SWATH-MS based proteomic profiling of Pancreatic Ductal Adenocarcinoma tumours reveals the interplay between the extracellular matrix and related intracellular pathways

Nweke EE[1], Naicker P[2], Aron S[3], Stoychev S[2], Devar J[1], Tabb DL[4], Jones OJ[1], Smith MD[1] and Candy GP[1]

[1]Department of Surgery, Faculty of Health Sciences, University of the Witwatersrand, Johannesburg 2193, South Africa.

[2]Department of Biosciences, Centre for Scientific and Industrial Research, Pretoria 0184, South Africa.

[3]Sydney Brenner Institute for Molecular Bioscience, University of the Witwatersrand, Johannesburg 2193, South Africa.

[4]Bioinformatics Unit, South African Tuberculosis Bioinformatics Initiative, Division of Molecular Biology and Human Genetics, Department of Biomedical Sciences, Stellenbosch University, Cape Town 7505, South Africa.

Short Title: Proteomic profiling of resectable PDAC tissues

Keywords: Pancreatic cancer, proteins, SWATH-MS, SNVs, Extracellular matrix

Corresponding author: Nweke E, Department of Surgery, Faculty of Health Sciences, 7 York Road, Parktown, 2193, Johannesburg, Republic of South Africa. Electronic address: ekene.nweke@wits.ac.za Tel.: +27735384461 Fax.: +27 864608582

Abstract

Pancreatic cancer accounts for 2.8% of new cancer cases worldwide and is projected to become the second leading cause of cancer-related deaths by 2030. Patients of African ancestry appear to be at an increased risk for pancreatic ductal adenocarcinoma (PDAC), with worse severity and outcomes. The purpose of this study was to map the proteomic and genomic landscape of a cohort of PDAC patients of African ancestry.

Thirty tissues (15 tumours and 15 normal adjacent tissues) were obtained from consenting South African PDAC patients. Optimisation of the sample preparation method allowed for the simultaneous extraction of high-purity protein and DNA for SWATH-MS and OncoArray SNV analyses.

We quantified 3402 proteins with 49 upregulated and 35 downregulated proteins at a minimum 2.1 fold change and FDR adjusted p-value (q-value) $\leq$ 0.01 when comparing tumour to normal adjacent tissue. Many of the upregulated proteins in the tumour samples are involved in extracellular matrix formation (ECM) and related intracellular pathways. Proteins such as EMIL1, KBTB2, and ZCCHV involved in the regulation of ECM proteins were observed to be dysregulated in pancreatic tumours. Approximately 11% of the dysregulated proteins, including ISLR, BP1, PTK7 and OLFL3, were predicted to be secretory proteins. Additionally, we identified missense mutations in some upregulated proteins, such as MYPN, ESTY2 and SERPINB8. These findings help in further elucidating the biology of PDAC and may aid in identifying future plausible markers for the disease.

Introduction

Pancreatic ductal adenocarcinoma (PDAC) originates from the ductal epithelial cells of the pancreas and accounts for 85% of all pancreatic cancers. Worldwide, the incidence and mortality rates of PDAC are rising compared to other cancers (1). Despite current therapeutic strategies, the survival rate is dismal, with a five-year survival rate of 8% (2). These poor statistics are chiefly due to late detection, allowing for invasion and metastasis of cancer, and therapeutic resistance. In multiracial countries such as the United States of America, African Americans have increased incidence and poorer survival rates compared to other ethnicities. Although this has been largely attributed to social factors such as smoking, alcohol consumption, obesity and diabetes mellitus, current studies have determined genetics as an underlying factor (3–5).

The tumour microenvironment (TME) of PDAC plays a role in enabling chemotherapeutic resistance, cancer cell proliferation, invasion, migration and metastasis. It is characterised by a dense extracellular matrix (ECM) structure, stromal cells, cancer-associated fibroblasts and immune cells (6–10). The overexpression of ECM proteins exacerbates PDAC tumourigenesis, fostering tumour growth. Pancreatic cancer cells secreting ECM proteins such as collagen and fibronectin, were found to have increased proliferation and were desensitized to chemotherapeutic drugs such as gemcitabine (11–14). Hence, there is a need for the identification of novel proteins involved in the ECM formation that could be potential chemotherapeutic targets and biomarkers of PDAC.

Mass spectrometry-based technologies have proven useful in the identification of proteins with important roles in PDAC progression (15). Sequential window acquisition of all theoretical fragment ion spectra (SWATH) is a label-free application of mass spectrometry (MS), suitable for discovery proteomics where hundreds or thousands of proteins are

quantified at high analyte throughput and reproducibility (Gillet et al, 2012, Ludwig et al., 2018).

In the present study, protein and DNA were simultaneously extracted from the same tissue samples, which allowed SWATH-MS and OncoArray single nucleotide variation (SNV) profiling of tissues obtained from resectable (stage I and II) South African PDAC patients who underwent a Whipple procedure. Several significantly differentiated proteins were identified, many of which were enriched in the extracellular matrix formation and related intracellular pathways. By querying databases, some of the identified proteins were shown to be secreted into the bloodstream which could make them potential biomarkers of PDAC.

Methods

Ethics statement

The Human Research Ethics Committee of the University of Witwatersrand approved this study (HREC- M150778). Each patient gave written consent for sample collection and recording of demographic and clinical data.

Patient recruitment and sample collection

Biopsy samples were obtained from 15 consenting South African patients (aged between 53-95 years old) with histologically confirmed PDAC at the Chris Hani Baragwanath Academic Hospital (Johannesburg, South Africa). All biopsies were obtained immediately after Whipple procedures to resect of the pancreas. One core tumour biopsy and one adjacent 'normal' tissue (> 2cm away from the tumour) were obtained (30 samples) and immediately frozen at -80°C (**Table S1**). The overview of the study workflow is shown in **Figure 1**.

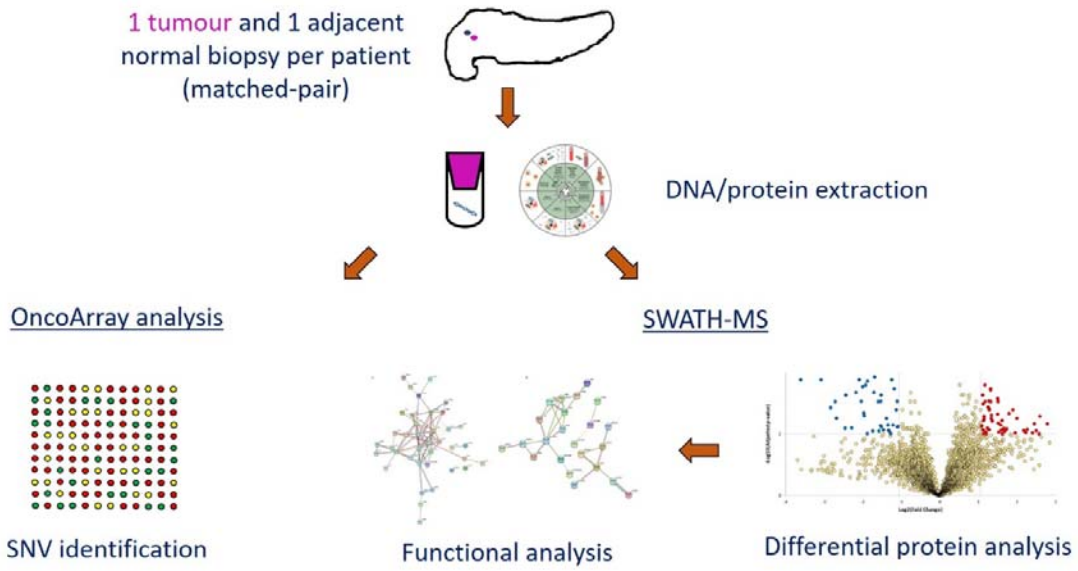


Figure 1: Overview of study workflow, from sample collection to functional analysis.

Tissue processing

Tissues (15-35mg) were homogenised in 600µl of ATL lysis buffer (Qiagen Hilden, Germany) supplemented with 40mM DTT, using a Tissue Ruptor (Qiagen). The homogenised solution was centrifuged at 14 000*g* for 3 mins to remove any micro-tissue particulate from the solution. From the recovered lysate 180µl was taken for DNA purification.

Protein preparation for mass spectrometry

The remainder of the lysate was incubated with four parts pre-chilled acetone for 30 min at -20°C. The mixture was centrifuged at 14 000*g* at 4°C for 10 min and the resultant pellet was washed with ice-cold ethanol. The pellet was resuspended in 50mM Tris-HCl pH 8.0 containing 4% sodium dodecyl sulphate (SDS) and a Roche complete™ EDTA-free protease inhibitor cocktail. The protein concentration was measured using the Pierce Bicinchoninic assay (Thermo Fisher Scientific, Massachusetts, USA). Aliquots of protein solution were then stored at -80°C until further processing.

Protein samples were thawed and 20µg per sample was reduced with 10mM dithiothreitol (VWR Pennsylvania, United States) at 37°C for 30 min and alkylated with 40mM iodoacetamide at room temperature in the dark for 30 min. The samples were then purified of detergents and salts using MagReSyn™ HILIC microparticles (ReSyn Biosciences Edenvale, South Africa) in a 96 deep-well plate for automated sample preparation of 12 samples in parallel using the KingFisher Duo™ system (Thermo Fisher Scientific), as previously described (17). Briefly, magnetic hydrophilic affinity microparticles (20µl, 200µg) were equilibrated in 200µl of 100mM ammonium acetate pH 4.5 containing 15% acetonitrile (MeCN). Microparticles were then transferred to 100µl of protein binding solution (each protein sample adjusted to 50µl using 4% SDS and added to 50ul of 200mM ammonium acetate pH 4.5 containing 30% MeCN) and mixed for 30 min at room temperature. The captured proteins were washed twice in 200µl of 95% MeCN and transferred to 200µl of 25mM ammonium bicarbonate containing 2µg of sequencing grade trypsin (Promega, Madison, USA) and mixed for 4 hrs digestion at 37°C. The automated on-bead protein capture, clean-up and digest protocol was programmed using BindIt software v4.0 (Thermo Fisher Scientific) and is available upon request (info@resynbio.com) and further described in **Figure S1**. Resultant peptides were dried and stored at -80°C before Liquid Chromatography-MS analysis.

Liquid chromatography-mass spectrometry (LC-MS) data acquisition

Approximately 1ug of tryptic peptides per sample was analysed using a Dionex Ultimate 3000 RSLC system coupled to an AB Sciex 6600 TripleTOF mass spectrometer (AB Sciex, Massachusetts, USA). For data-dependent acquisition (DDA, used for spectral library building), tumour and normal samples were each pooled in their respective groups and spiked with Biognosys iRT retention time peptide standards. Four injection replicates of each group were acquired (n = 8). For SWATH, each sample was injected once. Peptide samples were inline de-salted using an Acclaim PepMap C18 trap column (75µm × 2cm; 2 min at 5µl.min^{-1} using 2% Acetonitrile, ACN/0.2% FA). Trapped peptides were gradient eluted and separated on an Acclaim PepMap C18 RSLC column (75µm × 15cm, 2µm particle size) at a flow-rate of 0.5µl.min^{-1} with a gradient of 4-40% B over 60 min (A: 0.1% FA; B: 80% ACN/0.1% FA). For DDA, precursor (MS) scans were acquired from *m/z* 400-1500 (2^{+}-5^{+} charge states) using an accumulation time of 250ms followed by 80 fragment ion (MS/MS) scans, acquired from *m/z* 100-1800 with 25ms accumulation time each. For SWATH, precursor scans ranged from *m/z* 400 to 900 using 100 variable-width windows that overlapped by 0.5 Da, and fragment ions were acquired from *m/z* 100-1800 with 25ms accumulation time per window.

Liquid Chromatography-Mass Spectrometry (LC-MS) Data analysis

Raw DDA files (.wiff), n = 8, were included in a combined search using Protein Pilot software v5.0.2.0 (AB Sciex, Massachusetts, USA), against a database of human reference proteome contained in SwissProt (Accessed 9 June 2019) supplemented with sequences of common contaminants and iRT peptide retention time standards (Biognosys Schlieren, Switzerland). This searched data file was imported into Skyline software v19.1 (Maclean et al., 2010) with peptide features extracted to build a spectral library. A cut-off score of 0.994

(from Protein Pilot report) corresponding to 1% local peptide false discovery rate (FDR) was applied for a library building, with fixed carbamidomethylation of cysteine and variable N-terminal acetylation being the only allowable modifications. The minimum peptide length was 7 and the maximum 36 amino acids and enzyme for digestion set as trypsin with one allowable missed cleavage. The built spectral library contained 141 775 transitions matching to 19 861 peptides and 3413 proteins. Decoy entries were generated by shuffling the peptide sequence, with 141 775 decoy transitions and 19 861 decoy peptides added.

SWATH data files (.wiff), n = 30, were centroided and converted to mzML format using the msconvert tool in the ProteoWizard software suite v3.0.19217 (18). The converted data files were imported into Skyline and matched to the spectral library for correct assignment of the mass spectra. A fragment ion mass tolerance of 0.5m/z was employed for library matching, with a minimum of three and maximum of six of the most intense product ions being selected. The b and y ions with +1 and +2 charge states matching to peptides with +2, +3 and +4 charge states were allowed and product ions from b3 to the last ion were allowed for selection. The iRT peptides were used to align retention times across the dataset using the score to run a regression. A 2 min window around the predicted retention time was used to assist in identification. A peak scoring model was trained using mProphet (19) and the decoy peptides generated from the spectral library. All peaks were re-integrated using this model. The MSstats (20) group comparison plugin was implemented within Skyline to perform quantitative analysis. A normalisation method (equalise medians) was applied for the comparisons across the experimental groups. The MSstats design sample size plugin was used to determine the statistically significant fold change that could be observed based on the desired power, sample number and variability of the dataset. The filtered list of dysregulated proteins was manually inspected within Skyline to verify the quality of the precursors and

transitions used for quantitation. Principal component analysis (PCA) was performed using the peptide-level quantification of each sample within Perseus v1.6.5.0 (21).

DNA extraction and SNV array

DNA purification was performed as per the manufacturer's instruction using the Qiagen DNeasy Blood and Tissue kit (Qiagen Hilden, Germany). The total DNA was quantified using a NanoDrop ND-1000 UV spectrophotometer. A ratio of absorbance at 260nm:280nm of >1.9 and 260nm:230nm of >1.5 was observed in all samples. The Illumina Infinium® OncoArray-500k Beadchip (Ilumina California, United States) was used for single nucleotide variation (SNV) analysis using 200ng of DNA for each sample. The SNV array was conducted as per manufacturer's instruction. GenomeStudio (https://emea.illumina.com/techniques/microarrays/) was used to perform clustering and QC of the raw intensity data and the genotype calls were exported as a GenomeStudio report. For downstream analysis, the GenomeStudio report was converted into PLINK (22) format using the *topbottom* module of the H3ABioNet workflow (https://github.com/h3abionet/h3agwas). The original dataset consisted of 499 170 SNVs genotyped for 48 samples. Some samples were genotyped in replicate and the replicates were utilised to improve the genotype calling when converting the data to PLINK format.

Single nucleotide variation (SNV) analysis

The genotype data were subset to include only tumour-normal pairs and replicates were removed, as related samples cannot be used in the association tests. Quality control was conducted on the dataset to remove low-quality SNVs and samples. SNVs with greater than 1% missingness, lower than 1% minor allele frequency and deviations from Hardy-Weinberg

Equilibrium were removed. Tumour samples were coded as cases and normal samples coded as controls.

Functional analysis

PANTHER v14.1 (23) and REACTOME v70 (24) were applied to show pathway-enrichment analysis of identified dysregulated proteins. Cytoscape v3.8.0 (25) with the STRING (26) plug-in was used to show network interaction between dysregulated proteins. ShinyGO v0.61 (27) was used for graphical representation of enriched biological processes/pathways and the Ensembl Variant Effect Predictor tool (28) modelled the consequences of observed SNPs.

Results

Differential protein analysis in tumours

Using SWATH-MS, we matched 3413 proteins and 19 861 peptides to our in-house generated spectral library. Peptide quantification was performed on MS2 ion measurements using Skyline (Maclean et al., 2010). A principal component analysis (PCA) determined the maximum covariance among the samples based on the abundance of all peptides measured in the analysis. This allowed for unsupervised grouping of samples without any manual assignment of groups. The PCA plot (**Figure S2**) shows two defined groups of samples in agreement with the two clinically defined groups in the study. The normal adjacent group dataset shows wider distribution compared to the tumour group dataset.

An MSstats group comparison was performed with Skyline software to determine the observed differential abundance (fold change) of each quantifiable protein across the tumour and normal adjacent experimental groups. A minimum fold change ≥2.1 and maximum FDR adjusted p-value (q-value) ≤0.01 was used to filter proteins that were significantly different

between tumour and corresponding normal tissues. The fold change threshold was calculated using the MSstats design sample size plugin within Skyline, based on 0.8 power, 0.01 FDR and 15 replicates per group (**Figure S3**). We found 49 upregulated and 35 downregulated proteins in tumours compared to normal adjacent tissues (**Figure 2, Table S2**). Across the complete dataset of quantified proteins (3402), 318 are predicted to be secreted based on information from the Human Protein Altas (29) and Uniprot (30), both assessed on March 24th 2020. Among the dysregulated proteins, 78% of the observed differentially expressed proteins (DEPs) have been shown to have prognostic potential in various cancers like colorectal, renal, liver, lung, urothelial, thyroid and endometrial. AP2A2, PLST, PLSI, BLVRB and SLC2A1 were known intracellular prognostic markers of pancreatic cancer (**Table S2**). Additionally, about 11% were predicted to be secreted into the bloodstream, making them potential reporters for tumour progression and metastasis. These include proteins such as ISLR, OLFL3, PTK7, GBB1, HEM2, BP1, RL39, CAMP, PF4V, and LTBP1, which were reported as prognostic pancreatic cancer biomarkers in the Human Protein Atlas database.

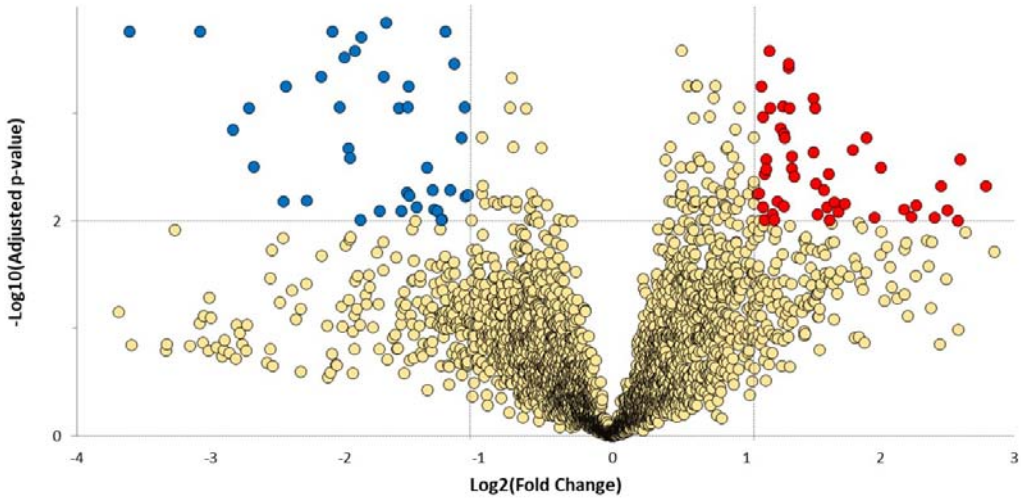


Figure 2: Volcano plot showing dysregulated proteins. Red and green nodes indicate downregulated and upregulated proteins group (based on a minimum fold change ≥2.1 and a maximum adjusted p-value (q-value) ≤0.01) in tumour compared to normal adjacent, respectively.

Functional and network analysis of differentially expressed proteins

We observed that dysregulated proteins (**Table S2**) share several key biological processes (**Figure 3**) and pathways (**Table 1, 2**). These were mostly linked to the extracellular matrix formation/organisation and related intracellular signalling pathways (**Figure 4**). Top pathways enriched for the upregulated proteins included recycling pathway of L1, cell-extracellular matrix interactions and cell junction organisation. Top downregulated pathways were involved in the take-up of oxygen and the release of carbon dioxide from erythrocytes.

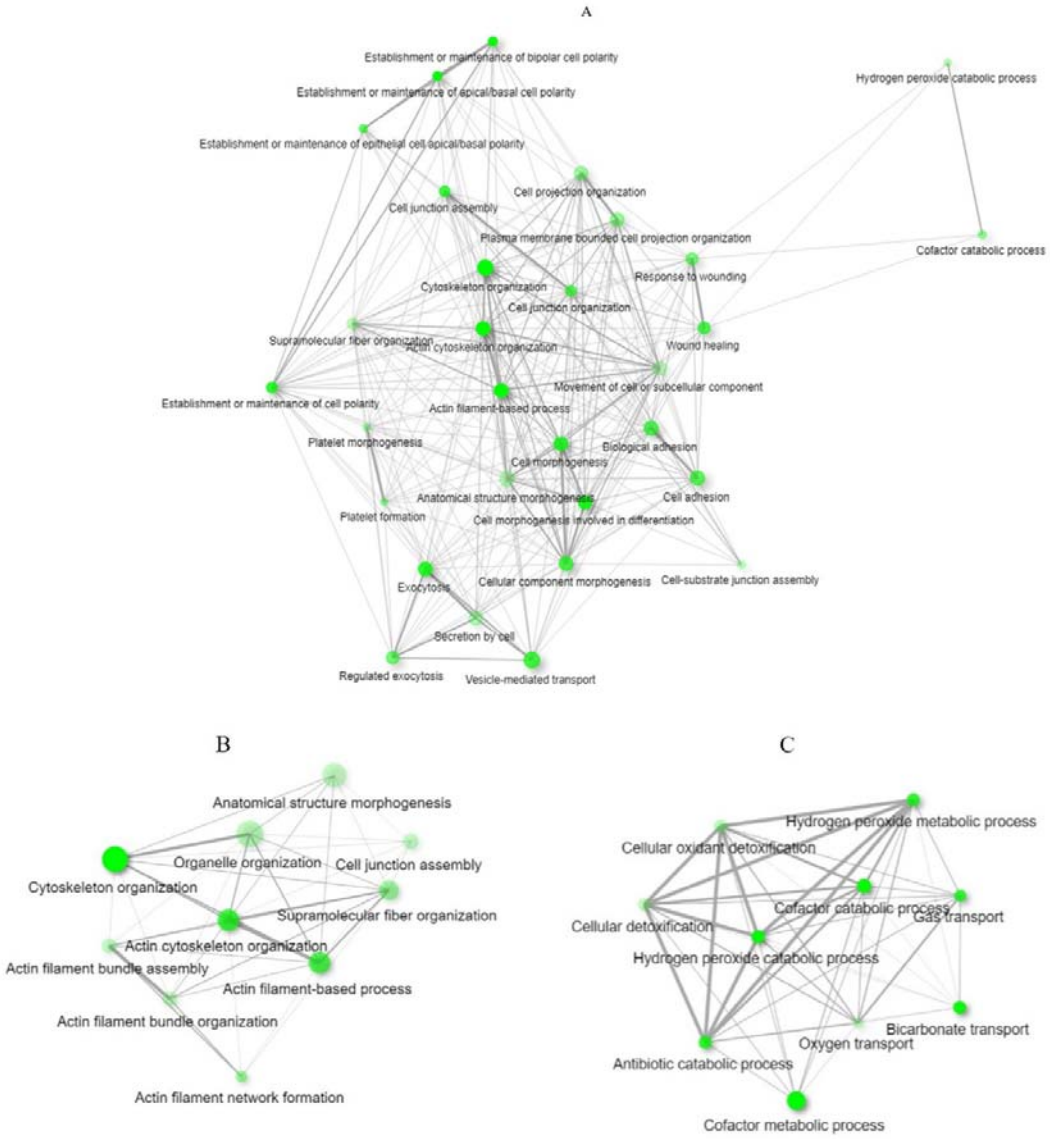


Figure 3: The interaction network analysis of the relationship between biological processes that are enriched by (a) dysregulated proteins. Separate analyses of (b) upregulated and (c) downregulated proteins are also shown. An edge indicates that two processes share 20% or more proteins. Thicker edges (lines) show that there are more overlapped edges. Bigger and darker nodes represent larger protein sets and more significantly enriched proteins, respectively. The plot was generated from ShinyGO.

Table 1: Top 10 significantly upregulated pathways in tumour samples. Generated from Reactome.

Pathway name	False discovery rate	Submitted proteins
Recycling pathway of L1	2.57E-05	AP2A2, TBB6, TBB2A, TBB8, ACTG, DPYL2
Cell-extracellular matrix interactions	2.02E-04	ACTG, ACTN1, LIMS1, ILK
RHO GTPases activate IQGAPs	9.95E-04	TBB6, TBB2A, TBB8, ACTG
Translocation of SLC2A4 (GLUT4) to the plasma membrane	9.95E-04	TBB6, TBB2A, TBB8
MHC class II antigen presentation	1.86E-03	AP2A2, TBB6, TBB2A, CATB, CATC
Cell junction organisation	2.30E-03	PLEC, ACTG, ACTN1, LIMS1, ILK
Hemostasis	5.73 E-03	WDR1, TBB6, TBB2A, SPB8, TBB8, GBB2, GBB1, CAP1, ACTN1, ISLR, AT2B4
Cell-Cell communication	5.73 E-03	PLEC, ACTG, ACTN1, LIMS1, ILK
Aggrephagy	1.21E-02	TBB6, TBB2A, TBB8
Platelet activation, signalling and aggregation	1.21E-02	WDR1, ACTN1, ISLR, GBB2, GBB1

Table 2: Significantly downregulated pathways in tumour samples. Generated from Reactome.

Pathway name	False discovery rate	Submitted proteins
Erythrocytes take up oxygen and release carbon dioxide	9.99E-09	CAH1, HBB, HBA, B3AT, CAH2
Erythrocytes take up carbon dioxide and release oxygen	2.51E-08	CAH1, HBB, HBA, B3AT, CAH2
O2/CO2 exchange in erythrocytes	2.51E-08	CAH1, HBB, HBA, B3AT, CAH2
Metabolism of porphyrins	4.17E-03	HEM2, BLVRB, HEM3
Reversible hydration of carbon dioxide	2.33E-02	CAH1, CAH2
Heme biosynthesis	3.03E-02	HEM2, HEM3

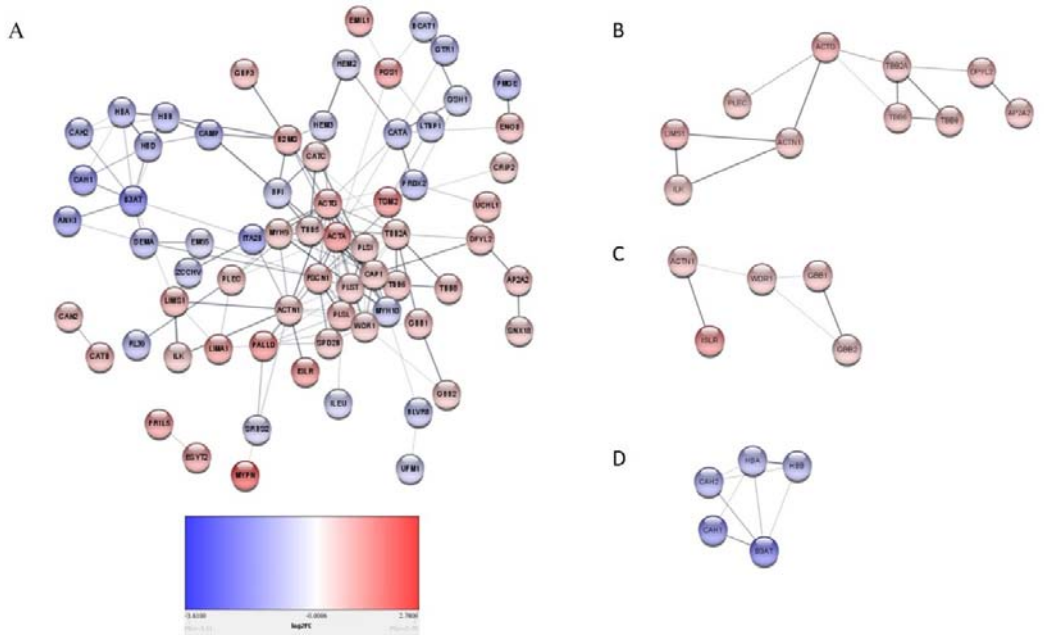


Figure 4: (A) The interaction network of the different dysregulated proteins. Proteins involved in Extracellular matrix formation/organisation (B) (Recycling pathway of L1, cell-extracellular matrix interactions, cell junction organisation, cell-cell communication, Platelet activation(C), and O_2/CO_2 transport(D) are also shown.

Single nucleotide variants observed in tumours

The quality-controlled dataset contained 369, 259 SNVs for the 15 paired samples. Although the study was underpowered for detecting statistically significant associations due to the small sample size, we performed a simple linear regression association test on the complete dataset (all tumour versus normal adjacent) to identify SNVs with a difference in allele frequency between the tumour and normal adjacent groups. Top associated results are listed with p-values, although none of these was significant after correcting for multiple testing (**Table S3**).

To assess the possible role of SNVs in the regulation of protein levels, the association analysis was re-run on a subset of SNVs extracted based on the gene regions for the previously identified DEPs. The association results were assessed to identify SNVs with a

difference in allele frequency between cases (tumour) and controls (normal adjacent). Extraction of SNVs in the gene regions identified from the proteomics results (significantly up and down-regulated proteins) resulted in 912 SNVs. A linear regression association test with this subset of SNVs was conducted and, although no significant difference in allele frequency was observed between the tumour and normal adjacent groups, SNVs were found within some DEPs (**Table S3**). We further annotated the types of SNVs (**Figure S4**). Sixty-nine percent of the coding SNVs found in DEPs were synonymous variants, while 31% were missense variants. Missense variants were found in *MYPN*, *SERPINB8* (SPB8) and *ESYT2*, which were among the top ten most upregulated proteins in the tumour group.

Discussion

In this study, we performed proteomic and genomic profiling and identified proteins dysregulated in resected pancreatic tumours obtained from patients of African ancestry. For the first time, to our knowledge, dysregulation of proteins such as EMIL1, KBTB2, and ZCCHV, that are responsible for the regulation of ECM components, were observed in pancreatic tumours. Additionally, the secretion of some differentiated proteins into the bloodstream allows for them to be validated in future studies as potential biomarkers. Pathway analysis further showed that the majority of the DEPs were enriched in pathways responsible for, and related to extracellular matrix formation (**Figure 2**).

Dysregulated proteins implicated in extracellular matrix formation

In this study, we observed that pathways involved in ECM formation and interactions were dysregulated corroborating the findings of several studies (31–34). The ECM modulates intracellular signalling pathways; consequently, aberrant ECM homeostasis and ECM

remodelling can induce and enhance tumorigenesis. Furthermore, there is evidence that the altered expression of ECM components can affect intracellular signalling (35,36).

The enrichment of ECM-related pathways (Recycling pathway of L1, cell-extracellular matrix interactions, cell junction organisation, cell-cell communication) was due to the upregulation of specific proteins, such as PLEC, ACTN1, LIMS1, and ILK (**Table 1, Figure 4(B)**). We recently showed that these proteins were also overexpressed at the mRNA level (37) with increased activity of the MAPK and P13K/AKT signalling pathways (37). Both the MAPK and P13K/AKT signalling pathways are regulated by FAK/SRC activation which results from the interaction of integrin with ECM molecules (38). Drawing from our studies and literature, **Figure 5** shows a schematic of how these pathways interplay in the ECM.

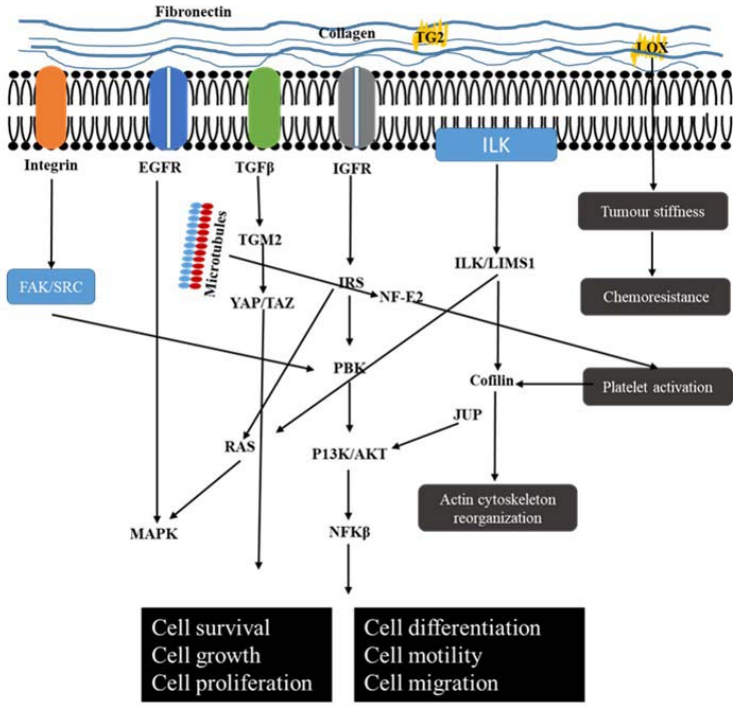


Figure 5: The schematic interplay between the extracellular matrix and intracellular signalling pathways.

The identified dysregulated proteins play important roles in tumorigenesis. PLEC (Plectin) was further found to be up-regulated in squamous cell carcinoma, aiding in the induction of malignant transformation from sinonasal inverted papilloma (39) and has been utilised in pancreatic cancer detection using targeted nanoparticles (40–42). High expression of ACTN1(Actinin Alpha 1) promotes cellular migration, invasion and metastasis, by weakening E-cadherin adhesion that ensures cellular integrity (43). Additionally, ACTN1 has been identified as a key regulator of PDAC progression, using a multidimensional systems-level analysis approach (44). Increased LIMS1 (LIM and senescent cell antigen-like-containing domain protein 1), a focal adhesion protein, has been associated with poor prognosis in laryngeal and pancreatic cancers (45,46). Also, during the knockdown of LIMS1, apoptosis was induced in breast cancer cells (47). In PDAC, it was found to promote cell survival under hypoxic conditions by activating the AKT/mTOR signalling pathway (45). ILK (Integrin-linked kinase) which interacts with LIMS1 (48) was also overexpressed in PDAC tumours. One study using neuroblastoma cell lines found that silencing the LIMS1/ILK pathway reduced cellular proliferation, highlighting its potential as a therapeutic target (49).

We also observed the dysregulation of several other proteins, such as TGM2, EMIL1, KBTB2, and ZCCHV, related to the ECM but not enriched in the top significantly dysregulated pathways. Compared to normal adjacent tissues, tumours showed a 6.04-fold increase in Transglutaminase (TGM2) expression, an enzyme involved in extracellular matrix stiffness, by cross-linking collagen 1 fibres (50,51). In PDAC tissues, transforming growth factor-beta induces TGM2 expression resulting in the activation of the YAP/TAZ signalling pathway, promoting cellular proliferation, invasion and metastasis (51,52). EMIL1 is an extracellular matrix glycoprotein that has been identified in both primary and metastatic colon tumours (53). KBTB2 (Kelch repeat and BTB domain-containing protein 2) is part of

the kelch-repeat superfamily that functions in mediating the ubiquitination and degradation of target proteins. Importantly, members of this family are known to mediate cellular adhesion (54,55). One member of the family, keap1, was shown to target ectoplasmic specializations indicating interactions with ECM components (56). ZCCHV (or ZAP) functions in inhibiting viral replication by preventing viral mRNA accumulation (57,58). The anti-viral function of ZAP is regulated by Matrin 3, a member of the matrix metalloproteins known to be involved in the degradation of ECM components during tissue and embryo development (59,60). This study also found missense mutations in several upregulated proteins, including ESTY2. ESTY2 plays a role in cytoskeleton organisation, which may impact ECM and potentially the overall biology of PDAC cells. In lung cancer cells (61), the short and long variants of ESYT2 were implicated in the cortical distribution of actin and endocytosis, respectively.

Furthermore, we observed that upregulation of platelet activation, signalling and aggregation in tumours (Table 1, Figure 4(C)). Platelet activation results in the release of pro-angiogenic factors (such as VEGF and FGF) and pro-inflammatory markers (such as IL-8 and members of the C–X–C motif ligand family), inducing angiogenesis and inflammation, respectively (62–64). Of interest, pro-inflammatory factors contribute to the strengthening of the extracellular matrix in the tumour microenvironment (65). One study showed that elevated mean platelet volume (MPV) correlated with poor survival in pancreatic cancer patients (66). A major characteristic of platelet activation is the reorganisation of the actin cytoskeleton by cofilin-1 induction (67). Importantly, WDR1, which was overexpressed in this study, enhances the function of cofilin-1 (68,69).

In the present study, the downregulation of pathways involved in oxygen transport indicative of hypoxia (**Table 2, Figure 4(D)**) was observed. Hypoxia has been linked to tumour stroma stiffness that is characteristic of the extracellular matrix of pancreatic tumours. Tumour stiffness induces epithelial to mesenchymal transition, promoting chemo-resistance. This

stiffness is primarily achieved when collagen fibres are cross-linked by lysyl oxidase (70), which is increased during hypoxia (Cox et al., 2013; Miller et al., 2015). Additionally, hypoxia is associated with poor survival in pancreatic cancer patients, promoting invasion and metastasis, even in the early stages of the disease, and enabling therapeutic resistance (74).

Upregulation of proteins involved in cell division and metastasis

Several β-tubulin subtypes including TBB2A, TBB6 and TBB8 were significantly overexpressed in tumour samples (**Table 1)**. β-tubulin combine with α-tubulin to form microtubules, which play crucial roles in mitotic cell division (75) and cell adhesion (76). Drugs binding β-tubulin, such as colchicine and paclitaxel, are known to kill cancer cells (77). Increased expression and mutations in β-tubulin subtypes exacerbate drug resistance in cancer (78–81). Additionally, β-tubulin is expressed in platelets, aiding in their formation via the NF-E2 pathway (82).

Conclusion

We have shown differentially expressed proteins in resected pancreatic ductal adenocarcinoma tumours obtained from patients of African ancestry using SWATH-MS. Many of these proteins are involved in cell proliferation and in the extracellular matrix formation/organisation, which plays a role in tumour progression and chemo-resistance. Targeting these proteins could be of effective therapeutic potential. Going forward, we plan to expand on the discovery dataset described here and further investigate the DEPs in a larger sample cohort. This would include analyses of biopsies at a single-cell level and plasma

proteomics. In particular, the detection of secretory proteins in the bloodstream of PDAC patients may be beneficial in prognosis of the disease.

Funding

The study was funded by the Council for Scientific and Industrial Research and the South African Medical Research Council through a grant awarded to the Wits Common Epithelial Cancer Research Centre. DLT was supported in part by a Strategic Health Innovation Partnership (SHIP) grant from the South African (SA) Department of Science and Technology (DST) and the SA Medical Research Council (SAMRC) to Gerhard Walzl. GC is funded by The Cancer Association of South Africa (CANSA).

Conflict of interest

The authors declare no conflict of interest.

Data availability

The data files described in this study have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository (83) with the dataset identifier (PXD019549).

References

1. Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians. 2018 Nov 1;68(6):394–424.

2. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68(1):7–30.

3. Khawja SN, Mohammed S, Silberfein EJ, Musher BL, Fisher WE, Van Buren GI. Pancreatic Cancer Disparities in African Americans. Pancreas. 2015 May;44(4):522–527.

4. Silverman DT, Hoover RN, Brown LM, Swanson GM, Schiffman M, Greenberg RS, et al. Why do Black Americans have a higher risk of pancreatic cancer than White Americans? Epidemiology. 2003 Jan;14(1):45–54.

5. Vick AD, Hery DN, Markowiak SF, Brunicardi FC. Closing the Disparity in Pancreatic Cancer Outcomes: A Closer Look at Nonmodifiable Factors and Their Potential Use in Treatment. Pancreas. 2019;48(2):242–9.

6. Chu GC, Kimmelman AC, Hezel AF, DePinho RA. Stromal biology of pancreatic cancer. J Cell Biochem. 2007 Jul 1;101(4):887–907.

7. Hwang HJ, Oh M-S, Lee DW, Kuh H-J. Multiplex quantitative analysis of stroma-mediated cancer cell invasion, matrix remodeling, and drug response in a 3D co-culture model of pancreatic tumor spheroids and stellate cells. J Exp Clin Cancer Res. 2019 Jun 14;38(1):258.

8. Murakami T, Hiroshima Y, Matsuyama R, Homma Y, Hoffman RM, Endo I. Role of the tumor microenvironment in pancreatic cancer. Ann Gastroenterol Surg. 2019 Jan 4;3(2):130–7.

9. Sommariva M, Gagliano N. E-Cadherin in Pancreatic Ductal Adenocarcinoma: A Multifaceted Actor during EMT. Cells [Internet]. 2020 Apr 22 [cited 2020 Jun 1];9(4). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7226001/

10. Uzunparmak B, Sahin IH. Pancreatic cancer microenvironment: a current dilemma. Clin Transl Med [Internet]. 2019 Jan 15 [cited 2020 May 28];8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6333596/

11. Amrutkar M, Aasrum M, Verbeke CS, Gladhaug IP. Secretion of fibronectin by human pancreatic stellate cells promotes chemoresistance to gemcitabine in pancreatic cancer cells. BMC Cancer. 2019 Jun 17;19(1):596.

12. Grzesiak JJ, Bouvet M. The alpha2beta1 integrin mediates the malignant phenotype on type I collagen in pancreatic cancer cell lines. Br J Cancer. 2006 May 8;94(9):1311–9.

13. Miyamoto H, Murakami T, Tsuchida K, Sugino H, Miyake H, Tashiro S. Tumor-stroma interaction of human pancreatic cancer: acquired resistance to anticancer drugs and proliferation regulation is dependent on extracellular matrix proteins. Pancreas [Internet]. 2004;28. Available from: http://dx.doi.org/10.1097/00006676-200401000-00006

14. Procacci P, Moscheni C, Sartori P, Sommariva M, Gagliano N. Tumor–Stroma Cross-Talk in Human Pancreatic Ductal Adenocarcinoma: A Focus on the Effect of the Extracellular Matrix on Tumor Cell Phenotype and Invasive Potential. Cells [Internet]. 2018 Oct 5 [cited 2020 Jun 1];7(10). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6209911/

15. Kim K, Ahn S, Lim J, Yoo BC, Hwang J-H, Jang W. Detection of Pancreatic Cancer Biomarkers Using Mass Spectrometry. Cancer Inform. 2015 Jan 6;13(Suppl 7):45–53.

16. Ludwig C, Gillet L, Rosenberger G, Amon S, Collins BC, Aebersold R. Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial. Mol Syst Biol. 2018 13;14(8):e8126.

17. Buthelezi SG, Dirr HW, Chakauya E, Chikwamba R, Martens L, Tsekoa TL, et al. The study of degradation mechanisms of glyco-engineered plant produced anti-rabies monoclonal antibodies E559 and 62-71-3. PLoS ONE. 2018;13(12):e0209373.

18. Chambers MC, Maclean B, Burke R, Amodei D, Ruderman DL, Neumann S, et al. A cross-platform toolkit for mass spectrometry and proteomics. Nat Biotechnol. 2012 Oct;30(10):918–20.

19. Reiter L, Rinner O, Picotti P, Hüttenhain R, Beck M, Brusniak M-Y, et al. mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nature Methods. 2011 May;8(5):430–5.

20. Choi M, Chang C-Y, Clough T, Broudy D, Killeen T, MacLean B, et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics. 2014 Sep 1;30(17):2524–6.

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. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007 Sep;81(3):559–75.

23. Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, et al. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Research. 2017 Jan 4;45(D1):D183–9.

24. Fabregat A, Jupe S, Matthews L, Sidiropoulos K, Gillespie M, Garapati P, et al. The Reactome Pathway Knowledgebase. Nucleic Acids Res. 2018 04;46(D1):D649–55.

25. P S, A M, O O, Ns B, Jt W, D R, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks [Internet]. Vol. 13, Genome research. Genome Res; 2003 [cited 2020 May 20]. Available from: https://pubmed.ncbi.nlm.nih.gov/14597658/

26. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017 Jan 4;45(D1):D362–8.

27. Ge SX, Jung D, Yao R. ShinyGO: a graphical enrichment tool for animals and plants. Bioinformatics. 2019 Dec 27;

28. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, et al. The Ensembl Variant Effect Predictor. Genome Biol. 2016 Jun 6;17(1):122.

29. Thul PJ, Lindskog C. The human protein atlas: A spatial map of the human proteome. Protein Sci. 2018;27(1):233–44.

30. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019 Jan 8;47(Database issue):D506–15.

31. Hosein AN, Brekken RA, Maitra A. Pancreatic cancer stroma: an update on therapeutic targeting strategies. Nature Reviews Gastroenterology & Hepatology. 2020 May 11;1–19.

32. Tian C, Öhlund D, Rickelt S, Lidström T, Huang Y, Hao L, et al. Cancer Cell-Derived Matrisome Proteins Promote Metastasis in Pancreatic Ductal Adenocarcinoma. Cancer Res. 2020 Apr 1;80(7):1461–74.

33. Tian C, Clauser KR, Öhlund D, Rickelt S, Huang Y, Gupta M, et al. Proteomic analyses of ECM during pancreatic ductal adenocarcinoma progression reveal different contributions by tumor and stromal cells. Proc Natl Acad Sci U S A. 2019 Sep 24;116(39):19609–18.

34. Weniger M, Honselmann KC, Liss AS. The Extracellular Matrix and Pancreatic Cancer: A Complex Relationship. Cancers (Basel) [Internet]. 2018 Sep 6 [cited 2020 Jan 7];10(9). Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6162452/

35. Hastings JF, Skhinas JN, Fey D, Croucher DR, Cox TR. The extracellular matrix as a key regulator of intracellular signalling networks. Br J Pharmacol. 2019;176(1):82–92.

36. Skhinas JN, Cox TR. The interplay between extracellular matrix remodelling and kinase signalling in cancer progression and metastasis. Cell Adh Migr. 2018;12(6):529–37.

37. Nweke E, Ntwasa M, Brand M, Devar J, Smith M, Candy G. Increased expression of plakoglobin is associated with upregulated MAPK and PI3K/AKT signalling pathways in early resectable pancreatic ductal adenocarcinoma. Oncology Letters [Internet]. 2020 Mar 23 [cited 2020 Mar 25]; Available from: http://www.spandidos-publications.com/10.3892/ol.2020.11473/abstract

38. Mitra SK, Hanson DA, Schlaepfer DD. Focal adhesion kinase: in command and control of cell motility. Nat Rev Mol Cell Biol. 2005 Jan;6(1):56–68.

39. Yang Z, Zhang Y, Wang X, Huang J, Guo W, Wei P, et al. Putative biomarkers of malignant transformation of sinonasal inverted papilloma into squamous cell carcinoma. J Int Med Res. 2019 Jun;47(6):2371–80.

40. Chen X, Zhou H, Li X, Duan N, Hu S, Liu Y, et al. Plectin-1 Targeted Dual-modality Nanoparticles for Pancreatic Cancer Imaging. EBioMedicine. 2018 Apr;30:129–37.

41. Kelly KA, Bardeesy N, Anbazhagan R, Gurumurthy S, Berger J, Alencar H, et al. Targeted nanoparticles for imaging incipient pancreatic ductal adenocarcinoma. PLoS Med. 2008 Apr 15;5(4):e85.

42. Yang L, Mao H, Cao Z, Wang YA, Peng X, Wang X, et al. Molecular imaging of pancreatic cancer in an animal model using targeted multifunctional nanoparticles. Gastroenterology. 2009 May;136(5):1514-1525.e2.

43. Kovac B, Mäkelä TP, Vallenius T. Increased α-actinin-1 destabilizes E-cadherin-based adhesions and associates with poor prognosis in basal-like breast cancer. PLoS ONE. 2018;13(5):e0196986.

44. Rajamani D, Bhasin MK. Identification of key regulators of pancreatic cancer progression through multidimensional systems-level analysis. Genome Med. 2016 03;8(1):38.

45. Huang C, Li Y, Li Z, Xu Y, Li N, Ge Y, et al. LIMS1 Promotes Pancreatic Cancer Cell Survival under Oxygen-Glucose Deprivation Conditions by Enhancing HIF1A Protein Translation. Clin Cancer Res. 2019 Jul 1;25(13):4091–103.

46. Tsinias G, Nikou S, Papadas T, Pitsos P, Papadaki H, Bravou V. High PINCH1 Expression in Human Laryngeal Carcinoma Associates with Poor Prognosis. Anal Cell Pathol (Amst). 2018;2018:2989635.

47. Bhatia A, Muthusamy S, Giridhar K, Goel S. Knockdown of PINCH-1 protein sensitizes the estrogen positive breast cancer cells to chemotherapy induced apoptosis. Pathol Res Pract. 2018 Feb;214(2):290–5.

48. Tu Y, Li F, Goicoechea S, Wu C. The LIM-only protein PINCH directly interacts with integrin-linked kinase and is recruited to integrin-rich sites in spreading cells. Mol Cell Biol. 1999 Mar;19(3):2425–34.

49. Saeki N, Saito A, Sugaya Y, Amemiya M, Ono H, Komatsuzaki R, et al. Chromatin Immunoprecipitation and DNA Sequencing Identified a LIMS1/ILK Pathway Regulated by LMO1 in Neuroblastoma. Cancer Genomics Proteomics. 2018 Jun;15(3):165–74.

50. Huang L, Xu A-M, Liu W. Transglutaminase 2 in cancer. American Journal of Cancer Research. 2015;5(9):2756–76.

51. Lee J, Condello S, Yakubov B, Emerson R, Caperell-Grant A, Hitomi K, et al. Tissue Transglutaminase Mediated Tumor-Stroma Interaction Promotes Pancreatic Cancer Progression. Clin Cancer Res. 2015 Oct 1;21(19):4482–93.

52. Verderio E a. M, Johnson T, Griffin M. Tissue transglutaminase in normal and abnormal wound healing: review article. Amino Acids. 2004 Jul;26(4):387–404.

53. Naba A, Clauser KR, Whittaker CA, Carr SA, Tanabe KK, Hynes RO. Extracellular matrix signatures of human primary metastatic colon cancers and their metastases to liver. BMC Cancer. 2014 Jul 18;14:518.

54. Heiska L, Carpén O. Src Phosphorylates Ezrin at Tyrosine 477 and Induces a Phosphospecific Association between Ezrin and a Kelch-Repeat Protein Family Member. J Biol Chem. 2005 Mar 18;280(11):10244–52.

55. Velichkova M, Guttman J, Warren C, Eng L, Kline K, Vogl AW, et al. A human homologue of Drosophila kelch associates with myosin-VIIa in specialized adhesion junctions. Cell Motility. 2002;51(3):147–64.

56. Spence HJ, Johnston I, Ewart K, Buchanan SJ, Fitzgerald U, Ozanne BW. Krp1, a novel kelch related protein that is involved in pseudopod elongation in transformed cells. Oncogene. 2000 Mar;19(10):1266–76.

57. Wang X, Tu F, Zhu Y, Gao G. Zinc-finger antiviral protein inhibits XMRV infection. PLoS One. 2012;7(6):e39159.

58. Zhu Y, Chen G, Lv F, Wang X, Ji X, Xu Y, et al. Zinc-finger antiviral protein inhibits HIV-1 infection by selectively targeting multiply spliced viral mRNAs for degradation. Proc Natl Acad Sci U S A. 2011 Sep 20;108(38):15834–9.

59. Erazo A, Goff SP. Nuclear matrix protein Matrin 3 is a regulator of ZAP-mediated retroviral restriction. Retrovirology. 2015 Jul 1;12(1):57.

60. Geurts N, Opdenakker G, Van den Steen PE. Matrix metalloproteinases as therapeutic targets in protozoan parasitic infections. Pharmacology & Therapeutics. 2012 Mar 1;133(3):257–79.

61. Miguel FJ de, Pajares MJ, Martínez-Terroba E, Ajona D, Morales X, Sharma RD, et al. A large□scale analysis of alternative splicing reveals a key role of QKI in lung cancer. Molecular Oncology. 2016 Nov;10(9):1437.

62. Gasic GJ, Gasic TB, Stewart CC. Antimetastatic effects associated with platelet reduction. Proc Natl Acad Sci USA. 1968 Sep;61(1):46–52.

63. Gay LJ, Felding-Habermann B. Contribution of platelets to tumour metastasis. Nat Rev Cancer. 2011 Feb;11(2):123–34.

64. Labelle M, Begum S, Hynes RO. Direct signaling between platelets and cancer cells induces an epithelial-mesenchymal-like transition and promotes metastasis. Cancer Cell. 2011 Nov 15;20(5):576–90.

65. Olsson AK, Cedervall J. The pro-inflammatory role of platelets in cancer. Platelets. 2018 Sep;29(6):569–73.

66. Yin J, Wang X, Zhang X, Liu L, Wang R. Mean platelet volume predicts survival in pancreatic cancer patients with synchronous liver metastases. Sci Rep [Internet]. 2018 Apr 16 [cited 2020 Jan 10];8. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5902615/

67. Dasgupta S, Le A, Haudek S, Entman M, Thiagarajan P. Cofilin-1 – Induced Actin Reorganization and Phosphatidylserine Exposure in Platelets. Blood. 2014 Dec 6;124(21):4153–4153.

68. Dasgupta SK, Le A, Da Q, Cruz M, Rumbaut RE, Thiagarajan P. Wdr1-Dependent Actin Reorganization in Platelet Activation. PLoS ONE. 2016;11(9):e0162897.

69. Rodal AA, Tetreault JW, Lappalainen P, Drubin DG, Amberg DC. Aip1p interacts with cofilin to disassemble actin filaments. J Cell Biol. 1999 Jun 14;145(6):1251–64.

70. Smith-Mungo LI, Kagan HM. Lysyl oxidase: properties, regulation and multiple functions in biology. Matrix Biol. 1998 Feb;16(7):387–98.

71. Cox TR, Bird D, Baker A-M, Barker HE, Ho MW-Y, Lang G, et al. LOX-mediated collagen crosslinking is responsible for fibrosis-enhanced metastasis. Cancer Res. 2013 Mar 15;73(6):1721–32.

72. Erler JT, Bennewith KL, Nicolau M, Dornhöfer N, Kong C, Le Q-T, et al. Lysyl oxidase is essential for hypoxia-induced metastasis. Nature. 2006 Apr 27;440(7088):1222–6.

73. Miller BW, Morton JP, Pinese M, Saturno G, Jamieson NB, McGhee E, et al. Targeting the LOX/hypoxia axis reverses many of the features that make pancreatic cancer deadly: inhibition of LOX abrogates metastasis and enhances drug efficacy. EMBO Mol Med. 2015 Aug;7(8):1063–76.

74. Erkan M, Kurtoglu M, Kleeff J. The role of hypoxia in pancreatic cancer: a potential therapeutic target? Expert Rev Gastroenterol Hepatol. 2016;10(3):301–16.

75. Parker AL, Kavallaris M, McCarroll JA. Microtubules and their role in cellular stress in cancer. Front Oncol. 2014;4:153.

76. Bershadsky A, Chausovsky A, Becker E, Lyubimova A, Geiger B. Involvement of microtubules in the control of adhesion-dependent signal transduction. Current Biology. 1996 Oct 1;6(10):1279–89.

77. Zhou J, Giannakakou P. Targeting microtubules for cancer chemotherapy. Curr Med Chem Anticancer Agents. 2005 Jan;5(1):65–71.

78. Huzil JT, Chen K, Kurgan L, Tuszynski JA. The roles of beta-tubulin mutations and isotype expression in acquired drug resistance. Cancer Inform. 2007 Apr 27;3:159–81.

79. Kelley MJ, Li S, Harpole DH. Genetic Analysis of the β-Tubulin Gene, TUBB, in Non-Small-Cell Lung Cancer. J Natl Cancer Inst. 2001 Dec 19;93(24):1886–8.

80. Leandro-García LJ, Leskelä S, Landa I, Montero-Conde C, López-Jiménez E, Letón R, et al. Tumoral and tissue-specific expression of the major human beta-tubulin isotypes. Cytoskeleton (Hoboken). 2010 Apr;67(4):214–23.

81. Nami B, Wang Z. Genetics and Expression Profile of the Tubulin Gene Superfamily in Breast Cancer Subtypes and Its Relation to Taxane Resistance. Cancers (Basel). 2018 Aug 18;10(8).

82. Lecine P, Italiano JE, Kim SW, Villeval JL, Shivdasani RA. Hematopoietic-specific beta 1 tubulin participates in a pathway of platelet biogenesis dependent on the transcription factor NF-E2. Blood. 2000 Aug 15;96(4):1366–73.

83. Vizcaíno JA, Côté RG, Csordas A, Dianes JA, Fabregat A, Foster JM, et al. The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. Nucleic Acids Res. 2013 Jan;41(Database issue):D1063-1069.

Supplementary figures

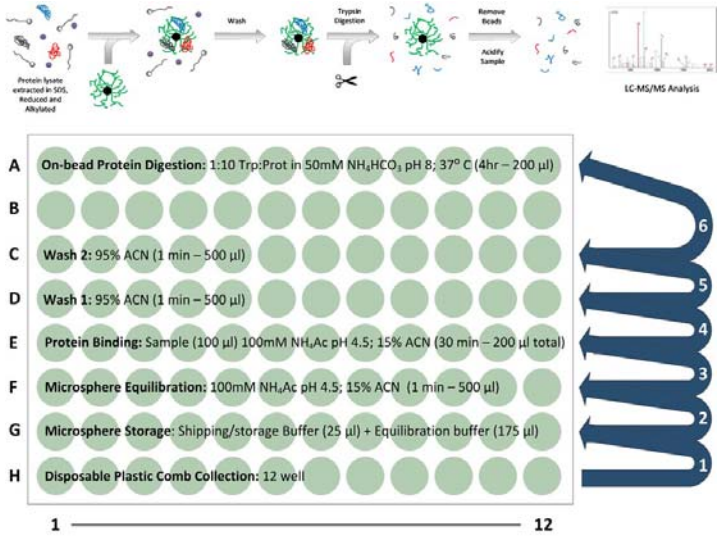


Figure S1: On-bead based protein capture, clean-up and digestion: Top: proteins are captured on magnetic hydrophilic affinity microparticles, followed by a high-organic wash to remove contaminants and on-bead digestion by addition of sequencing grade trypsin**.** Bottom: Plate set-up for automated sample processing in a Thermo Fisher Scientific KingFisher™ Duo Magnetic handling station.

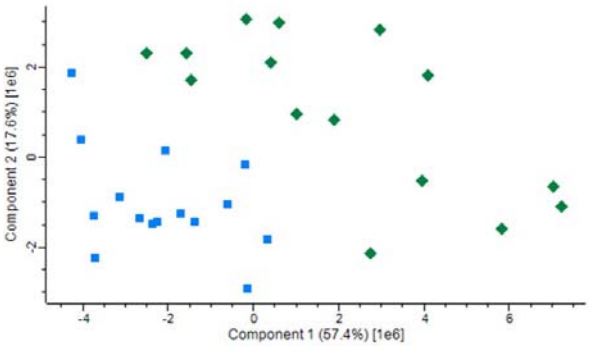


Figure S2: Principal component analysis (PCA) showing tumour samples as blue squares and normal adjacent samples as green diamonds. The PCA plot was generated using peptide abundance data of all peptides analysed per sample.

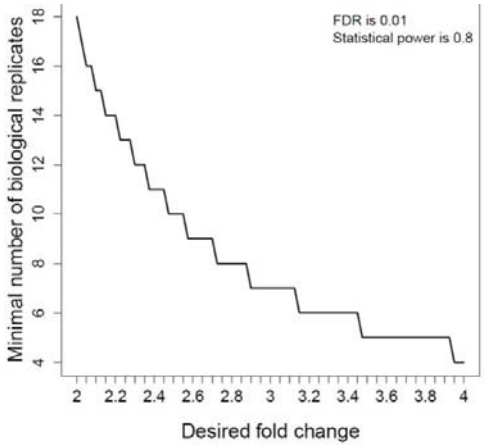
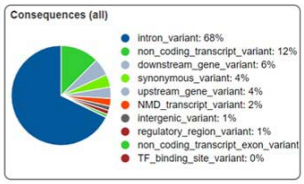
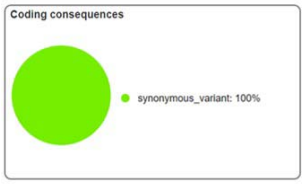


Figure S3: Sample size plot showing the relationship between the number of biological replicates and protein level fold change that can be distinguished based on the current dataset. A 2.1 fold change can be distinguished in the current study (biological replicates = 15) with statistical power set to 0.8 and false discovery rate (FDR) set to 0.01.

A

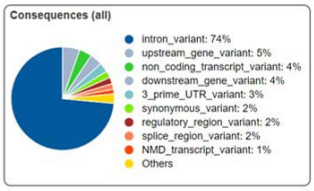
Category	Count
Variants processed	53
Variants filtered out	0
Novel / existing variants	0 (0.0) / 53 (100.0)
Overlapped genes	87
Overlapped transcripts	565
Overlapped regulatory features	10





B

Category	Count
Variants processed	185
Variants filtered out	0
Novel / existing variants	0 (0.0) / 185 (100.0)
Overlapped genes	76
Overlapped transcripts	376
Overlapped regulatory features	55



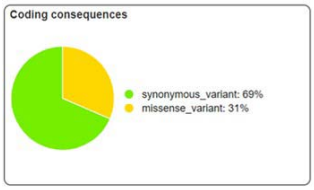


Figure S4: Types of variants observed from the list of SNVs found in (A) All Tumour vs Normal adjacent and (B) Tumour vs Normal adjacent in differentially expressed proteins only

Supplementary tables

Table S1: Characteristics of patients recruited for the study

Patient	Gender	Staging	TNM
1	Female	IIA	T2N1M0
2	Male	IIA	T2N1M0
3	Male	IIA	T2N1M0
4	Male	IA	T1N0N0
5	Female	IIB	T3N1M0
6	Male	IIA	T2N1M0
7	Male	IIA	T2N1M0
8	Male	IIB	T3N1M0
9	Female	IA	T1N0N0
10	Male	IIA	T2N1M0
11	Male	IIB	T3N1M0
12	Female	IIA	T2N1M0
13	Female	IIB	T3N1M0
14	Female	IIB	T3N1M0
15	Female	IIB	T3N1M0

TNM: TNM Classification of Malignant Tumors

Table S2: Differentially expressed proteins (fold change ≥2.1) measured in group comparison. Positive fold change describes proteins upregulated in the tumour group and negative fold change describes proteins downregulated in tumour compared to the normal adjacent group.

Uni-prot ID	Uniprot Name	Fold change	Adjusted p-value	Location	Cancer (prognostic) biomarker status	Protein atlas reference link
Q86TC9	MYPN_HUMAN	6,90	4,71E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000138347-MYPN
P21980	TGM2_HUMAN	6,04	2,63E-03	Intracellular	Unfavourable prognostic marker for renal and pancreatic cancers	https://www.proteinatlas.org/ENSG00000198959-TGM2
P62736	ACTA_HUMAN	5,96	9,81E-03	Intracellular	Unfavourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000107796-ACTA2
Q8WX93	PALLD_HUMAN	5,65	7,93E-03	Intracellular	Unfavourable prognostic marker for urothelial cancer	https://www.proteinatlas.org/ENSG00000129116-PALLD
P50452	SPB8_HUMAN	5,47	4,71E-03	Intracellular	Unfavourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000166401-SERPINB8

O14498	ISLR_HUMAN	5,29	9,15E-03	Secreted	Unfavourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000129009-ISLR
P21810	PGS1_HUMAN	4,81	7,05E-03	Intracellular	Unfavourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000182492-BGN
A0AVI2	FR1L5_HUMAN	4,69	9,02E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000249715-FER1L5
Q9UHB6	LIMA1_HUMAN	4,52	7,80E-03	Intracellular	Prognostic marker for head and neck cancers (unfavourable) and for renal cancer (favourable)	https://www.proteinatlas.org/ENSG00000050405-LIMA1
A0FGR8	ESYT2_HUMAN	4,00	3,16E-03	Membrane	Unfavourable prognostic marker for urothelial and pancreatic cancers	https://www.proteinatlas.org/ENSG00000117868-ESYT2
Q9Y6C2	EMIL1_HUMAN	3,88	9,16E-03	Intracellular	Unfavourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000138080-EMILIN1
P63261	ACTG_HUMAN	3,72	1,66E-03	Intracellular	Favourable prognostic marker for colorectal cancer	https://www.proteinatlas.org/ENSG00000184009-ACTG1
Q9NRN5	OLFL3_HUMAN	3,47	2,16E-03	Intracellular/Secreted	Unfavourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000116774-OLFML3
P48059	LIMS1_HUMAN	3,33	6,84E-03	Intracellular	Unfavourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000169756-LIMS1
P61769	B2MG_HUMAN	3,22	8,22E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000166710-B2M
P09936	UCHL1_HUMAN	3,15	6,58E-03	Intracellular	Unfavourable prognostic marker for urothelial and endometrial cancers	https://www.proteinatlas.org/ENSG00000154277-UCHL1
P23634	AT2B4_HUMAN	3,07	9,73E-03	Intracellular	Unfavourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000058668-ATP2B4
Q16658	FSCN1_HUMAN	3,05	3,61E-03	Intracellular	Unfavourable prognostic marker for renal, lung, head and neck cancers	https://www.proteinatlas.org/ENSG00000075618-FSCN1

P07858	CATB_HUMAN	3,03	7,37E-03	Intracellular	Prognostic marker for thyroid (favourable) and urothelial (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000164733-CTSB
P13796	PLSL_HUMAN	2,99	5,14E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000136167-LCP1
Q13308	PTK7_HUMAN	2,89	8,64E-03	Intracellular/Secreted/ Membrane	Not prognostic	https://www.proteinatlas.org/ENSG00000112655-PTK7
Q9H0R5	GBP3_HUMAN	2,86	4,44E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000117226-GBP3
Q16555	DPYL2_HUMAN	2,85	8,79E-04	Intracellular/Membrane	Unfavourable prognostic marker for urothelial cancer	https://www.proteinatlas.org/ENSG00000092964-DPYSL2
P13929	ENOB_HUMAN	2,83	7,11E-04	Intracellular	Unfavourable prognostic marker for colorectal cancer	https://www.proteinatlas.org/ENSG00000108515-ENO3
Q14141	SEPT6_HUMAN	2,83	2,27E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000125354-SEPT6
Q3ZCM7	TBB8_HUMAN	2,56	3,83E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000261456-TUBB8
Q14764	MVP_HUMAN	2,53	2,47E-03	Intracellular	Favourable prognostic marker for renal and breast cancers	https://www.proteinatlas.org/ENSG00000013364-MVP
P17655	CAN2_HUMAN	2,52	3,21E-03	Intracellular	Prognostic marker for renal (favourable), urothelial and pancreatic (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000162909-CAPN2
P12814	ACTN1_HUMAN	2,49	8,79E-04	Intracellular	Unfavourable prognostic marker for renal, lung, urothelial, head and neck cancers	https://www.proteinatlas.org/ENSG00000072110-ACTN1
P62873	GBB1_HUMAN	2,49	3,41E-04	Intracellular/Secreted	Unfavourable prognostic marker for liver cancer	https://www.proteinatlas.org/ENSG00000078369-GNB1
O94973	AP2A2_HUMAN	2,48	3,71E-04	Intracellular	Prognostic marker for liver (favourable) and pancreatic (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000183020-AP2A2

Q9UPN3	MACF1_HUMAN	2,43	1,66E-03	Intracellular	Favourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000127603-MACF1
Q9NZU5	LMCD1_HUMAN	2,43	7,16E-03	Intracellular	Unfavourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000071282-LMCD1
Q9BUF5	TBB6_HUMAN	2,43	1,53E-03	Intracellular	Unfavourable prognostic marker for renal and urothelial cancers	https://www.proteinatlas.org/ENSG00000176014-TUBB6
Q15149	PLEC_HUMAN	2,42	8,49E-04	Intracellular	Unfavourable prognostic marker for lung, colorectal and renal cancers	https://www.proteinatlas.org/ENSG00000178209-PLEC
Q13885	TBB2A_HUMAN	2,38	1,35E-03	Intracellular	Prognostic marker for renal (favourable), urothelial (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000137267-TUBB2A
Q01518	CAP1_HUMAN	2,34	6,56E-03	Intracellular	Unfavourable prognostic marker for liver cancer	https://www.proteinatlas.org/ENSG00000131236-CAP1
P35579	MYH9_HUMAN	2,31	9,71E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000100345-MYH9
P53634	CATC_HUMAN	2,29	8,65E-03	Intracellular	Prognostic marker for renal (favourable) and liver (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000109861-CTSC
P62879	GBB2_HUMAN	2,26	8,79E-04	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000172354-GNB2
O75083	WDR1_HUMAN	2,25	2,59E-04	Intracellular	Favourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000071127-WDR1
P13797	PLST_HUMAN	2,22	3,29E-03	Intracellular	Unfavourable prognostic marker for renal, thyroid, pancreatic, urothelial, head and neck cancers	https://www.proteinatlas.org/ENSG00000102024-PLS3
A1X283	SPD2B_HUMAN	2,21	2,64E-03	Intracellular	Unfavourable prognostic marker for liver, renal, urothelial, cancers	https://www.proteinatlas.org/ENSG00000174705-SH3PXD2B

Q13418	ILK_HUMAN	2,20	3,63E-03	Intracellular	Favourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000166333-ILK
P52943	CRIP2_HUMAN	2,20	9,71E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000182809-CRIP2
Q96RF0	SNX18_HUMAN	2,18	7,37E-03	Intracellular	Favourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000178996-SNX18
Q14651	PLSI_HUMAN	2,18	1,06E-03	Intracellular	Prognostic marker for renal (favourable) and pancreatic (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000120756-PLS1
Q9Y4F1	FARP1_HUMAN	2,16	5,55E-04	Intracellular	Favourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000152767-FARP1
P07437	TBB5_HUMAN	2,13	5,44E-03	Intracellular	Unfavourable prognostic marker for liver and renal cancers	https://www.proteinatlas.org/ENSG00000196230-TUBB
Q00013	EM55_HUMAN	-2,15	5,92E-03	Intracellular/Membrane	Favourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000130830-MPP1
P13716	HEM2_HUMAN	-2,15	8,59E-04	Intracellular/Secreted	Prognostic marker for liver, renal, endometrial (favourable) and colorectal (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000148218-ALAD
Q8N183	MIMIT_HUMAN	-2,19	1,66E-03	Intracellular	Unfavourable prognostic marker for liver and renal cancers	https://www.proteinatlas.org/ENSG00000164182-NDUFAF2
Q8IZ83	A16A1_HUMAN	-2,28	3,41E-04	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000161618-ALDH16A1
O94875	SRBS2_HUMAN	-2,32	5,14E-03	Intracellular	Favourable prognostic marker for endometrial, liver and renal cancer	https://www.proteinatlas.org/ENSG00000154556-SORBS2
P48506	GSH1_HUMAN	-2,38	1,70E-04	Intracellular	Prognostic marker for renal (favourable) and thyroid (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000001084-GCLC

P61960	UFM1_HUMAN	-2,42	9,71E-03	Intracellular	Prognostic marker for endometrial (favourable) and head and neck (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000120686-UFM1
P30740	ILEU_HUMAN	-2,42	9,71E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000021355-SERPINB1
P17213	BPI_HUMAN	-2,47	8,05E-03	Intracellular/Secreted	Not prognostic	https://www.proteinatlas.org/ENSG00000101425-BPI
P16219	ACADS_HUMAN	-2,52	7,80E-03	Intracellular	Favourable prognostic marker for renal, endometrial, liver cancers	https://www.proteinatlas.org/ENSG00000122971-ACADS
P08397	HEM3_HUMAN	-2,54	5,14E-03	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000256269-HMBS
P54687	BCAT1_HUMAN	-2,87	5,81E-03	Intracellular	Unfavourable prognostic marker for renal, stomach, urothelial, head and neck cancers	https://www.proteinatlas.org/ENSG00000060982-BCAT1
P35580	MYH10_HUMAN	-2,88	5,55E-04	Intracellular	Prognostic marker for renal (favourable) and urothelial (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000133026-MYH10
Q7Z2W4	ZCCHV_HUMAN	-2,90	5,40E-03	Intracellular	Prognostic marker for stomach (favourable) and liver (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000105939-ZC3HAV1
Q14766	LTBP1_HUMAN	-2,98	8,05E-03	Intracellular/Secreted	Unfavourable prognostic marker for renal and urothelial cancers	https://www.proteinatlas.org/ENSG00000049323-LTBP1
P10412	H14_HUMAN	-3,02	8,79E-04	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000168298-HIST1H1E
P30043	BLVRB_HUMAN	-3,23	1,40E-04	Intracellular	Unfavourable prognostic marker for pancreatic cancer	https://www.proteinatlas.org/ENSG00000090013-BLVRB
P04040	CATA_HUMAN	-3,28	4,46E-04	Intracellular	Favourable prognostic marker for renal and liver cancer	https://www.proteinatlas.org/ENSG00000121691-CAT

P62891	RL39_HUMAN	-3,35	8,05E-03	Intracellular/Secreted	Unfavourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000198918-RPL39
P68871	HBB_HUMAN	-3,67	1,92E-04	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000244734-HBB
P32119	PRDX2_HUMAN	-3,81	2,59E-04	Intracellular	Favourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000167815-PRDX2
Q08495	DEMA_HUMAN	-3,90	2,56E-03	Intracellular	Prognostic marker for renal (favourable) and glioma(unfavourable) cancers	https://www.proteinatlas.org/ENSG00000158856-DMTN
P11166	GTR1_HUMAN	-3,94	2,08E-03	Intracellular/Membrane	Unfavourable prognostic marker for renal, liver, lung, pancreatic and urothelial cancers	https://www.proteinatlas.org/ENSG00000117394-SLC2A1
P69905	HBA_HUMAN	-4,01	2,98E-04	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000206172-HBA1
P07738	PMGE_HUMAN	-4,12	8,59E-04	Intracellular	Favourable prognostic marker for renal and cervical cancers	https://www.proteinatlas.org/ENSG00000172331-BPGM
P02042	HBD_HUMAN	-4,27	1,70E-04	Intracellular/Membrane	Not prognostic	https://www.proteinatlas.org/ENSG00000223609-HBD
P00918	CAH2_HUMAN	-4,52	4,46E-04	Intracellular/Membrane	Favourable prognostic marker for renal cancer	https://www.proteinatlas.org/ENSG00000104267-CA2
P49913	CAMP_HUMAN	-4,89	6,40E-03	Secreted	Favourable prognostic marker for cervical cancer	https://www.proteinatlas.org/ENSG00000164047-CAMP
Q8IY47	KBTB2_HUMAN	-5,42	5,55E-04	Intracellular	Unfavourable prognostic marker for liver and cervical cancers	https://www.proteinatlas.org/ENSG00000170852-KBTBD2
Q96C19	EFHD2_HUMAN	-5,49	6,56E-03	Intracellular	Prognostic marker for urothelial, thyroid (favourable), lung and renal (unfavourable) cancers	https://www.proteinatlas.org/ENSG00000142634-EFHD2
P08514	ITA2B_HUMAN	-6,43	3,09E-03	Intracellular/Membrane	Not prognostic	https://www.proteinatlas.org/ENSG00000005961-ITGA2B

P16157	ANK1_HUMAN	-6,59	8,82E-04	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000029534-ANK1
P10720	PF4V_HUMAN	-7,14	1,40E-03	Secreted	Not prognostic	https://www.proteinatlas.org/ENSG00000109272-PF4V1
P00915	CAH1_HUMAN	-8,47	1,70E-04	Intracellular	Not prognostic	https://www.proteinatlas.org/ENSG00000133742-CA1
P02730	B3AT_HUMAN	-12,20	1,70E-04	Intracellular/Membrane	Not prognostic	https://www.proteinatlas.org/ENSG00000004939-SLC4A1

Table S3: Top single nucleotide polymorphisms (SNVs) and associated genes

Genes with SNVs (All tumour vs Normal adjacent)				**Genes with SNVs (significant differentially expressed proteins)**			
Associated Gene Name	**Chromo-some Name**	**Gene Start (bp)**	**Gene End (bp)**	**Associated Gene Name**	**Chromosome Name**	**Gene Start (bp)**	**Gene End (bp)**
PIN1P1	1	70385005	70386000	ESYT2	7	158523686	158622944
NFASC	1	204797779	204991950	SH3PXD2B	5	171752185	171881527
PITHD1	1	24104895	24114722	ISLR	15	74466012	74469213
LRRC7	1	70034081	70617628	WDR1	4	10075963	10118573
MTR	1	236958610	237067281	SORBS2	4	186506598	186877806
POU2F1	1	167190066	167396582	AP2A2	11	924894	1012239
PRKG1	10	52750945	54058110	CA1	8	86239837	86291243
AC005037.3	2	201827986	201873825	CA2	8	86376081	86393722
NDUFB3	2	201936156	201950473	HBD	11	5253908	5256600
AC079145.4	2	20189978	20203971	SLC4A1	17	42325753	42345509
DPP10	2	115199876	116603328	CAT	11	34460472	34493609
ACOXL	2	111490150	111875799	TUBB	HSCHR6_MHC_APD	30696223	30701448
CDCA7	2	174219548	174233725	TUBB	HSCHR6_MHC_QBL	30677512	30682738
TANC1	2	159825146	160089170	TUBB	HSCHR6_MHC_DBB	30678239	30683464
TTC7A	2	47143296	47303276	TUBB	HSCHR6_MHC_MANN	30732675	30737900
ARHGEF4	2	131594489	131804836	TUBB	HSCHR6_MHC_SSTO	30679427	30684652
ATG7	3	11313995	11599139	TUBB	HSCHR6_MHC_MCF	30766464	30771689
EPHB1	3	134316643	134979309	TUBB	6	30687978	30693203
GRM7	3	6811688	7783215	TUBB	HSCHR6_MHC_COX	30677842	30683067
PRKCI	3	169940153	170023769	BPGM	7	134331560	134364565

NRG1	8	31496902	32622548	CTSB	8	11700033	11726957
MVB12B	9	129089128	129269320	HMBS	11	118955576	118964259
CNTLN	9	17134980	17503921	HMBS	HG299_PATCH	118955575	118964258
				ITGA2B	17	42449548	42466873
				UCHL1	4	41258430	41270472
				HIST1H1E	6	26156559	26157343
				PF4V1	4	74718906	74719872
				SLC2A1	1	43391052	43424530
				ACTN1	14	69340860	69446157
				ALAD	9	116148597	116163613
				LCP1	13	46700055	46786006
				PLS3	X	114795501	114885181
				PLS3	HG1462_PATCH	114797644	114887331
				ENO3	17	4851387	4860426
				ANK1	8	41510739	41754280
				ACADS	12	121163538	121177811
				BPI	20	36888551	36965907
				CAPN2	1	223889295	223963720
				BGN	X	152760397	152775012
				BGN	HG1497_PATCH	152663411	152678026
				TGM2	20	36756863	36794980
				ATP2B4	1	203595689	203713209
				BLVRB	19	40953696	40971747
				SERPINB1	6	2832566	2842240
				PRDX2	19	12907634	12912694
				MYH9	22	36677327	36784063
				MYH10	17	8377523	8534079
				LIMS1	2	109150857	109303702
				GCLC	6	53362139	53481768
				CAMP	3	48264837	48266981
				SERPINB8	18	61637159	61672278
				CRIP2	14	105939299	105946499
				CRIP2	HG1592_PATCH	105939299	105946499
				CTSC	11	88026760	88070955

	BCAT1	12	24964295	25102393
	B2M	15	45003675	45011075
	UFM1	13	38923986	38937140
	ACTA2	10	90694831	90751147
	GNB1	1	1716729	1822495
	GNB2	7	100271154	100276797
	RPL39	X	118920467	118925606
	ACTG1	HG271_PATCH	79483771	79497647
	ACTG1	17	79476997	79490873
	HBB	11	5246694	5250625
	HBA2	16	222846	223709
	HBA1	16	226679	227521
	MPP1	X	154006959	154049282
	MPP1	HG1497_PATCH	153947152	153989475
	CAP1	1	40505905	40538321
	DMTN	8	21906506	21940038
	PTK7	6	43044006	43129457
	ILK	11	6624961	6632102
	TUBB2A	6	3153903	3157760
	44080	X	118749687	118827333
	PLS1	3	142315229	142432506
	MVP	16	29831715	29859355
	LTBP1	2	33172039	33624576
	PLEC	HG104_HG975_PATCH	144997317	145058904
	PLEC	8	144989321	145050902
	DPYSL2	8	26371791	26515694
	FSCN1	7	5632439	5646286
	TUBB8	HG905_PATCH	96892	100113
	TUBB8	10	92828	120103
	ZC3HAV1	7	138728266	138794465
	MYPN	10	69865912	69971774
	KBTBD2	7	32907784	32933743
	ALDH16A1	19	49956426	49974305
	NDUFAF2	5	60240956	60448853

	PALLD	4	169418217	169849608
	EFHD2	1	15736391	15756839
	SNX18	5	53813589	53842415
	TUBB6	18	12307668	12344319
	GBP3	1	89472349	89488577
	OLFML3	1	114522063	114524876
	LMCD1	3	8543393	8609805
	LIMA1	12	50569571	50677329
	MACF1	1	39546988	39952849
	FARP1	13	98794816	99102027
	EMILIN1	2	27301435	27309271