Published in final edited form as:

Proteomics. 2012 November ; 12(22): 3393–3402. doi:10.1002/pmic.201100663.

Validation of a robust proteomic analysis carried out on formalin-fixed paraffin-embedded tissues of the pancreas obtained from mouse and human

Kyoko Kojima^{1,2}, Gregory J. Bowersock², Chinatsu Kojima², Christopher A. Klug³, William E. Grizzle⁴, and James A. Mobley^{1,2}

¹Department of Surgery, University of Alabama at Birmingham, Birmingham, AL, USA

²The Comprehensive Cancer Center, University of Alabama at Birmingham, Birmingham, AL, USA

³Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL, USA

⁴Department of Pathology, University of Alabama at Birmingham, Birmingham, AL, USA

Abstract

A number of reports have recently emerged with focus on extraction of proteins from formalin-fixed paraffin-embedded (FFPE) tissues for MS analysis; however, reproducibility and robustness as compared to flash frozen controls is generally overlooked. The goal of this study was to identify and validate a practical and highly robust approach for the proteomics analysis of FFPE tissues. FFPE and matched frozen pancreatic tissues obtained from mice ($n = 8$) were analyzed using 1D-nanoLC-MS(MS)² following work up with commercially available kits. The chosen approach for FFPE tissues was found to be highly comparable to that of frozen. In addition, the total number of unique peptides identified between the two groups was highly similar, with 958 identified for FFPE and 1070 identified for frozen, with protein identifications that corresponded by approximately 80%. This approach was then applied to archived human FFPE pancreatic cancer specimens ($n = 11$) as compared to uninvolved tissues ($n = 8$), where 47 potential pancreatic ductal adenocarcinoma markers were identified as significantly increased, of which 28 were previously reported. Further, these proteins share strongly overlapping pathway associations to pancreatic cancer that include estrogen receptor α . Together, these data support the validation of an approach for the proteomic analysis of FFPE tissues that is straightforward and highly robust, which can also be effectively applied toward translational studies of disease.

Keywords

Animal proteomics; FPE; LC-MS; Nontagged proteomics; Pancreatic cancer; Systems biology

1 Introduction

Pancreatic ductal adenocarcinoma (PDAC) has an epithelial phenotype, and is accompanied and characterized by expansion of surrounding stroma [1–3]. PDAC is one of the most fatal

© 2012 WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim

Correspondence: Dr. James A. Mobley, University of Alabama at Birmingham, 1900 University Ave., Tinsley Harrison Tower Room 521, Birmingham, AL 35294, USA, mobleyja@uab.edu, **Fax:** +1-205-996-6940.

Colour Online: See the article online to view Fig. 1 in colour.

The authors have declared no conflict of interest.

malignancies in the United States, accounting for approximately 44 000 new cases and 37 000 deaths each year, with the recurrence rate of 50–85% following surgical resection [4]. The high death rate is due to the asymptomatic nature of the disease resulting in delayed diagnoses at an advanced stage when the cancer is resistant to conventional therapies [5]. Surgical resection at an early stage provides the only cure, but there are no early diagnostic tools currently available for PDAC. This strongly supports the need for novel diagnostics in addition to personalized targeted therapies.

FFPE (formalin-fixed paraffin-embedded) is a cost- and space-effective method to preserve tissues long term and as such, it has been a preferred technique to store tissues in clinical and basic science laboratories for decades. This has resulted in large archives of FFPE specimen biorepositories representing various diseases, which are ideal resources for biomarker discovery studies. Until recently, investigations of proteins in the FFPE tissues had been primarily carried out through antibody-based approaches that are highly specific yet are not suitable for studies involving the identification of previously unknown proteins associated with a disease. On the other hand, with the ability to identify hundreds of proteins in a medium- to high-throughput manner, MS has become the tool of choice for analysis of complex biological samples for the identification of potential disease-specific markers [6]. Unfortunately, the application of mass spectrometric technologies to FFPE tissues has been a challenge due to the high level of covalent protein–protein and/or protein– nucleotide cross-links caused by the formalin fixation process [7–10].

In the early 1990s, heat antigen retrieval (HAR) techniques were developed to increase protein antigenicity in FFPE tissues [11]. Although the mechanism of HAR is not completely understood, it has been suggested that it involves reversal of protein–protein and protein–nucleotide cross-links [12]. HAR was originally developed for and has been primarily applied to immunohistochemical staining of tissues. The technique has since been regularly utilized in other FFPE tissue-based applications including RNA/DNA extraction [13–15], whole tissue lysate analyses via SDS-PAGE, Western blotting [16], and more recently by MS [17–26]. Most of the latter studies focused on comparing protein extraction efficiencies in addition to MS compatibility. However, reproducibility and robustness is often not assessed, and comparison to serially obtained flash frozen controls is also generally overlooked.

In our current study, the focus was to verify a practical mass spectrometric/proteomic workflow that is straightforward, highly reproducible, and applicable to FFPE specimens. A commercially available kit that involves HAR was incorporated in the workflow for its ease of use, time effectiveness, and consistency. A highly reproducible nanoLC-MS(MS)² instrument was also utilized, which allowed for the analysis of complex biological samples in nanogram quantities for analysis of what is often a very limited amount of tissue derived from clinical specimens. In order to carry out this type of study, we utilized an animal model where fresh pancreatic tissues were halved and either flash frozen or formalin fixed followed by paraffin embedding in an established pathology laboratory. We then followed up by analyzing small sections of human FFPE specimens containing PDAC versus healthy adjacent tissues as proof-of-principle. This analysis was then compared to known protein markers associated with PDAC, as well as known pathways that together served to further verify the approach. The outcome of these experiments serves to validate this approach for MS-driven proteomic analysis of FFPE tissue utilized for clinical studies of disease.

2 Material and methods

2.1 Materials

Jupiter C18 5 μm 300 $\text{\AA}$ (Cat.# 04A-4053 Phenomenex). Fused-silica capillaries 100 μm id (Cat.# TSP100375 Polymicro). M-PER mammalian protein extraction reagent (Cat.# 78501 Thermo Scientific). Qproteome FFPE Tissue kit (Cat.# 37623 Qiagen). Novex[®] 4–12% Tris-Glycine gels (Cat.# EC6035BOX Invitrogen) and Colloidal Blue staining kit (Cat.# LC6025 Invitrogen). Dithiothreitol (Cat.# D9779 Sigma Aldrich) and iodoacetamide (Cat.# I1149 Sigma Aldrich). Trypsin Gold Mass Spectrometry Grade (Cat.# V5280 Promega). ACN (Cat.# A996SK Fisher Scientific). MilliQ purified water (MilliQ Synthesis System, Millipore).

2.2 Sample collection and preparation

Pancreatic tissues were excised from age-matched C57BL/6 mice ($n = 8$) at 3 months of age. Each tissue was cut in half with one-half immediately frozen in a cryovial on dry ice and kept at -80°C until sectioning, and the other half fixed in 4% paraformaldehyde in PBS for 4 h at room temperature (RT) and embedded in paraffin on the following day. Frozen and FFPE tissues were then sectioned at $2 \times 10 \mu\text{m}$ each and prepared for analysis. Human pancreatic tissues were collected in accordance with a protocol approved by the University of Alabama Institutional Review Board. Informed consent was obtained from all the patients. Following overnight fixation with 10% neutral buffered formalin, tissues were embedded in paraffin and stored at RT until use. Two 10- μm sections were prepared and processed for analyses by MS as described below in Section 2.3. All of the FFPE tissue processing and sectioning, as well as the standard H&E staining (separate 4- μm sections) were performed at the UAB Tissue Procurement Core Facility and evaluated by a trained GI pathologist (W.E.G.). For the human studies, serial sections were defined through careful histological examination by a pathologist as either containing $>50\%$ cancer ($n = 11$) or uninvolved tissues adjacent to pancreatitis ($n = 8$) (Supporting Information Table 1). The uninvolved tissues were obtained from pancreatitis patients in this study, due to the unavailability of uninvolved specimens from the matched cancer patients. The histology of the uninvolved tissue specimens resembled normal pancreas and did not contain the elements of pancreatitis (Supporting Information Fig. 1).

2.3 Protein extraction

Pancreatic tissues from mice were processed via two separate protocols: (i) frozen tissues by Mammalian Protein Extraction Reagent (M-PER) (frozen) and (ii) FFPE tissues by a Qproteome FFPE Tissue kit, both following the manufacturers' instructions. As an additional control, matched frozen tissues were similarly processed with Qproteome FFPE Tissue kit, without applying the deparaffinization step. Human FFPE pancreatic tissues were also processed using the Qproteome FFPE Tissue kit. Briefly, on the tissue processing of the FFPE specimens, two $3 \text{ mm} \times 3 \text{ mm} \times 10 \mu\text{m}$ sections per sample were deparaffinized with three changes of xylene (1 mL per wash) and then rehydrated with three changes each of 100%, 96%, and 70% ethanol (1 mL per wash) prior to protein extraction via HAR. Each wash consisted of 10-s vortexing, 10-min incubation at RT, and 2-min centrifugation at $14\,000 \times g$ at RT to remove the supernatant. HAR was performed in extraction buffer provided in the kit at 100°C for 20 min followed by 80°C for 2 h. Samples were centrifuged at $14\,000 \times g$, 4°C for 15 min to obtain the protein extracts in the supernatant. Extracts were quantified with a BCA Protein Assay Kit (Thermo Scientific).

2.4 Sample clean up with 1D PAGE and in-gel tryptic digestion

Five micrograms of protein extracts were denatured and reduced using Novex Tris-Glycine SDS sample buffer, and separated on 4–12% Novex Tris-Glycine gels at constant voltage of 125 V for 10 min, similar in fashion to an earlier study [27]. Gels were stained overnight with a Colloidal Blue Staining kit at RT with gentle agitation, and destained in MilliQ water for more than 3 h at RT the following day. Each lane was excised as one gel piece (Supporting Information Fig. 2), and in-gel tryptic digestion with the Trypsin Gold was performed following the manufacturer's instruction. Digests were reduced in volume to near dryness in a SpeedVac concentrator (Savant) and brought up to 25 μ L using 5% ACN/0.1% formic acid (FA).

2.5 NanoLC-MS(MS)² and protein IDs

Peptide digests (3 μ L = approximately 300 ng, controls calculated at ~50% efficiency in digestion and peptide recovery) were injected onto a Surveyor HPLC plus (Thermo Scientific, San Jose, CA, USA) using a split flow configuration on the back end of a 100 micron id $\times$ 13 cm pulled tip C-18 column (Jupiter C-18 300 Å, 5 micron, Phenomenex). This system ran in-line with a Thermo LTQ XL ion trap mass spectrometer equipped with a nanoelectrospray source (Thermo Scientific), and all data were collected in CID mode. Buffers used were solvent A (0.1% FA in ddH₂O) and solvent B (0.1% FA in ACN). The instrument setting was as follows: 0–10 min, 2% solvent B; 10–15 min, 12% solvent B; 15–31 min, 30% solvent B; 31–36 min, 36% solvent B; 36–41 min, 45% solvent B; 41–48 min, 65% solvent B; 48–56 min, 95% solvent B; 56–57 min, 2% solvent B; 57–82 min, 2% solvent B. The LTQ collected data on the top three ions using a triple play method from 10 to 57 min, with a flow rate of 0.6 μ L/min. Before and after the analysis window, the spray voltage was 0.0 kV and the flow rate was 3 μ L/min. During data collection, the LTQ was configured as follows: spray voltage 2.2 kV, capillary temperature 170°C, 1 microscan with a maximum inject time of 25 ms for all modes. For the triple play setup, the SIM scans were obtained in zoom mode using a $\pm$ 2.5 Da window, and triggered on a signal threshold of 15 000 counts. The MS/MS scans were obtained in normal mode with a minimum signal threshold of 500 counts based on the SIM scan. The activation settings were charge state 5, isolation width 5.0 m/z , normalized collision energy 30.0, activation Q 0.250, and activation time 50.000 ms. For the dependent scans, charge state screening was enabled, and the dynamic exclusion was enabled with the following settings: repeat count 3, repeat duration 20.0 s, exclusion list size 500, and exclusion duration 90.0 s.

The XCalibur RAW files were collected in profile mode, centroided and converted to MzXML using ReAdW v. 3.5.1. The mgf files were then created using MzXML2Search (included in TPP v. 3.5) for all scans with a precursor mass between 400 and 20 000 Da. The data were searched using Mascot (v2.1, .mgf files), which was set for two missed cleavages, a precursor mass window of 0.45 Da, trypsin digestion, static modification C at 57.0293, and a variable modification M at 15.9949. For the fragment-ion mass tolerance, the Mascot default of 0.8 Da was used. Searches were performed with a mouse or human subset of the UniRef100 database, which included common contaminants such as digestion enzymes and human keratin.

2.6 MS data analysis

In the mouse study, LC-MS(MS)² raw data were first analyzed with Expressionist Refiner MS software (Genedata AG, Basel, Switzerland) using TIC. This allowed a visual inspection of individual LC-MS(MS)² runs by mass and time to determine their quality. In addition, a peak (mass, time, and intensity) list (ABS file) was exported and used to generate the reproducibility scatter plots by plotting natural log of individual ion intensity (x -axis) versus natural log of average ion intensity (y -axis) using Origin v.7.5 (OriginLab, Northampton,

MA, USA). Pearson's r coefficient (R^2) was also calculated between ion intensities for each LC-MS(MS)² run versus the average for each group. Separately, a list of protein IDs was generated based on Mascot search results, which were filtered based on calculated peptide and protein probabilities [28], followed by grouping of top scoring protein IDs with ProteoIQ (NuSep, Athens, GA, USA). The unique and shared proteins, as compared between the frozen and FFPE groups, were determined and visualized using strict Venn diagrams.

In the human study, ProteoIQ was similarly applied in order to group all protein IDs while also generating normalized spectral counts (SC) for label-free quantification across all samples for each protein and used for relative quantification purposes [29]. The filter cut-off values were set with peptide length (>4 AAs), peptide probability (>0.25), peptides per protein (± 2 peptides), and protein probability (>0.5), which resulted in a list of protein IDs with 95% confidence. Peptides with +1 charge were also excluded from any further analyses.

2.7 Statistical analysis and systems biology

Comparative statistical analysis between Control and PDAC group was carried out by filtering the ProteoIQ output data by significance analysis of microarray (SAM): $SAM = [(PDAC \text{ averaged normalized SC}) - (control \text{ averaged normalized SC})] / [(PDAC \text{ standard deviation}) + (control \text{ standard deviation})]$ [30, 31] with a cut-off set at ± 0.8 , and fold changes with a cut-off set at ± 2 , in addition to Wilcoxon rank sum test with $p < 0.05$. Heat maps were generated on the statistically significant features using PermutMatrix [32].

Those proteins that were detected in >50% of one arm but none in the other arm were considered as “all-or-nothing” protein hits, and described as either “Control only” or “PDAC only.” The list of statistically significant as well as all-or-nothing protein IDs generated from this process was further analyzed with the systems analyses described below.

The systems biology (pathway analysis) analyses were carried out with MetaCore (GeneGO Inc., St. Joseph, MI, USA). In the reproducibility study, both unique and shared proteins identified in either the FFPE or frozen group were categorized by cellular localization and molecular function using STRAP (SourceForge), and pie charts were generated using Excel [33]. In the human study, proteins in the final list were analyzed using the autoexpand algorithm within MetaCore with the default setting (i.e. expanded by 50 nodes). In parallel, the expand-by-one algorithm was used to identify connections to the neighboring proteins and their correlation to PDAC. Interactions identified within MetaCore were manually correlated using full text articles. Detailed algorithms have been described previously [34].

3 Results

3.1 Similarity and reproducibility of FFPE versus frozen tissues by MS

In this study, we have applied a straightforward workflow for the robust and highly reproducible analysis of FFPE tissues by MS that was compared side-by-side to a conventional frozen lysate approach (Fig. 1). The amount of proteins obtained in FFPE tissues ($28 \pm 13 \mu\text{g}$) was comparable to that of matched frozen tissues ($30 \pm 8 \mu\text{g}$). The extraction buffer supplied in the FFPE tissue kit contained various compounds including SDS as indicated in MSDS, which interfered strongly with the MS analysis and yet could not be adequately removed by common SPE chemistries. Therefore, denatured protein extracts were run partway into a 1D acrylamide gel, and fixed overnight using 20% methanol (as part of the staining kit) to remove contaminants prior to tryptic digestion and LC-MS(MS)² analysis. This process was also found to be more efficient and reproducible

for low nano- to microgram quantities of proteins than molecular weight cut-off dialysis membranes or spin-column centrifugation (data not shown).

To demonstrate the experimental reproducibility of the workflow, matched frozen, and FFPE pancreatic tissues from eight C57BL/6 mice were analyzed. A TIC overview of LC-MS(MS)² analyses with the m/z (mass/charge) versus RT (retention time) plots illustrated highly similar ion profiles for all 16 samples (eight FFPE and eight frozen), regardless of the protein extraction method utilized (Fig. 2A). The outcome of these results suggests high experimental reproducibility for this FFPE proteomic approach. The $\ln$ versus $\ln$ plot based on TIC further supported this observation with linear correlation coefficients (R^2 values) ranging 0.96–0.98 for the FFPE group, which was highly comparable to the frozen group whose R^2 values ranged from 0.97–0.99 (Fig. 2B).

To determine whether the FFPE proteomics approach was similar to the conventional frozen lysate approach, comparative analyses between the FFPE and the frozen groups were carried out. Approximately 80% of the total number of identified proteins overlapped between the FFPE and frozen groups (Fig. 3A). While approximately 20% of the protein IDs were found to be different between groups, the cellular localization and molecular function for those proteins were highly similar (Supporting Information Fig. 3). Furthermore, the total number of unique peptides identified in frozen and in FFPE was similar, with 958 identified for FFPE and 1070 identified for frozen (Fig. 3B). Of the peptides identified, approximately 63% overlapped between the two groups and an additional approximately 17% of the no matching peptides were found to map to the same protein group. We have illustrated this point with an example of a highly expressed protein, pancreatic amylase shown in Fig. 3C. Interestingly, based on repeat identifications for each peptide, the FFPE group presented with either increased or decreased efficiencies as compared to controls, yet overall peptides identified were nearly the same with 26 of 33 peptides overlapping (~80%).

3.2 Analysis of human pancreatic cancer versus healthy adjacent FFPE tissues

In this experiment, 11 PDAC tissues were compared to eight healthy adjacent (control) tissues (Supporting Information Tables 1 and 2). With the use of normalized SC, we applied a minimum protein probability cut-off of >95% confidence whereby 162 proteins were consistently identified across both PDAC and Control groups, with 16 proteins only detected in the Control group, and 36 proteins only detected in the PDAC group (Fig. 4A; Supporting Information Table 2 and 3). Among the 162 shared proteins identified, 37 were found to be significantly different in abundance, of which 15 were increased and 22 were decreased in the PDAC group (Fig. 4B; Supporting Information Tables 2 and 3). Among the 15 proteins increased in cancer, alpha 3 type VI collagen isoform 3 (COL6A3, a stroma marker) and keratin type I cytoskeletal 19 (KRT19, a duct marker) were identified with fold increases of 33- and 5-fold, respectively (Supporting Information Table 2). Systems biology, pathway analysis was carried out with a focus on associations between these two proteins (COL6A3 and KRT19) and pancreatic cancer, where an additional 12 differentially abundant proteins identified in this confirmation study showed tight associations through a potentially important pathway, which involves estrogen receptor α (ESR1) (Fig. 5A and B).

4 Discussion

One important aspect of FFPE tissue preparation for use in MS analysis is reversal of formalin-induced protein crosslinks, which has been addressed by various groups whereby the usefulness of applying HAR to FFPE specimens is established [17–20, 22–26]. However, reproducibility and robustness is often not assessed, and comparison to serially obtained flash frozen controls is also overlooked.

As high reproducibility is crucial in both targeted and biomarker discovery studies, we evaluated this workflow using matched FFPE and flash frozen mouse pancreatic tissues where all animals shared the same genetic background ($n = 8$). We observed highly consistent LC-MS(MS)² results regardless of the extraction method (FFPE, $R^2 = 0.97 \pm 0.01$; frozen $R^2 = 0.98 \pm 0.01$). In addition, the total number of unique peptides identified between the two groups was highly similar, with 958 for FFPE and 1070 identified for frozen, which correlated with an average of 254 proteins that overlapped by approximately 80%. The significance of the unique proteins identified between the two groups is not clear; however, we found it interesting that the cellular localization and molecular function profiles were highly similar. Therefore, the sampling of intact peptide and protein number, in addition to overall protein class was essentially the same. This point was highlighted with amylase as a representative protein for this study, where peptides were found to be differentially observed between FFPE and frozen specimens, yet approximately 80% of the peptides were consistently identified between both groups. This difference may be due to the more exposed and reactive amino acid side chains expected to be less reversible after intramolecular formalin cross-linking [35]. As a control, we analyzed a separate set of frozen sections ($n = 8$) whereby we applied the FFPE extraction methodology, and observed a approximately 12% change in proteins identified as compared to the standard frozen tissue methodology (Supporting Information Fig. 4). However, most of these proteins were identified in the FFPE tissue samples, indicating that the extraction methodology does have a modest yet notable effect on the proteins that are liberated regardless of storage type.

These data strongly suggest that the combination of HAR and trypsin digestion can effectively and reproducibly liberate peptides from FFPE tissues similar in number and type to frozen tissues. Our work is distinct from previous studies in that we utilized matched tissue specimens obtained from multiple littermate mice to directly compare FFPE with frozen tissue sections. In addition, the amount of protein utilized in this study is in-line with commonly attainable clinical specimens from a majority of human biorepositories, including analysis of tissue arrays and micro-dissected tissue specimens.

In order to further validate this approach, we analyzed a small human cohort of archived FFPE specimens representing pancreatic cancer ($n = 11$) versus uninvolved tissues as controls ($n = 8$). We then applied a SC method to measure relative differences in protein abundance between these two groups following LC-MS. The use of SC-based analysis on label-free shotgun proteomics data was recently shown to be highly effective in quantifying the relative changes in protein abundance [36–38]. This approach is based on the observation that protein abundance is directly correlated with the number of peptides and resultant MS/MS spectra [39, 40]. This pilot study led to the identification of approximately 200 proteins identified across both PDAC and control groups, where 162 of these proteins were the same, and 47 were either significantly increased as compared to controls or only detected in the PDAC group. Of those proteins found to be increased in PDAC, over 50% (28) have been previously reported as PDAC markers. These include AAT (alpha1-anti-trypsin), Annexin A1, Mucin 5AC, COL6A3, and KRT19, which strongly implies that the SC-based comparative analysis effectively yielded a reliable set of proteins differentially abundant in PDAC obtained from the archived FFPE tissues [41].

The analysis of highly complex omics data generated from biological specimens are often modeled and made sense of using systems biology [42, 43]. From this perspective, the model system in this study, pancreatic cancer, is characterized by a pronounced desmoplasia [2, 3, 44]. This phenotype likely involves crosstalk between the stroma and neoplastic ducts within the pancreas [3, 45]. Two proteins known to be involved in this type of cytopathology were identified in our study as significantly increased in PDAC, and include COL6A3, a stromal marker [46, 47], and KRT19, expressed in pancreatic ducts [48]. Systems biology

pathway analysis was carried out with a focus on pancreatic cancer where COL6A3 and KRT19, and an additional 12 identified proteins, showed tight associations through ESR1. However, the role of ESR1 in pancreatic tumorigenesis is controversial at present [49], therefore additional work is needed to determine its role in PDAC development and/or progression.

Together, these data support the validation of an approach for the proteomic analysis of FFPE tissues that is straightforward and highly robust, which can also be effectively applied toward translational studies of disease. Additionally, these studies were designed using commonly attainable clinical specimens with the concept of extending further studies to tissue microarrays or a histo-guided laser-captured microdissection approach.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank the staff in the UAB Comprehensive Cancer Center Mass Spectrometry/Proteomics Shared Facility research team for their excellent technical support (P30CA13148-38). The financial support for this study was provided by the UAB section of Pancreatic SPORE (5P50CA101955-07) awarded in part to C.A.K., W.E.G., and J.A.M. K.K. is currently supported by the UAB NRSA Trauma and Inflammation Research Fellowship (5T32GM063490).

Abbreviations

COL6A3	alpha 3 type VI collagen isoform 3
ESR1	estrogen receptor α
FA	formic acid
FFPE	formalin-fixed paraffin-embedded
KRT19	keratin type I cytoskeletal 19
HAR	heat antigen retrieval
PDAC	pancreatic ductal adenocarcinoma
RT	room temperature
SAM	significance analysis of microarray
SC	spectral counts

References

1. Hong SM, Park JY, Hruban RH, Goggins M. Molecular signatures of pancreatic cancer. Arch. Pathol. Lab. Med. 2011; 135:716–727. [PubMed: 21631264]
2. Korc M. Pancreatic cancer-associated stroma production. Am. J. Surg. 2007; 194:S84–S86. [PubMed: 17903452]
3. Mahadevan D, Von Hoff DD. Tumor-stroma interactions in pancreatic ductal adenocarcinoma. Mol. Cancer Ther. 2007; 6:1186–1197. [PubMed: 17406031]
4. Siegel R, Ward E, Brawley O, Jemal A. Cancer statistics, The impact of eliminating socioeconomic and racial disparities on premature cancer deaths. CA Cancer J. Clin. 2011; 61:212–236. [PubMed: 21685461]
5. Danovi SA, Wong HH, Lemoine NR. Targeted therapies for pancreatic cancer. Br. Med. Bull. 2008; 87:97–130. [PubMed: 18753179]

6. Nilsson T, Mann M, Aebersold R, Yates JR 3rd, et al. Mass spectrometry in high-throughput proteomics: ready for the big time. *Nat. Methods*. 2010; 7:681–685. [PubMed: 20805795]
7. Otali D, Stockard CR, Oelschlager DK, Wan W, et al. Combined effects of formalin fixation and tissue processing on immunorecognition. *Biotech. Histochem*. 2009; 84:223–247. [PubMed: 19886759]
8. Bancroft, JD., editor. *Theory and Practice of Histological Techniques*. Philadelphia: Churchill Livingstone; 2007.
9. Ahram M, Flaig MJ, Gillespie JW, Duray PH, et al. Evaluation of ethanol-fixed, paraffin-embedded tissues for proteomic applications. *Proteomics*. 2003; 3:413–421. [PubMed: 12687609]
10. Fowler CB, O'Leary TJ, Mason JT. Modeling formalin fixation and histological processing with ribonuclease A: effects of ethanol dehydration on reversal of formaldehyde cross-links. *Lab. Invest*. 2008; 88:785–791. [PubMed: 18490897]
11. Shi SR, Key ME, Kalra KL. Antigen retrieval in formalin-fixed paraffin-embedded tissues: an enhancement method for immunohistochemical staining based on microwave oven heating of tissue sections. *J. Histochem. Cytochem*. 1991; 39:741–748. [PubMed: 1709656]
12. Yamashita S, Okada Y. Mechanisms of heat-induced antigen retrieval: analyses in vitro employing SDS-PAGE and immunohistochemistry. *J. Histochem. Cytochem*. 2005; 53:13–21. [PubMed: 15637334]
13. Steg A, Vickers SM, Eloubeidi M, Wang W, et al. Hedgehog pathway expression in heterogeneous pancreatic adenocarcinoma: implications for the molecular analysis of clinically available biopsies. *Diagn. Mol. Pathol*. 2007; 16:229–237. [PubMed: 18043287]
14. Ribeiro-Silva A, Zhang H, Jeffrey SS. RNA extraction from ten year old formalin-fixed paraffin-embedded breast cancer samples: a comparison of column purification and magnetic bead-based technologies. *BMC Mol. Biol*. 2007; 8:118. [PubMed: 18154675]
15. Sepp R, Szabo I, Uda H, Sakamoto H. Rapid techniques for DNA extraction from routinely processed archival tissue for use in PCR. *J. Clin. Pathol*. 1994; 47:318–323. [PubMed: 8027368]
16. Azimzadeh O, Barjaktarovic Z, Aubele M, Calzada-Wack J, et al. Formalin-fixed paraffin-embedded (FFPE) proteome analysis using gel-free and gel-based proteomics. *J. Proteome Res*. 2010; 9:4710–4720. [PubMed: 20604508]
17. Groseclose MR, Massion PP, Chaurand P, Caprioli RM. High-throughput proteomic analysis of formalin-fixed paraffin-embedded tissue microarrays using MALDI imaging mass spectrometry. *Proteomics*. 2008; 8:3715–3724. [PubMed: 18712763]
18. Hood BL, Darfler MM, Guiel TG, Furusato B, et al. Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol. Cell. Proteomics*. 2005; 4:1741–1753. [PubMed: 16091476]
19. Jiang X, Feng S, Tian R, Ye M, Zou H. Development of efficient protein extraction methods for shotgun proteome analysis of formalin-fixed tissues. *J. Proteome Res*. 2007; 6:1038–1047. [PubMed: 17266348]
20. Nirmalan NJ, Harnden P, Selby PJ, Banks RE. Development and validation of a novel protein extraction methodology for quantitation of protein expression in formalin-fixed paraffin-embedded tissues using western blotting. *J. Pathol*. 2009; 217:497–506. [PubMed: 19156775]
21. Ostasiewicz P, Zielinska DF, Mann M, Wisniewski JR. Proteome phosphoproteome and N-glycoproteome are quantitatively preserved in formalin-fixed paraffin-embedded tissue and analyzable by high-resolution mass spectrometry. *J. Proteome Res*. 2010; 9:3688–3700. [PubMed: 20469934]
22. Ronci M, Bonanno E, Colantoni A, Pieroni L, et al. Protein unlocking procedures of formalin-fixed paraffin-embedded tissues: application to MALDI-TOF imaging MS investigations. *Proteomics*. 2008; 8:3702–3714. [PubMed: 18704906]
23. Shi SR, Liu C, Balgley BM, Lee C, Taylor CR. Protein extraction from formalin-fixed, paraffin-embedded tissue sections: quality evaluation by mass spectrometry. *J. Histochem. Cytochem*. 2006; 54:739–743. [PubMed: 16399996]
24. Sprung RW Jr, Brock JW, Tanksley JP, Li M, et al. Equivalence of protein inventories obtained from formalin-fixed paraffin-embedded and frozen tissue in multidimensional liquid chromatography-tandem mass spectrometry shotgun proteomic analysis. *Mol. Cell. Proteomics*. 2009; 8:1988–1998. [PubMed: 19467989]

25. Wang L, Clark ME, Crossman DK, Kojima K, et al. Abundant lipid and protein components of drusen. *PLoS One*. 2010; 5:e10329. [PubMed: 20428236]
26. Xiao Z, Li G, Chen Y, Li M, et al. Quantitative proteomic analysis of formalin-fixed and paraffin-embedded nasopharyngeal carcinoma using iTRAQ labeling, two-dimensional liquid chromatography, and tandem mass spectrometry. *J. Histochem. Cytochem.* 2010; 58:517–527. [PubMed: 20124091]
27. Thakur D, Rejtar T, Wang D, Bones J, et al. Microproteomic analysis of 10,000 laser captured microdissected breast tumor cells using short-range sodium dodecyl sulfate-polyacrylamide gel electrophoresis and porous layer open tubular liquid chromatography tandem mass spectrometry. *J. Chromatogr. A*. 2011; 1218:8168–8174. [PubMed: 21982995]
28. Keller A, Nesvizhskii AI, Kolker E, Aebersold R. Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal. Chem.* 2002; 74:5383–5392. [PubMed: 12403597]
29. Bantscheff M, Lemeier S, Savitski MM, Kuster B. Quantitative mass spectrometry in proteomics: critical review update from 2007 to the present. *Anal. Bioanal. Chem.* 2012; 404:939–965. [PubMed: 22772140]
30. Golub TR, Slonim DK, Tamayo P, Huard C, et al. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science*. 1999; 286:531–537. [PubMed: 10521349]
31. Xu BJ, Shyr Y, Liang X, Ma LJ, et al. Proteomic patterns and prediction of glomerulosclerosis and its mechanisms. *J. Am. Soc. Nephrol.* 2005; 16:2967–2975. [PubMed: 16079267]
32. Caraux G, Pinloche S. PermutMatrix: A graphical environment to arrange gene expression profiles in optimal linear order. *Bioinformatics*. 2005; 21:1280–1281. [PubMed: 15546938]
33. Bhatia VN, Perlman DH, Costello CE, McComb ME. Software tool for researching annotations of proteins: open-source protein annotation software with data visualization. *Anal. Chem.* 2009; 81:9819–9823. [PubMed: 19839595]
34. Ekins S, Bugrim A, Brovold L, Kirillov E, et al. Algorithms for network analysis in systems-ADME/Tox using the MetaCore and MetaDrug platforms. *Xenobiotica*. 2006; 36:877–901. [PubMed: 17118913]
35. Metz B, Kersten GF, Hoogerhout P, Brugghe HF, et al. Identification of formaldehyde-induced modifications in proteins: reactions with model peptides. *J. Biol. Chem.* 2004; 279:6235–6243. [PubMed: 14638685]
36. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics*. 2005; 4:1487–1502. [PubMed: 15979981]
37. Prokai L, Stevens SM Jr, Rauniyar N, Nguyen V. Rapid label-free identification of estrogen-induced differential protein expression in vivo from mouse brain and uterine tissue. *J. Proteome Res.* 2009; 8:3862–3871. [PubMed: 19545149]
38. Zhang B, VerBerkmoes NC, Langston MA, Uberbacher E, et al. Detecting differential and correlated protein expression in label-free shotgun proteomics. *J. Proteome Res.* 2006; 5:2909–2918. [PubMed: 17081042]
39. Liu H, Sadygov RG, Yates JR 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* 2004; 76:4193–4201. [PubMed: 15253663]
40. Washburn MP, Wolters D, Yates JR 3rd. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol.* 2001; 19:242–247. [PubMed: 11231557]
41. Harsha HC, Kandasamy K, Ranganathan P, Rani S, et al. A compendium of potential biomarkers of pancreatic cancer. *PLoS Med.* 2009; 6:e1000046. [PubMed: 19360088]
42. Auffray C, Imbeaud S, Roux-Rouquie M, Hood L. From functional genomics to systems biology: concepts and practices. *C. R. Biol.* 2003; 326:879–892. [PubMed: 14744094]
43. Gehlenborg N, O'Donoghue SI, Baliga NS, Goessmann A, et al. Visualization of omics data for systems biology. *Nat. Methods*. 2010; 7:S56–S68. [PubMed: 20195258]
44. Bardeesy N, DePinho RA. Pancreatic cancer biology and genetics. *Nat. Rev. Cancer*. 2002; 2:897–909. [PubMed: 12459728]

45. Ricci F, Kern SE, Hruban RH, Iacobuzio-Donahue CA. Stromal responses to carcinomas of the pancreas: juxtatumoral gene expression conforms to the infiltrating pattern and not the biologic subtype. *Cancer Biol. Ther.* 2005; 4:302–307. [PubMed: 15876873]
46. Binkley CE, Zhang L, Greenson JK, Giordano TJ, et al. The molecular basis of pancreatic fibrosis: common stromal gene expression in chronic pancreatitis and pancreatic adenocarcinoma. *Pancreas.* 2004; 29:254–263. [PubMed: 15502640]
47. Hustinx SR, Cao D, Maitra A, Sato N, et al. Differentially expressed genes in pancreatic ductal adenocarcinomas identified through serial analysis of gene expression. *Cancer Biol. Ther.* 2004; 3:1254–1261. [PubMed: 15477757]
48. Osborn M, van Lessen G, Weber K, Kloppel G, Altmannsberger M. Differential diagnosis of gastrointestinal carcinomas by using monoclonal antibodies specific for individual keratin polypeptides. *Lab. Invest.* 1986; 55:497–504. [PubMed: 2429069]
49. Satake M, Sawai H, Go VLW, Satake K, et al. Estrogen receptors in pancreatic tumors. *Pancreas.* 2006; 33:119–127. [PubMed: 16868476]

Figure 1. FFPE tissue workflow for proteomic analysis. Workflow for the proteomic analysis of the frozen and FFPE tissues.

Figure 2.

LC-MS analysis of FFPE tissue proteome with high reproducibility. (A) Top, representative LC-MS ion chromatogram plotted as m/z (x -axis) versus retention time (RT, y -axis). Bottom, one randomly chosen area within the plot (arrow) was expanded across all LC-MS runs as an example. This process was used to illustrate the quality and reproducibility of the individual LC-MS runs (rep = replicate). (B) Scatter plot of $\ln$ individual normalized ion intensity (x -axis) versus natural log ($\ln$) average normalized ion intensities (y -axis). The lines correlate with ± 2 - and ± 4 -fold changes. The Pearson's correlation coefficient was calculated for each of the eight FFPE samples with an average and standard deviation of $R^2 = 0.97 \pm 0.01$ and $R^2 = 0.98 \pm 0.01$ for the frozen samples.

Figure 3.

Comparison of FFPE tissue and matched frozen tissue. (A) Venn diagram of the total number of protein groups identified from eight animals per group with two or more peptides per protein and >95% confidence from two (10 mm × 10 mm × 10 μm) tissue sections. Greater than 80% of the protein IDs overlapped between two groups. (B) Venn diagram of all peptides identified with >95% confidence. (C) Top, pancreatic amylase amino acid sequence. Red letters: peptides identified in this study. Yellow highlights: peptides identified in both FFPE and frozen groups. Cyan highlights: peptides identified only in the frozen group. Bottom, table of peptide sequences identified in this study for pancreatic amylase with number of repeated identification of the specific peptide indicated.

Figure 4. Markers of pancreatic ductal adenocarcinoma (PDAC) via proteomic analysis of human FFPE tissues. (A) Venn diagram of identified protein groups with 95% confidence in control and PDAC. (B) Heat map of normalized spectral counts in control and PDAC, corresponding to those proteins that were shared by both groups that also met the criteria of SAM $>|\pm 0.8|$, fold change $>|\pm 2|$, and Wilcoxon rank sum test < 0.05 .

Figure 5. Pathway analysis of human pancreatic cancer FFPE tissue proteome. (A) Potential relationship between COL6A3 and KRT19 by systems biology pathway analysis (red arrows). Pink to red double circle: increase in pancreatic ductal adenocarcinoma (PDAC). Light blue double circle: decrease in PDAC. Cyan highlight: upstream regulators. (B) Bar charts representing the normalized spectral counts with standard deviation for 14 statistically significant proteins comparing the control (blue) and PDAC (red) groups.