Initial Development and Validation of a Novel Extraction Method for Quantitative Mining of the Formalin-Fixed, Paraffin-Embedded Tissue Proteome for Biomarker Investigations

Niroshini J. Nirmalan,[†,‡] Christopher Hughes,[§] Jianhe Peng,[†] Therese McKenna,[§] James Langridge,[§] David A. Cairns,[†] Patricia Harnden,[†] Peter J. Selby,[†] and Rosamonde E. Banks*[,†]

Cancer Research UK Centre, Leeds Institute of Molecular Medicine, St. James's University Hospital, Beckett Street, Leeds LS9 7TF, United Kingdom, Waters Corporation, Manchester M23 9LP, United Kingdom, and University of Salford, Manchester, M5 4WT, United Kingdom

Abstract: Annotated formalin-fixed, paraffin-embedded (FFPE) tissue archives constitute a valuable resource for retrospective biomarker discovery. However, proteomic exploration of archival tissue is impeded by extensive formalin-induced covalent cross-linking. Robust methodology enabling proteomic profiling of archival resources is urgently needed. Recent work is beginning to support the feasibility of biomarker discovery in archival tissues, but further developments in extraction methods which are compatible with quantitative approaches are urgently needed. We report a cost-effective extraction methodology permitting quantitative proteomic analyses of small amounts of FFPE tissue for biomarker investigation. This surfactant/heat-based approach results in effective and reproducible protein extraction in FFPE tissue blocks. In combination with a liquid chromatography−mass spectrometry-based label-free quantitative proteomics methodology, the protocol enables the robust representative and quantitative analyses of the archival proteome. Preliminary validation studies in renal cancer tissues have identified typically 250−300 proteins per 500 ng of tissue with 1D LC−MS/MS with comparable extraction in FFPE and fresh frozen tissue blocks and preservation of tumor/normal differential expression patterns (205 proteins, $r = 0.682$; $p < 10^{-15}$). The initial methodology presented here provides a quantitative approach for assessing the potential suitability of the vast FFPE tissue archives as an alternate resource for biomarker discovery and will allow exploration of methods to increase depth of coverage and investigate the impact of preanalytical factors.

* Corresponding author: Email: r.banks@leeds.ac.uk, Tel: +44-113-2064927, Fax: +44-113-2429886.
† St. James's University Hospital.
‡ University of Salford.
§ Waters Corporation.

Introduction

Acquiring fresh tissue biopsies requisite for the discovery/validation phases of biomarker investigation is challenging, particularly for rarer diseases.[1,2] Formalin-fixed, paraffin-embedded (FFPE) tissue archives represent a valuable alternate resource. Significant efforts are now being made to overcome the effects of formalin cross-linkage on protein analysis.[3−10] While most studies are Western blot, reverse-phase array or surface enhanced laser desorption ionization-time-of-flight (SELDI-TOF)-based,[7,11−13] recently developed mass spectrometry-compatible extraction methods have expanded our ability to profile the archival proteome.[6,9,10,14−17]

Studies exploring extracted archival proteins using high-throughput shotgun proteomics methodologies are few, due to incompatibility of the detergent buffer components which are needed for effective solubilization with mass spectrometry.[8,10,15,18,19] The current leader methodology uses a commercially available buffer (Liquid Tissue buffer (LTB), Expression Pathology) to successfully extract proteins from FFPE tissues.[6,17,18] Using LTB, comparable protein numbers have been reported for matched frozen and FFPE (168 and 143 proteins. respectively. using 1D-LC−MS) in renal tissue extracts, with 68% overlap between the two groups. Using a relative quantitative approach of normalized spectral counting, grade-dependent correlations with glycolytic and amino acid synthetic pathways in renal cell carcinoma were described.[17] Similar extraction efficiencies and subcellular representations have also been achieved using LTB in frozen versus FFPE rat liver tissues.[15] However, patent restrictions on the LTB constituents restrict the necessary further protocol refinement required at the early phase of method development. Furthermore, its high cost limits the widespread use of the buffer, particularly for large-scale biomarker applications. An alternative approach combining ammonium bicarbonate buffer, sonication, and heating to analyze colon adenoma tissues resulted in the identification of almost 400 proteins by direct LC−MS/MS, comparable to data from our study. A second experiment using a combined peptide isoelectric focusing (IEF) and reverse phase LC−MS/MS approach identified over 2000 proteins with an overlap of 67%.[8] The higher number of protein identities achieved

due to the additional layer of fractionation afforded by IEF separation of peptides into 20 initial fractions is impressive, but implementing this degree of fractionation has implications for both quantitation and application to the large cohorts typically encountered in biomarker investigation due to the analysis time involved. The study additionally examined the effect of preanalytic variables (fixation times, storage duration) on protein extraction in FFPE tissues. These initial studies highlight the potential utility of the application of high-throughput MS techniques to FFPE tissue. However, further protocol refinements and standardization are needed to objectively evaluate the extraction methodologies and the suitability of the archival resources for biomarker investigation, particularly innovations permitting quantitative comparisons.[20−22]

The efficacy of sodium dodecyl sulfate (SDS) in early heat-induced antigen retrieval (HIAR) protocols for the reversal of formalin-induced cross-links and subsequent recognition of proteins by antibodies in immunohistochemistry is attributable to its dual role of detergent and denaturant.[23] However, interference with trypsin digestion, incompatibility due to ion suppression, and adduct formation have necessitated SDS cleanup steps prior to mass spectrometric (MS) analysis, resulting in lower throughput and unpredictable protein losses, with implications for downstream detection and quantitation. MS-compatible surfactants offer a viable alternative. Of particular interest is an acid-labile surfactant developed by Waters Corporation as a substitute for SDS in polyacrylamide gel electrophoresis. The analogue, sodium 3-[(2-methyl-2-undecyl-1,3-dioxolan-4-yl)methoxy]-1-propanesulfonate (RapiGest SF) has detergent, denaturant, and electrophoretic properties comparable to SDS and actually enhances trypsin digestion.[24] The acid-labile character of the detergent results in its degradation on acidification, permitting direct MS analysis without further sample processing resulting in minimal sample loss or introduced bias for quantitiative measurements. Furthermore, it increases the solubility and proteolytic cleavage of hydrophobic proteins, increasing peptide recovery.

In this study, we report a novel, robust, cost-effective approach using a RapiGest extraction buffer to investigate the archival proteome in a robust and high-throughput manner. The workflow is coupled to an LC−MS based label-free quantitative approach (Waters Corporation, Manchester, U.K.)[25,26] to achieve reproducible, representative quantitation of archival tissue proteins from FFPE blocks. In this work, we apply multiple LC−MS quantitative strategies to the analysis of extracted FFPE tissues to assess parameters including; extraction efficiency, depth of coverage, and analytical reproducibility. The resulting extraction and quantitative methodology will facilitate further detailed objective assessment of the suitability of formalin-fixed archival tissue as an alternate resource for novel biomarker discovery and validation.

Experimental Procedure

Materials. General chemicals were purchased from Sigma (Poole, U.K.), VWR (Leicester, U.K.), and MP Biomedicals (London, U.K.). Complete mini protease inhibitor tablets were purchased from (Roche, U.K.); RapiGest SF from Waters Corporation, U,K,; and the Micro BCA Protein Quantitation Kit from Pierce. Modified sequencing grade trypsin was obtained from Promega and all HPLC-grade solvents for mass spectrometry, for example, formic acid and acetonitrile were from Merck (U.K.).

Tissue Samples. Informed consent was obtained from all patients and the research was reviewed and approved by the local research ethics committee. Tissue blocks (~125 mm³) from patients undergoing nephrectomy were placed in 10 mL of ice-cold RPMI buffer containing a Complete mini protease inhibitor tablet, embedded in OCT compound, and frozen in liquid nitrogen. In parallel, the nephrectomy specimens were fixed in formalin by the hospital Pathology Department for routine diagnostic use. For the latter, tissue blocks were fixed overnight in 10% (v/v) neutral buffered formaldehyde, dehydrated in graded ethanol, permeated in xylene, paraffin-embedded, and stored at room temperature.

The preliminary workup experiments were carried out on frozen and formalin-fixed matched normal renal cortex and renal cell carcinoma (RCC) tissue samples from a total of 5 patients undergoing nephrectomy (R330N/T, R320N/T, R319N/T, R316N, R318N) with grade 3 or 4, and TNM stages II or III clear cell RCC. Matched frozen and FFPE tumor and normal renal cortex from two patient samples (R320 for 1D analysis, and R330 for 2D analysis) were taken forward for more stringent label-free quantitation studies and technical replicate analyses. Sections from all tissue blocks were histologically reviewed by a Consultant Pathologist (P.H.). All samples used in the study had been stored for between 6−24 months to minimize possible variation introduced by long-term storage.

Method Development Overview. In the early protocol optimization phase, extraction solutions including RapiGest buffer (0.2% (w/v) RapiGest, 5 mM dithiothreitol (DTT) in 50 mM NH$_4$HCO$_3$), modified RIPA buffer (pH 7.0),[11] 30% v/v acetonitrile (ACN) in 100 mM NH$_4$HCO$_3$,[3,4] and SDS extraction buffer[7] were compared and assessed in terms of protein extraction and compatibility with LC−MS analysis. Briefly, three serial FFPE tissue sections of 10 μm thickness (each ~5 × 5 mm area) were deparaffinized and rehydrated as described below, and extracted with the different buffers, using an extended temperature protocol (105 °C for 30 min, 80 °C for 2 h). Extracts were reduced, alkylated, and trypsin digested as detailed below. Initial experiments examining extraction efficiency and technical reproducibility were carried out using a Q-STAR mass spectrometer (ABI, Foster City, CA) by assessing the number of proteins identified and amino acid sequence coverage. The resulting optimized RapiGest extraction protocol was subsequently selected as the method to pursue. The optimized workflow for the extraction and quantitation of proteins from archival tissue blocks is described further below and summarized in Figure 1. The compatibility with downstream quantification was assessed using label-free LC−MS quantitation methodology with analysis on a Synapt HDMS quadrupole-oa-TOF (Waters Corporation, U.K.) and using matched frozen and FFPE tissues. To define technical reproducibility of the extraction methodology, four independently extracted lysates from a normal renal FFPE tissue block were each assessed using triplicate replicate injections by LC−MS. A further comparative experiment used online multidimensional chromatographic separations prior to MS analysis to evaluate the depth of proteome coverage in extracts from matched frozen and FFPE samples.

Protein Extraction from Fresh Frozen Tissue Samples. Three serial sections of 10 μm thickness (each ~5 × 5 mm area) of fresh normal renal or RCC tissue (blocks were trimmed of surrounding OCT prior to sectioning) were collected into an Eppendorf tube. Flanking 5 μm sections on slides were routinely stained with hematoxylin and eosin (H&E) to verify tissue
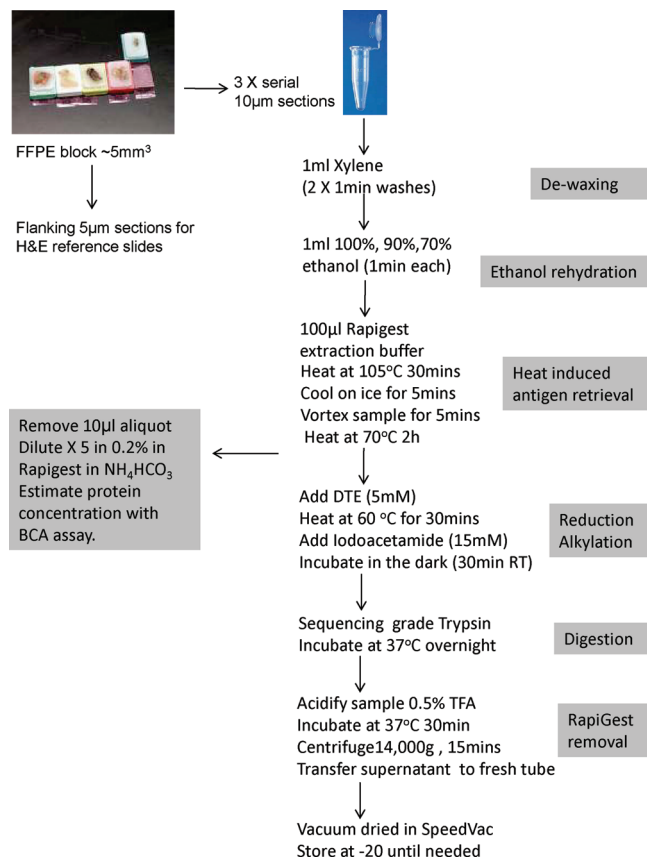
Figure 1. Summary of optimized surfactant/heat-based protein extraction protocol for formalin-fixed, paraffin-embedded tissue blocks.

integrity. A total of 100 μL of 0.2% (w/v) RapiGest in 50 mM ammonium bicarbonate was added to each tube prior to heating at 100 °C for 5 min. Samples were cooled for 5 min on ice.

Protein Extraction from FFPE Tissue Blocks. Three serial sections of 10 μm thickness (each ~5 × 5 mm area) from FFPE tissue blocks were collected in Eppendorf tubes. Samples were deparaffinized by washing tissue slices in 1 mL of xylene (2 × 1 min) followed by rehydration in a graded ethanol series (1 mL of 100%, 90% (v/v), 70% (v/v), 1 min each). The supernatant was discarded after each wash and the tissues were microcentrifuged at the end of the washes to remove excess ethanol. A total of 100 μL of 0.2% (w/v) RapiGest in 50 mM ammonium bicarbonate and 5 mM DTT was added and tubes were sealed with plastic clips (Qiagen, Germany) prior to heating at 105 °C for 30 min. The samples were cooled for 5 min on ice, vortexed vigorously, and heated at 80 °C for a further 2 h before cooling on ice for 5 min.

Reduction, Alkylation, and Trypsin Digestion of Frozen and FFPE Protein Extracts. A 10 μL aliquot of the protein extract was removed for the estimation of protein concentration. Samples were reduced in 5 mM dithioerythritol (DTE) for 30 min on a 60 °C heating block followed by alkylation with incubation in 15 mM iodoacetamide for a further 30 min in the dark at room temperature. Modified sequencing grade trypsin (16000 U/mg, Promega, U.K.) reconstituted in 50 mM ammonium bicarbonate was added to the extract at an enzyme/substrate ratio of 1:50 (w/w), and incubated overnight at 37 °C.

Postdigestion Protocol. To facilitate the hydrolysis of the RapiGest detergent for compatibility with mass spectrometry, the samples were acidified in 0.5% (v/v) trifluoroacetic acid (TFA) and incubated at 37 °C for 30 min. Samples were microfuged for 15 min at 12 000g to remove the insoluble white precipitate and any cellular debris, with the supernatant dried in a SpeedVac (Thermo Electron, U.K.) and stored at −20 °C until used. Estimation of protein concentrations was carried out using a microBCA protein quantitation kit (Pierce) according to the manufacturer's instructions with 0.2% (w/v) RapiGest in 50 mM ammonium bicarbonate as the diluent and albumin for the construction of standard curves. Tissue extracts were required to be diluted 5-fold in the RapiGest diluent to prevent any interference with the assay from residual DTE.

SELDI-TOF Analysis. A SELDI-TOF instrument was used in preliminary optimization experiments to allow the rapid comparison of extraction and digestion efficiencies of different buffers with fresh and formalin-fixed renal tissue. Digests from frozen and FFPE extractions were resuspended in 5% (v/v) ACN/0.05% TFA and 1 μg of protein extract was spotted on to a NP20 chip (Ciphergen Biosystems, Inc.). After drying, spots were overlaid with 2 μL of freshly prepared matrix (saturated sinapinic acid in 50% (v/v) ACN/0.05% (v/v) TFA) for acquisition of spectra. Chips were analyzed using a PBS-II SELDI-TOF (Ciphergen, Freemont, CA) using protocols with low and high mass acquisition of 500 Da−25 000 kDa, optimized center of 2.5−6.5 kDa, laser intensity of 175−195, and detector sensitivity of 9, collecting 50 transients per spot.

1D LC/MS/MS and 2D LC/MS/MS Qualitative Analyses. For 1D LC/MS/MS analyses, the tryptic digests from FFPE and frozen tissue extracts were analyzed using an Agilent 1100 series nano-LC system coupled online to an Applied Biosystems QSTAR XL quadrupole TOF hybrid mass spectrometer. Mobile phases were 0.1% (v/v) formic acid (A) and ACN, 0.1% (v/v) formic acid (B). Samples (1.0 μg) were loaded at a 20 μL/min flow rate into a trap column (5 mm ×0.3 mm i.d., 5 μm particle size, Zorbax 300SB C18, Agilent Technologies) in eluent A. The trap column was switched in-line with the analytical column (150 mm ×75 μm i.d., 5 μm particle size, Vydac 218MS5.07515, Grace) and the peptides were separated on an 8−38% linear gradient of eluent B over 45 min, with a 200 nL/min flow rate. The eluate was ionized and subjected to data-dependent mass spectrometry: an MS scan of 400−1800 m/z was performed for 500 ms and signals below 380 m/z were electronically suppressed. The three strongest doubly and triply charged ions with intensities over 40 counts were selected for CID using nitrogen, and MS/MS accumulation from 90−1800 m/z was performed for 1 s per peptide. The precursor ion was excluded for 200 s.

As a preliminary comparison at the level of 2D LC/MS/MS analysis, ~50 μg of the tryptic digests from FFPE and frozen tissue extracts was first fractionated by strong cation exchange chromatography followed by reverse phase nano-HPLC/MS/MS analysis. The separation column was Polysulfethyl A (150 × 1.0 mm, 5 μm, PolyLC). Samples were fractionated using a two buffer system: SCX1 (10 mM KH$_2$PO$_4$, pH3) and SCX2 (10 mM KH$_2$PO$_4$, 500 mM KCl, pH3) employing a linear gradient from 0% SCX2 to 40% SCX2 in 40 min with 20 fractions collected. Each fraction was then analyzed in 3 separate injections by nano-HPLC/MS/MS analysis as described above.

The 1D LC/MS/MS data was extracted (Analyst ver. r2.0, Mascot ver. b21, Applied Biosystems) and submitted to the Mascot server using Mascot Daemon (ver. 2.2, Matrix Science,

London, U.K.) to search against the NCBI database restricted to human entries (208 155 entries, 04.09.2009) using the following parameters: enzyme, trypsin; missed cleavages, 1; variable modifications, methionine oxidation and carbamidomethylation; peptide charge, 2^+ and 3^+; peptide tolerance, 0.15 Da; MSMS tolerance, 0.10 Da. In all probabilistic database searches, proteins with a minimum of 2 significant peptide(s) ($p < 0.05$) were considered identified. The false positive rate was 2.29% determined by selecting the decoy option of the Mascot search. The 2D LC/MS/MS data was analyzed by Protein Pilot (v1, Applied Biosystems) using the same database and the same parameters as the Mascot search. Proteins from multiple injection analysis were automatically grouped to eliminate the redundancy. The false positive rate was estimated at 3.31% using the decoyed database generated by Emboss (ver. 6.2.0).

NanoAcquity UPLC, LC−MS^E and Label-Free Quantitative Analyses. A nanoAcquity UPLC system operating in trapping mode was used to separate peptides and introduce the samples to the mass spectrometer. The trapping column (180 μm i.d. × 20 mm) contained Symmetry C18 5 μm material with trapping performed at 15 μL/min for 1 min using 100% solvent A (aqueous 0.1% formic acid). The BEH130 C18 1.7 μm analytical column (75 μm i.d. × 200 mm) was operated at a flow rate of 300 nL/min with a linear gradient from 1 to 40% solvent B (ACN, 0.1% formic acid) over 90 min.

While preliminary optimization and qualitative validation of the Rapigest protocol was carried out using an ABI Qstar mass spectrometer, the quantitative validation of the extraction methodology and its technical reproducibility were carried out on a Waters Synapt HDMS system operating in the "V" mode with a resolution of 10 000 fwhm. The system was operated in LC−MS^e mode where alternate scans are performed throughout the LC separation, one at low collision energy to provide peptide precursor mass information and the next at an elevated collision energy to produce fragments. Lock mass reference scans were acquired every 60 s. Each of the samples was run in triplicate with blank injections between each set to ensure minimal carryover. Vacuum-dried protein digests were resuspended in 10% (v/v) ACN to stock concentrations of 1 μg/μL and were used as internal standards. For routine injection, the sample was diluted 10-fold in 50 mM NH$_4$HCO$_3$ and an internal standard of yeast alchohol dehydrogenase (ADH) was spiked into the injection solution for a final concentration of 10 fmol/μL (4 μL of sample was injected on column containing 400 ng of protein extract and 40 fmol of yeast alcohol dehydrogenase).

In addition, two-dimensional chromatography using reversed phase-reversed phase at differential pH coupled with LC−MS^E was also used for the analysis of extracted FFPE and frozen tissue as a preliminary examination of the increase in protein identifications. The 2D separation was performed using RP chromatography on a 150 μm × 5 cm XBridge C$_{18}$ (5 μm) column with a discontinuous step gradient at high pH in the first dimension, followed by a low pH gradient separation on a 75 μm × 15 cm BEH C$_{18}$ (1.7 μm) column in the second dimension. To maximize sample recovery, an aqueous flow was mixed with the eluted organic-containing fractions prior to trapping on the second-dimension trap column. The released peptides were analyzed in both a 5 stage and 10 stage separation loading 2.4 and 4.8 μg of total protein, respectively. Each separation was processed separately, to create a monoisotopic peak list of all the eluting peptides and their associated fragment ions, and these were subsequently merged prior to

database searching with Protein Lynx Global Server (PLGS 2.4) against a nonredundant and randomized database.

The quantitation method was based on areas of the deconvoluted chromatographic and mass spectral peaks for peptide precursor-ions across injections and between samples. This was analyzed automatically using the PLGS v2.4 informatics package (Waters, U.K.).[8,9] The approach compares two or more experiments by defining the normalized mass spectrometric signal intensity for any and all given peptides, as determined from the LC−MS^E experiment as a measure of their respective amounts in the sample. In addition, concentrations of the proteins in the sample (expressed in nanograms or femtomole equivalents) were determined by referencing the average ion intensity of the three best ionizing peptides for each protein to those of the known concentration internal standard spike, as has been previously described in detail.[26]

Protein identification was performed by database searching using a previously published algorithm contained within the PLGS software.[27] The data was searched against a Uniprot Human database (version 14.6) with a 1× reverse sequence database appended. In brief, time-aligned precursor ions and their associated fragment ions were deconvoluted and deisotoped and then searched against the databank with 10 and 20 ppm tolerances for the peptide and their fragments, respectively. For positive acceptance of a protein, the following needed to be met: three fragment ions matched per peptide, seven fragment ions per protein with at least one unique peptide sequence per protein. The protein false discovery rate (FDR) was set to 4% such that protein hits were returned until the 4% of the total protein hits returned were random. A further requirement placed upon the protein level results was that protein sequences were needed to replicate in at least two of three replicate injections, and as such, this took the protein FDR to below 1%.

Statistical Analysis. Reproducibility in the quantitation for proteins common to all biological and technical replicates was assessed using the coefficient of variation expressed as a percentage. Spearman's rank correlation was used to assess the association between quantitation (expressed as a fold-change between tumor and normal tissue) in frozen and FFPE tissue extracts.

Results

Protocol Optimization. Initial method optimization experiments directly comparing buffers and reproducibility in fresh-frozen and FFPE samples showed that maximum extraction efficiency as assessed by numbers of proteins identified and sequence coverage was consistently achieved with the RapiGest buffer protocol, followed by the ACN−NH$_4$HCO$_3$ buffer in both 1D and 2DLC−MS/MS analyses (data not shown). SELDI-TOF analyses of other buffers studied failed to produce data comparable to RapiGest and ACN−NH$_4$HCO$_3$. A preliminary comparison of the latter two buffers using SELDI-TOF as a screening platform in postextraction (protein) and postdigestion extracts (peptides) is shown in Figure 2, clearly illustrating the enhanced tryptic digestion in Rapigest as reported previously.[24] The optimum RapiGest concentration was selected as 0.2% (w/v) in 50 mM NH$_4$HCO$_3$ as informed by titration experiments (data not shown). Routine extractions from FFPE blocks (3 × 10 μm thick slices, 5 × 5 mm each) in 100 μL extraction buffer volumes yielded total protein concentrations of 1.36 ± 0.60 μg/μL ($n = 8$). Extracted total protein concentra-

(a) Post-extraction protein spectra
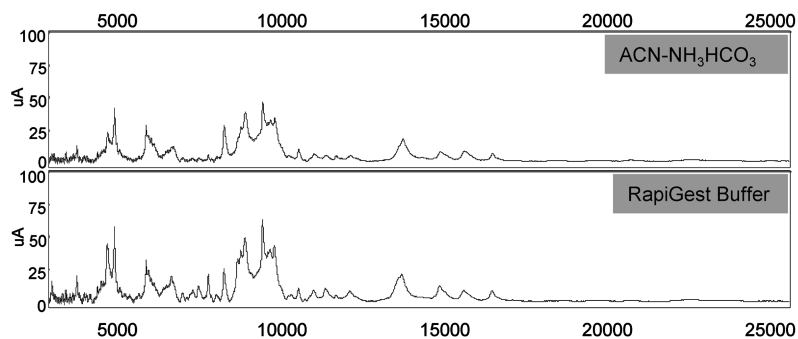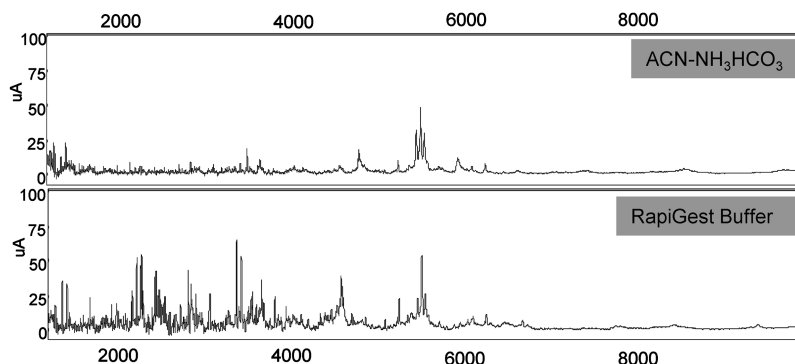

(b) Post-digestion peptide spectra

Figure 2. Comparison of representative protein extraction efficiencies for FFPE normal renal tissues with Rapigest and $ACN-NH_3HCO_3$ buffers using SELDI-TOF (Ciphergen). Spectral acquisition was carried out on 1 μg protein extracts spotted on NP20 chips. (a) The postextraction protein spectra and (b) the post-tryptic digestion peptide spectra for the two extraction buffers.

tions were similar to matched frozen tissue extracts of a similar tissue volume (1.08 + 0.25 μg/μL; $n = 8$, $p = 0.17$).

Comparison of Extraction of Frozen and FFPE Tissue Extracts. Initial experiments using 8 samples from 2 patients (matched normal/tumor and FFPE/frozen) were carried out comparing extraction efficacies by 1D-LC−MS/MS using the QSTAR mass spectrometer. Following this, one set of four samples were further analyzed using the Waters Synapt HDMS quadruple-oa-Tof mass spectrometer. Triplicate samples of fresh and FFPE extracts spiked with an internal standard to normalize protein loading between experiments were analyzed in LC−MS^e mode in a 1D-UPLC-HDMS experiment and proteins replicating in at least two out of three injections for each sample were determined and used for the comparative analysis (Supplementary Data 1). A total of 283 proteins were identified from the normal renal tissue FFPE extract compared with 268 from the matched fresh-frozen extract. A total of 350 unique protein were identified; of these, 201 (70−75% of each data set and 58% of total proteins identified) were common to both extracts with 82 (23%) proteins being unique to the FFPE data and a further 67 (19%) proteins identified only in the fresh normal tissue extract (Figure 3a). Analysis of the matched tumor samples showed similar results (Figure 3b). An analysis of the subcellular localization pattern of individual proteins in this data set derived using the Ingenuity Systems pathway analysis software (Ingenuity Systems, Inc., version 8.0) shows broadly similar distributions of proteins in the cellular compartments

(a) Normal renal cortex
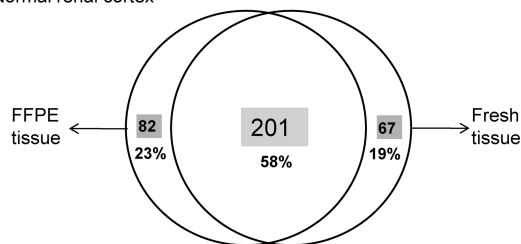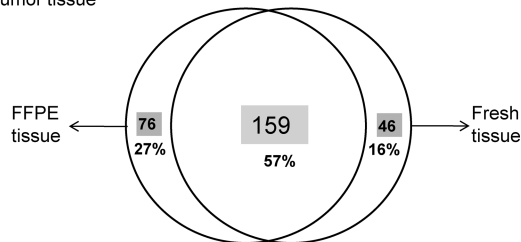

(b) Tumor tissue

Figure 3. Representative example of a comparison of protein extracts (400 ng) from frozen and FFPE extracts in (a) normal renal cortical tissue and (b) RCC tissue, analyzed using a NanoAcquity 1D-UPLC-HDMS quadrupole-oa-Tof (Waters U.K.). Each extract was run in triplicate and proteins appearing in at least 2 of three replicates were included in the analysis.

(Figure 4) irrespective of whether they were form frozen or FFPE extracts.

Assessing Technical Reproducibility of the RapiGest Extraction Method. In each of the technical replicates (i.e., independently prepared and analyzed extracts TR1 to TR4),

Frozen Normal

Frozen Tumor

FFPE Normal

FFPE Tumor

Cytoplasm
Unknown
Plasma membrane
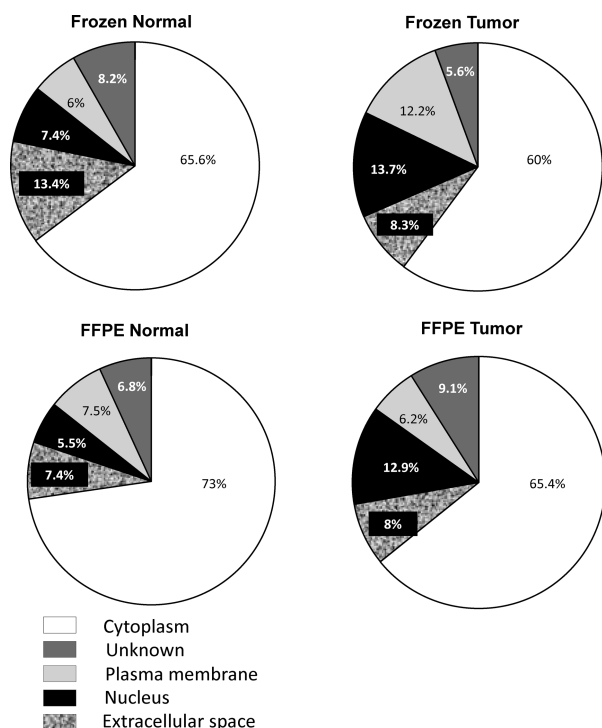Nucleus
Extracellular space

Figure 4. Subcellular localization analyses (Ingenuity IPA software (version 8.0)) of the proteins identified from matched normal renal cortex and RCC tissue extract (FFPE and frozen tissue; Figure 3) using the optimized Rapigest surfactant protocol. The proteins were identified at high confidence using a NanoAcquity UPLC-HDMS quadrupole-oa.Tof with protein loading normalized using a spiked external standard.

there were 254, 259, 272, and 246 proteins, respectively, found in at least 2 out of the three determinations, that is, (254/383) = 66%, 63%, 65%, and 61.5% of the total number of proteins across all injections (Figure 5a). Similar results were also obtained using the QSTAR in this and one further sample (data not shown). A comparative analysis of the 4 data sets showed 191 proteins conserved across all four technical replicate data sets (i.e., (191/254) = 75.2% of TR1, 73.7% of TR2, 70.2% of TR3, 77.6% of TR4), showing reproducible qualitative and quantitative metrics at the protein and peptide level (Supplementary Data 2). The overall analysis of protein overlap between the 4 replicates is shown in Figure 5b.

Absolute quantitation amounts could be determined for 150 proteins common to all replicates. The average CV (%) of nanograms amounts calculated on column were determined as 19.1% for TR1, 18.4% for TR2, 18.3% for TR3, and 20.5% for TR4. Combining all technical replicates, the average CV for nanograms amount on column was 28.8%. The reproducibility of quantitation and the stringency of isoform assignment is illustrated in Figure 6 for two selected proteins: fructose biphosphate aldolase (A-muscle, B-liver and C-brain types) and lactate dehydrogenase (A-muscle, B-heart, and C-testis subunits).

Comparing Fresh and FFPE Extraction Using 2D LC−MS-MS. Preliminary 2D-LC−MS-MS experiments to compare overlap in identified proteins between the fresh and FFPE extracts from matched normal and tumor were carried out using the Applied Biosystems QSTAR XL quadrupole TOF hybrid mass spectrometer as detailed in the Eperimental Procedure section.
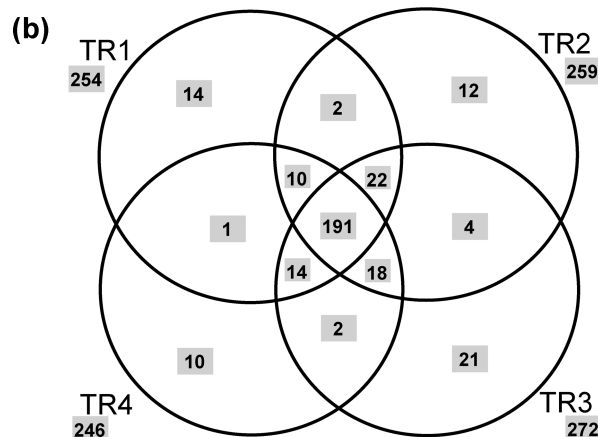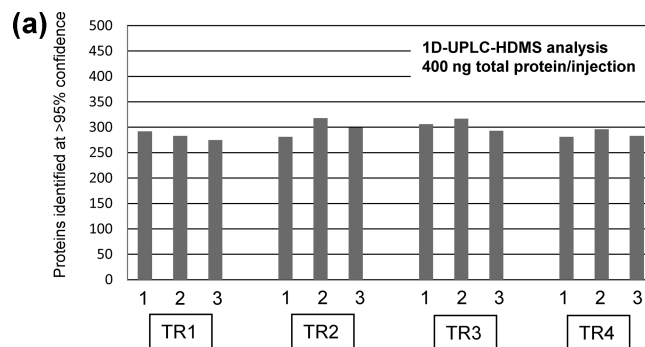
(a)

(b)

Figure 5. Technical replicate analysis of FFPE normal renal tissue extracts. (a) Proteins identified at high stringency across triplicate injections (1, 2, 3; 400 ng each injection) of each of 4 independently extracted replicates (TR1, TR2, TR3, TR4) from a single normal renal cortex FFPE tissue block; (b) shows the overall distribution patterns of the proteins within the replicate sets analyzed by 1D-LC−MS/MS. A total of 191 proteins were common to all four replicates with TR1, TR2, TR3 and TR4 extracts having 254, 259, 272, and 246 proteins identified, respectively, at high stringency.

Fresh frozen and FFPE extract from matched tumor tissue yielded 490 and 477 proteins, respectively (Supplementary Data 3). However, only 234, that is, approximately 50% of the proteins in each set, were common to both, with a total of 678 unique proteins overall. Data from normal frozen and FFPE tissue identified 622 and 490 proteins, respectively, with 305 (i.e., 50−60% of each set) being common to both and a total of 807 unique proteins.. Further, 2D experiments comparing expansion in profiling capabilities to 1D analysis were carried out using 2D-UPLC-MS[E] on the Waters Synapt HDMS quadrupole-oa-Tof mass spectrometer and are detailed below.

Comparing 1D versus 2D LC−MS/MS. Reverse phase (RP) and multidimensional (RP-RP) LC−MS/MS separations of FFPE extracts from two samples, matched tumor and normal renal tissue, were directly compared to provide a preliminary assessment of the expansion in protein profiling capabilities with increasing multidimensionality. For the 2D data processing, merging of individual fractions and searching showed an increase in number of proteins identified from approximately 250 (2/3) typically in a 1D experiments to >450 proteins in 2D 5 step experiments. With more separation steps, an increased number of low-abundance species were reproducibly sampled due to the reduced interference from highly abundant species. The increase in protein identifications can be correlated to the
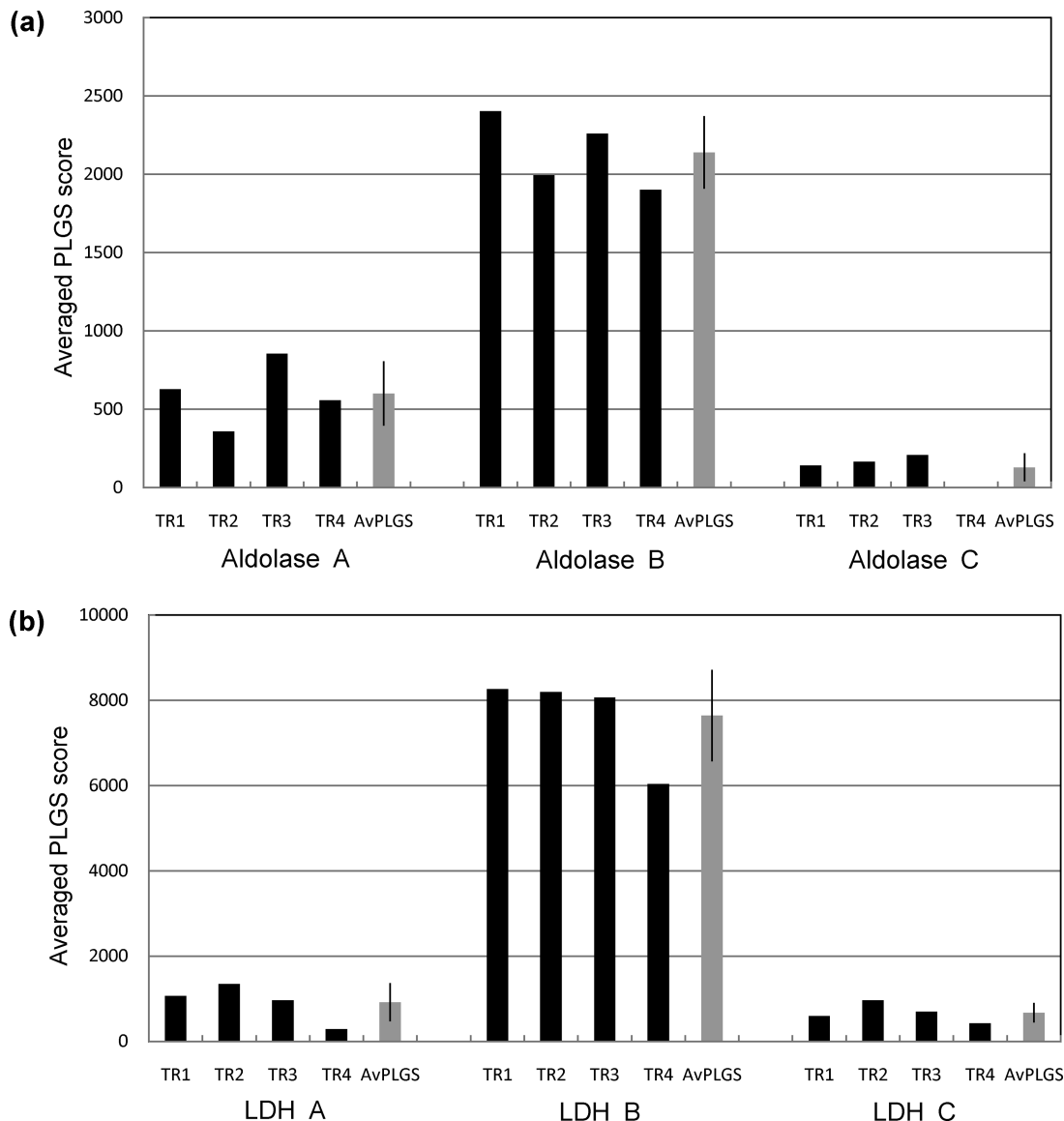
(a)

(b)

Figure 6. Quantitative analysis of reproducibility of isoform assignment in a series of 4 technical replicates from normal renal cortical FFPE extracts. PLGS scores for the replicates are shown for two selected protein examples, (a) aldolase A, B, C and (b) LDH A, B and C. The quantitation shows averaged PLGS scores from triplicate injections for each protein compared in the 4 technical replicates with standard deviations for the replicate series.

total protein load applied to the column. The protein load increases in direct proportion to the number of first-dimension fractions, or steps, that are used, with the optimum load of protein being approximately 500 ng for each first-dimension step. An expansion in peptide numbers and sequence coverage was observed which correlates to the increased protein load on column between the single and two-dimensional separation approaches (data not shown). A comparison of the overall expansion of the number of proteins identified in normal and tumor FFPE extracts using 1D and 2D approaches is summarized in Figure 7 (Supplementary Data 4). For normal FFPE tissue extracts, 284 proteins were identified at high confidence (2/3) in triplicate injection 1D-UPLC-MS^E runs and 505 proteins were identified using a 5 step 2D-UPLC-MS^E. For tumor samples, 236 proteins (2/3) were identified in the 1D run and an expansion to 699 proteins observed with 2D-UPLC-MS^E. The

total number of proteins identified in the 1D triplicate injections was 418 for the normal and 370 for the tumor.

Preliminary Comparative Analysis of Differential Expression in Frozen/FFPE RCC Tissue Extracts by 1D LC−MS/MS. Absolute label-free quantitation utilizing the intensity of the top three peptides from a yeast alchohol dehydrogenase internal standard was used to compare the four samples to assess tumor-normal differential expression profiles for the FFPE and frozen tissue extracts from a single patient. All samples were run in triplicate and tumor/normal expression ratios were based on the averaged concentrations of individual proteins in comparison with the internal standard (Supplementary Data 1). All protein concentrations less than 1 fmol were given an arbitrary value of 1 for the calculation of normal/tumor fold changes. Fold changes plotted for a total of 205 proteins in each data set are shown in Figure 8. The Spearman
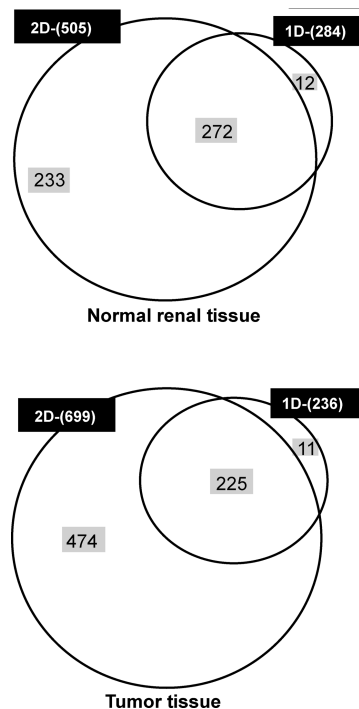
Figure 7. Comparison of the expansion in proteins identified between 1D (400 ng protein) and 2D LC–MS/MS (2 μg) experiments on normal renal and RCC FFPE extracts using the Nano-Acquity UPLC-HDMS quadrupole-oa Tof (Waters, U.K.).

rank correlation of $r = 0.682$; $p < 10^{-15}$ indicates a highly significant moderately strong positive correlation.

Specific examples of proteins which we and others have previously reported as being differentially expressed in RCC tissue[27–29] were examined in these results as a preliminary validation of the overall method (Figure 9). Normal/tumor differential expression patterns previously seen and described here in frozen tissue were also maintained in FFPE tissue extracts, further validating the approach.

Discussion

Current tissue-based proteomic applications, while offering a realistic chance of yielding novel markers, are limited by problems in procuring adequate numbers of tissue samples in its optimum fresh or frozen form, particularly for less common diseases or subtypes.[1,2] Annotated archival tissue banks constitute an untapped resource for biomarker-driven investigations,[16,18,21,30] particularly in the validation phase when large sample cohorts are required. However, covalent formalin-induced modifications in proteins impair immune-reactivity and extraction, making FFPE tissues largely intractable for proteomic analyses.[30,31] The potential implications of harnessing this vast alternate tissue resource has, more recently, triggered a keen research focus on developing detergent-based, heat-induced cross-link reversal strategies to permit effective protein extraction for proteomic applications.[4,6,7,21,32,33] Although still in its infancy, a notable bottleneck in the whole exercise continues to be the lack of a high-throughput extraction methodology which is reproducibly quantitative and cost-effective. Such an option is critical if the method is to be useful in harnessing the potential of the archival proteome for large-scale biomarker discovery and validation research. Reproducible quantitative metrics will enable the comparison and
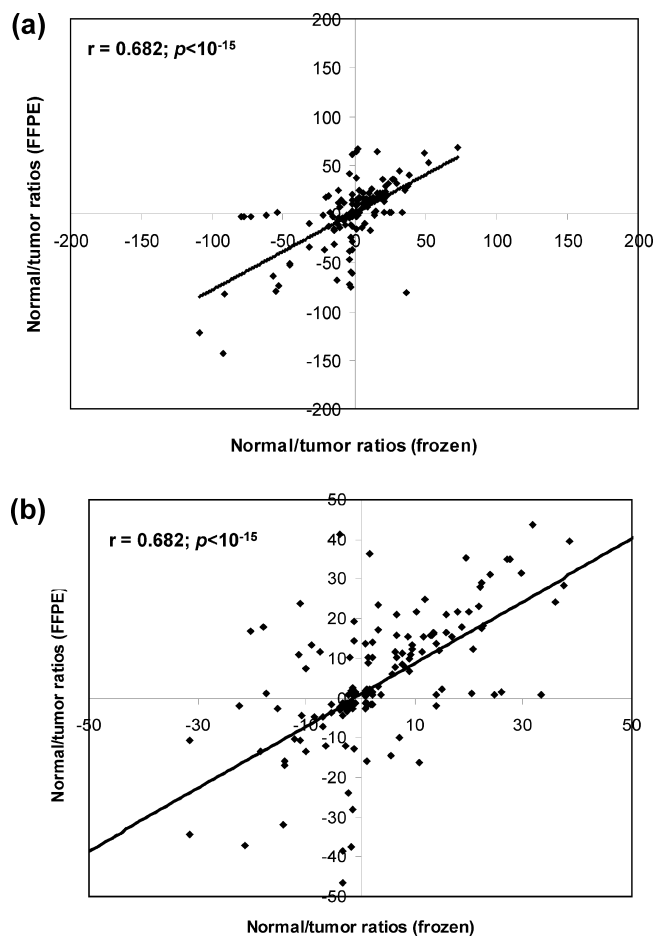

Figure 8. Comparison of normal/RCC tumor protein expression ratios (positive values = higher in tumor, negative = higher in normal tissue) derived from the label-free quantitative analyses of frozen versus FFPE tissue extracts (205 proteins in total). (a) All proteins plotted; (b) zoomed-in version of graph showing the region between 50:1 and 1:50.

refinement of published extraction protocols in this early phase of method development. More importantly, it would allow the quantitative evaluation of a range of preanalytical variables in FFPE extracts (e.g., age of blocks, fixation methods, duration of fixation, etc.) that could potentially interfere with data interpretation, permitting the objective assessment of the usefulness of the vast archival resources.

The surfactant/heat-based extraction methodology reported in the paper offers a cost-effective, user-friendly, single tube protein extraction option for FFPE tissue. Its compatibility with high-throughput mass spectrometry and label-free quantitation make it a powerful tool to begin to further quantitatively explore the archival tissue proteome. A systematic review of detergent options identified the SDS substitute RapiGest as the most effective extraction buffer for archival tissue proteomics. The surfactant was selected on the basis of its suitability for mass spectrometry and trypsin digestion, its heat resistance, and strong detergent/denaturant properties. The increased protein extraction observed is in part attributable to the enhancement of trypsin digestion by the RapiGest buffer and its compatibility with HPLC and MS, obviating the need for detergent cleanup and the sample loss thereby incurred.

Protein extraction and identification comparable to matched fresh frozen tissue extracts were achieved reproducibly, mir-
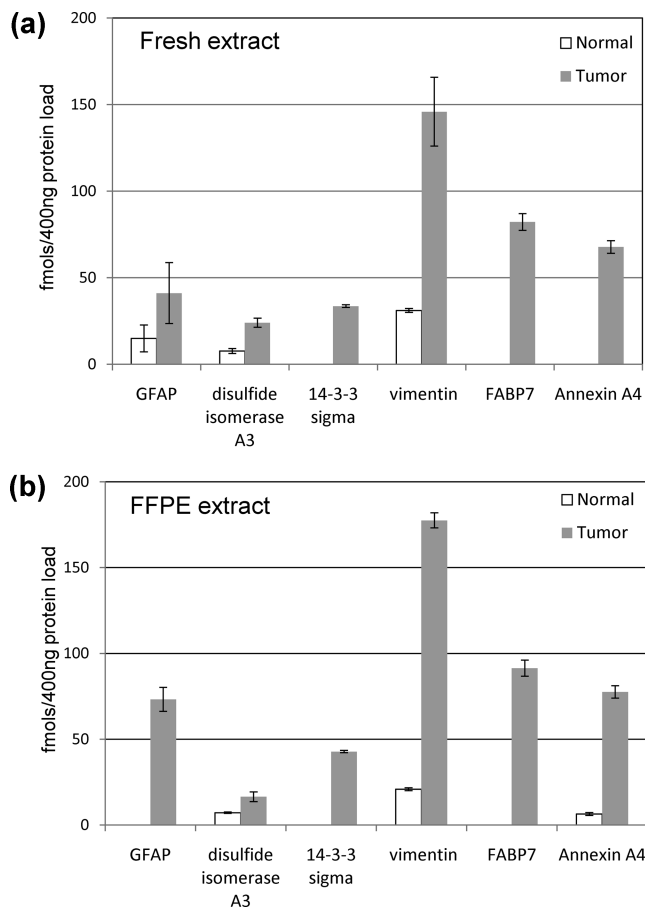
(a)

(b)

Figure 9. Quantitative comparison of normal and tumor expression of several previously described proteins shown to be differentially expressed[27−29] in (a) fresh and (b) FFPE tissue extracts. Ratios calculated from deduced concentrations with reference to an internal standard in femtomoles/400 ng protein load using 1D-LC−MS are averaged from 3 injections and presented with their respective standard deviations.

roring recently published data using the commercially available Liquid Tissue buffer protocols for 1D LC−MS analysis.[15,17] Total protein yields from FFPE and frozen blocks were comparable (and similar to values from the Liquid Tissue and ammonium bicarbonate sonication protocols).[8,15] Percentage overlap between frozen and FFPE protein identifications were 58% (normal renal extract) and 57% (RCC tumor extract) of total proteins identified in our study, which is comparable with 57−67% in published results when similar search parameters were used.[8] In preliminary experiments using the 2D approach, only 50−60% of proteins in each FFPE and frozen tissue data set were common representing only 35−37% of total proteins identified. This highlights the challenges faced particularly with regard to reproducibility of detection of the lower abundance proteins when these types of analyses are extended through further fractionation steps.

Compared to the current leader methodologies that use an MS platform to mine the archival proteome, the RapiGest surfactant protocol, while exhibiting equivalence in protein inventories, has the significant advantage of being a low cost, easy to use quantitative option for suitable for application to large sample cohorts. The slightly higher proportion of proteins identified in our FFPE extract when compared to frozen extract could be attributed to the use of OCT embedded tissue blocks

and the possible interference with LC−MS analysis of the OCT tissue embedding medium.[5,15] Preliminary analysis of FFPE extracts shows proteins with a broad range of molecular functions from all subcellular compartments, including membrane proteins. Analysis of technical reproducibility in 4 parallel separate FFPE extracts from the same tissue block, each injected as a triplicate, showed preservation of 70.2 − 77.6% of the total number of proteins across all injections.

The lack of studies using quantitative methodology in current published literature makes it difficult to compare our protocol with other published protocols, highlighting the relevance of such protocols in the early phase of method development. A similar argument can be applied to lack of adherence to standard criteria for publishing proteomic data in many published papers. The resulting variation in protein numbers attributed to different methodologies pose significant challenges for method comparison and protocol refinement. In a detailed study investigating the equivalence of protein inventories across FFPE and frozen samples, Sprung et al. report a yield of ~400 confident protein identities from 200 μg total protein extract in FFPE tissue, reporting 90% protein equivalence both in numbers of proteins identified and in identities of the protein inventories.[8] Our 1D-LC−MS approach yielded ~235−285 proteins from a considerable smaller protein quantity of 500 ng. The lower number could be attributed to the use of high stringency 2/3 selection criteria from the 3 replicate runs. Total unique protein numbers identified across all three replicates in the experiments were ~418 for normal and 370 for the tumor sample. Comparable approaches with the Liquid Tissue proprietary buffer identified ~179 proteins from FFPE liver extracts.[15]

Label-free quantitative approaches are becoming increasingly popular in mass spectrometry and informatics due to their ability to overcome limitations associated with other labeling techniques like cost, complexity, and time. In addition, there is in principle essentially no limit to the number of experiments that can be compared. In experiments designed to evaluate the label-free quantitation strategy, differential expression profiles for the data sets studied showed comparable expression patterns between fresh and FFPE normal/tumor pairs with a statistically significant preservation in directional regulation for most proteins. Of the proteins that failed to follow this pattern, a significant proportion constituted cytoskeleton derived proteins (e.g., alpha actinin, tropomyosin alpha 4 chain, cytokeratin 18, tubulin alpha 18, myosin light polypeptide 6, tubulin alpha 3 chain, and myosin 9 heavy chain). Further work needs to be done to ascertain if this observation is attributable to an artifact of processing. Furthermore, a modest disparity was also observed in a group of membrane proteins (e.g., moesin, amine oxidase, and peroxiredoxin 5 mitochondrial precursor), corroborating previous reports of aggregation and poor extraction of some membrane proteins when using extended heat protocols.[7] The study also identified a range of previously validated renal cell carcinoma tissue markers from both FFPE and frozen extracts (e.g., vimentin, annexin 4, GFAP, FABP7, HSP27, gamma enolase, superoxide dismutase, etc.).[27−29]

Archival proteomics is still in its infancy and continuous method development to increase depth of tissue mining, sequence coverage, and reliable quantitation is essential. Critical to advancing such studies further are the more detailed examination of its technical use in a larger-scale biomarker discovery experiment with extensive validation of findings, for example, by immunohistochemistry or Western blotting. Fur-

thermore, as with fresh frozen tissue, the contributory effects of a range of preanalytic variables associated with tissue collection, processing, and storage (including surgical ischemia, processing protocols, fixatives and fixation length, and length of storage) that could potentially confound data interpretation need to be quantitatively defined. The label-free surfactant based extraction strategy presented in this paper constitutes a powerful and necessary tool to initiate the objective assessment of this alternate tissue resource and harness the significant opportunities made available by the maturation of proteomic technologies.

Conclusions

The potential contribution of the vast formalin-fixed archival collections to the biomarker discovery and validation initiative cannot be underestimated, particularly in rarer diseases where the acquisition of tissue in its optimal fresh frozen form is increasingly challenging. We report the initial development and validation of an efficient, cost-effective, quantitative extraction methodology permitting the quantitative analysis of archival tissues. Further studies will investigate the variation induced as a consequence of the fixation and storage process and will employ the method in a full-scale comparative biomarker discovery project, but at this stage, only at the 1D LC fractionation level to fully explore its potential.

Supporting Information Available: Protein identified and comparison analysis; comparative analysis of the 4 data sets. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Cho, W. C. Contribution of oncoproteomics to cancer biomarker discovery. *Mol. Cancer* **2007**, *6*, 25.

(2) Rifai, N.; Gillette, M. A.; Carr, S. A. Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* **2006**, *24* (8), 971–983.

(3) Bagnato, C.; Thumar, J.; Mayya, V.; Hwang, S. I.; Zebroski, H.; Claffey, K. P.; Haudenschild, C.; Eng, J. K.; Lundgren, D. H.; Han, D. K. Proteomics analysis of human coronary atherosclerotic plaque - A feasibility study of direct tissue proteomics by liquid chromatography and tandem mass spectrometry. *Mol. Cell. Proteomics* **2007**, *6* (6), 1088–1102.

(4) Hwang, S. I.; Thumar, J.; Lundgren, D. H.; Rezaul, K.; Mayya, V.; Wu, L.; Eng, J.; Wright, M. E.; Han, D. K. Direct cancer tissue proteomics: a method to identify candidate cancer biomarkers from formalin-fixed paraffin-embedded archival tissues. *Oncogene* **2007**, *26* (1), 65–76.

(5) Palmer-Toy, D. E.; et al. Efficient method for the proteomic analysis of fixed and embedded tissues. *J. Proteome Res.* **2005**, *4* (6), 2404–2411.

(6) Prieto, D. A.; Hood, B. L.; Darfler, M. M.; Guiel, T. G.; Lucas, D. A.; Conrads, T. P.; Veenstra, T. D.; B. Krizman, D. B. Liquid Tissue: proteomic profiling of formalin fixed tissues. *Mass Spectrom.* **2005**, *38* (S32−S35), 32−35.

(7) Nirmalan, N. J.; et al. 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* (4), 497–506.

(8) Sprung, R. W.; 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* (8), 1988–1998.

(9) Huang, S. K.; et al. LC/MS-based quantitative proteomic analysis of paraffin-embedded archival melanomas reveals potential proteomic biomarkers associated with metastasis. *PLoS One* **2009**, *4* (2), -.

(10) Balgley, B. M.; et al. Evaluation of archival time on shotgun proteomics of formalin-fixed and paraffin-embedded tissues. *J. Proteome Res.* **2009**, *8* (2), 917–25.

(11) Ikeda, K.; Monden, T.; Kanoh, T.; Tsujie, M.; Izawa, H.; Haba, A.; Ohnishi, T.; Sekimoto, M.; Tomita, S.; Shiozaki, H.; Monden, M. Extraction and analysis of diagnostically useful proteins from formalin-fixed, paraffin-embedded tissue sections. *J. Histochem. Cytochem.* **1998**, *46* (3), 397–404.

(12) Becker, K. F.; et al. Quantitative protein analysis from formalin-fixed tissues: implications for translational clinical research and nanoscale molecular diagnosis. *J. Pathol.* **2007**, *211* (3), 370–8.

(13) Addis, M. F.; et al. Generation of high-quality protein extracts from formalin-fixed, paraffin-embedded tissues. *Proteomics* **2009**, *9* (15), 3815–23.

(14) Hood, B. L.; et al. Proteomic analysis of formalin-fixed prostate cancer tissue. *Mol. Cell. Proteomics* **2005**, *4* (11), 1741–1753.

(15) Scicchitano, M. S.; Dalmas, D. A.; Boyce, R. W.; Thomas, H. C.; Frazier, K. S. Protein extraction from formalin-fixed, paraffin-embedded tissue enables robust proteomic profiles by mass spectrometry. *J. Histochem. Cytochem.* **2009**, *57* (9), 849–860.

(16) Crockett, D. K.; Lin, Z. S.; Vaughn, C. P.; Lim, M. S.; Elenitoba-Johnson, K. S. J. Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. *Lab. Invest.* **2005**, *85* (11), 1405–1415.

(17) Perroud, B.; et al. Grade-dependent proteomics characterization of kidney cancer. *Mol. Cell. Proteomics* **2009**, *8* (5), 971–985.

(18) Hood, B. L.; Darfler, M. M.; Guiel, T. G.; Furusato, B.; Lucas, D. A.; Ringeisen, B. R.; Sesterhenn, I. A.; Conrads, T. P.; Veenstra, T. D.; Krizman, D. B. Proteomic analysis of formalin-fixed prostate cancer tissue. *Molecular & Cellular Proteomics* **2005**, *4* (11), 1741–1753.

(19) Crockett, D. K.; et al. Identification of proteins from formalin-fixed paraffin-embedded cells by LC-MS/MS. *Lab. Invest.* **2005**, *85* (11), 1405–1415.

(20) Hood, B. L.; Conrads, T. P.; Veenstra, T. D. Mass spectrometric analysis of formalin-fixed paraffin-embedded tissue: Unlocking the proteome within. *Proteomics* **2006**, *6* (14), 4106–4114.

(21) Nirmalan, N. J.; et al. Mining the archival formalin-fixed paraffin-embedded tissue proteome: opportunities and challenges. *Mol. BioSyst.* **2008**, *4* (7), 712–720.

(22) Blonder, J.; Veenstra, T. D. Clinical proteomic applications of formalin-fixed paraffin-embedded tissues. *Clin. Lab. Med.* **2009**, *29* (1), 101–113.

(23) Shi, S. R.; Key, M. E.; Kalra, K. L. 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.

(24) Chen, E. I.; et al. Optimization of mass spectrometry-compatible surfactants for shotgun proteomics. *J. Proteome Res.* **2007**, *6* (7), 2529–2538.

(25) Silva, J. C.; et al. Absolute quantification of proteins by LCMSE - A virtue of parallel MS acquisition. *Mol. Cell. Proteomics* **2006**, *5* (1), 144–156.

(26) Vissers, J. P. C.; et al. Quantitative label-free nanoscale LC-MS analysis of HIV drug-treated human T-cell cultures. *Mol. Cell. Proteomics* **2006**, *5* (10), S13–S13.

(27) Unwin, R. D.; et al. Proteomic changes in renal cancer and co-ordinate demonstration of both the glycolytic and mitochondrial aspects of the Warburg effect. *Proteomics* **2003**, *3* (8), 1620–32.

(28) Unwin, R. D.; et al. Serological and proteomic evaluation of antibody responses in the identification of tumor antigens in renal cell carcinoma. *Proteomics* **2003**, *3* (1), 45–55.

(29) Lichtenfels, R.; Dressler, S. P.; Zobawa, M.; Recktenwald, C. V.; Ackerman, A.; Atkins, D.; Kersten, M.; Hesse, A.; Puttkammer, M.; Lottspeich, F.; Seliger, B. Systematic comparative protein expression profiling in clear cell renal carcinoma: A pilot study based on the separation of tissue specimens by two dimensional gel electrophoresis. *Mol. Cell. Proteomics* **2009**, *8*, p. 2827–2842.

(30) Shi, S. R.; et al. Protein extraction from formalin-fixed, paraffin-embedded tissue sections: Quality evaluation by mass spectrometry. *J. Histochem. Cytochem.* **2006**, *54* (6), 739–743.

(31) Shi, S. R.; Liu, C.; Taylor, C. R. Standardization of immunohistochemistry for formalin-fixed, paraffin-embedded tissue sections based on the antigen-retrieval technique: From experiments to hypothesis. *J. Histochem. Cytochem.* **2007**, *55* (2), 105–109.

(32) Becker, K. F.; Schott, C.; Hipp, S.; Metzger, V.; Porschewski, P.; Beck, R.; Nahrig, J.; Becker, I.; Hofler, H. Quantitative protein analysis from formalin-fixed tissues: implications for translational clinical research and nanoscale molecular diagnosis. *J. Pathol.* **2007**, *211* (3), 370–378.

(33) Guo, T.; Wang, W. J.; Rudnick, P. A.; Song, T.; Li, J.; Zhuang, Z. P.; Weil, R. J.; DeVoe, D. L.; Lee, C. S.; Balgley, B. M. Proteome analysis of microdissected formalin-fixed and paraffin-embedded tissue specimens. *J. Histochem. Cytochem.* **2007**, *55* (7), 763–772.