker for positive adjuvant chemotherapy response.

3.8. WT1 expression is correlated to poor survival and chemoresistance

Kaplan-Meier analysis revealed that patients in the high WT1 expression (Score 3) group had significantly shorter OS compared to those in the low WT1 expression (Score 0, 1, and 2) group (median survival, 22.2 vs. 25.7 months, respectively, $p = .028$, Fig. 4d). Further univariable Cox regression analysis demonstrated that besides high WT1 expression ($p = .029$), other factors such as smoking history ($p = .012$), presenting symptoms at diagnosis ($p = .049$), and high pathological grades ($p = .035$) were also associated with OS. In

Fig. 2. Targeted proteomics study and bioinformatic analysis of candidate protein biomarkers. **a.** Heat map of 45 Parallel Reaction Monitoring (PRM) verified biomarkers. An apparent discrimination between PC and HC can be observed, the verified protein biomarkers are listed to the right side. **b.** PRM verification, PRM transitions used for targeted verification of BASP1 (upper), box-plot showing relative expression levels of BASP1 in PC patients and matched healthy controls (under). **c.** Ingenuity Pathway Analysis (IPA) of all proteins from the pancreatic adenocarcinoma signalling pathway that have direct with biological relationships with BASP1 or WT1. **d.** All proteins with biological relationships with WT1 or BASP1 that are differentially expressed in pancreatic cancer vs healthy controls (Blue: proteins up-regulated in pancreatic cancer. Red: proteins down-regulated in pancreatic cancer).

Table 3
Clinicopathological variables stratified by BASP1 and WT1 expression.

Factors	BASP1 cohort				WT1 cohort			
	Total	Negative	Positive	*p* value	Total	Low	High	*p* value
	N = 141	*N* = 15	*N* = 126		N = 139	*N* = 77	*N* = 62	
Age (> 65 years)	93 (66)	9 (60)	84 (66.7)	0.58	91 (65.5)	54 (70.1)	37 (59.7)	0.214
Female gender	67 (47.5)	7 (46.7)	60 (47.6)	1	68 (48.9)	35 (45.5)	33 (53.2)	0.397
BMI (>25 kg/m^2)	57 (42.9)	9 (60)	48 (40.7)	0.175	56 (42.7)	30 (41.7)	26 (44.1)	0.86
Smoking history	66 (47.1)	6 (40)	60 (48)	0.596	65 (47.1)	38 (50)	27 (43.5)	0.495
Diabetes mellitus	33 (23.6)	1 (7.1)	32 (25.4)	0.188	33 (23.9)	20 (26)	13 (21.3)	0.553
Symptoms at diagnosis	132 (96.4)	14 (100)	118 (95.9)	1	130 (96.3)	73 (96.1)	57 (96.6)	1
Tumour location (head)	117 (83)	15 (100)	102 (81)	0.074	116 (83.5)	63 (81.8)	53 (85.5)	0.649
Tumour size (>2 cm)	118 (84.3)	13 (86.7)	105 (84)	1	115 (83.3)	64 (83.1)	51 (83.6)	1
T-stage (≥T2)	122 (87.1)	14 (93.3)	108 (86.4)	0.693	119 (86.2)	66 (85.7)	53 (86.9)	1
N-stage (≥N1)	106 (76.3)	10 (66.7)	96 (77.4)	0.349	103 (75.2)	62 (81.6)	41 (67.2)	0.073
AJCC 8th edition (≥IIA)	114 (82)	12 (80)	102 (82.3)	0.734	111 (81)	66 (86.8)	45 (73.8)	0.078
Histological grade (≥3)	82 (59)	10 (66.7)	72 (58.1)	0.589	83 (60.6)	47 (61.8)	36 (59)	0.86
Resection margin (≥R1)	55 (39.3)	6 (40)	49 (39.2)	1	53 (38.4)	28 (36.4)	25 (41)	0.601
Adjuvant chemotherapy	115 (84.6)	12 (80)	103 (85.1)	0.703	112 (83.6)	61 (83.6)	51 (83.6)	1
Recurrence of disease	102 (79.7)	12 (85.7)	90 (78.9)	0.734	102 (81)	55 (78.6)	47 (83.9)	0.5

Data were incomplete for some variables. Abbreviations: AJCC, American joint committee on cancer; BMI, body mass index; N-stage, nodal stage; T-stage, tumour stage.

multivariable Cox regression analysis, high WT1 expression was identified as an independent factor associated with OS (HR 1.636, 95% CI 1.083–2.473, $p = .019$, Table 4).

Interestingly, pancreatic cancer patients with strong expression of WT1, adjuvant chemotherapy displayed no significant impact on OS ($p = .335$, Fig. 4e). Of note, pancreatic cancer patients with weak-to-moderate WT1 expression, who received adjuvant chemotherapy presented significantly extended OS compared to patients that did not receive chemotherapy (median survival, 24.5 vs. 16.9 months, respectively, $p = .006$, Fig. 4f). These findings indicate that WT1 expression is correlated with chemoresistance in pancreatic cancer.

3.9. Patients with negative BASP1 and high WT1 expression have the poorest outcome

To examine the potential biological cross-talk between BASP1 and WT1 in terms of patient survival, we performed subgroup functionality analysis of these prognostic markers. For patients with negative expression of BASP1, the high WT1 expression group had significantly reduced OS compared to the low WT1 expression group (median survival, 9.4 vs. 20.4 months, respectively, $p = .022$, Fig. 5a). Interestingly, no significant difference in OS between high and low WT1 groups was observed in patients with positive expression of BASP1 ($p = .065$, Fig. 5b). These data suggested that BASP1 can potentially relieve the oncogenic effect of WT1 in pancreatic cancer patients.

Moreover, for patients with high WT1 expression, the positive BASP1 expression group presented significantly prolonged OS compared to the BASP1 negative group (median survival, 25.8 vs. 9.4 months, respectively, $p = .00012$, Fig. 5c). In addition, patients with high WT1 and positive BASP1 expression presented a similar survival pattern as the group of patients with low WT1 and negative BASP1 expression ($p = .822$, Fig. 5d). The results confirm the possibility that the protective role of BASP1 can impede the tumour promoting function of WT1. Finally, the best prognosis was seen in patients with positive

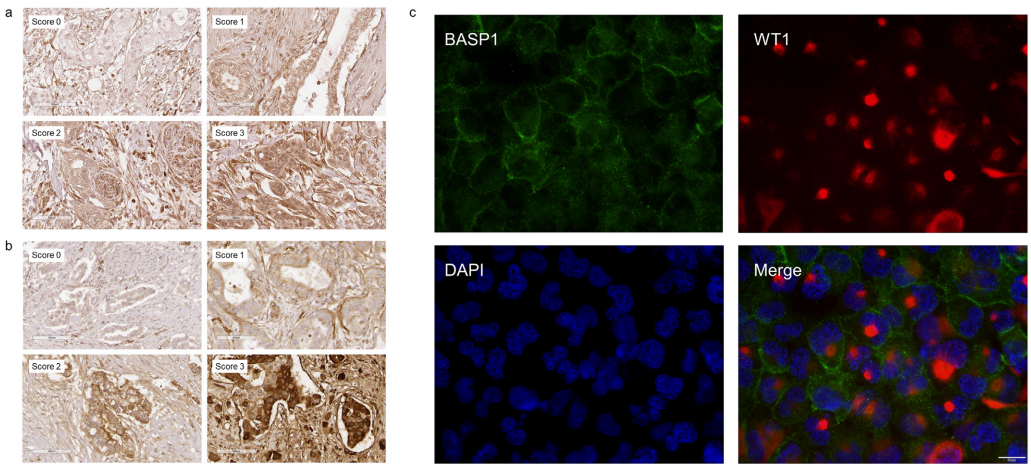


Fig. 3. Immunohistochemistry (IHC) and Immunofluorescence (IF) analysis of BASP1/WT1. **a.** Representative photomicrograph showing different levels of IHC staining of BASP1 expression in pancreatic cancer tissue samples. **b.** Representative photomicrograph showing different levels of IHC staining of WT1 expression in pancreatic cancer tissue samples. **c.** PANC1 cancer cells were labeled with antibodies for BASP1, WT1, and DAPI; green represents BASP1 (mostly expressed in cytoplasm and plasma membrane), red represents WT1 (detected in cytoplasm and mainly perinuclearly localised), and blue represents nuclear DNA staining by DAPI.

expression of BASP1 and low expression of WT1, whereas patients with negative BASP1 expression and high WT1 expression had the poorest outcome (median survival, 25.7 vs. 9.4 months, $p = .0001$, Fig. 5e). The multivariable Cox regression analysis, highlighted negative BASP1 expression and high WT1 expression as an independent factor associated with significantly shortened OS (HR 3.536, 95% CI 1.336–9.362, $p = .011$). These data suggest that BASP1 may act as a tumour suppressor rescuing the oncogenic effect of overexpressed WT1.

4. Discussion

In this study, we have employed nanoflow LC-MS/MS analysis to explore global protein expression patterns of fresh frozen pancreatic cancer tissues and healthy pancreas controls, and successfully identified a novel panel of potential protein biomarkers. This is the first report, presenting the functional role of BASP1 as a protein marker for prognosis and response to chemotherapy in pancreatic cancer. Importantly, we demonstrated that BASP1 could interact with WT1, providing valuable information for future research and clinical practice.

Currently, the gold standard for predicting outcomes in pancreatic cancer is the TNM classification system [27]. However, the TNM staging system is relatively non-discriminatory and tumours of the same stage may have different clinical behaviour in terms of prognosis and treatment response [28], which may lead to under- or overtreatment. Thus, tremendous efforts have been put into finding novel, reliable biomarkers for predicting clinical outcomes for pancreatic cancer patients [29–31]. Many interesting biomarkers have been proposed, however, few of those markers have been introduced into clinical practice, mainly due to the lack of sufficient validation [32]. To overcome this limitation, we constructed a TMA comprising large, clinically well-characterised pancreatic cancer cohort, and performed IHC analysis for assessing the validity of BASP1 as a targeted prognostic biomarker candidate.

BASP1 (also known as CAP-23 or NAP22) was originally identified as a cytoplasmic and plasma membrane-bound protein from brain extracts. It is known to be involved in axon regeneration and neuronal plasticity [33,34]. Recently, BASP1 was found to be a potential tumour suppressor and implicated in many cancers [35,36]. For instance, in hepatocellular carcinoma (HCC), aberrant promoter methylation of the

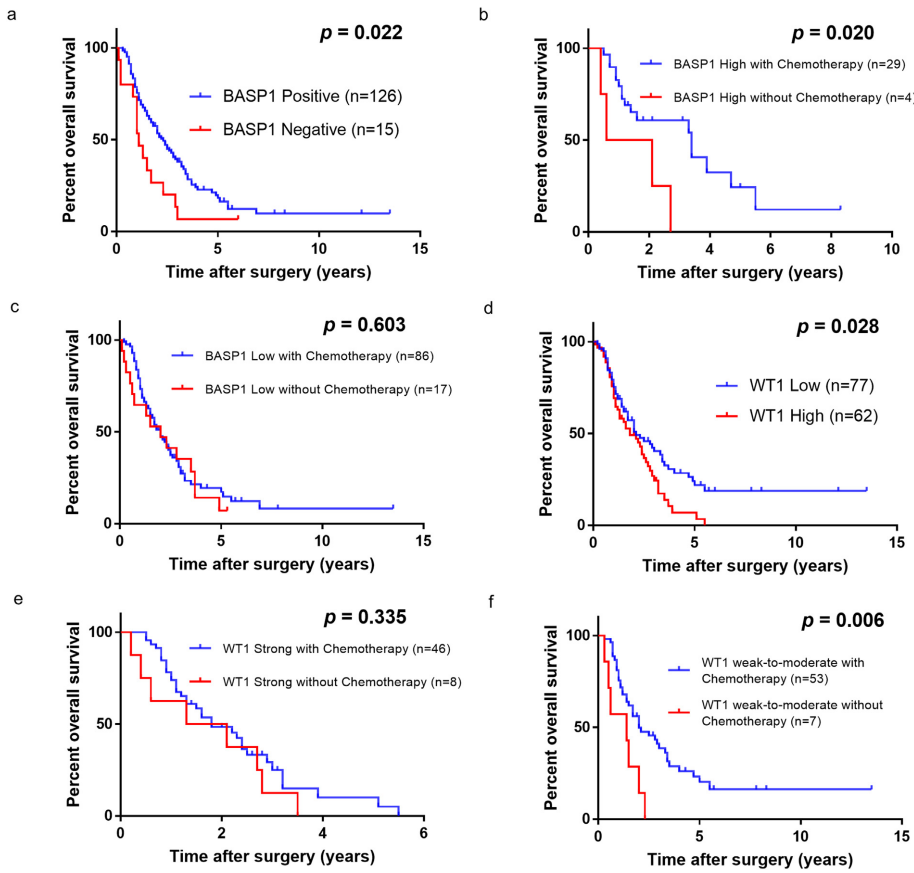


Fig. 4. Survival analysis of BASP1/WT1 in tissue microarray samples. **a.** Kaplan-Meier analyses for BASP1 (log-rank tests). **b.** When BASP1 exhibited high expression, adjuvant chemotherapy could significantly improve OS. **c.** Patients with low expression of BASP1 adjuvant chemotherapy showed no significance in improving OS. **d.** Kaplan-Meier analyses for WT1 (log-rank tests). **e.** When WT1 was strongly expressed, adjuvant chemotherapy displayed no significant impact on OS. **f.** When WT1 was weakly-to-moderately expressed, adjuvant chemotherapy significantly prolonged the OS.

BASP1 gene resulted in down-regulation of BASP1 protein expression, considered as useful finding for early detection of HCC [37]. In breast cancer, BASP1 interacts with the estrogen receptor α and enhances the anti-cancer effects of tamoxifen treatment. Additionally, high expression of BASP1 in breast cancer tissue is associated with better patient survival [36].

In the present study, the MS discovery and verification phases showed that the BASP1 protein was significantly overexpressed in pancreatic cancer tissues compared to healthy pancreas. These results are in accordance with previous findings from breast cancer studies, after IHC analysis, which indicated BASP1 as significantly up-regulated in malignant tissue compared to normal tissue [36]. However, the reason for this up-regulation remains unknown. In agreement with previous BASP1 studies, our current study highlights the tumour suppressor function of BASP1, supported by the link to favourable prognosis for pancreatic cancer patients after surgery. Additionally, pancreatic cancer patients with high BASP1 expression levels in tumour tissue showed a significantly enhanced benefit from adjuvant chemotherapy. This finding may aid clinicians to individualise chemotherapeutic treatment, hopefully improving patient outcomes and their respective survival patterns.

Interestingly, our bioinformatic analyses demonstrated that BASP1 shares a close interaction network with WT1 in pancreatic cancer. WT1 is a zinc finger transcription factor, which is a confirmed oncogenic factor. Overexpression of WT1 is associated with worse prognosis in patients with hematologic malignancies and various solid tumours [38], such as acute myeloid leukemia [39], breast cancer [40], and hepatocellular carcinoma [41]. Notably, WT1 was suggested as the most promising tumour-associated antigen for cancer immunotherapy by the National Cancer Institute [42], and many preclinical studies and clinical trials have demonstrated that WT1-targeted cancer vaccines have the potential to treat patients with pancreatic cancer [43–45]. Few studies have studied expression patterns of WT1 in pancreatic cancer [46,47]. However, only one recent study evaluated the prognostic value of WT1 in pancreatic ductal adenocarcinoma, indicating that cytoplasmic overexpression of WT1 correlated with unfavourable prognosis for the patient [48]. Notwithstanding, these findings were based on a small cohort of only 50 patients and additional studies would be necessary to confirm these results.

Consistent with other studies, we propose that WT1 may play an oncogenic role in pancreatic cancer, promoting tumour progression and being correlated to short-term relapse and poor survival. Furthermore, WT1 may be associated with chemotherapy resistance, which has not been reported previously.

There is evidence suggesting that BASP1 regulates and silences WT1 transcriptional activation [49]. Additional genome wide analysis indicated that the expression of BASP1 in leukemia cells leads to the transcriptional repression of >90% of the WT1 target genes [50]. Moreover, BASP1 and WT1 were found together in large complexes from cell lines and showed transcriptional repression activities [51,52]. It may be speculated that modulation of BASP1 in pancreatic cancer cells may facilitate WT1 targeted immunotherapy to achieve improved response rates.

A particular strength of our study was that healthy pancreas biopsies were used as control systems in the biomarker discovery phase. These unique and rare specimens were acquired from organ donors. Previous proteomic studies commonly use histologically normal tissue adjacent to the tumour as a control [53,54]. However, the regions adjacent to tumours have been found to have many aberrant morphologic and phenotypic alterations as predicted by the "field cancerisation theory" by Slaughter et al. [55,56]. The choice of healthy tissue as a comparative material for identification and further development of a discriminative biomarker is therefore preferable. Another important feature of the present study was that we used Trypsin/Lys-C Mix for protein digestion. >20% of cleavage sites may be missed by regular Trypsin [57]. The Trypsin/Lys-C Mix misses fewer lysine cleavage sites and enhances overall proteolytic efficiency as compared to Trypsin alone [58].

There are some potential limitations in our study that must be acknowledged. The fresh frozen samples used in the discovery phase were limited in number. The tissue microarray samples were accrued over a long time period with potential changes in histopathological characterisation, treatment and follow-up. However, all tissue specimens were re-evaluated by a dedicated pancreas pathologist to confirm diagnosis and uniformity of histopathological evaluation. Chemotherapy regimens varied during the study period, but most patients received gemcitabine-based chemotherapy.

5. Conclusion

We have demonstrated the feasibility of MS-based proteomic profiling of patient derived tissue specimens for biomarker development in pancreatic cancer. The proteomic strategy identified BASP1 as a promising biomarker candidate. The independent prognostic importance of BASP1 was validated in a large series of pancreatic cancer patients, together with its interaction partner WT1. We believe that our findings support that BASP1 and its putative interaction partner WT1 can be used as biomarkers for predicting the outcomes of pancreatic cancer patients, why further studies examining the function of BASP1 are warranted.

Table 4
Univariable and multivariable Cox regression analyses of overall survival.

Variables	BASP1 cohort				WT1 cohort			
	Univariable HR (95% CI)	*p* value	Multivariable HR (95% CI)	*p* value	Univariable HR (95% CI)	*p* value	Multivariable HR (95% CI)	*p* value
Age (>65 years)	1.028 (0.682–1.547)	0.896			1.035 (0.685–1.565)	0.871		
Female gender	0.785 (0.530–1.162)	0.227			0.769 (0.519–1.139)	0.19		
BMI (>25 kg/m^2)	1.346 (0.897–2.020)	0.151			1.273 (0.846–1.914)	0.247		
Smoking history	1.633 (1.100–2.422)	0.015*	1.626 (1.071–2.471)	0.023*	1.664 (1.116–2.480)	0.012*	1.727 (1.128–2.644)	0.012*
Diabetes mellitus	0.851 (0.526–1.377)	0.511			0.854 (0.522–1.395)	0.527		
Symptoms at diagnosis	0.353 (0.128–0.973)	0.044*	0.404 (0.143–1.142)	0.087	0.361 (0.131–0.995)	0.049*	0.459 (0.163–1.294)	0.141
Tumour location (head)	0.667 (0.394–1.128)	0.131			0.657 (0.389–1.110)	0.117		
Tumour size (>2 cm)	1.107 (0.663–1.849)	0.697			1.082 (0.654–1.788)	0.759		
T-stage (≥T2)	1.173 (0.684–2.011)	0.562			1.139 (0.673–1.927)	0.629		
N-stage (≥N1)	1.472 (0.916–2.366)	0.11			1.454 (0.911–2.321)	0.117		
AJCC 8th edition (≥IIA)	1.443 (0.855–2.436)	0.17			1.421 (0.852–2.370)	0.179		
Histological grade (≥3)	1.536 (1.018–2.317)	0.041*	1.647 (1.077–2.518)	0.021*	1.564 (1.033–2.368)	0.035*	1.696 (1.105–2.601)	0.016*
Resection margin (≥R1)	1.479 (0.986–2.219)	0.058			1.506 (0.996–2.277)	0.052		
Adjuvant chemotherapy	0.713 (0.431–1.180)	0.188			0.712 (0.435–1.166)	0.177		
BASP1 (positive)	0.523 (0.297–0.921)	0.025*	0.468 (0.257–0.852)	0.013*				
WT1 (high)					1.561 (1.047–2.328)	0.029*	1.636 (1.083–2.473)	0.019*

Abbreviations: AJCC, American joint committee on cancer; BMI, body mass index; CI, confidence interval; HR, Hazard ratio; N-stage, nodal stage; T-stage, tumour stage. Variables with $p <$ .05 are marked with asterisk (*), variables with $p <$ .05 in univariable analysis were included in multivariable analysis.

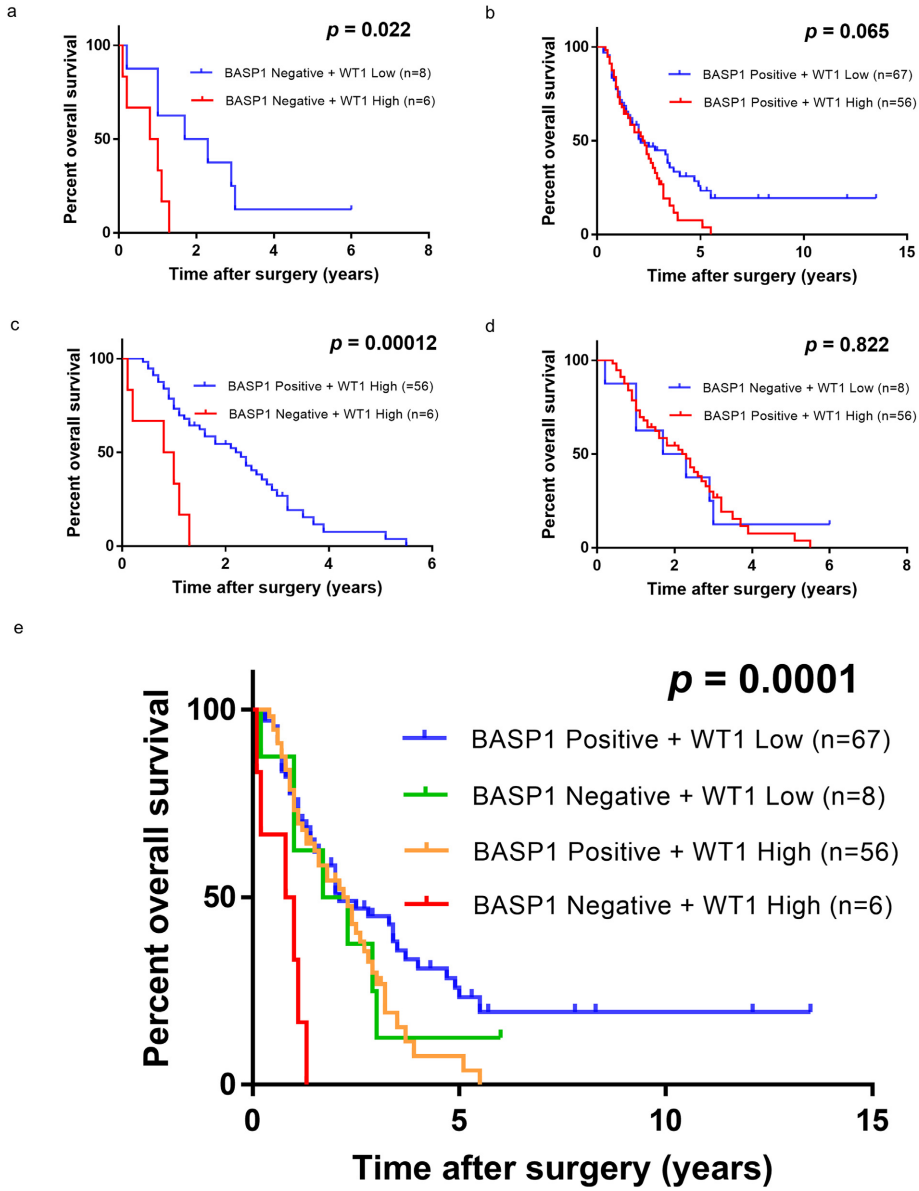


Fig. 5. Subgroup analyses of BASP1/WT1 expression and their correlation with overall survival (OS) in pancreatic cancer patients. **a.** When BASP1 was negative, high WT1 expression was associated with significantly reduced OS. **b.** When BASP1 was positive, high WT1 had no significant impact on OS. **c.** When WT1 was highly expressed, positive BASP1 correlated with significantly prolonged OS. **d.** When WT1 was highly expressed and BASP1 was positive, OS was similar to that of WT1 low expression level patients. **e.** Patients with positive BASP1 and low WT1 expression showed the best prognosis, whereas patients with negative BASP1 and high WT1 expression presented the poorest OS.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ebiom.2019.04.008.

Ethics approval and consent to participate

This study was performed in compliance with the Helsinki Declaration on ethical principles for handling human tissue specimens, with all EU and national regulations and requirements. Written informed consent was obtained from participants. Ethical permission for the study was granted by the Ethics Committee at Lund University (Ref 2010/684, 2012/661, 2015/266, 2017/320).

Consent for publication

Consent for publication was obtained from included participants.

Availability of data and material

The datasets generated and/or analysed during the current study are available from the corresponding author on reasonable request.

Funding sources

This work was supported by the Magnus Bergvall Foundation (2017-02189), the Inga and John Hain Foundation for Medical Research (2016-09-07/DA, 2017-09-07/DA), the Clas Groschinsky Foundation (M1741, M18207), the Gunnar Nilsson Foundation (GN-2018-1-90), the Gyllenstiernska Krapperup Foundation (2017-0055), the Erik and Angelica Sparre Research Foundation (2016-11-09/DA, 2018-10-28/DA), the Emil and Wera Cornell Foundation (2018-06-18/DA), the Crafoord Foundation (20170555), Governmental Funding of Clinical Research within the National Health Service (ALF, 2018-YF0012) and Sweden's Innovation Agency (Vinnova, 2019-00715). The funding sources had no role in the design and conduct of this study; the analysis and interpretation of data; or the preparation and submission of the manuscript. The corresponding author had full access to all the data in the study and had final responsibility for the decision to submit for publication.

Declaration of interests

RA, GMV and DA have filed a patent related to the findings presented in this manuscript. They are board members of Reccan Diagnostics. The other authors declare that they have no conflicts of interests.

Author contributions

QZ, TK, and DA conceived the original idea and designed the study with MZ, FL, GMV, and RA. QZ, DH, MB, TK, AS, KSH, and DA collected the data for the study, which were analysed by QZ, KP, and IPP. The data interpretation and manuscript drafting were performed by QZ and DA. The manuscript was revised by MB, KP, MZ, FL, and RA. All authors reviewed the manuscript and gave the final approval for submission.

Acknowledgements

We thank Aniel Sanchez Puente, Jeovanis Gil Valdes, Lazaro Hiram Betancourt, and Melinda Rezeli for technical support. Thermo Fisher Scientific, San Jose, is greatly acknowledged for their generous support.

References

[1] Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin 2019;69(1):7–34.
[2] Wu W, He X, Yang L, Wang Q, Bian X, Ye J, et al. Rising trends in pancreatic cancer incidence and mortality in 2000-2014. Clin Epidemiol 2018;10:789–97.
[3] Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res 2014;74(11):2913–21.
[4] Zuckerman DS, Ryan DP. Adjuvant therapy for pancreatic cancer: a review. Cancer 2008;112(2):243–9.
[5] Sperti C, Pasquali C, Piccoli A, Pedrazzoli S. Recurrence after resection for ductal adenocarcinoma of the pancreas. World J Surg 1997;21(2):195–200.
[6] Barhli A, Cros J, Bartholin L, Neuzillet C. Prognostic stratification of resected pancreatic ductal adenocarcinoma: past, present, and future. Dig Liver Dis 2018;50(10):979–90.
[7] Huang Z, Ma L, Huang C, Li Q, Nice EC. Proteomic profiling of human plasma for cancer biomarker discovery. Proteomics 2017;17(6).
[8] Peng L, Cantor DI, Huang C, Wang K, Baker MS, Nice EC. Tissue and plasma proteomics for early stage cancer detection. Mol Omics 2018;14(6):405–23.
[9] Park J, Choi Y, Namkung J, Yi SG, Kim H, Yu J, et al. Diagnostic performance enhancement of pancreatic cancer using proteomic multimarker panel. Oncotarget 2017;8(54):93117–30.
[10] Potjer TP, Mertens BJ, Nicolardi S, van der Burgt YE, Bonsing BA, Mesker WE, et al. Application of a serum protein signature for pancreatic cancer to separate cases from controls in a pancreatic surveillance cohort. Transl Oncol 2016;9(3):242–7.
[11] Takadate T, Onogawa T, Fukuda T, Motoi F, Suzuki T, Fujii K, et al. Novel prognostic protein markers of resectable pancreatic cancer identified by coupled shotgun and targeted proteomics using formalin-fixed paraffin-embedded tissues. Int J Cancer 2013;132(6):1368–82.
[12] Ansari D, Andersson R, Bauden MP, Andersson B, Connolly JB, Welinder C, et al. Protein deep sequencing applied to biobank samples from patients with pancreatic cancer. J Cancer Res Clin Oncol 2015;141(2):369–80.
[13] Crockett DK, Lin Z, Vaughn CP, Lim MS, Elenitoba-Johnson KS. Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. Lab Invest 2005;85(11):1405–15.
[14] Metz B, Kersten GF, Hoogerhout P, Brugghe HF, Timmermans HA, de Jong A, et al. Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. J Biol Chem 2004;279(8):6235–43.
[15] Bauden M, Kristl T, Andersson R, Marko-Varga G, Ansari D. Characterization of histone-related chemical modifications in formalin-fixed paraffin-embedded and fresh-frozen human pancreatic cancer xenografts using LC-MS/MS. Lab Invest 2017;97(3):279–88.
[16] McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Statistics Subcommittee of the NCIEWGoCD: REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer 2005;93(4):387–91.
[17] Henderson CM, Shulman NJ, MacLean B, MacCoss MJ, Hoofnagle AN. Skyline performs as well as vendor software in the quantitative analysis of serum 25-hydroxy vitamin D and vitamin D binding globulin. Clin Chem 2018;64(2):408–10.
[18] Tabb DL. The SEQUEST family tree. J Am Soc Mass Spectrom 2015;26(11):1814–9.
[19] Chen C, Huang H, Wu CH. Protein bioinformatics databases and resources. Methods Mol Biol 2017;1558:3–39.
[20] Hu D, Ansari D, Zhou Q, Sasor A, Hilmersson KS, Bauden M, et al. Calcium-activated chloride channel regulator 1 as a prognostic biomarker in pancreatic ductal adenocarcinoma. BMC Cancer 2018;18(1):1096.
[21] Tyanova S, Temu T, Sinitcyn P, Carlson A, Hein MY, Geiger T, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 2016;13(9):731–40.
[22] Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci 2001;98(9):5116–21.
[23] Mi H, Dong Q, Muruganujan A, Gaudet P, Lewis S, Thomas PD. PANTHER version 7: improved phylogenetic trees, orthologs and collaboration with the Gene Ontology Consortium. Nucleic Acids Res 2010;38(suppl_1):D204–10.
[24] Team RC. R: A language and environment for statistical computing; 2013.
[25] Metsalu T, Vilo J. ClustVis: a web tool for visualizing clustering of multivariate data using Principal Component Analysis and heatmap. Nucleic Acids Res 2015;43(W1):W566–70.
[26] Zhou Q, Andersson R, Hu D, Bauden M, Sasor A, Bygott T, Pawlowski K, Pla I, Marko-Varga G, Ansari D. Alpha-1-acid glycoprotein 1 is a diagnostic and prognostic biomarker for pancreatic cancer; 2019.
[27] Edge SB. Cancer AJCo: AJCC cancer staging handbook: from the AJCC cancer staging manual. New York: Springer; 2010.
[28] Helm J, Centeno BA, Coppola D, Melis M, Lloyd M, Park JY, et al. Histologic characteristics enhance predictive value of American Joint Committee on cancer staging in resectable pancreas cancer. Cancer 2009;115(18):4080–9.
[29] Gronborg M, Kristiansen TZ, Iwahori A, Chang R, Reddy R, Sato N, et al. Biomarker discovery from pancreatic cancer secretome using a differential proteomic approach. Mol Cell Proteomics 2006;5(1):157–71.
[30] Nie S, Lo A, Wu J, Zhu J, Tan Z, Simeone DM, et al. Glycoprotein biomarker panel for pancreatic cancer discovered by quantitative proteomics analysis. J Proteome Res 2014;13(4):1873–84.
[31] Ansari D, Aronsson L, Sasor A, Welinder C, Rezeli M, Marko-Varga G, et al. The role of quantitative mass spectrometry in the discovery of pancreatic cancer biomarkers for translational science. J Transl Med 2014;12:87.
[32] Ansari D, Rosendahl A, Elebro J, Andersson R. Systematic review of immunohistochemical biomarkers to identify prognostic subgroups of patients with pancreatic cancer. Br J Surg 2011;98(8):1041–55.
[33] Maekawa S, Maekawa M, Hattori S, Nakamura S. Purification and molecular cloning of a novel acidic calmodulin binding protein from rat brain. J Biol Chem 1993;268(18):13703–9.
[34] Bomze HM, Bulsara KR, Iskandar BJ, Caroni P, Skene JH. Spinal axon regeneration evoked by replacing two growth cone proteins in adult neurons. Nat Neurosci 2001;4(1):38–43.

[35] Hartl M, Nist A, Khan MI, Valovka T, Bister K. Inhibition of Myc-induced cell transformation by brain acid-soluble protein 1 (BASP1). Proc Natl Acad Sci U S A 2009;106(14):5604–9.
[36] Marsh LA, Carrera S, Shandilya J, Heesom KJ, Davidson AD, Medler KF, et al. BASP1 interacts with oestrogen receptor alpha and modifies the tamoxifen response. Cell Death Dis 2017;8(5):e2771.
[37] Moribe T, Iizuka N, Miura T, Stark M, Tamatsukuri S, Ishitsuka H, et al. Identification of novel aberrant methylation of BASP1 and SRD5A2 for early diagnosis of hepatocellular carcinoma by genome-wide search. Int J Oncol 2008;33(5):949–58.
[38] Qi XW, Zhang F, Wu H, Liu JL, Zong BG, Xu C, et al. Wilms' tumor 1 (WT1) expression and prognosis in solid cancer patients: a systematic review and meta-analysis. Sci Rep 2015;5(8924).
[39] Lapillonne H, Renneville A, Auvrignon A, Flamant C, Blaise A, Perot C, et al. High WT1 expression after induction therapy predicts high risk of relapse and death in pediatric acute myeloid leukemia. J Clin Oncol 2006;24(10):1507–15.
[40] Qi X-W, Zhang F, Yang X-H, Fan L-J, Zhang Y, Liang Y, et al. High Wilms' tumor 1 mRNA expression correlates with basal-like and ERBB2 molecular subtypes and poor prognosis of breast cancer. Oncol Rep 2012;28(4):1231–6.
[41] Sera T, Hiasa Y, Mashiba T, Tokumoto Y, Hirooka M, Konishi I, et al. Wilms' tumour 1 gene expression is increased in hepatocellular carcinoma and associated with poor prognosis. Eur J Cancer 2008;44(4):600–8.
[42] Cheever MA, Allison JP, Ferris AS, Finn OJ, Hastings BM, Hecht TT, et al. The prioritization of cancer antigens: a national cancer institute pilot project for the acceleration of translational research. Clin Cancer Res 2009;15(17):5323–37.
[43] Kaida M, Morita-Hoshi Y, Soeda A, Wakeda T, Yamaki Y, Kojima Y, et al. Phase 1 trial of Wilms tumor 1 (WT1) peptide vaccine and gemcitabine combination therapy in patients with advanced pancreatic or biliary tract cancer. J Immunother 2011;34(1):92–9.
[44] Koido S, Okamoto M, Shimodaira S, Sugiyama H. Wilms' tumor 1 (WT1)-targeted cancer vaccines to extend survival for patients with pancreatic cancer. Immunotherapy 2016;8:1309–20.
[45] Koido S, Homma S, Okamoto M, Takakura K, Mori M, Yoshizaki S, et al. Treatment with chemotherapy and dendritic cells pulsed with multiple Wilms' tumor 1 (WT1)-specific MHC class I/II-restricted epitopes for pancreatic cancer. Clin Cancer Res 2014;20(16):4228–39.
[46] Oji Y, Nakamori S, Fujikawa M, Nakatsuka S, Yokota A, Tatsumi N, et al. Overexpression of the Wilms' tumor gene WT1 in pancreatic ductal adenocarcinoma. Cancer Sci 2004;95(7):583–7.
[47] Nakatsuka S, Oji Y, Horiuchi T, Kanda T, Kitagawa M, Takeuchi T, et al. Immunohistochemical detection of WT1 protein in a variety of cancer cells. Mod Pathol 2006;19(6):804–14.
[48] Kanai T, Ito Z, Oji Y, Suka M, Nishida S, Takakura K, et al. Prognostic significance of Wilms' tumor 1 expression in patients with pancreatic ductal adenocarcinoma. Oncol Lett 2018;16(2):2682–92.
[49] Carpenter B, Hill KJ, Charalambous M, Wagner KJ, Lahiri D, James DI, et al. BASP1 is a transcriptional cosuppressor for the Wilms' tumor suppressor protein WT1. Mol Cell Biol 2004;24(2):537–49.
[50] Goodfellow SJ, Rebello MR, Toska E, Zeef LA, Rudd SG, Medler KF, et al. WT1 and its transcriptional cofactor BASP1 redirect the differentiation pathway of an established blood cell line. Biochem J 2011;435(1):113–25.
[51] Green LM, Wagner KJ, Campbell HA, Addison K, Roberts SG. Dynamic interaction between WT1 and BASP1 in transcriptional regulation during differentiation. Nucleic Acids Res 2009;37(2):431–40.
[52] Toska E, Shandilya J, Goodfellow SJ, Medler KF, Roberts SG. Prohibitin is required for transcriptional repression by the WT1-BASP1 complex. Oncogene 2014;33(43):5100–8.
[53] Kosanam H, Prassas I, Chrystoja CC, Soleas I, Chan A, Dimitromanolakis A, et al. Laminin, gamma 2 (LAMC2): a promising new putative pancreatic cancer biomarker identified by proteomic analysis of pancreatic adenocarcinoma tissues. Mol Cell Proteomics 2013;12(10):2820–32.
[54] Coleman O, Henry M, O'Neill F, Roche S, Swan N, Boyle L, et al. A comparative quantitative LC-MS/MS profiling analysis of human pancreatic adenocarcinoma, adjacent-normal tissue, and patient-derived tumour xenografts. Proteomes 2018;6(4).
[55] Slaughter DP, Southwick HW. Smejkal W: field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer 1953;6(5):963–8.
[56] Aran D, Camarda R, Odegaard J, Paik H, Oskotsky B, Krings G, et al. Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat Commun 2017;8(1):1077.
[57] Saveliev S, Bratz M, Zubarev R, Szapacs M, Budamgunta H, Urh M. Trypsin/Lys-C protease mix for enhanced protein mass spectrometry analysis. Nat Methods 2013;10(1134).
[58] Glatter T, Ludwig C, Ahrne E, Aebersold R, Heck AJ, Schmidt A. Large-scale quantitative assessment of different in-solution protein digestion protocols reveals superior cleavage efficiency of tandem Lys-C/trypsin proteolysis over trypsin digestion. J Proteome Res 2012;11(11):5145–56.

Paper III

Zhou et al. J Transl Med (2020) 18:77
https://doi.org/10.1186/s12967-020-02254-7

Journal of Translational Medicine

RESEARCH Open Access

YAP1 is an independent prognostic marker in pancreatic cancer and associated with extracellular matrix remodeling

Qimin Zhou[1,2], Monika Bauden[2], Roland Andersson[2], Dingyuan Hu[2], György Marko-Varga[3], Jianfeng Xu[4], Agata Sasor[5], Hua Dai[6], Krzysztof Pawłowski[7,8], Katarzyna Said Hilmersson[2], Xi Chen[1] and Daniel Ansari[2*]

Abstract

Background: Pancreatic cancer is a major cause of cancer-related mortality. The identification of effective biomarkers is essential in order to improve management of the disease. Yes-associated protein 1 (YAP1) is a downstream effector of the Hippo pathway, a signal transduction system implicated in tissue repair and regeneration, as well as tumorigenesis. Here we evaluate the biomarker potential of YAP1 in pancreatic cancer tissue.

Methods: YAP1 was selected as a possible biomarker for pancreatic cancer from global protein sequencing of fresh frozen pancreatic cancer tissue samples and normal pancreas controls. The prognostic utility of YAP1 was evaluated using mRNA expression data from 176 pancreatic cancer patients in The Cancer Genome Atlas (TCGA), as well as protein expression data from immunohistochemistry analysis of a local tissue microarray (TMA) cohort comprising 140 pancreatic cancer patients. Ingenuity Pathway Analysis was applied to outline the interaction network for YAP1 in connection to the pancreatic tumor microenvironment. The expression of YAP1 target gene products was evaluated after treatment of the pancreatic cancer cell line Panc-1 with three substances interrupting YAP–TEAD interaction, including Super-TDU, Verteporfin and CA3.

Results: Mass spectrometry based proteomics showed that YAP1 is the top upregulated protein in pancreatic cancer tissue when compared to normal controls (log2 fold change 6.4; p = 5E−06). Prognostic analysis of YAP1 demonstrated a significant correlation between mRNA expression level data and reduced overall survival (p = 0.001). In addition, TMA and immunohistochemistry analysis suggested that YAP1 protein expression is an independent predictor of poor overall survival [hazard ratio (HR) 1.870, 95% confidence interval (CI) 1.224–2.855, p = 0.004], as well as reduced disease-free survival (HR 1.950, 95% CI 1.299–2.927, p = 0.001). Bioinformatic analyses coupled with in vitro assays indicated that YAP1 is involved in the transcriptional control of target genes, associated with extracellular matrix remodeling, which could be modified by selected substances disrupting the YAP1-TEAD interaction.

Conclusions: Our findings indicate that YAP1 is an important prognostic biomarker for pancreatic cancer and may play a regulatory role in the remodeling of the extracellular matrix.

Keywords: Pancreatic cancer, YAP1, Transcriptomics, Proteomics, Prognosis, Extracellular matrix remodeling, Cancer

Background

Pancreatic cancer is one of the most aggressive malignancies with a dismal 5-year survival rate of 9% [1]. It has surpassed breast cancer to become the third leading cause of cancer-related death and is estimated to rise to

*Correspondence: daniel.ansari@med.lu.se
[2] Department of Surgery, Clinical Sciences Lund, Lund University and Skåne University Hospital, 221 85 Lund, Sweden
Full list of author information is available at the end of the article

the second leading cause by 2030 [2]. Multiple factors, such as late diagnosis and resistance to conventional therapies, contribute to the overall poor prognosis.

The ability to identify subgroups of patients that may benefit from specific clinical management is considered central to modern precision oncology. For that purpose, large-scale genomic studies have been performed to determine molecular subtypes of pancreatic cancer requiring individualized treatments [3–6]. Such studies have massively increased our understanding of pancreatic cancer at the molecular level.

Proteomics is a valuable complement to genetic studies. Mass spectrometry (MS)-based proteomics profiling of patient-derived samples has been suggested as an effective approach for the discovery of biomarkers and detection of suitable therapeutic targets in many cancers [7–10].

Yes-associated protein 1 (YAP1) is a downstream effector of the Hippo signaling pathway, which is involved in tissue repair and regeneration, as well as tumorigenesis. Activation of the Hippo pathway leads to inactivation of YAP1 by cytoplasmic retention or proteolytic degradation [11, 12]. YAP1 in its active form, on the other hand, functions as a transcriptional co-activator predominantly mediated by an interaction with TEAD transcription factors [13]. Active YAP1 is also recognized as a potent oncogene closely linked to the progression of several cancer types [14, 15]. However, the role of the YAP1-TEAD interaction in regulating the expression of target genes in pancreatic cancer has not been completely explored.

In a previous study [10], we identified YAP1 as a differentially expressed protein between pancreatic cancer and normal controls using MS-based proteomics profiling. In the present study, we investigate the prognostic utility and the biological significance of YAP1 in pancreatic cancer using large and clinically well-annotated cohorts, complemented by bioinformatics and in vitro experimental analyses.

Materials and methods

Patient samples

For the MS-based proteomics, fresh frozen pancreatic cancer tissues (n = 10) were collected from patients with pancreatic ductal adenocarcinoma undergoing pancreaticoduodenectomy between July 2013 and April 2015 at the Department of Surgery, Skåne University Hospital, Lund, Sweden. Written informed consent was obtained from the patients included in the study. Age and gender-matched, fresh frozen, normal pancreatic biopsies (n = 10) were assessed from organ donors and obtained from the national consortium Excellence of Diabetes Research in Sweden and Lund University Diabetes center (LUDC).

The immunohistochemical (IHC) target verification was performed using tissue microarrays (TMA) from archival formalin-fixed paraffin-embedded (FFPE) resection specimens from 140 patients with pancreatic ductal adenocarcinoma who underwent curative intent pancreatic surgery from 1995 to 2017 at Skåne University Hospital, Lund and Malmö, Sweden.

All samples were histopathologically verified and selected by a specialized surgical pathologist prior to analysis. Ethical permission for the study was granted by the Ethical Committee at Lund University (Ref 2010/684, 2012/661, 2015/266, 2017/320). The REMARK guidelines were followed where applicable [16].

MS-based proteomics

Sample processing and LC–MS/MS analysis were performed as reported previously [10]. Briefly, proteins extracted from fresh frozen pancreas specimens were reduced, alkylated and digested into peptides using Lys-C and trypsin. The peptides were analyzed using a high-performance liquid chromatography system, EASY-nLC 1000 connected to Q Exactive quadrupole-Orbitrap mass spectrometer equipped with a nanospray ion source (Thermo-Fisher Scientific, Bremen, Germany). To identify the detected proteins, the acquired MS/MS data were managed using Proteome Discoverer software, version 1.4 (Thermo Fisher).

mRNA expression data

Publicly available transcriptomics data were retrieved from 176 pancreatic cancer patients from The Cancer Genome Atlas (TCGA) [17–19]. RNA-seq data were analyzed as the number of Fragments Per Kilobase of exon per Million reads (FPKM).

Tissue microarray

The TMA was constructed from FFPE pancreatic tumors by a trained biomedical technician using an automated tissue array device (Minicore® 3, Alphelys, Plaisir, France). A set of 4 cores with a diameter of 2 mm were extracted from each specimen and fixed into a new paraffin block. The completed blocks were then sectioned into 3 µm thick sections and mounted on glass slides.

Immunohistochemistry

IHC analysis was performed as described previously [20]. Briefly, deparaffinization, rehydration and antigen-retrieval were performed using the automated PT Link system (Dako, Agilent Technologies, Glostrup, Denmark). TMA-slides were then incubated with monoclonal rabbit anti-human primary antibody against YAP1

(dilution 1:200; Cell Signaling) followed by biotinylated goat anti-rabbit secondary antibody (dilution 1:200; Vector Laboratories, Burlingame, CA). Avidin–biotin–peroxidase complex (Vectastain Elite ABC-HRP Kit, Vector Laboratories, Burlingame, CA) was used for signal amplification. The color was developed using chromogen diaminobenzidine (DAB) (Vector Laboratories). The nuclei were colored with hematoxylin. The immunostaining was evaluated by three independent pathologists, blinded to clinical information. H-score was applied as a semiquantitative approach [21, 22]. The intensity of YAP1 staining was scored as [0] (negative), [1+] (weak), [2+] (moderate), or [3+] (strong) and the percentage of cells at each staining intensity level was recorded. The H-scores were calculated by following formula:

$$\text{H-score} = 0 \times (\%\ \text{cells}\ [0]) + 1 \times (\%\ \text{cells}\ [1]) + 2 \times (\%\ \text{cells}\ [2]) + 3 \times (\%\ \text{cells}\ [3]).$$

Bioinformatics

Ingenuity Pathway Analysis software (IPA, Qiagen, Inc. Redwood City, CA, USA) was used for bioinformatic analysis of networks involving the biological relationship between YAP1 and pancreatic cancer. A network involving all direct interactors of these proteins was built and analyzed for pathway enrichment and functional annotations.

Cell culture

The patient derived pancreatic cancer cell line Panc-1 (ATCC-LGC Standards, Manassas, VA, USA) was used for the in vitro experiments. The cells were maintained in DMEM supplemented with 10% fetal bovine serum, 100 U/ml penicillin and 100 µg/ml streptomycin and kept in a humified atmosphere, in 5% CO_2 at 37 °C. Prior experiment, the cells were observed using phase contrast microscope to ensure the condition of the cells including morphological characteristics and vitality.

Immunofluorescence based Cellomics

To assert the YAP1 expression profile, the cells were seeded in 6 well plates with the density of fifty thousand cells per well. After 48 h, the cells were fixed with 4% paraformaldehyde (Histolab, Västra Frölunda, Sweden) and stained with primary rabbit anti-human YAP1 (dilution 1: 250, Cell Signaling) followed by Alexa Fluor 488 conjugated donkey-anti-rabbit secondary antibody (dilution 1:200, Invitrogen, USA). The nucleus was marked using DAPI (NucBlue®, Molecular probes, Life technologies, USA). Cellomics ArrayScan platform VTI HCS (ThermoScientific, Rockford, IL, USA) reader connected to Bioapplication software was thereafter used for image processing.

In each well, a cell population consisting of two thousand cells was analysed using multiparameter fluorescent microscopic imaging system designed for high content screening. The processed data obtained from automatically acquired images were quantified as fluorescence intensity for the selected channel (Alexa 488). The accessed images were visualized using automated fluorescence microscopy.

YAP1 target gene expression

To evaluate the expression of selected YAP1 target genes, the cells were seeded in 6-well plates with a concentration of thirty thousand cells per well. After one cell cycle, the cells were incubated with a maximal tolerable dose of three substances interrupting YAP–TEAD interaction; Super-TDU (500 nM), Verteporfin (100 nM) and CA3 (100 nM) or complete medium. After 48 h, the cell lysates and conditioned medium from respective well and plate were collected. All experiments were executed in triplicates. Expression levels of YAP1 targets genes, including amphiregulin (AREG), connective tissue growth factor (CTGF), cysteine-rich angiogenic inducer 61 (CYR61), fibroblast growth factor 1 (FGF1) and mesothelin (MSLN), were selected from the Ingenuity Pathway Analysis and measured in each sample using enzyme-linked immunosorbent assay (ELISA). 100 µg protein from respective sample was analyzed in each assay according to the manufacturer's instructions. AREG, CTGF, CYR61, FGF1 were purchased from Nordic Biosite AB, Täby, SE and MSLN from Biolegend, San Diego, CA, USA.

Statistical analysis

The correlation between YAP1 expression levels and clinicopathological parameters was estimated using the Mann–Whitney U test for continuous variables and Fisher's exact test or χ^2 for categorical variables. The Kaplan–Meier method was used to model the cumulative probability of overall survival (OS) and disease-free survival (DFS) and statistical differences were assessed using the log-rank test. Univariable and multivariable survival analysis were also performed using Cox proportional hazards regression modeling.

One-way ANOVA parametric test was applied to compare the concentrations of secreted YAP target genes measured in condition medium obtained from Panc-1 cells subjected to three substances interrupting YAP1 transcriptional activity or untreated cells.

Statistical evaluation was conducted with SPSS version 23.0 (SPSS Inc., Chicago, IL, USA) and GraphPad Prism

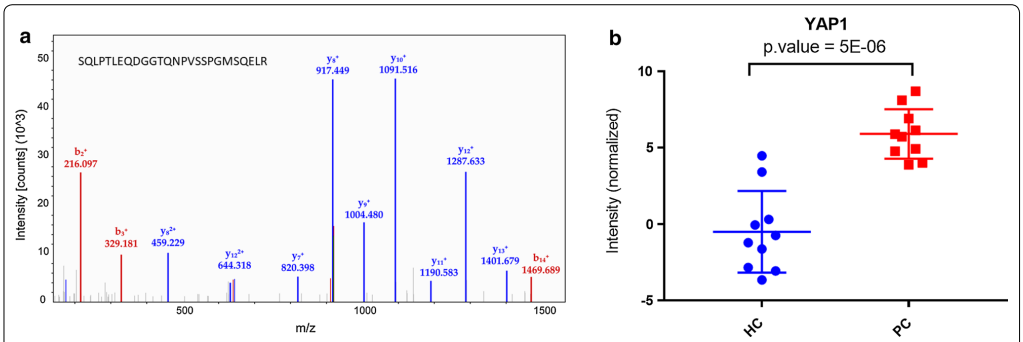


Fig. 1 Selection of the YAP1 protein for validation. **a** Label-free quantitative MS spectra of YAP1 (based on peptide SQLPTLEQDGGTQNPVSSPGMSQELR). **b** Box-plot showing relative expression levels of YAP1 in pancreatic cancer (PC) and healthy controls (HC)

Table 1 Characteristics of the TCGA cohort (n = 176)

Variable	N = 176
Median age (range), years	65 (35–88)
Female gender	50 (45.5%)
AJCC-stage	
I	21 (11.9%)
II	145 (82.4%)
III	3 (1.7%)
IV	4 (2.3%)
Unknown	3 (1.7%)
Median FPKM (range)	19.0 (0.5-46.6)

FPKM fragments per kilobase of exon per million reads

v.8.0.1 (La Jolla, CA, USA). A p-value < 0.05 was considered statistically significant.

Results

YAP1 is the top upregulated protein in pancreatic cancer

Fresh frozen biopsies from pancreatic tumors (n = 10) and healthy pancreatic tissue (n = 10), were analyzed using label-free quantitative proteomics to discover differentially expressed proteins. In total, 4138 proteins were identified, and 2950 proteins were quantified based on one or more unique peptides. 165 candidates were subsequently determined as potential biomarkers for pancreatic cancer, as previously reported [10]. Characterized by six unique peptides, YAP1 was annotated as the top upregulated protein in pancreatic tumor specimens (log2 fold change 6.4; p = 5E−06) (Fig. 1a, b).

mRNA expression levels of YAP1 as a prognostic marker

To assess the prognostic significance of YAP1, we analyzed mRNA expression level data and patient survival based on 176 pancreatic cancer patients included in TCGA (Table 1). The median FPKM value was 19.0, ranging from 0.5 to 46.6. The median FPKM value was used to divide the cohort into a low (FPKM ≤ 19) and a high expression group (FPKM > 19). The Kaplan–Meier plots revealed that high YAP1 mRNA expression was significantly correlated with poorer OS when compared with low mRNA YAP1 expression, as illustrated in Fig. 2 (median survival 17 months vs. 23 months, respectively, p = 0.001).

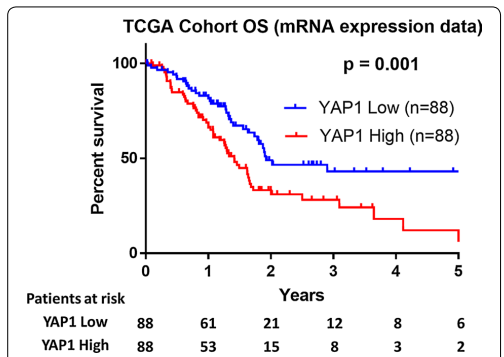


Fig. 2 Kaplan–Meier survival curves stratified by YAP1 mRNA expression levels in the TCGA cohort. Patients were categorized based on the median number of fragments per kilobase of exon per million reads (FPKM) into low expression (≤ 19) and high expression groups (> 19)

YAP1 protein expression levels and prognosis

The protein expression levels of YAP1 were analyzed using immunohistochemistry staining on TMA sections constructed from 140 pancreatic tumors. The antibody

staining specific for YAP1 was detected in the nucleus or in the nucleus and cytoplasm of tumor cells. The median H-score was 170 (range, 59–289). Based on the median H-score (170), a low (H-score ≤ 170) and a high expression group (H-score > 170) were created (Fig. 3a). No significant differences in clinicopathological features were identified between high and low YAP1 expression groups (Table 2).

Kaplan–Meier analysis revealed that high YAP1 protein expression was significantly correlated with shorter OS when compared with low YAP1 protein expression (median survival, 17.9 vs. 34.3 months, respectively, p = 0.001, log-rank test; Fig. 3b). Furthermore, patients exhibiting high YAP1 protein expression had significantly reduced DFS when compared to the low YAP1 protein expression group (median DFS, 10.7 vs. 17.5 months, respectively, p = 0.005, log-rank test; Fig. 3c).

The univariable Cox regression analysis of OS identified smoking history (p = 0.04), symptoms at diagnosis (p = 0.05), histopathological grade (p = 0.03), and high expression of YAP1 (p = 0.001) as factors associated with shorter OS. In multivariable Cox regression analysis, high YAP1 protein expression was identified as an independent risk factor for poor OS (hazard ratio (HR) 1.870, 95% confidence interval (CI) 1.224–2.855, p = 0.004). Moreover, univariable Cox regression analysis of DFS determined histopathological grade (p = 0.028), resection margin ≥ R1 (p = 0.028), and high expression of YAP1 (p = 0.006) as factors associated with decreased DFS. Multivariable Cox regression analysis confirmed the results, indicating that high YAP1 protein expression is an independent risk factor for reduced DFS (HR 1.950, 95% CI 1.299–2.927, p = 0.001) (Table 3).

We thus interpret that YAP1 may function as a marker for poor prognosis and disease relapse in pancreatic cancer patients.

YAP1 is connected to mediators promoting remodeling of the extracellular matrix

Subsequently, we explored the biological background of the obtained results with the aim to identify the most significant networks and relationships associated with YAP1 expression in pancreatic cancer. Bioinformatic analysis using the IPA software revealed that YAP1 is

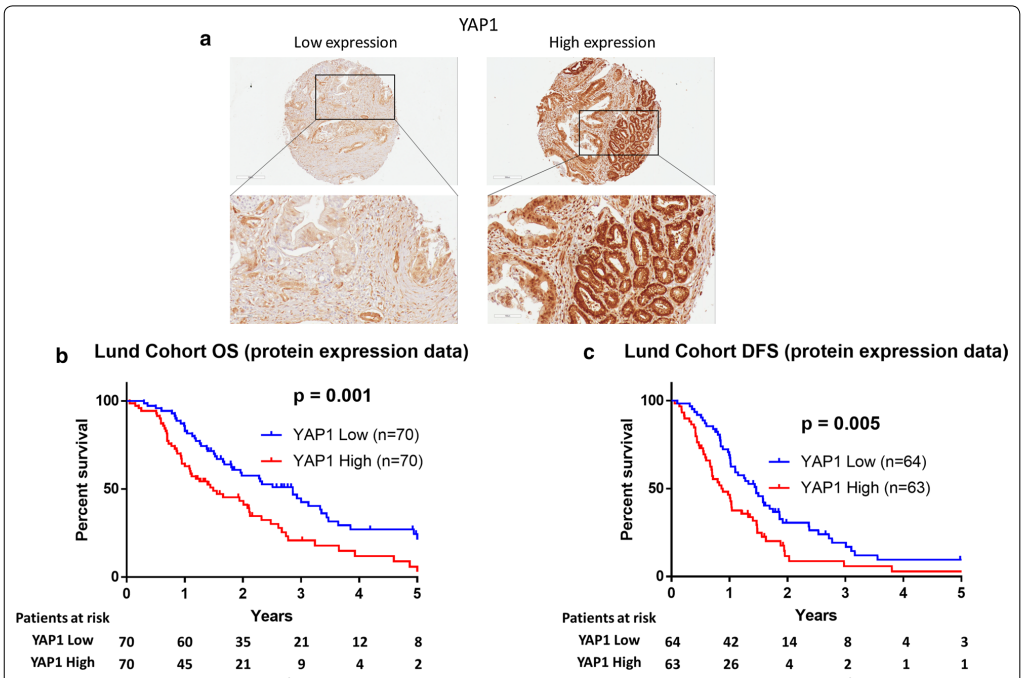


Fig. 3 Immunohistochemical analysis of YAP1 protein expression in the tissue microarray cohort. **a** Representative images of YAP1 immunostaining in low and high expression groups using the median H-score (170) as cut-off. **b** Kaplan–Meier survival curves for overall survival stratified by YAP1 protein expression. **c** Kaplan–Meier survival curves for disease-free survival stratified by YAP1 protein expression

Table 2 Characteristics of the TMA cohort (n = 140)

Variable	N	All patients (n = 140)	Low YAP1 protein expression (n = 70)	High YAP1 protein expression (n = 70)	*p*
Age > 65 years	140	93 (66.4)	48 (68.6)	45 (64.3)	0.721
Female gender	140	66 (47.1)	35 (50)	31 (44.3)	0.612
BMI > 25 kg/m^2	132	57 (43.2)	32 (47.1)	25 (39.1)	0.383
Smoking history	139	67 (48.2)	28 (40.6)	39 (55.7)	0.09
Diabetes mellitus	139	33 (23.7)	19 (27.1)	14 (20.3)	0.426
Symptoms at diagnosis	136	131 (96.3)	68 (100)	63 (92.6)	0.058
Tumor location (head)	140	117 (83.6)	62 (88.6)	55 (78.6)	0.17
Tumor size > 2 cm	139	117 (84.2)	60 (87)	57 (81.4)	0.487
T-stage ≥ T2	139	121 (87.1)	60 (87)	61 (87.1)	1
N-stage ≥ N1	138	104 (75.4)	53 (76.8)	51 (73.9)	0.844
AJCC-stage ≥ II	138	112 (81.2)	56 (81.2)	56 (81.2)	1
Histological grade ≥ 3	138	83 (60.1)	38 (55.9)	45 (64.3)	0.385
Positive resection margin	139	55 (39.6)	28 (40.6)	27 (38.6)	0.863
Adjuvant chemotherapy	135	113 (83.7)	60 (87)	53 (80.3)	0.355
Recurrence of disease	127	103 (81.1)	51 (79.7)	52 (82.5)	0.821

N, number of non-missing values. Qualitative data are expressed as n (%)

AJCC American Joint Committee on Cancer, *BMI* body mass index, *N-stage* nodal stage, *T-stage* tumor stage

Table 3 Univariable and multivariable Cox regression analysis in the TMA cohort (n = 140)

Variable	OS				DFS			
	Univariable HR (95% CI)	*p*	Multivariate HR (95% CI)	*p*	Univariable HR (95% CI)	*p*	Multivariable HR (95% CI)	*p*
Age (> 65)	0.994 (0.658–1.501)	0.977			0.760 (0.506–1.144)	0.189		
Female gender	0.825 (0.557–1.221)	0.336			0.675 (0.453–1.005)	0.053		
BMI (> 25 kg/m^2)	1.250 (0.832–1.876)	0.283			1.372 (0.913–2.061)	0.128		
Smoking history	1.510 (1.019–2.239)	0.04*	1.319 (0.868–2.003)	0.195	1.268 (0.852–1.887)	0.242		
Diabetes	0.782 (0.479–1.277)	0.326			0.927 (0.567–1.515)	0.762		
Symptoms at diagnosis	0.363 (0.132–1.000)	0.05*	0.548 (0.193–1.559)	0.260	0.620 (0.227–1.693)	0.351		
Tumor location (head)	0.658 (0.390–1.112)	0.118			1.143 (0.625–2.092)	0.664		
Tumor size (> 2 cm)	1.090 (0.653–1.819)	0.741			1.215 (0.710–2.079)	0.478		
T-stage (≥ T2)	1.152 (0.672–1.973)	0.607			1.429 (0.795–2.571)	0.233		
N-stage (≥ N1)	1.474 (0.924–2.352)	0.104			1.316 (0.829–2.088)	0.244		
AJCC-stage (≥ II)	1.426 (0.855–2.379)	0.174			1.345 (0.814–2.222)	0.248		
Histological grade (≥ 3)	1.580 (1.045–2.390)	0.03*	1.728 (1.123–2.657)	0.013*	1.592 (1.050–2.413)	0.028*	1.628 (1.072–2.472)	0.022*
Resection margin (≥ R1)	1.388 (0.926–2.080)	0.112			1.585 (1.050–2.394)	0.028*	1.716 (1.127–2.613)	0.012*
Adjuvant chemotherapy	0.712 (0.435–1.166)	0.177			1.632 (0.887–3.002)	0.115		
YAP1 protein expression (High)	1.917 (1.288–2.854)	0.001*	1.870 (1.224–2.855)	0.004*	1.752 (1.178–2.608)	0.006*	1.950 (1.299–2.927)	0.001*

Variables with p ≤ 0.05 were marked with asterisk (*), variables with p ≤ 0.05 in univariable analysis were included in multivariable analysis

AJCC American Joint Committee on Cancer, *BMI* body mass index, *CI* confidence interval, *DFS* disease free survival, *HR* hazard ratio, *N-stage* nodal stage, *OS* overall survival, *T-stage* tumor stage

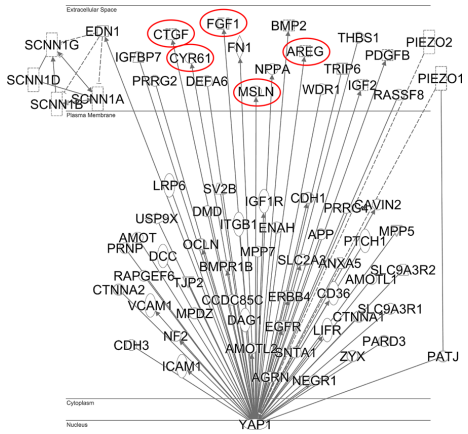


Fig. 4 Ingenuity Pathway Analysis showing the plasma membrane and extracellular proteins directly related to YAP1. The relation to proteins involved in mechanotransduction include the cell membrane protein PATJ (crumbs cell polarity complex component), which is directly related to YAP1 and is also interacting with PIEZO1, the Piezo type mechanosensitive ion channel component 1. YAP1 is also an indirect regulator of both PIEZO1 and PIEZO2. Further, the cytokine endothelin 1 (EDN1) is directly related to YAP1 and is also a regulator of the degenerin/epithelial sodium channels (DEG/ENaC, here marked as SCNN1A, SCNN1B, SCNN1G, SCNN1D). Tight junction signaling proteins related to YAP1 include CTNNA1, MPDZMPP5, OCLN, PATJ, TJP2. Epithelial adherens junction signaling proteins related to YAP1 include CDH1, CTNNA1, CTNNA2, EGFR, FGF1, PARD3, ZYX. Examples of secreted proteins involved in creating a pro-fibrotic microenvironment include AREG, CTGF, CYR61, FGF1, and MSLN and these YAP1 target genes are highlighted and were chosen for further in vitro confirmation

directly related to proteins involved in mechanotransduction, such as PATJ and PIEZO1, and the cytokine EDN1 (Fig. 4). Tight junction signalling proteins related to YAP1 include CTNNA1, MPDZMPP5, OCLN, PATJ, and TJP2, while epithelial adherens junction signaling proteins related to YAP1 include CDH1, CTNNA1, CTNNA2, EGFR, FGF1, PARD3, and ZYX. Examples of secreted proteins involved in creating a pro-fibrotic microenvironment include AREG, CTGF, CYR61, FGF1, and MSLN and these YAP1 target genes were chosen for further in vitro confirmation.

YAP1 protein expression in a patient derived cell line

We performed immunofluorescence based Cellomics to evaluate the protein expression profile of YAP1 in Panc-1 cells. In accordance with the TMA/IHC patient data, a positive YAP1 staining was detected in both nucleus and cytoplasm of Panc-1 cells. The majority of positively stained cells showed a strong fluoresce intensity located in the nucleus (Fig. 5a).

YAP1 participates in the transcription of target genes involved in profibrotic tumor microenvironment

Next, we investigated co-transcriptional activity of YAP1 in synthesis of secreted proteins associated with remodeling of the tumor microenvironment in pancreatic cancer. First, Panc-1 cells were cultured under standard conditions to assess the expression levels of proteins ascertained by the IPA analysis. All investigated proteins, AREG, CTGF, CYR61, FGF1, and MSLN were considered as low abundant and detected in low concentrations (pg/ml) in lysates of Panc-1 cells cultured under standard conditions. As presented in Fig. 5b, the expression levels corresponded to at a maximum 0.2‰ of the total cellular protein amount.

Next, the collected conditioned medium from the Panc-1 cells was analyzed for the presence of selected proteins. AREG, CTGF, CYR61, and MSLN were identified and the secretion pattern was further investigated. Panc-1 cells were subjected to substances inhibiting YAP1 transcriptional activity and the concentrations of the determined secreted proteins were measured. Levels of secreted AREG, CTGF, CYR61, and MSLN were significantly lower ($p = 0.0001$) or undetectable in conditioned medium after the treatment (Fig. 5c). Based on the obtained results, we suggest that YAP1 is involved in the transcription of genes associated with remodeling of the pancreatic tumor microenvironment.

Discussion

In this transcriptome- and proteome-based study, we identified YAP1 as an indicator of poor OS and DFS in patients with pancreatic cancer.

The American Joint Committee on Cancer (AJCC) tumor-node-metastasis (TNM) classification system is currently the gold standard for pancreatic cancer prognostication [23]. However, the AJCC TNM system is only concerned with the anatomical extent of the disease though patients within the same stage may exhibit different outcomes [24]. Such evaluation may lead to either over- or undertreatment. Improved staging systems, considering molecular factors are necessary in order to enhance individual prognostication and utilization of precision therapies.

The prognostic significance of YAP1 protein expression has only been evaluated in one previous small study by Allende et al. [25]. However, YAP1 protein expression did not reach statistical significance in their Kaplan–Meier analysis, likely due to the small cohort size (64 patients). Only when conducting subgroup analyses, stratifying

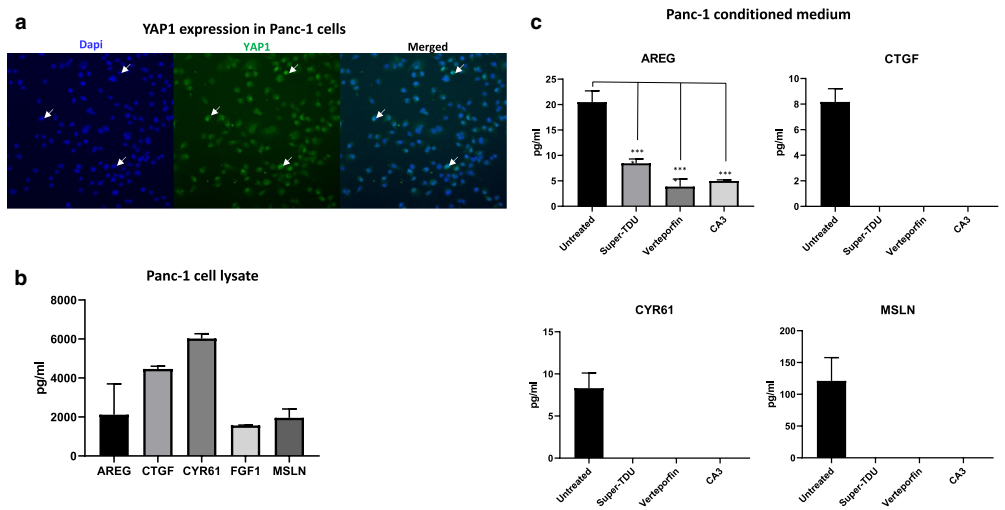


Fig. 5 In vitro analysis of YAP1 and selected target genes in Panc-1 cells. **a** YAP1 protein expression in Panc-1 cells. The image represents an immunofluorescence staining of endogenous YAP1 in Panc-1 cells, plated in 6 well plates and cultivated for 48 h under standard conditions. The arrows indicate an exemplification of YAP1 nuclear accumulation. **b** Concentrations of YAP1 target genes in lysates obtained from Panc-1 cells cultivated under standard conditions. C) Concentrations of YAP1 target genes in conditioned medium obtained from Panc-1 cells that were subjected to maximal tolerable doses (MTD) of substances blocking the YAP1/TEAD interaction

survival into groups of patients surviving more than or less than 30 months, it was shown that patients with high YAP1 expression had worse survival. Therefore, to clarify the prognostic role of YAP1 protein expression in pancreatic cancer, additional studies based on larger cohorts are needed. The TMA/immunohistochemistry analysis based on 140 patients in our study revealed that overexpression of YAP1 is an independent factor for unfavorable outcome and disease recurrence. These findings are in agreement with the public mRNA dataset from the TCGA, which illustrate that high expression of YAP1 significantly correlates with poor survival in pancreatic cancer patients. The agreement between the transcriptome- and proteome-based survival analyses in the present study strengthens the clinical significance of YAP1 as a prognostic variable. However, it is important to note that knowledge about mRNA abundances can only partially predict protein abundances, with a large fraction of the variance also being explained by other factors such as post-transcriptional and translational regulation, as well as protein degradation [26].

To understand the biological role of YAP1 in pancreatic cancer, we performed bioinformatic analyses of protein networks. The results revealed that YAP1 is directly connected to secreted AREG, CTGF, CYR61, FGF1 and MSLN that are involved in fibrosis and other key signaling pathways involved in the tumor-stroma interactions [27–31].

Pancreatic cancer progression is generally associated with a dense fibrotic stroma characterized by an extensive deposition of extracellular matrix components surrounding the cancer cells [32, 33]. The desmoplastic extracellular matrix, mainly produced by activated cancer associated fibroblasts, accounts for up to 80% of entire tumor mass [33]. The fibrotic environment is known to undergo an extensive remodeling connected to the stiffening of tumor tissue. Such stromal reshaping presumably modifies the crosstalk between residual cells within the tumor and directs the tumor progression towards an aggressive phenotype [33–35]. The increased stiffness of matricellular tumor microenvironment also activates YAP1 to further modulate the behavior of cancer cells on the transcriptional level [36, 37].

YAP1 itself, however, lacks DNA-binding activity and requires an interaction with DNA-binding transcription factors such as TEAD to activate target genes [38]. AREG, CTGF and CYR61 account for the most acknowledged target genes for YAP1/TEAD [39–41]. The YAP1/TEAD interactions are also reported to regulate the expression of FGF1 and MSLN [42–44].

We hypothesized that the secreted YAP1/TEAD target gene products contribute to the enhanced fibrotic

reaction and intra-tumoral stiffening which consecutively promote YAP1 transcriptional activity. Such paracrine loop would further affect the tumor microenvironment and maintain the aggressive course of the disease.

Using the patient derived pancreatic cancer cell line Panc-1, we evaluated the effect of substances designed to inhibit the YAP1/TEAD mediated gene transcription. We showed that the disruption of YAP1/TEAD complex significantly reduced the presence of the selected YAP1/TEAD target gene products in the conditioned medium. Suppression of YAP1 oncogenic activity with a subsequent modification of the tumor microenvironment may thus be an advantageous approach to control tumor growth and improve prognosis. Although the clinical utilization for such treatment remains to be determined, YAP1 as a biomarker may aid in the individual prognostication of patients diagnosed with pancreatic cancer and the selection of precision therapy.

Conclusions

We demonstrate that YAP1 is an independent prognostic marker associated with recurrence and unfavorable survival in pancreatic cancer. We also show that inhibition of YAP1/TEAD interaction interferes with the expression of AREG, CTGF, CYR61, and MSLN suggesting that YAP1 transcriptional activity may affect the development and persistence of a fibrotic tumor microenvironment. YAP1 is thus considered as a clinically and biologically relevant biomarker derived from pancreatic cancer tissue.

Abbreviations

AJCC: American Joint Committee on Cancer; AREG: Amphiregulin; BMI: Body mass index; CI: Confidence interval; CTGF: Connective tissue growth factor; CYR61: Cysteine-rich angiogenic inducer 61; DFS: Disease-free survival; ELISA: Enzyme-linked immunosorbent assay; FFPE: Formalin-fixed paraffin-embedded; FGF1: Fibroblast growth factor 1; FPKM: Fragments per kilobase of exon per million reads; HR: Hazard ratio; IHC: Immunohistochemistry; IPA: Ingenuity Pathway Analysis; MS: Mass spectrometry; MSLN: Mesothelin; OS: Overall survival; TCGA: The Cancer Genome Atlas; TMA: Tissue microarray; TNM: Tumor-node-metastasis; YAP1: Yes-associated protein 1.

Acknowledgements

The mRNA results published here are based upon data generated by the TCGA Research Network (https://www.cancer.gov/tcga) and the Human Protein Atlas program (http://www.proteinatlas.org/pathology). We thank Indira Pla, Aniel Sanchez Puente, Jeovanis Gil valdes and Lazaro Hiram Betancourt for technical support with the proteomics analysis in this manuscript.

Authors' contributions

QZ and DA conceived the original idea and designed the study with GMV and RA. QZ, MB, JX, AS, DH, HD, XC, KSH and DA collected the data for the study, which were analyzed by QZ, MB and KP. The data interpretation and manuscript drafting were performed by QZ, MB and DA. The manuscript was revised by GMV and RA. All authors reviewed the manuscript. All authors read and approved the final manuscript.

Funding

Open access funding provided by Lund University. This work was supported by the Magnus Bergvall Foundation, the Royal Physiographic Society of Lund, the Tore Nilsson Foundation, the Inga and John Hain Foundation for Medical Research, the Clas Groschinsky Foundation, the Gunnar Nilsson Foundation, the Gyllenstiernska Krapperup Foundation, the Bengt Ihre Foundation, the Emil and Wera Cornell Foundation, the Crafoord Foundation, Governmental Funding of Clinical Research within the National Health Service (ALF) and Sweden´s Innovation Agency (Vinnova).

Availability of data and materials

The datasets generated and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

This study was performed in compliance with the Helsinki Declaration on ethical principles for handling human tissue specimens, with all EU and national regulations and requirements. Written informed consent was obtained from participants. Ethical permission for the study was granted by the Ethics Committee at Lund University (Ref 2010/684, 2012/661, 2015/266, 2017/320).

Consent for publication

Consent for publication was obtained from included participants.

Competing interests

The authors declare that they have no competing interests.

Author details

[1] The Eye Hospital, School of Ophthalmology and Optometry, Wenzhou Medical University, Wenzhou, Zhejiang, China. [2] Department of Surgery, Clinical Sciences Lund, Lund University and Skåne University Hospital, 221 85 Lund, Sweden. [3] Clinical Protein Science and Imaging, Biomedical Centre, Department of Biomedical Engineering, Lund University, Lund, Sweden. [4] Department of Surgery, The Second Affiliated Hospital and Yuying Children's Hospital of Wenzhou Medical University, Wenzhou, Zhejiang, China. [5] Department of Pathology, Skåne University Hospital, Lund, Sweden. [6] Department of Pathology, The First Affiliated Hospital of Nanchang University, Nanchang, China. [7] Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences, Warsaw, Poland. [8] Department of Translational Medicine, Lund University, Malmö, Sweden.

Received: 10 September 2019 Accepted: 1 February 2020
Published online: 13 February 2020

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019;69:7–34.
2. Rahib L, Smith BD, Aizenberg R, Rosenzweig AB, Fleshman JM, Matrisian LM. Projecting cancer incidence and deaths to 2030: the unexpected burden of thyroid, liver, and pancreas cancers in the United States. Cancer Res. 2014;74:2913–21.
3. Jones S, Zhang X, Parsons DW, Lin JC, Leary RJ, Angenendt P, Mankoo P, Carter H, Kamiyama H, Jimeno A, et al. Core signaling pathways in human pancreatic cancers revealed by global genomic analyses. Science. 2008;321:1801–6.
4. Collisson EA, Sadanandam A, Olson P, Gibb WJ, Truitt M, Gu S, Cooc J, Weinkle J, Kim GE, Jakkula L, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med. 2011;17:500–3.
5. Bailey P, Chang DK, Nones K, Johns AL, Patch AM, Gingras MC, Miller DK, Christ AN, Bruxner TJ, Quinn MC, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature. 2016;531:47–52.
6. Moffitt RA, Marayati R, Flate EL, Volmar KE, Loeza SG, Hoadley KA, Rashid NU, Williams LA, Eaton SC, Chung AH, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet. 2015;47:1168–78.
7. Wilhelm M, Schlegl J, Hahne H, Gholami AM, Lieberenz M, Savitski MM, Ziegler E, Butzmann L, Gessulat S, Marx H, et al. Mass-spectrometry-based draft of the human proteome. Nature. 2014;509:582–7.
8. Zhou Q, Andersson R, Hu D, Bauden M, Sasor A, Bygott T, Pawłowski K, Pla I, Marko-Varga G, Ansari D. Alpha-1-acid glycoprotein 1 is upregulated in

pancreatic ductal adenocarcinoma and confers a poor prognosis. Transl Res. 2019;212:67–79.
9. Melo SA, Luecke LB, Kahlert C, Fernandez AF, Gammon ST, Kaye J, LeBleu VS, Mittendorf EA, Weitz J, Rahbari N, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. Nature. 2015;523:177–82.
10. Zhou Q, Andersson R, Hu D, Bauden M, Kristl T, Sasor A, Pawlowski K, Pla I, Hilmersson KS, Zhou M, et al. Quantitative proteomics identifies brain acid soluble protein 1 (BASP1) as a prognostic biomarker candidate in pancreatic cancer tissue. EBioMedicine. 2019;43:282–94.
11. Hao Y, Chun A, Cheung K, Rashidi B, Yang X. Tumor suppressor LATS1 is a negative regulator of oncogene YAP. J Biol Chem. 2008;283:5496–509.
12. Maugeri-Sacca M, De Maria R. The Hippo pathway in normal development and cancer. Pharmacol Ther. 2018;186:60–72.
13. Stein C, Bardet AF, Roma G, Bergling S, Clay I, Ruchti A, Agarinis C, Schmelzle T, Bouwmeester T, Schubeler D, Bauer A. YAP1 exerts its transcriptional control via TEAD-mediated activation of enhancers. PLoS Genet. 2015;11:e1005465.
14. Guo L, Chen Y, Luo J, Zheng J, Shao G. YAP1 overexpression is associated with poor prognosis of breast cancer patients and induces breast cancer cell growth by inhibiting PTEN. FEBS Open Bio. 2019;9:437–45.
15. Werneburg N, Gores GJ, Smoot RL. The Hippo pathway and YAP signaling: emerging concepts in regulation, signaling, and experimental targeting strategies with implications for hepatobiliary malignancies. Gene Expr. 2019. https://doi.org/10.3727/105221619X15617324583639.
16. McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM. Statistics Subcommittee of the NCIEWGoCD: REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer. 2005;93:387–91.
17. Uhlen M, Zhang C, Lee S, Sjostedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu Z, Edfors F, et al. A pathology atlas of the human cancer transcriptome. Science. 2017;357:eaan2507.
18. The Human Protein Atlas. http://www.proteinatlas.org/pathology. Accessed 20 Jan 2020.
19. The Cancer Genome Atlas Program. https://www.cancer.gov/tcga. Accessed 20 Jan 2020.
20. Hu D, Ansari D, Zhou Q, Sasor A, Hilmersson KS, Bauden M, Jiang Y, Andersson R. Calcium-activated chloride channel regulator 1 as a prognostic biomarker in pancreatic ductal adenocarcinoma. BMC Cancer. 2018;18:1096.
21. Ishibashi H, Suzuki T, Suzuki S, Moriya T, Kaneko C, Takizawa T, Sunamori M, Handa M, Kondo T, Sasano H. Sex steroid hormone receptors in human thymoma. J Clin Endocrinol Metab. 2003;88:2309–17.
22. John T, Liu G, Tsao MS. Overview of molecular testing in non-small-cell lung cancer: mutational analysis, gene copy number, protein expression and other biomarkers of EGFR for the prediction of response to tyrosine kinase inhibitors. Oncogene. 2009;28(Suppl 1):S14–23.
23. Amin MB, Edge SB, Greene FL, et al., editors. AJCC cancer staging manual. 8th ed. New York: Springer; 2017.
24. Helm J, Centeno BA, Coppola D, Melis M, Lloyd M, Park JY, Chen DT, Malafa MP. Histologic characteristics enhance predictive value of American Joint Committee on Cancer staging in resectable pancreas cancer. Cancer. 2009;115:4080–9.
25. Salcedo Allende MT, Zeron-Medina J, Hernandez J, Macarulla T, Balsells J, Merino X, Allende H, Tabernero J, Ramon YCS. Overexpression of yes associated protein 1, an independent prognostic marker in patients with pancreatic ductal adenocarcinoma, correlated with liver metastasis and poor prognosis. Pancreas. 2017;46:913–20.
26. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 2012;13:227–32.
27. Argani P, Iacobuzio-Donahue C, Ryu B, Rosty C, Goggins M, Wilentz RE, Murugesan SR, Leach SD, Jaffee E, Yeo CJ, et al. Mesothelin is overexpressed in the vast majority of ductal adenocarcinomas of the pancreas: identification of a new pancreatic cancer marker by serial analysis of gene expression (SAGE). Clin Cancer Res. 2001;7:3862–8.
28. Chu CY, Chang CC, Prakash E, Kuo ML. Connective tissue growth factor (CTGF) and cancer progression. J Biomed Sci. 2008;15:675–85.
29. Ding L, Liu T, Wu Z, Hu B, Nakashima T, Ullenbruch M, Gonzalez De Los Santos F, Phan SH. Bone marrow CD11c+ cell-derived amphiregulin promotes pulmonary fibrosis. J Immunol. 2016;197:303–12.
30. Kurundkar AR, Kurundkar D, Rangarajan S, Locy ML, Zhou Y, Liu RM, Zmijewski J, Thannickal VJ. The matricellular protein CCN1 enhances TGF-beta1/SMAD3-dependent profibrotic signaling in fibroblasts and contributes to fibrogenic responses to lung injury. FASEB J. 2016;30:2135–50.
31. Li JT, Liao ZX, Ping J, Xu D, Wang H. Molecular mechanism of hepatic stellate cell activation and antifibrotic therapeutic strategies. J Gastroenterol. 2008;43:419–28.
32. Erkan M, Hausmann S, Michalski CW, Fingerle AA, Dobritz M, Kleeff J, Friess H. The role of stroma in pancreatic cancer: diagnostic and therapeutic implications. Nat Rev Gastroenterol Hepatol. 2012;9:454–67.
33. Neesse A, Algul H, Tuveson DA, Gress TM. Stromal biology and therapy in pancreatic cancer: a changing paradigm. Gut. 2015;64:1476–84.
34. Laklai H, Miroshnikova YA, Pickup MW, Collisson EA, Kim GE, Barrett AS, Hill RC, Lakins JN, Schlaepfer DD, Mouw JK, et al. Genotype tunes pancreatic ductal adenocarcinoma tissue tension to induce matricellular fibrosis and tumor progression. Nat Med. 2016;22:497–505.
35. Zhan HX, Zhou B, Cheng YG, Xu JW, Wang L, Zhang GY, Hu SY. Crosstalk between stromal cells and cancer cells in pancreatic cancer: new insights into stromal biology. Cancer Lett. 2017;392:83–93.
36. Dupont S. Role of YAP/TAZ in cell-matrix adhesion-mediated signalling and mechanotransduction. Exp Cell Res. 2016;343:42–53.
37. Lee J, Condello S, Yakubov B, Emerson R, Caperell-Grant A, Hitomi K, Xie J, Matei D. Tissue transglutaminase mediated tumor-stroma interaction promotes pancreatic cancer progression. Clin Cancer Res. 2015;21:4482–93.
38. Kim MK, Jang JW, Bae SC. DNA binding partners of YAP/TAZ. BMB Rep. 2018;51:126–33.
39. Choi HJ, Zhang H, Park H, Choi KS, Lee HW, Agrawal V, Kim YM, Kwon YG. Yes-associated protein regulates endothelial cell contact-mediated expression of angiopoietin-2. Nat Commun. 2015;6:6943.
40. Jia J, Li C, Yang J, Wang X, Li R, Luo S, Li Z, Liu J, Liu Z, Zheng Y. Yes-associated protein promotes the abnormal proliferation of psoriatic keratinocytes via an amphiregulin dependent pathway. Sci Rep. 2018;8:14513.
41. Zhao B, Ye X, Yu J, Li L, Li W, Li S, Yu J, Lin JD, Wang CY, Chinnaiyan AM, et al. TEAD mediates YAP-dependent gene induction and growth control. Genes Dev. 2008;22:1962–71.
42. Hucl T, Brody JR, Gallmeier E, Iacobuzio-Donahue CA, Farrance IK, Kern SE. High cancer-specific expression of mesothelin (MSLN) is attributable to an upstream enhancer containing a transcription enhancer factor dependent MCAT motif. Cancer Res. 2007;67:9055–65.
43. Koontz LM, Liu-Chittenden Y, Yin F, Zheng Y, Yu J, Huang B, Chen Q, Wu S, Pan D. The Hippo effector Yorkie controls normal tissue growth by antagonizing scalloped-mediated default repression. Dev Cell. 2013;25:388–401.
44. Ren YR, Patel K, Paun BC, Kern SE. Structural analysis of the cancer-specific promoter in mesothelin and in other genes overexpressed in cancers. J Biol Chem. 2011;286:11960–9.

Publisher's Note

Paper IV

ORIGINAL ARTICLE

Alpha-1-acid glycoprotein 1 is upregulated in pancreatic ductal adenocarcinoma and confers a poor prognosis

QIMIN ZHOU, ROLAND ANDERSSON, DINGYUAN HU, MONIKA BAUDEN, AGATA SASOR, THOMAS BYGOTT, KRZYSZTOF PAWŁOWSKI, INDIRA PLA, GYÖRGY MARKO-VARGA, and DANIEL ANSARI

LUND, AND MALMÖ, SWEDEN; WENZHOU, CHINA; CAMBRIDGE, UK; AND WARSAW, POLAND

Pancreatic cancer is an aggressive malignancy that carries a high mortality rate. A major contributor to the poor outcome is the lack of effective molecular markers. The purpose of this study was to develop protein markers for improved prognostication and noninvasive diagnosis. A mass spectrometry (MS)-based discovery approach was applied to pancreatic cancer tissues and healthy pancreas. In the verification phase, extracellular proteins with differential expression were further quantified in targeted mode using parallel reaction monitoring (PRM). Next, a tissue microarray (TMA) cohort including 140 pancreatic cancer resection specimens was constructed, in order to validate protein expression status and investigate potential prognostic implications. The levels of protein candidates were finally assessed in a prospective series of 110 serum samples in an accredited clinical laboratory using the automated Cobas system. Protein sequencing with nanoliquid chromatography tandem MS (nano-LC-MS/MS) and targeted PRM identified alpha-1-acid glycoprotein 1 (AGP1) as an upregulated protein in pancreatic cancer tissue. Using TMA and immunohistochemistry, AGP1 expression was significantly associated with shorter overall survival (HR = 2.22; 95% CI 1.30–3.79, *P* = 0.004). Multivariable analysis confirmed the results (HR = 1.87; 95% CI 1.08–3.24, *P* = 0.026). Circulating levels of AGP1 yielded an area under the curve (AUC) of 0.837 for the discrimination of resectable pancreatic cancer from healthy controls. Combining AGP1 with CA 19-9 enhanced the diagnostic performance, with an AUC of 0.963. This study suggests that AGP1 is a novel prognostic biomarker in pancreatic cancer tissue. Serum AGP1 levels may be useful as part of a biomarker panel for early detection of pancreatic cancer but further studies are needed. (Translational Research 2019; 212:67–79)

From the Department of Surgery, Clinical Sciences Lund, Lund University, Skåne University Hospital, Lund, Sweden; School of Ophthalmology and Optometry, Eye Hospital, Wenzhou Medical University, Wenzhou, China; Department of Clinical Genetics and Pathology, Labmedicin Skåne, Lund, Sweden; Cambridge Personal Genomics, Cambridge, UK; Department of Experimental Design and Bioinformatics, Warsaw University of Life Sciences, Warsaw, Poland; Department of Translational Medicine, Lund University, Malmö, Sweden; Clinical Protein Science and Imaging, Biomedical Centre, Department of Biomedical Engineering, Lund University, Lund, Sweden.

Submitted for Publication January 20, 2019; received submitted June 22, 2019; accepted for publication June 26, 2019.

Reprint requests: Daniel Ansari, Department of Surgery, Clinical Sciences Lund, Lund University and Skåne University Hospital, SE-221 85 Lund, Sweden. E-mail address: daniel.ansari@med.lu.se.

1931-5244/$ - see front matter

https://doi.org/10.1016/j.trsl.2019.06.003

Abbreviations: AGP1 = alpha-1-acid glycoprotein; AUC = area under the curve; CA 19-9 = carbohydrate antigen 19-9; DDA = data-dependent analysis; FFPE = formalin-fixed paraffin-embedded; IPA = ingenuity pathway analysis; IPMN = intraductal papillary mucinous neoplasm; MS = mass spectrometry; nano-LC-MS/MS = nano-liquid chromatography tandem MS; PRM = parallel reaction monitoring (PRM); ROC = receiver operating characteristic; TMA = tissue microarray

AT A GLANCE COMMENTARY

Zhou Q, et al.

Background

Pancreatic adenocarcinoma is often detected at late stages and the prognosis is poor. Molecular markers are needed for early diagnosis, tumor stratification, and treatment selection. Using proteomic technology and bioinformatic tools, we provide an integrative approach from initial biomarker discovery, to targeted verification and subsequent clinical validation. The results show that alpha-1-acid glycoprotein 1 (AGP1) is an independent prognostic biomarker in pancreatic cancer tissue. Furthermore, circulating levels of AGP1 together with CA 19-9 provide high accuracy (AUC 0.96) for diagnosing operable pancreatic cancer against healthy controls.

Translational Significance

This is the first study describing the prognostic role of AGP1 in pancreatic cancer tissue. Additional data are needed to support whether AGP1 can be included in a multiplex serum panel for early diagnosis of pancreatic cancer. The biomarker workflow presented in this paper may be applied also in other tumor forms.

INTRODUCTION

Pancreatic cancer is the most lethal human malignancy with a median survival of only 4.6 months.[1] Due to the late presentation, molecular heterogeneity, and aggressive tumor biology, the development of effective diagnostic and therapeutic strategies for pancreatic cancer remains elusive.[2-5] Pancreatectomy is still the single most effective treatment modality. Unfortunately, only 15%−20% of patients will present with a tumor that is deemed operable by traditional criteria. With state-of-the-art therapy with surgery and chemotherapy, a median survival ranging from 25 up to 54 months is achievable.[6,7]

Serum carbohydrate antigen 19-9 (CA 19-9) is the sole biomarker for pancreatic cancer that is in routine clinical use. Serial measurement of CA 19-9 levels is useful to monitor patients after potentially curative surgery or for those who are receiving chemotherapy for advanced disease. However, CA 19-9 is not recommended as a screening tool due to inadequate sensitivity and specificity.[8] Hence, there is an unmet need for new molecular markers that can aid in diagnosis, as well as prognostication and potential therapy selection.

Gene expression analyses of pancreatic tumor tissues have shown that pancreatic cancer can be subgrouped into clusters expressing different genetic configurations and outcome.[9-11] However, a deeper understanding of the underlying molecular pathophysiology of the disease is needed to advance the development of effective early detection and therapeutic strategies. Proteomics refers to the systematic analysis of protein profiles expressed by the genome. This makes proteomics an essential tool not only for understanding disease pathology, but also for identifying diagnostic and prognostic markers. In this context, investigating the extracellular proteome of tumor tissue may be of particular interest. Proteins secreted by tumor cells into the extracellular space may potentially function as tissue markers, as well as noninvasive diagnostic markers detectable in blood as liquid biopsies.

Mass spectrometry (MS) has become the method of choice for proteomic studies in biomarker research.[12,13] Liquid chromatography tandem MS (LC−MS/MS) can be used to identify up to thousands of proteins in clinical samples. However, markers derived from such studies need to be verified and validated before they can be introduced in clinical practice. Parallel reaction monitoring (PRM) is an MS technique recently adapted to the field of proteomics.[14,15] PRM offers the opportunity to verify multiple biomarker candidates simultaneously, once the target proteins are known. Thus, a strategy combining LC−MS/MS for a discovery phase and PRM for targeted verification of discovered proteins might be effective for the development of reliable biomarker candidates.

In this study, we applied an MS-based approach for both discovery and targeted verification of proteins in pancreatic cancer biomarker research, with subsequent validation in a large cohort using an orthogonal technique in tissue and serum.

MATERIALS AND METHODS

Patients and samples. Primary operable, nonpretreated tissue specimens were included from 140 patients with pancreatic ductal adenocarcinoma who were subjected

to surgery between 1996 and 2017 at the Department of Surgery, Skåne University Hospital, Sweden. Serum samples were prospectively collected from 110 individuals, including 52 patients with resectable pancreatic cancer, 24 patients with benign pancreatic disease, and 34 healthy controls between 2012 and 2017 at the same institution. Patient sera were obtained at diagnosis, prior to surgical treatment. Healthy control sera were obtained from donors at the local blood donation center. Blood samples were collected in BD SST II Advance tubes (product no. 368498; Becton Dickinson, Franklin Lakes, NJ). The minimum clotting time was 30 minutes. The samples were centrifuged at 2000 × g for 10 minutes at 25°C, and serum was collected and stored at −80°C until further analysis.

Tumors were classified according to the revised 8th edition of the AJCC/UICC tumor-node-metastasis (TNM) staging system.[16] Ten fresh frozen healthy pancreas biopsies were obtained from organ donors and acquired through the Lund University Diabetes Centre (LUDC), a part of the national consortium Excellence of Diabetes Research in Sweden. As additional controls, histologically normal pancreas tissues were obtained from 10 patients who underwent surgical resection for benign pancreatic lesions, including 7 serous cystadenomas, 1 mucinous cystadenoma, 1 pancreatic pseudotumor, and 1 thrombosed splenic artery aneurysm. The REMARK[17] and STARD guidelines[18] for biomarker research were followed where possible. Written informed consent was obtained from the study participants. The study was performed in accordance with the Declaration of Helsinki. Ethical approval was granted by the local ethics committee at Lund University (Ref 2010/684, 2012/661, 2015/266, 2017/320).

Discovery study. ***Tissue processing and MS.*** The individual fresh frozen tissue specimens were pulverized in liquid N_2 and crudely homogenized in extraction buffer containing 500 mM Tris-Cl, [pH 8] and 6 M guanidine-HCl in 50 mM AMBIC and digestive enzyme inhibitors, as previously reported.[19] The protein extraction was accelerated by cellular disruption applying four freeze-thaw cycles and ultrasonic bath treatment. Further, the soluble proteins were reduced with 15 mM DTT, alkylated using 50 mM IAA, precipitated with ice cold 99.5% ethanol and digested using Mass Spec Grade Trypsin/Lys-C Mix (Promega, Madison, WI). The concentration of obtained peptides was measured using the Pierce quantitative colorimetric peptide assay. Thermo Scientific Pierce Peptide Retention Time Calibration Mixture consisting of 15 peptides was finally added to each sample to enable a normalization and control of the chromatographic performance.

Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis was performed using high-performance liquid chromatography system, EASY-nLC 1000, coupled to Q Exactive quadrupole Orbitrap mass spectrometer equipped with a nanospray ion source. Samples containing 1 μg peptide mixture were individually injected at a flow rate of 300 nL/min and separated with a 132 minute gradient of 5%−22% acetonitrile (ACN) in 0.1% formic acid (FA), followed by an 18 minute gradient of 22%−38% ACN in 0.1 % FA. For the separation, a two-column setup was used, including the EASY-Spray analytical column (25 cm × 75 μm ID, particle size 2 μm, pore size 100 Å, PepMap C18) and the Acclaim precolumn (2cm × 75 μm ID, particle size 3 μm, pore size 100 Å, PepMap C18). Each sample was measured in duplicate in a random order. The raw files acquired from the two measurements were combined and evaluated using Proteome Discoverer targeting high confident peptides only.

The Orbitrap instrument was operated in the positive data-dependent acquisition mode to automatically shift between the full scan MS and MS/MS acquisition. For the peptide identification, a full MS survey scan was performed in the Orbitrap detector. Fifteen data-dependent higher energy collision dissociation MS/MS scans were performed on the most intense precursors. The MS scans with a resolution of 70 000 were collected using an automatic gain control (AGC) target value of 1×10^6 with a maximum injecting time of 100 ms over a mass range of 400.0 and 1 600.0 m/z. The resolution of the data dependent MS/MS scans was fixed at 17,500. The data were collected using the AGC target value of 5×10^5 with maximum injection time of 80 ms. The normalized collision energy was set at 27.0% for all scans.

MS data analysis. The acquired MS/MS raw data files acquired from the combined randomized measurements were processed with Proteome Discoverer software, Version1.4 (Thermo Fisher), to identify and quantify the extracted tissue proteins. The selection of spectra was based on the following settings: min precursor mass 350 Da; max precursor mass 5 000 Da; s/n threshold 1.5. Parameters for Sequest HT database searches included the following: precursor mass tolerance 10 ppm; product mass tolerance 0.02 Da; trypsin as enzyme; 1 missed cleavage site was allowed; UniProt human database; variable modifications: acetyl (K), methyl (K,R), dimethyl (K,R), trimethyl (K,R), glygly (K) and oxidation (M,P) static modification: carbamidomethyl (C). The false discovery rate was set to 0.01. The precursor ions area detector was applied in the search engine for the quantification of peptides. Extracellular proteins were annotated manually using the European Bioinformatics Institute (EMBL-EBI), UniProt Gene Ontology

Annotation database, and QuickGO software (GO: 0005576) and extracted for targeted analysis.

Targeted study. The PRM analysis was conducted to verify the differentially expressed proteins. One or two unique peptides of each targeted protein were selected from the MS spectra acquired from the discovery measurements and a spectral library was created. The individual samples were analyzed using the RPM time-scheduled acquisition method with a retention time +/− 5 minutes and a resolution of 35,000 (AGC target 5×10^5, 50 ms maximum injection time). The chromatographic peak width is 30 seconds. The normalized collision energy was 26 and the isolation window 2 m/z. Skyline software was used for a relative quantification in the PRM study.

Ingenuity pathway analysis. Differentially expressed proteins were submitted to functional analysis with Ingenuity Pathway Analysis (IPA).[20] The software generated causal networks based on more than 40,000 nodes representing mammalian genes and their products (transcripts, proteins, miRNAs) as well as 1,480,000 interactions between them. IPA enables the identification of relationships between proteins in the data set and enrichment of canonical pathways.

Tissue microarray and immunohistochemistry. Tissue microarrays (TMAs) were constructed using an automated tissue arraying device (Minicore 3, Alphelys, Plaisir, France). A standard set of four 2-mm tissue cores was acquired from each of the 140 formalin-fixed, paraffin-embedded tumors and mounted in a new recipient block. For immunohistochemical analysis, 3-μm TMA sections attached on individual glass slides were pretreated using the automated PT Link system (Dako, Agilent Technologies, Glostrup, Denmark) for 20 minutes at 97°C in EnVisionFlex retrieval solution, low pH (Dako, Agilent Technologies). The sections were incubated with a polyclonal antibody targeting alpha-1-acid glycoprotein 1 (AGP1, Atlas Antibodies, dilution 1: 50) followed by a biotinylated secondary antibody (Vector Laboratories, Burlingame, CA BA-1000, dilution 1:200). Thereafter, avidin-biotin-peroxidase complex (Vectastain Elite ABC-HRP Kit, Vector Laboratories) was applied to amplify the signal. The antibody−antigen complex was visualized with chromogen diaminobenzidine (Vector Laboratories). Nuclei were counterstained with Mayer's hematoxylin (Histolab, Gothenburg, Sweden) for 30 seconds. Finally, the sections were dehydrated in graded alcohol, cleared in xylene, and mounted using Pertex (Histolab). TMA slides were scanned for evaluation using an Aperio scanscope scanner (Leica Biosystems, Wetzlar, Germany).

The expression of AGP1 was denoted as negative (0), weak (1), moderate (2) and strong (3). Tumors showing >10% stained cancer cells were considered positive. Staining of AGP1 was evaluated by two independent observers (A. Sasor and M. Bauden) who were blinded to clinical and outcome data.

Serum analysis. AGP1 levels in serum samples were measured with an immunoturbidimetric method in an accredited clinical laboratory (Department of Clinical Chemistry and Pharmacology, University and Regional Laboratories Region Skåne, Sweden). The c501 module of the Cobas 6000 analyzer (Roche diagnostics, Mannheim, Germany) was used according to the IFCC-IUPAC-coding system NPU19873. The formed AGP1−antibody complex was measured bichromatically at 340 and 660 nm. The standardized method was calibrated with the CFAS protein calibrator (Roche Diagnostics), traceable to the international protein calibrator CRM 470. The detection range was defined between 0.1 g/L and 6.0 g/L and the total coefficient of variation (CV) for the imprecision of the measurement was 4.9% at 0.37 g/L and 3.6% at 0.95 g/L.

CA19-9 concentrations were obtained from immunometric sandwich analysis using an electrochemiluminescence immunoassay detection technique based on Ruthenium (Ru) derivate. The measurements were carried out on a Cobas immunoanalyzer according to the IFCC-IUPAC-coding system NPU01450 in an accredited clinical laboratory (Department of Clinical Chemistry and Pharmacology, University and Regional Laboratories Region Skåne, Sweden). The method was standardized as stated in the CA19-9 immunoassay reference CA19-9 Cobas, 11776193 122, 2016-11, V23 (Roche diagnostics). The detection range was defined between 0.6 kU/Land 10 000 kU/L and the total coefficient of variation (CV) for the imprecision of the measurement was 5% at 20 kU/L and 4% at 80 kU/L.

Statistical analysis. The MS intensities were log2 transformed and normalized by subtracting the median intensity of all the proteins per sample. The proteins detected in less than 5 samples in respective group were excluded. Missing values were replaced from a normal distribution by using following settings width: 0.3 and down shift 0. A Two-Sample test (Student's t test) was applied to define the differentially expressed proteins. The settings included S0:2 on both sides and false discovery rate 0.01.

For the immunohistochemistry analysis, AGP1 expression was dichotomized into negative or positive. The Wilcoxon-Mann-Whitney test and $\chi 2$ test were used for comparison of AGP1 expression and relevant clinicopathologic characteristics. Kaplan-Meier analysis and the log-rank test were used to illustrate differences in overall survival according to AGP1 expression. Cox regression proportional hazards models were used for estimation of hazard ratios (HRs) for death according to AGP1 expression in both uni- and multivariable analysis, adjusted for age, gender, TNM status, differentiation grade, resection margin status, and adjuvant chemotherapy.

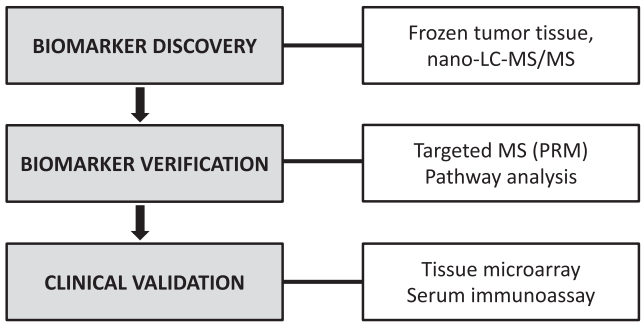


Fig 1. Scheme of the discovery-to-validation pipeline to identify potential protein biomarkers for pancreatic cancer. MS, mass spectrometry; nano-LC-MS/MS, nano-liquid chromatography tandem MS; PRM, parallel reaction monitoring (PRM).

For the immunoassay, differences in serum levels of AGP1 and CA 19-9 between groups were assessed by the Wilcoxon rank-sum (Mann-Whitney) test. Potential correlations between AGP1 and CA 19-9 levels were tested with the Spearman rank order correlation. Logistic Regression (LR) was conducted, using the R function lrm, from the package rms.[21] Receiver operator characteristic (ROC) curves, areas under the curve (AUCs), and sensitivities/specificities were calculated for discrimination between cancer, benign pancreatic disease, and healthy controls using the R package pROC.[22] To find the optimal cut-off concentrations for AGP1 in the detection of pancreatic cancer from healthy and benign controls, Youden's index (J) was employed and calculated as J = sensitivity + specificity − 1.[23]

Statistical analyses were performed using Perseus software version 1.6.0.7,[24] GraphPad Prism 8, STATA/MP 14.2, SPSS Statistics 25, and R programming language version 3.5.0.[25]

RESULTS

Fig 1 shows the methodological workflow used in this study. Mass spectrometry was used for discovery and targeted verification of protein biomarkers for pancreatic cancer. The validation phase was conducted in a larger cohort with antibody-based technology in tissue and serum. Demographic and clinical characteristics are shown in Table 1.

Mass spectrometry-based identification of differentially expressed proteins. In this biomarker discovery phase, 10 representative pancreatic cancer tissues and 10 healthy donor pancreas tissues were analyzed with LC-MS/MS. After stringent filtering, the LC-MS/MS analyses resulted in the identification of 4138 proteins and 28 781 peptides. A total of 165 proteins were found to be differentially expressed comparing pancreatic

Table 1. Patient and tumor characteristics

	No. of patients*
(A) Tissue cohort, pancreatic cancer	**N = 140**
Age (years)†	69 (63-73)
Female sex	67 (47.9)
Tumor category	
T1	19 (13.6)
T2	94 (67.1)
T3	26 (18.6)
T4	1 (0.7)
Node status	
0	33 (23.6)
1	54 (38.6)
2	52 (37.1)
Unknown	1 (0.7)
Poorly differentiated/anaplastic	84 (60)
R1 resection	56 (40)
Adjuvant chemotherapy	113 (80.7)
(B) Serum cohort	**N = 110**
Pancreatic cancer	N = 52
Age (years)†	69 (64-73)
Female sex	23 (44.2)
Tumor category	
T1	2 (3.9)
T2	26 (50)
T3	24 (46.2)
Node status	
0	7 (17.5)
1	11 (27.5)
2	22 (55)
Benign pancreatic disease	N = 24
Age (years)†	69 (61-74)
Female sex	9 (37.5)
Histopathology	
Chronic pancreatitis	11
Serous cystadenoma	6
IPMN, low-grade dysplasia	7
Healthy controls	N = 34
Age (years)†	61 (56-63)
Female sex	12 (35.3)

*With percentages in parenthesis unless indicated otherwise; †values are median (i.q.r.). IPMN, intraductal papillary mucinous neoplasm.

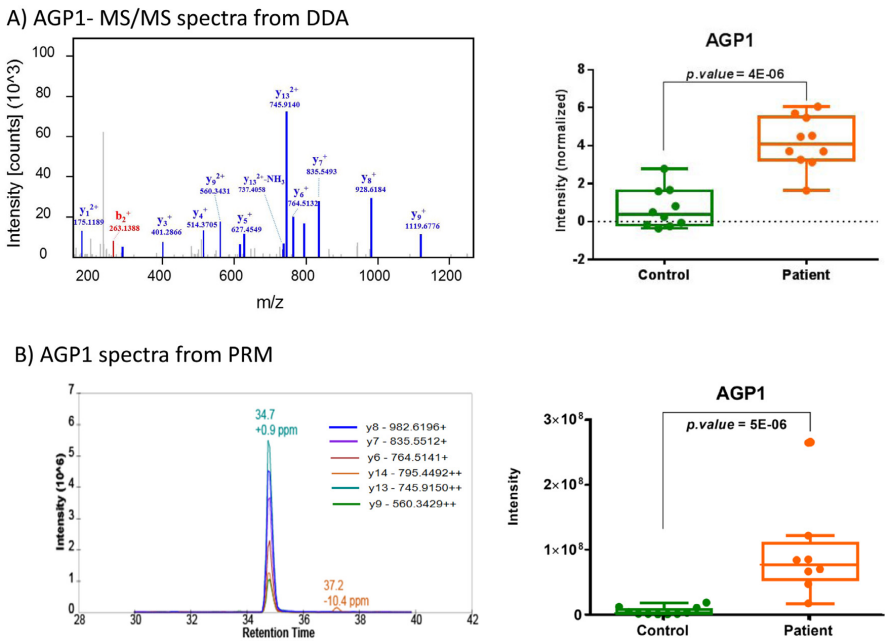


Fig 2. Proteomic discovery and development of a targeted PRM assay. (A) MS/MS spectrum is used in the discovery mode for the confident identification of the AGP1 peptide (YVGGQEHFAHLLILR). MS spectra of AGP1 at label-free quantitative MS discovery phase (left), box-plot showing relative expression levels of AGP1 in pancreatic cancer patients and matched controls (right). (B) The fragment ions with highest intensities in the MS/MS spectrum are used for PRM quantification. PRM transitions for targeted verification of AGP1 (left), box-plot showing relative expression levels of AGP1 in pancreatic cancer patients and matched controls (right). AGP1, alpha-1-acid glycoprotein 1; DDA, data-dependent analysis; MS, mass spectrometry; PRM, parallel reaction monitoring.

cancer and healthy pancreas, including 52 extracellular proteins that were selected for PRM, as reported previously by us.[19]

Verification of differentially expressed extracellular proteins. A total of 16 extracellular proteins could be verified with targeted label-free PRM: S100A6, TF, FBLN1, HYOU1, PNLIP, P4HB, AHSG, PLA2G1B, AGP1, PRSS1, PRSS2, APOA1, ALB, SERPINA1, CLPS and COL14A1, confirming differential expression between pancreatic cancer and healthy pancreas. Based on quantification of the following unique peptides: YVGGQEHFAHLLILR and SDVVYTDWK, AGP1 was found to be a top-ranked protein ($P = 5E-06$) and thus selected as a potential pancreatic cancer biomarker for further evaluation (Fig 2).

Pathway analysis. To better understand the functional importance of AGP1 in pancreatic cancer, we performed a series of pathway analyses using IPA. The IPA bioinformatics tool, which uses a proprietary database of curated, literature-derived functional relationships between proteins, showed a network of proteins physically and/or functionally interacting with AGP1 (Fig 3). The network included a number of extracellular proteins, notably several interleukins, TNF, and IFNG. However, AGP1 was also found to be involved in the MAPK signaling pathway (including p38 and MAPK14), and the AGP1-centred network was regulated by a number of known cancer-associated transcriptional regulators, including p53, YY1, Creb, and HNF4A. Among the proteins in the AGP1-centred relationship network, some were poorly characterized, eg, the TMEM37 transmembrane protein.

For the relationship network centered on AGP1, the IPA analysis provided top canonical pathways, which were significantly enriched among proteins belonging to the network. Noteworthy was the well-known cancer-related p38 MAPK signaling pathway (enrichment P value 1E-10) and IL-10 and IL-6 signaling pathways. Most significantly enriched was the acute phase response signaling pathway (enrichment P value 1E-21). Interestingly, microenvironmental signatures were

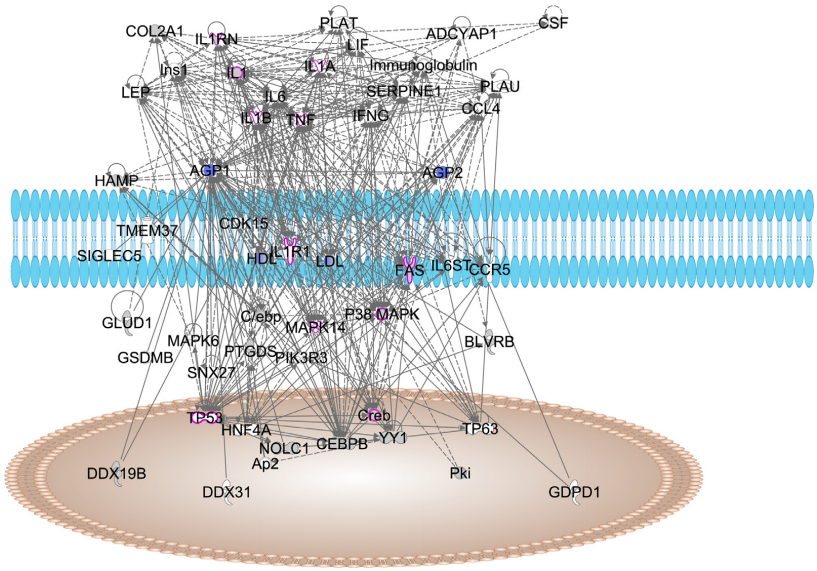


Fig 3. Network of proteins physically and/or functionally interacting with AGP1 (Ingenuity Pathway Analysis, IPA). Cell nucleus shown as brown oval. Cell membrane shown in blue. Solid lines: direct relationships (eg, protein−protein interactions). Dashed lines: indirect relationships (eg, regulation of expression). Proteins belonging to the canonical MAPK signaling pathway marked by magenta contours. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.).

also enriched in the AGP1 network, including stellate cell activation (enrichment *P* value 1E-13) and dendritic cell maturation (enrichment *P* value 1E-10).

AGP1 as a biomarker in tissue biopsies. Following antibody optimization and staining, AGP1 expression could be evaluated in 140 pancreatic tumors represented in the TMA. A total of 112 tumors (80%) were positive for AGP1, including weak staining in 31 tumors (22.1%), moderate staining in 48 tumors (34.3%) and strong staining in 33 tumors (23.6%). AGP1 was expressed in pancreatic tumor cells as well as inflammatory cells (mostly lymphocytes) and fibroblasts in the stromal compartment (Fig 4). In 28 tumors (20%), no AGP1 expression could be visualized. Analysis of the relationship between AGP1 expression and established clinicopathologic parameters revealed no correlation between AGP1 and age at diagnosis, gender, T-stage, N-stage, differentiation grade or resection margin status. AGP1 expression was also evaluated in 10 histologically normal pancreas tissues obtained from patients with benign pancreatic lesions. The pancreatic parenchyma was analyzed, not the lesion. Most of the pancreatic parenchyma samples (8 out of 10) had negative AGP1 expression, while only 2 out of 10 samples were weakly positive for AGP1.

AGP1 expression is associated with shorter survival. Kaplan-Meier analysis demonstrated that AGP1 tissue expression correlated with a significantly worse overall survival ($P = 0.003$, log-rank test), as depicted in Fig 5. This association was confirmed in Cox univariable analysis (HR = 2.22; 95% CI 1.30−3.79, $P = 0.004$) and remained significant in multivariable analysis adjusted for age, gender, TNM status, differentiation grade, resection margin status, and adjuvant chemotherapy (HR = 1.87; 95% CI 1.08−3.24, $P = 0.026$), as shown in Table 2.

AGP1 as a biomarker in liquid biopsies. Circulating levels of AGP1 in serum were measured by an automated immunoassay to assess the potential use of AGP1 as a noninvasive biomarker for the diagnosis of pancreatic cancer. Results indicated that the levels of AGP1 in serum were significantly higher in patients with pancreatic cancer compared with the healthy controls, $P < 0.001$ (Fig 6, A). There were no significant differences in AGP1 levels between pancreatic cancer patients and those with benign pancreatic disease. CA 19-9 levels were significantly higher in pancreatic cancer patients compared to the benign pancreatic disease group ($P < 0.001$), as well as the healthy control group ($P < 0.001$) (Fig 6, B).

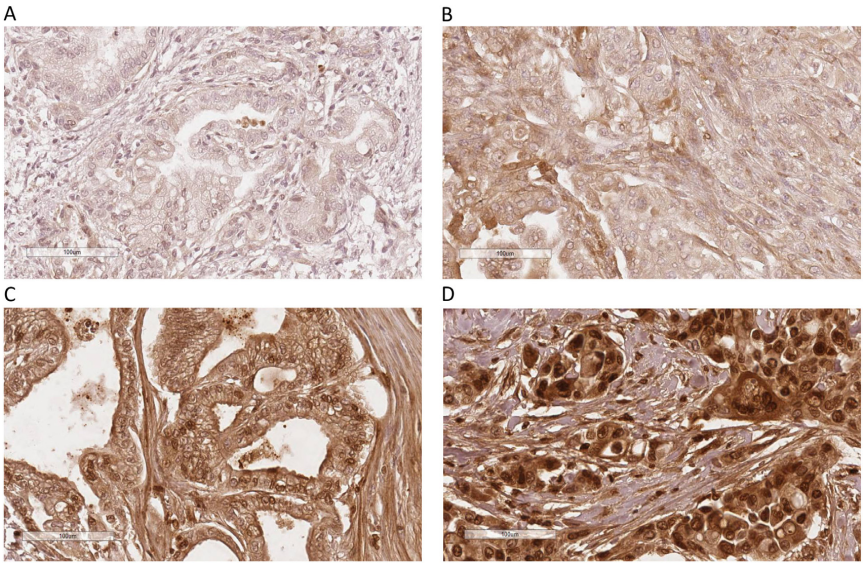


Fig 4. Immunohistochemical images of AGP1 staining representing pancreatic tumors with (A) negative, (B) weak, (C) moderate, and (D) strong staining in a varying proportion of tumor cells. All images are captured at x20 magnification.

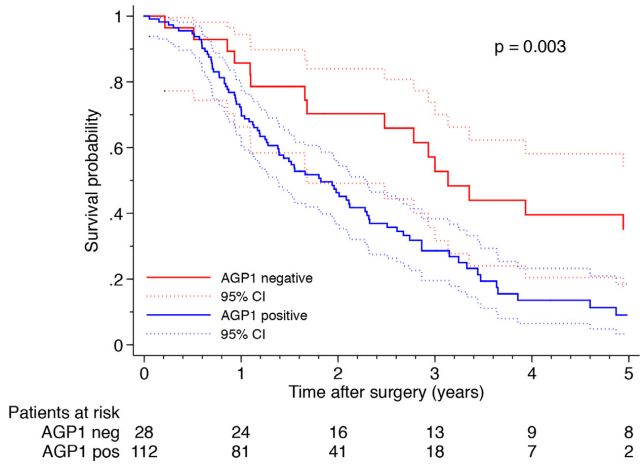


Fig 5. Kaplan-Meier survival curves for all-cause mortality according to AGP1 expression in tumor tissue in patients undergoing surgical resection for pancreatic cancer ($P = 0.003$, log-rank test). The blue solid line shows the observed cumulative survival in the AGP1 positive group and the blue dotted line show the 95% confidence interval (estimated with Kaplan-Meier survival function). The red solid line shows the observed cumulative survival for patients in the AGP1 negative group and the red dotted line shows the 95% confidence interval (estimated with Kaplan-Meier survival function). AGP1, alpha-1-acid glycoprotein 1. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.).

Table 2. Multivariable Cox proportional hazards regression analysis

Variables	HR	95% CI	*P* value
AGP1 positive vs negative, unadjusted	2.22	1.30-3.79	0.004
AGP1 positive vs negative, adjusted*	1.87	1.08-3.24	0.026

*Adjusted for age, gender, T-stage, N-stage, differentiation grade, resection margin status and adjuvant chemotherapy. AGP1, alpha-1-acid glycoprotein 1.

Diagnostic accuracy of circulating AGP1 levels. The Spearman correlation between AGP1 and CA19-9 was 25.3%, indicating that the two markers complement one another well. The cumulative performance of AGP1 alone and in combination with CA 19-9 in terms of AUC, sensitivity, and specificity is shown in Table 3. AGP1 displayed an AUC of 0.837 for the discrimination of pancreatic cancer from healthy controls, with a sensitivity of 86.5% at 82.4% specificity. CA 19-9 provided a lower sensitivity at 75% but with a specificity of 100% (AUC 0.919), when the standard cutoff of 37 kU/L was used. The maximum Youden's Index was applied to determine the optimal cut-off concentration for AGP1, which was found to be 0.74 g/L. Combining AGP1 with CA 19-9 increased the AUC to 0.963 (Fig 7). For discrimination of pancreatic cancer from healthy and benign groups, AGP1 alone provided an AUC of 0.678, which was increased to 0.798, when AGP1 was combined with CA 19-9.

DISCUSSION

This translational, proteomic study identified AGP1 as a biomarker for pancreatic cancer. Overexpression of AGP1 in pancreatic tumor tissue was found to be an independent factor of poor prognosis. Furthermore, we demonstrated that circulating AGP1 levels can be combined with CA 19-9 to distinguish resectable pancreatic cancers from healthy controls.

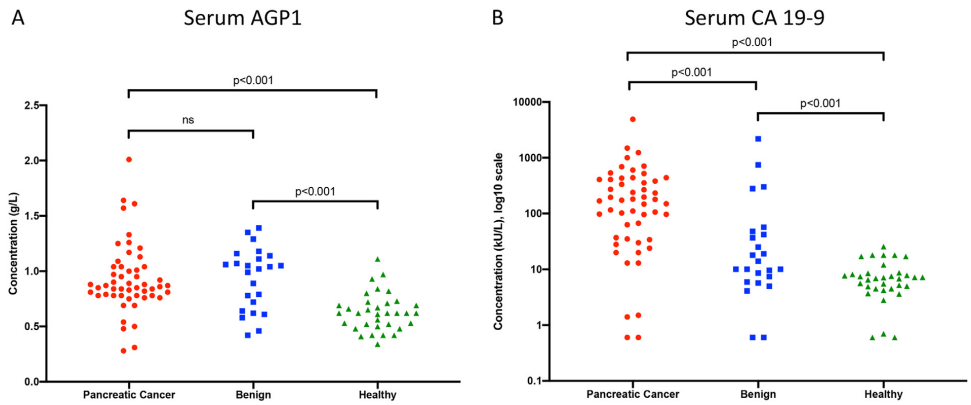


Fig 6. Scatter dot plots of serum AGP1 and CA 19-9 levels in the different groups. (A) Serum AGP1 levels in pancreatic cancer, benign pancreatic disease and healthy controls. (B) Serum CA 19-9 levels in pancreatic cancer, benign pancreatic disease and healthy controls. Wilcoxon rank-sum (Mann-Whitney) test. AGP1, alpha-1-acid glycoprotein 1.

Table 3. Performance of AGP1 and CA 19-9 in distinguishing pancreatic cancer from healthy and benign controls

Clinical question	Biomarker	AUC	Sensitivity %	Specificity %	Youden's index	Cutoff value
PC vs Healthy	AGP1	0.837	86.5	82.4	0.689	0.74[b]
	CA19-9	0.919	75	100	0.75	37[c]
	AGP1+CA19-9[a]	**0.963**	88.5	97.1	0.855	
PC vs Healthy + Benign	AGP1	0.678	86.5	60.3	0.467	0.74[b]
	CA19-9	0.855	75	86.2	0.612	37[c]
	AGP1+CA19-9[a]	**0.798**	88.5	65.5	0.540	

[a]Sensitivity and specificity of the combination of AGP1 and CA19-9;
[b]Predicted optimal cutoff value from the maximum Youden's index;
[c]Clinical reference value. PC, pancreatic cancer.

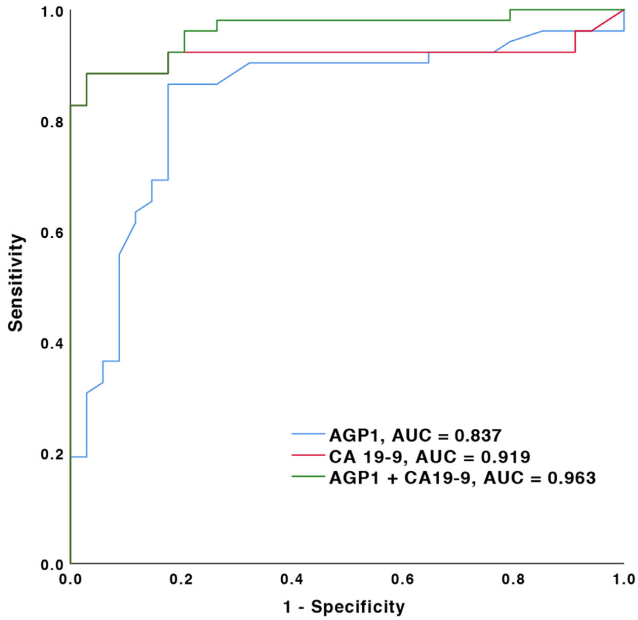


Fig 7. Diagnostic performance of serum AGP1 and CA 19-9 levels for detection of pancreatic cancer against healthy individuals. AGP1, alpha-1-acid glycoprotein 1; AUC, area under the curve.

The prognostication of pancreatic cancer is mainly based on the TNM staging system. However, tumors of the same stage may behave differently in terms of prognosis and treatment response. Therefore, much research focus has been directed toward developing new and better prognostic markers. Genomic[26] and proteomic[27] studies of pancreatic tumor tissue have led to the identification of several potential prognostic factors. Immunohistochemistry is commonly utilized to validate candidate protein biomarkers, due to the clinical feasibility and low costs. Over the years, multiple immunohistochemical biomarkers have been proposed for pancreatic cancer.[4] However, most markers are derived from small series and no prognostic marker has yet been introduced in clinical practice due to the lack of sufficient validation. To overcome these problems, we constructed a TMA using a large, clinically well-annotated pancreatic cancer patient cohort. We performed high-quality, immunohistochemical annotation and adjusted for traditional prognostic variables, allowing for a realistic assessment of the prognostic potential of AGP1 in pancreatic cancer.

The function of AGP1 in pancreatic cancer is not well understood. Our pathway analysis revealed involvement of AGP1 in central signaling cascades related to pancreatic cancer cell proliferation, migration, and invasion, including MAPK,[28] p53,[29] and YY1 signalling.[30] Among the proteins in AGP1-centred relationship networks, some are less known, eg, the TMEM37 transmembrane protein, which has only recently been proposed as a prognostic gene in colon cancer.[31] We also identified functional relationships with the tumor stroma, particularly stellate cells[32] and dendritic cells,[33] providing further support for the potential involvement of AGP1 in pancreatic carcinogenesis. Increased AGP1 levels have also been reported in hepatocellular carcinoma,[34] laryngeal cancer,[35] and lung cancer,[36] suggesting that AGP1 alone cannot be used to diagnose pancreatic cancer, necessitating a panel of biomarkers to enhance specificity.

The present study provides a foundation for development of early detection strategies. A useful biomarker for pancreatic cancer needs to be minimally invasive or noninvasive and have high diagnostic accuracy. CA 19-9 may be used for disease monitoring in pancreatic cancer patients, but is not recommended for population-based screening.[37] Other types of blood markers have been reported for pancreatic cancer, such as protein biomarker panels,[38] circulating tumor DNA,[39] exosomes,[40] microRNAs,[41] and cell-free nucleosomes.[42] However, none of the biomarkers have yet entered routine clinical practice. This may be due to the lack of independent validation, but also because of the complexity of the

methodology used to measure the markers, both in terms of equipment and the demands for multiplex analysis of many markers. We reasoned that single proteins secreted from the tumor into the extracellular space may be developed into diagnostic markers and measured by simple immunoassays as a noninvasive screening test. We found that AGP1 could discriminate pancreatic cancer from healthy controls with an AUC of 0.837. As a combination test, AGP1 and CA 19-9 provided an impressive AUC of 0.963 for diagnosis of pancreatic cancer against healthy controls.

To our knowledge, this is the largest clinical evaluation of AGP1 in pancreatic cancer. Increased AGP1 levels have been reported in serum samples from patients with pancreatic cancer, but these studies were conducted in small series.[43,44] In the study by Hashimoto et al. [43] only 11 patients with pancreatic cancer were included. Balmaña et al[44] reported altered glycosylation of AGP1 in 19 patients with pancreatic cancer. Thus, the present study represents the most comprehensive evaluation of AGP1 as a biomarker for pancreatic cancer, providing large-scale profiling both in tissue and serum, as well as pathway characterization.

A particular strength of our study was that biomarker candidates were derived from comparison to healthy pancreas controls, and not tissues adjacent to the tumors, as such regions may have many aberrant morphologic and phenotypic alterations.[45,46] The inclusion of healthy tissue as a control group is therefore essential when it comes to development of a diagnostic biomarker candidate (disease vs healthy). Furthermore, all serum analyses were performed in an accredited clinical laboratory, ensuring accuracy and reproducibility of the data generated.

Limitations of this study need to be addressed. We demonstrated that the AGP1/CA 19-9 panel can detect resectable pancreatic cancer with good precision and thereby may improve early detection of the disease. However, benign diseases of the pancreas, such as chronic pancreatitis, may cause elevated levels of AGP1, as shown in this study. Additional protein markers are investigated by our group that can be added into the serum panel in order to further improve performance. In the present study, total levels of AGP1 were measured. The glycosylation pattern of AGP1 was not evaluated. In the future, with the advancement of glycoproteomics, simple and effective methods may be developed that can accurately measure specific glycosylations associated with cancer. As detection of late-stage pancreatic cancer is of little clinical value, all subjects included in this study were selected from operable disease. In further validation tests, late-stage disease may be included to assess the correlation between AGP1 and tumor burden.

CONCLUSIONS

The present study identified AGP1 as a potential biomarker for pancreatic cancer by combinatorial mass spectrometry, pathway analysis, and antibody-based validation in tissue and serum. The general workflow presented in this work may be applicable to identify biomarkers for other complex and multifaceted diseases.

ACKNOWLEDGMENTS

Conflict of Interest: All authors have read the journal's policy on disclosure of potential conflict of interest. R. Andersson, G. Marko-Varga, and D. Ansari are board members of Reccan Diagnostics and have filed a patent related to the findings presented in this manuscript. The other authors declare no potential conflicts of interest.

Authorship Agreement: All authors have read the journal's authorship agreement. The article has been reviewed by and approved by all authors.

This work was supported by the Magnus Bergvall Foundation, the Royal Physiographic Society of Lund, the Tore Nilsson Foundation, the Inga and John Hain Foundation for Medical Research, the Clas Groschinsky Foundation, the Gunnar Nilsson Foundation, the Gyllenstiernska Krapperup Foundation, the Bengt Ihre Foundation, the Emil and Wera Cornell Foundation, the Crafoord Foundation, Governmental Funding of Clinical Research within the National Health Service (ALF), and Swedeńs Innovation Agency (Vinnova).

The authors would also like to acknowledge Maria Jönsson and Mohibullah Hotak at the Department of Clinical Chemistry and Pharmacology, University and Regional Laboratories Region Skåne, Sweden, for the excellent technical support. The authors would also like to thank the Lund University Diabetes Center (LUDC), a part of the national consortium on Excellence in Diabetes Research, for providing the control tissues. Thermo Fisher Scientific, San Jose, is greatly acknowledged for their generous support.

AUTHOR CONTRIBUTIONS

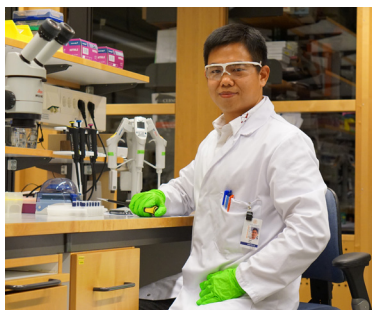
Q. Zhou and D. Ansari conceived the original idea and designed the study with G. Marko-Varga and R. Andersson. Q. Zhou, M. Bauden, A. Sasor, D. Hu, and D. Ansari collected the data for the study, which were analyzed by Q. Zhou, T. Bygott, K. Pawlowski, and I. Pla. The data interpretation and manuscript drafting were performed by Q. Zhou and D. Ansari. The manuscript was revised by M. Bauden, T. Bygott, K. Pawlowski, and R. Andersson. All authors reviewed the manuscript and gave the final approval for submission.

REFERENCES

1. Carrato A, Falcone A, Ducreux M, et al. A systematic review of the burden of pancreatic cancer in Europe: real-world impact on survival, quality of life and costs. J Gastrointest Cancer 2015;46:201–11.
2. Ansari D, Bauden M, Bergstrom S, et al. Relationship between tumour size and outcome in pancreatic ductal adenocarcinoma. Br J Surg 2017;104:600–7.
3. Samuel N, Hudson TJ. The molecular and cellular heterogeneity of pancreatic ductal adenocarcinoma. Nat Rev Gastroenterol Hepatol 2011;9:77–87.
4. Ansari D, Rosendahl A, Elebro J, Andersson R. Systematic review of immunohistochemical biomarkers to identify prognostic subgroups of patients with pancreatic cancer. Br J Surg 2011;98:1041–55.
5. Olivares O, Mayers JR, Gouirand V, et al. Collagen-derived proline promotes pancreatic ductal adenocarcinoma cell survival under nutrient limited conditions. Nat Commun 2017;8:16031.
6. Conroy T, Hammel P, Hebbar M, et al. FOLFIRINOX or gemcitabine as adjuvant therapy for pancreatic cancer. N Engl J Med 2018;379:2395–406.
7. Neoptolemos JP, Palmer DH, Ghaneh P, et al. Comparison of adjuvant gemcitabine and capecitabine with gemcitabine monotherapy in patients with resected pancreatic cancer (ESPAC-4): a multicentre, open-label, randomised, phase 3 trial. Lancet 2017;389:1011–24.
8. Ducreux M, Cuhna AS, Caramella C, et al. Cancer of the pancreas: ESMO clinical practice guidelines for diagnosis, treatment and follow-up. Ann Oncol 2015;26(Suppl 5):v56–68.
9. Bailey P, Chang DK, Nones K, et al. Genomic analyses identify molecular subtypes of pancreatic cancer. Nature 2016;531:47–52.
10. Collisson EA, Sadanandam A, Olson P, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med 2011;17:500–3.
11. Moffitt RA, Marayati R, Flate EL, et al. Virtual microdissection identifies distinct tumor- and stroma-specific subtypes of pancreatic ductal adenocarcinoma. Nat Genet 2015;47:1168–78.
12. Kim MS, Pinto SM, Getnet D, et al. A draft map of the human proteome. Nature 2014;509:575–81.
13. Wilhelm M, Schlegl J, Hahne H, et al. Mass-spectrometry-based draft of the human proteome. Nature 2014;509:582–7.
14. Doerr A. Targeting with PRM. Nat Methods 2012;9:950.
15. Peterson AC, Russell JD, Bailey DJ, Westphall MS, Coon JJ. Parallel reaction monitoring for high resolution and high mass accuracy quantitative, targeted proteomics. Mol Cell Proteomics 2012;11:1475–88.
16. Brierley J, Gospodarowicz M, Wittekind C. TNM classification of malignant tumours, 8th ed. John Wiley & Sons; 2017.
17. McShane LM, Altman DG, Sauerbrei W, et al. REporting recommendations for tumour MARKer prognostic studies (REMARK). Br J Cancer 2005;93:387–91.
18. Bossuyt PM, Reitsma JB, Bruns DE, et al. Towards complete and accurate reporting of studies of diagnostic accuracy: the STARD initiative. Standards for reporting of diagnostic accuracy. Clin Chem 2003;49:1–6.
19. Zhou Q, Andersson R, Hu D, et al. Quantitative proteomics identifies brain acid soluble protein 1 (BASP1) as a prognostic biomarker candidate in pancreatic cancer tissue. EBioMedicine 2019;43:282–94.
20. Kramer A, Green J, Pollard J. Jr., Tugendreich S. Causal analysis approaches in ingenuity pathway analysis. Bioinformatics 2014;30:523–30.
21. Harrell Jr, F.RMS: Regression Modeling Strategies. R package version 5.1-2.https://CRAN.R-project.org/package=rms.
22. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinf 2011;12:77.
23. Youden WJ. Index for rating diagnostic tests. Cancer 1950;3:32–5.
24. Tyanova S, Temu T, Sinitcyn P, et al. The Perseus computational platform for comprehensive analysis of (prote)omics data. Nat Methods 2016;13:731–40.
25. R Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing, *Vienna, Austria.*https://www.R-project.org/, 2017.
26. Oshima M, Okano K, Muraki S, et al. Immunohistochemically detected expression of 3 major genes (CDKN2A/p16, TP53, and SMAD4/DPC4) strongly predicts survival in patients with resectable pancreatic cancer. Ann Surg 2013;258:336–46.
27. Ansari D, Aronsson L, Sasor A, et al. The role of quantitative mass spectrometry in the discovery of pancreatic cancer biomarkers for translational science. J Transl Med 2014;12:87.
28. Furukawa T. Impacts of activation of the mitogen-activated protein kinase pathway in pancreatic cancer. Front Oncol 2015;5:23.
29. Weissmueller S, Manchado E, Saborowski M, et al. Mutant p53 drives pancreatic cancer metastasis through cell-autonomous PDGF receptor beta signaling. Cell 2014;157:382–94.
30. Yuan P, He XH, Rong YF, et al. KRAS/NF-kappaB/YY1/miR-489 signaling axis controls pancreatic cancer metastasis. Cancer Res 2017;77:100–11.
31. Li C, Shen Z, Zhou Y, Yu W. Independent prognostic genes and mechanism investigation for colon cancer. Biol Res 2018;51:10.
32. Duner S, Lopatko Lindman J, Ansari D, Gundewar C, Andersson R. Pancreatic cancer: the role of pancreatic stellate cells in tumor progression. Pancreatology 2010;10:673–81.
33. Deicher A, Andersson R, Tingstedt B, et al. Targeting dendritic cells in pancreatic ductal adenocarcinoma. Cancer Cell Int 2018;18:85.
34. Bachtiar I, Santoso JM, Atmanegara B, et al. Combination of alpha-1-acid glycoprotein and alpha-fetoprotein as an improved diagnostic tool for hepatocellular carcinoma. Clin Chim Acta 2009;399:97–101.
35. Uslu C, Taysi S, Akcay F, Sutbeyaz MY, Bakan N. Serum free and bound sialic acid and alpha-1-acid glycoprotein in patients with laryngeal cancer. Ann Clin Lab Sci 2003;33:156–9.
36. Bruno R, Olivares R, Berille J, et al. Alpha-1-acid glycoprotein as an independent predictor for treatment effects and a prognostic factor of survival in patients with non-small cell lung cancer treated with docetaxel. Clin Cancer Res 2003;9:1077–82.
37. Ansari D, Tingstedt B, Andersson B, et al. Pancreatic cancer: yesterday, today and tomorrow. Fut Oncol 2016;12:1929–46.
38. Mellby LD, Nyberg AP, Johansen JS, et al. Serum biomarker signature-based liquid biopsy for diagnosis of early-stage pancreatic cancer. J Clin Oncol 2018;36:2887–94.
39. Cohen JD, Javed AA, Thoburn C, et al. Combined circulating tumor DNA and protein biomarker-based liquid biopsy for the earlier detection of pancreatic cancers. Proc Natl Acad Sci USA 2017;114:10202–7.
40. Melo SA, Luecke LB, Kahlert C, et al. Glypican-1 identifies cancer exosomes and detects early pancreatic cancer. Nature 2015;523:177–82.
41. Schultz NA, Dehlendorff C, Jensen BV, et al. MicroRNA biomarkers in whole blood for detection of pancreatic cancer. JAMA 2014;311:392–404.
42. Bauden M, Pamart D, Ansari D, et al. Circulating nucleosomes as epigenetic biomarkers in pancreatic cancer. Clin Epigen 2015;7:106.

43. Hashimoto S, Asao T, Takahashi J, et al. Alpha1-acid glycoprotein fucosylation as a marker of carcinoma progression and prognosis. Cancer 2004;101:2825–36.
44. Balmana M, Gimenez E, Puerta A, et al. Increased alpha1-3 fucosylation of alpha-1-acid glycoprotein (AGP) in pancreatic cancer. J Proteomics 2016;132:144–54.
45. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer 1953;6:963–8.
46. Aran D, Camarda R, Odegaard J, et al. Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat Commun 2017;8:1077.

About the author



Qimin Zhou, M.D., is currently enrolled as a surgical resident at the Department of Plastic & Reconstructive Surgery, Shanghai Ninth People's Hospital. He has a special interest in translational research within the field of surgery.

Printed by Media-Tryck, Lund 2020 NORDIC SWAN ECOLABEL 3041 0903

LUND UNIVERSITY

FACULTY OF MEDICINE

Department of Surgery
Clinical Sciences Lund

Lund University, Faculty of Medicine
Doctoral Dissertation Series 2020:47
ISBN 978-91-7619-908-4
ISSN 1652-8220