Published in final edited form as:

J Chromatogr A. 2011 November 11; 1218(45): 8168–8174. doi:10.1016/j.chroma.2011.09.022.

Microproteomic Analysis of 10,000 Laser Captured Microdissected Breast Tumor Cells Using Short-Range Sodium Dodecyl Sulfate-Polyacrylamide Gel Electrophoresis (SDS-PAGE) and Porous Layer Open Tubular (PLOT) LC-MS/MS

Dipak Thakur¹, Tomas Rejtar¹, Dongdong Wang¹, Jonathan Bones¹, Sangwon Cha¹, Buffie Clodfelder-Miller², Elizabeth Richardson³, Shemeica Binns³, Sonika Dahiya³, Dennis Sgroi³, and Barry L. Karger^{1,*}

¹Barnett Institute and Dept. of Chemistry and Chemical Biology, Northeastern University, Boston, MA 02115

²Cellular and Molecular Neuropathology Core, University of Alabama, Birmingham, AL 35294

³Molecular Pathology Unit, Massachusetts General Hospital, Charlestown, MA 02129

Abstract

Precise proteomic profiling of limited levels of disease tissue represents an extremely challenging task. Here, we present an effective and reproducible microproteomic workflow for sample sizes of only 10,000 cells that integrates selective sample procurement *via* laser capture microdissection (LCM), sample clean up and protein level fractionation using short-range SDS-PAGE, followed by ultrasensitive LC-MS/MS analysis using a 10 μ m i.d. porous layer open tubular (PLOT) column. With 10,000 LCM captured mouse hepatocytes for method development and performance assessment, only 10% of the in-gel digest, equivalent to ~1000 cells, was needed per LC-MS/MS analysis. The optimized workflow was applied to the differential proteomic analysis of 10,000 LCM collected primary and metastatic breast cancer cells from the same patient. More than 1100 proteins were identified from each injection with >1700 proteins identified from three LCM samples of 10,000 cells from the same patient (1123 with at least two unique peptides). Label free quantitation (spectral counting) was performed to identify differential protein expression between the primary and metastatic cell populations. Informatics analysis of the resulting data indicated that vesicular transport and extracellular remodeling processes were significantly altered between the two cell types. The ability to extract meaningful biological information from limited, but highly informative cell populations demonstrates the significant benefits of the described microproteomic workflow.

Keywords

microproteomics; porous layer open tubular column; laser capture microdissection; breast cancer; sample preparation; low cell numbers

© 2011 Elsevier B.V. All rights reserved.

*Corresponding author: b.karger@neu.edu, Tel: +1 617 373 2867, Fax: +1 617 373 2855.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

1.0 INTRODUCTION

Laser capture microdissection (LCM) is an indispensable tool for the isolation of homogeneous populations of biologically distinct cell subtypes from tissue samples [1–4]. Collection of a large number of cells using LCM, however, can be a time consuming and labor intensive process. Furthermore, the number of cells of a specific pathological subtype may be limited. Therefore, a need exists for a specific and sensitive workflow to facilitate comprehensive proteomic profiling from a limited number of biologically important cells.

Due to the low levels of protein present, optimized sample preparation is a critical component in this analysis. A significant analytical challenge is to achieve maximum protein recovery. The addition of SDS to cells facilitates rapid cell lysis and protein solubilization; however, removal of SDS is required prior to further downstream proteomic analysis by LC-MS [5]. Filter aided sample preparation (FASP) methods have recently been described for SDS removal and subsequent protein digestion [6]. An alternative to the FASP based methods is short-range SDS-PAGE separation on a highly cross-linked gel [7, 8]. The short-range separation and the highly concentrated gel (1) help minimize protein losses, (2) facilitate SDS removal in the leading front and (3) beneficially result in more concentrated protein gel plugs for subsequent in-gel digestion than if traditional SDS-PAGE separation distances are employed [7, 8]. Moreover, the short-range separation should potentially reduce the accumulation of keratin impurities in the gel.

Following sample preparation, the chromatographic separation coupled on-line to the mass spectrometer requires optimization. High sensitivity and high resolving power are desired to separate the low quantities and highly complex peptide mixtures. Currently, 75 μm i.d. capillary columns packed with C_{18} stationary phases are commonly used for the analysis of in-gel digested proteins with operating flow rates of 100 nL/min or higher. As the electrospray ionization efficiency increases with decreasing mobile phase flow rate [9–13], operation at 20 nL/min or lower can greatly enhance the electrospray signal [9, 14]. Furthermore, due to reduced droplet size when using such low flow rates, ion suppression is also minimized. To operate efficiently at such low flow rates while maintaining high separation performance, significantly narrower LC column IDs are required. While packed beads and monolithic columns have been previously used [15, 16], we have selected to use porous layer open tubular (PLOT) columns of 10 μm ID. As previously shown, PLOT columns, due to their open tubular structure, allow column lengths of 3 meters or greater and are compatible with standard LC pumps (6000 psi) [14]. Attomole level sensitivity and high resolving power (peak capacities in excess of 400) are readily achieved using PLOT LC-MS [14].

Here, an optimized sample preparation platform using short-range Tricine SDS-PAGE on a 16% cross-linked gel followed by ultrasensitive PLOT-LC-MS is described. To demonstrate the feasibility of the approach, microproteomic analysis of 10,000 LCM collected malignant breast cells from primary and metastatic sites in a given patient was conducted. Greater than 1700 proteins have been identified per cell type (>1100 with 2 or more unique peptides) from three LCM samples of 10,000 cells from the same patient, with only ~1000 cells or a few hundred nanograms of total protein consumed per injection. Evaluation of differential protein expression following quantitative analysis (spectral counting) revealed biologically significant alterations in the levels of proteins involved in vesicular transport and extracellular matrix remodeling between the two cell subtypes. The described analytical platform facilitates microproteomic analysis, thus allowing biological insight into limited but highly informative LCM captured cell populations.

2.0 MATERIALS AND METHODS

2.1 Chemicals

LC-MS Optima grade solvents used throughout this study were obtained from Thermo Fisher Scientific (Fairlawn, NJ). All chemicals were from Sigma–Aldrich (St. Louis, MO) and were of the highest available purity. Fused-silica capillary tubing was purchased from Polymicro Technologies (Phoenix, AZ). Pico Clear tee and union fittings were received from New Objective (Woburn, MA). Sequencing grade trypsin was obtained from Promega (Madison, WI). Reducing agent (10X), Tricine gels (16%T), Tricine SDS sample buffer (2X) and Tricine SDS running buffer (10X) were all purchased from Invitrogen (Carlsbad, CA).

2.2 Clinical Specimens and Laser Capture Microdissection

Breast tumor and axillary lymph node tissue samples were obtained by surgical excision at Massachusetts General Hospital from a patient diagnosed with metastatic breast cancer. The collected tissues were frozen fresh within 20 minutes of surgery and stored at -80°C . The study was approved by the Massachusetts General Hospital human research committee in accordance with National Institute of Health human research study guidelines. The tissue samples were prepared as previously described without eosin tissue staining [17]. Enrichment of malignant epithelial cells from breast (primary) and lymph node tissue (metastatic) was performed using IR ($\lambda = 810\text{ nm}$) laser capture (PixCell IIe LCM Molecular Devices, Mountain View, CA) [18]. Approximately 10,000 cells were collected per tissue sample onto a single LCM cap. Three individual samples of 10,000 cells were collected for both sample types from the same patient and subjected to microproteomic analysis. For preliminary method development and optimization, approximately 10,000 hepatocyte cells from mouse liver tissue sections were collected using the Arcturus Veritas LCM system (Molecular Devices) at the laser microdissection facility at the University of Alabama.

2.3 Cell Lysis, Short-Range SDS-PAGE and In-Gel Digestion

An ExtracSure device (MDS Analytical Technologies, Sunnyvale, CA) was placed on the LCM cap containing 10,000 cells and cell lysis was immediately performed upon application of 10 μL of lysis buffer (Novex 2X Tricine SDS sample buffer: NuPAGE reducing agent: water, 5:1:4). The cell lysate was transferred to a microcentrifuge tube. The LCM cap was rinsed with two aliquots of lysis buffer (7.5 μL each), and the combined cell lysate was then incubated at 85°C for 2 minutes. The cell lysates from the six samples were separately loaded onto a 16% Tricine gel and electrophoresis was performed using 125 volts for a period of 20 minutes leading to approximately 2.5 cm migration of the buffer front down the gel.

After staining the gel with Coomassie Blue (SimplyBlue Safestain, Invitrogen), each gel lane was cut into three sections, as shown in Figure 1, and each section was in-gel digested, as previously described [19]. Briefly, the gel section was cut into cubes of approximate 1 mm^3 with subsequent reduction with dithiothreitol to ensure complete disulphide bond reduction followed by alkylation with iodoacetamide in the dark. Following extensive washing with ammonium bicarbonate and acetonitrile, in-gel digestion using 6 ng/ μL of a trypsin solution (in 50 mM ammonium bicarbonate solution) was performed at 37°C for 12 hours. The supernatant was removed and preserved, and the remaining digested peptides were extracted by hydration of the gel using 5% formic acid, followed by complete dehydration of the gel pieces by acetonitrile. The extracts were combined and reduced to dryness under vacuum using a centrifugal evaporator.

2.4 Nano LC-ESI-MS with 10 μ m i.d. PLOT Column

The on-line micro-SPE-PLOT LC-MS system was similar to that previously described [20]. Briefly, the PLOT column was prepared as follows: a 10 μ m i.d. capillary that had been pretreated with 3-(trimethoxysilyl) propyl methacrylate was filled with a degassed solution containing 5 mg of AIBN, 200 μ L of styrene, 200 μ L of divinylbenzene and 600 μ L ethanol. Both ends of the capillary were sealed, and polymerization was performed for 16 hours at 74°C. The column was washed overnight with acetonitrile and stored with water until further use.

The in-gel digest of each SDS-PAGE gel section was dissolved in 15 μ L of mobile phase A (0.1% formic acid in water). 1.5 μ L of the reconstituted digest (10% of the total volume) was loaded on the micro-SPE column (50 μ m i.d. PS-DVB monolithic micro-SPE columns) at a flow rate of approximately 1 μ L/min for desalting. An Ultimate 3000 nanoLC (Dionex, Sunnyvale, CA) was used to deliver a linear gradient of 2% to 30% solvent B (0.1% formic acid in acetonitrile) over 180 minutes. The mass spectrometer (LTQ XL, Thermo Fisher Scientific, San Jose, CA) was operated in the data-dependent mode using one full MS scan, followed by MS/MS acquisition for the 8 most intense ions with a normalized collision energy of 35%. Dynamic exclusion was employed with no repeat count for selected precursor ions with an exclusion duration of 60 seconds.

2.5 Protein Identification

Acquired MS/MS scans were converted into DTA files by Extract-MSn (version 4.0, Thermo Fisher Scientific) and searched against the SwissProt human database (release 2010_06 downloaded in July 2010) combined with a database containing reversed sequences using the Sequest algorithm (cluster version 27, rev.12, Thermo Fisher Scientific). The search results were stored in a CPAS system (version 9.10 LabKey, Seattle, WA) [21]. The peptide mass search tolerance was set to 1.4 Da, and the fragment ion mass tolerance was 1.0 Da. Full tryptic enzyme specificity was selected with up to 2 missed cleavage sites allowed. Cysteine carbamidomethylation was considered as a fixed modification. The search results (identified peptides) were filtered by Xcorr $\geq$ 1.9 for charge state +1, $\geq$ 2.2 for charge state +2 and $\geq$ 3.8 for charge state +3 and by PeptideProphet (Institute for System Biology, Seattle, WA) using a peptide probability $\geq$ 0.95. Since some peptides identified multiple proteins, ProteinProphet (Institute for System Biology, Seattle, WA) was used to assign peptides to protein groups with acceptance criteria specified using a probability $\geq$ 0.9 resulting in the false discovery rate (FDR) $<$ 2% at the protein level [22–24]. For the murine hepatocytes used for the preliminary experiments, the SwissProt murine database (release 2010_06 downloaded in July 2010) was used, with all other pre- and post-search parameters similar.

2.6 Quantitation by Spectral Counting

Comparison of differential protein abundance between the primary and metastatic breast cancer cell populations was determined based on spectral counts, the total number of MS/MS spectra assigned to a particular protein, in individual samples [25]. Proteins with two or more unique peptide identifications were considered for identifying differentially abundant proteins. Spectral count data from triplicate analyses of individual primary and metastatic breast cancer samples were compared using the TFold analysis tool contained within PatternLab software [26]. Differentially abundant proteins were determined using a *t*-test with a statistical significance threshold $p < 0.05$ and a Benjamini-Hochberg false discovery rate (BH-FDR) [26] of less than 10%.

2.7 Reproducibility of the Replicate Analyses of Primary and Metastatic Breast Cancer Samples

For each identified protein, average spectral counts and their relative standard deviations were calculated from the three replicate runs. From the list of total proteins, those proteins identified by an average spectral count ≥ 2 were selected for reproducibility analysis. The selected proteins were divided into three bins, such that each bin contained equal numbers of proteins. These bins were assigned as “low”, “middle” and “high” based on the average spectral count values. The protein groups obtained from replicates of primary and metastatic breast cancer samples were then compared. Statistical calculations to compare variability associated with replicate analysis of the primary and metastatic breast cancer cells were performed using R, (www.r-project.org).

2.8 Gene Ontology Annotation with DAVID

Functional annotation tools in DAVID [27, 28] were used to extract gene ontology terms for all the identified proteins and to determine significantly enriched biological characterizations for differentially abundant proteins. Lists of up and down regulated proteins in metastatic samples were separately processed against a background list of all the proteins that were submitted to the PatternLab software. The annotation clusters with an enrichment score greater than 1.3 and an FDR $< 20\%$ were selected to explore the biological differences between the two sample types [28].

3.0 RESULTS AND DISCUSSION

Laser capture microdissection facilitates the collection of homogenous populations of cells of a particular pathology from which highly informative biological meaning can potentially be extracted. A complication with the use of LCM is the low levels of biological material available. Unlike genomic methods where amplification technology exists, proteomic analysis of low number but information rich cell types requires the availability of ultrasensitive methods for the in-depth qualitative and quantitative analysis of these LCM collected populations. Here, we present the development of an optimized sample preparation approach, tailored for use with low cell number samples and PLOT LC-MS analysis. LCM collected murine hepatocytes were used for the development and optimization of the sample preparation approach. Once optimized, the technology was then applied for the quantitative proteomic profiling of primary and metastatic cancer cells collected from resected tissues from a patient with breast cancer.

3.1 Microproteomic workflow optimization

With the motivation of maximum protein recovery by means of minimal sample handling, an integrated sample preparation workflow for the manipulation of the LCM captured cell populations prior to ultrasensitive PLOT LC-MS analysis was developed. The collected cell population was lysed *in situ* on the LCM cap using only 10 μL of a 4% SDS based sample reducing buffer as such sample buffers have been previously recommended for complete proteome solubilization [6]. Lysis on the cap was rapid, with complete disappearance of cellular material noted within 3 minutes when visually monitoring the process with an optical microscope. The use of a small volume of lysis buffer facilitated subsequent washing of the LCM cap to result in a total sample volume suitable for direct loading on to the SDS-PAGE gel without further sample manipulation.

The benefits of SDS-PAGE for protein fractionation and sample cleanup are well known; however, for limited sample amounts, performing SDS-PAGE using standard gels and typical separation distances can result in less than optimum conditions. The composition of and the separation distance within the SDS-PAGE gel was investigated. Fig. 1 depicts the

separation of 1 and 5 μg of mouse hepatocyte-derived proteins as a model system using commercially available a 4–12% Bis-Tris gradient with a 10 cm separation distance in comparison to 16% Tricine SDS-PAGE gels using only a 2.5 cm separation distance, both followed by Coomassie Blue staining. As visually apparent from Fig. 1, the shorter separation distance on the highly cross-linked Tricine SDS-PAGE gel resulted in more intense protein staining (higher concentration) as a result of the confinement of the sample constituents into a more limited gel space as compared to the gradient gel. This tighter geometric confinement within the highly cross-linked gel aided in reducing protein loss. To further investigate these observations, both the gradient and the highly cross-linked gel were cut into three identical size sections and subjected to in-gel tryptic digestion and LC-MS analysis using a standard 0.075×100 mm i.d. capillary column packed with $3 \mu\text{m}$ C_{18} particles. It can be seen in Table 1 that the number of peptides and proteins identified using the short-range Tricine SDS-PAGE gel was approximately four times higher than the number obtained on either the short or long run 4–12% gradient gel. The higher number of proteins identified from the 16% Tricine gel is thought to arise due to effective entrapment of proteins within the highly crosslinked gel matrix. Considering the separation distance on the gradient gel, the percentage of acrylamide encountered by the proteins over the course of the electrophoresis distance is only ~4–6%. There is a high probability that medium to low molecular weight proteins may be eluted from this larger pore gel matrix during manipulation of the gel following electrophoresis and prior to digestion. Therefore, the concentration of proteins in the small highly crosslinked gel volume provides improved proteomic coverage when the sample amount is limited. High molecular proteins that may not electrophorese into the 16% Tricine gel will still be collected in the stacking gel that was part of gel section #1, see Figure 1.

With optimum parameters in place for sample preparation, fractionation and tryptic digestion, the final analytical parameter requiring optimization for microproteomic analysis was the scale of the LC separation. Initial evaluation of the complete workflow was performed using 10,000 LCM collected murine hepatocytes as a model system. Aliquots equivalent to 3%, 10% and 20% of the total sample were analyzed by PLOT LC-MS/MS on a LTQ-XL linear ion trap mass spectrometer using a gradient time of 3 hours to determine the optimum sample load on the PLOT column. To facilitate desalting of the digested sample, a PS-DVB monolith SPE pre-column was employed prior to injection [20]. The number of proteins identified in each of the different sample loading amounts is presented in Fig. 2(A). As expected, an increase in the number of identified proteins was observed with an increase in sample loading up to 10% of the total sample, (equivalent to 1000 cells). As no increase in the number of proteins was observed upon further increasing the sample load, 10% sample introduction was selected for all subsequent proteomic analysis. Importantly, this low sample consumption can allow multiple injections of a given sample even considering the limited amount of protein.

The LC gradient time was next optimized. 1, 3, 5 and 7 hour gradient times were evaluated by injecting 10% of the murine hepatocytes digest in each run. The number of identified proteins for each gradient are depicted in Fig. 2(B). As expected, a significant increase in the number of identified proteins was observed by increasing the gradient time from 1 to 3 hours. Longer times only resulted in a minor increase in the number of identified proteins. As a result, a gradient time of 3 hours was selected for all subsequent proteomic analyses.

Using the optimized parameters, a triplicate PLOT LC-MS/MS analysis was performed for each of the three gel sections excised from the short-range SDS-PAGE separation of the lysate from 10,000 LCM collected murine hepatocytes. The number of proteins identified per gel section per sample is presented in Table 2. More than 1000 proteins were consistently identified from each LC-MS/MS replicate of the three SDS-PAGE sections,

approximately 700 proteins per replicate were identified with two or more unique peptides. The total number of proteins found from the three replicate runs was close to 1500, with 1100 identified with at least two unique peptides. It needs to be emphasized that these numbers are based on using 1D LC on a linear ion trap mass spectrometer. Larger numbers of identified proteins can be expected if 2D LC and a faster ion trap (e.g. Velos) coupled to a high resolution MS (e.g. Orbitrap) were employed. It was also found that the level of protein identification degeneracy of the three gel sections was on average less than 10%. This low level suggests that a reasonable level of protein separation can be achieved using the short-range electrophoretic run on the highly cross-linked Tricine gel. The results in Table 2 demonstrate the effectiveness of the workflow for the proteomic profiling of a limited number of LCM collected cells.

3.2 Proteomic Analysis of 10,000 LCM Captured Breast Cancer Cells

10,000 primary and metastatic breast cancer cells were collected by LCM from resected tissue obtained from the same patient for analysis using the optimized microproteomic workflow. Triplicate samples (3 primary and 3 metastatic) were independently analyzed and subjected to quantitative comparison based on spectral counting. This experimental design was adopted to evaluate the complete sample preparation process. From each analysis of the primary breast cancer samples close to 1100 proteins were identified as shown in Table 3. As an example of the PLOT column performance, an overlay chromatogram of the base peak traces for each gel fraction from one of the replicates is shown in Fig. 4. Taken together, the three samples of the LCM collected primary breast cancer cells from the single patient identified more than 1700 unique proteins, with more than 1100 with at least two unique peptides, again using 1D LC and the LTQ linear ion trap. A similar number of proteins were identified for the LCM collected metastatic cells (data not shown).

Since the three primary and metastatic samples were from the same patient, the variation associated with the total microproteomic workflow (LCM to LC-MS) was examined. A simple statistical comparison based on the relative standard deviation (RSD) of spectral counts obtained from the analyses of the primary and metastatic samples was performed. For comparison, proteins with an average spectral count ≥ 2 were selected and divided into three equal sized groups as outlined in the Experimental Section. As shown in the box-plots presented in Fig. 3, the median values of the RSD of the spectral counts for each associated protein group (low, medium and high) were 0.27, 0.20 and 0.21, respectively, for the primary tumor cell samples, whereas the median values for the metastatic cell samples were 0.31, 0.22 and 0.20, respectively. In addition, the interquartile distances for protein groups between two samples types revealed similar trends in the spectral count variability. Since these RSDs represent the variation in the total analytical process, the workflow is demonstrated to be quite reproducible. Furthermore, the results were obtained over a period of 9 days analysis time demonstrating stable performance of the PLOT LC-MS/MS platform. The biological information available from the analysis of such a limited number of cells of the two populations was next examined.

To study differential protein expression between the primary and metastatic cell samples, the TFC module of the PatternLab software platform was employed. The criteria for differential protein abundance were based on a *t*-test statistical significance threshold $p < 0.05$ and a Benjamini-Hochberg false discovery rate (BH-FDR) of less than 10% [26]. Using these criteria, a total of 109 proteins were found to be altered between the two sample types. Eighty five proteins were found to be up-regulated in the metastatic samples, whereas 24 proteins were found to be down-regulated at the statistically significant level.

The 109 differentially expressed proteins were then submitted to DAVID for gene ontology (GO) analysis. GO terms with enrichment scores > 1.3 and a false discovery rate $< 20\%$ were

selected as overrepresented functional categories [28]. The eighty five proteins that were up-regulated in the metastatic breast cancer cells were linked to 47 overrepresented GO terms clustering into 4 functional categories, as listed in Table 4. A number of proteins up-regulated in the metastatic samples, such as coatamer subunits (gamma-2, epsilon, alpha and beta), transmembrane emp24 domain-containing protein 10 and Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1 are associated with vesicles targeting, transport and localization functions. It is known that metastatic cancer cells implement various exocytic routes to communicate important information for growth, migration and matrix degradation [31]. Proteins belonging to the 'cytoplasmic vesicle' and 'signal peptide' functional categories include endoplasmic reticulum protein ERp29, endoplasmin, Ras-related protein Rab-14, cathepsin D, clusterin, heat shock cognate 71 kDa protein, dihydrolipoyl dehydrogenase, mitochondrial and fibronectin. Proteins belonging to the cytoplasmic vesicle category are responsible for mediating vesicular transport between organelles of secretory and endocytic system. Additionally, changes in the regulation of extracellular matrix proteins in invasive breast cancer tumors has been observed previously [32]. As found in this study, the up-regulation of proteins such as three chains of collagen VI (alpha-1, alpha-2 and alpha-3) and tenascin likely represent remodeling of extracellular matrix in metastatic cancer cells as compared to primary cancer cells. Finally, glycolysis, an important biological function associated with energy metabolism, was found to be down-regulated in the metastatic samples. Proteins identified with altered expression levels associated with glycolysis were triose phosphate isomerase, phosphoglycerate kinase 1 and alpha-enolase. Taken together these findings are suggestive, as expected, of changes in the extracellular matrix, vesicle pathways and cellular metabolism of metastatic cells in comparison to the primary breast cancer cells. It can be concluded that the analytical scheme developed in this paper can successfully allow effective proteomic analysis of limited sample amounts to aid in understanding biological processes related to disease.

4.0 CONCLUSIONS

A microproteomic workflow featuring LCM of cells from tissue, *in situ* lysis, short-range SDS-PAGE on a highly cross-linked gel and ultrasensitive PLOT LC-MS/MS has been described for the comparative proteomic analysis of 10,000 LCM collected primary and metastatic breast cancer cells. The use of short-range SDS-PAGE, relative to longer gel migration (10 cm), permitted the identification of increased numbers of proteins in such limited samples using the 10 μ m ID PLOT column for LC-MS/MS analysis. Optimization of the sample loading amount on the PLOT column revealed maximum performance when the lysate equivalent to approximately 1000 cells was injected, resulting in the identification of more than 1000 proteins, even with 1D separation and use of a LTQ linear ion trap MS. Furthermore, this low sample consumption enabled multiple LC-MS/MS analysis on the same limited sample. When applied for the comparative proteomic analysis of cancer cells enriched from the primary and metastatic sites in the same patient, the optimized workflow demonstrated good performance with >1700 proteins identified per cell type. Using spectral counting for a quantitative comparison of differential protein expression between the two sites, increases in protein abundances related to expected biologically relevant categories such as vesicular transport, extracellular matrix remodeling and glycolysis were observed for the metastatic cell samples. Examining three samples of the two cell types from the same patient demonstrated the good quantitative reproducibility possible for the full analytical scheme from LCM to LC-MS. Work is continuing on further optimization of the platform both in terms of the sample preparation steps, enhanced separation (2D LC) and the PLOT column. These advances will be coupled to a hybrid LTQ-FT mass spectrometer for increased depth of proteomics.

Acknowledgments

This work was supported by the Susan G. Komen Breast Cancer Foundation (BLK and DCS), NIH grant GM15847 (BLK), NIH R01-CA112021 (DCS), NSF NS57098 (BCM) and the Avon Foundation (DCS). Contribution number 961 from the Barnett Institute.

6.0 REFERENCES

1. Bonner RF, EmmertBuck M, Cole K, Pohida T, Chuaqui R, Goldstein S, Liotta LA. Science. 1997; 278:1481. [PubMed: 9411767]
2. EmmertBuck MR, Bonner RF, Smith PD, Chuaqui RF, Zhuang ZP, Goldstein SR, Weiss RA, Liotta LA. Science. 1996; 274:998–1001. [PubMed: 8875945]
3. Fend F, Raffeld M. J Clin Pathol. 2000; 53:666–672. [PubMed: 11041055]
4. Johann DJ, Rodriguez-Canales J, Mukherjee S, Prieto DA, Hanson JC, Emmert-Buck M, Blonder J. J Proteome Res. 2009; 8:2310–2318. [PubMed: 19284784]
5. Nagaraj N, Lu AP, Mann M, Wisniewski JR. J Proteome Res. 2008; 7:5028–5032. [PubMed: 18839980]
6. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Nat Meth. 2009; 6:359–362.
7. Liebler DC, Ham AJL. Nat Meth. 2009; 6:785–785.
8. Lapierre LA, Avant KM, Caldwell CM, Ham AJL, Hill S, Williams JA, Smolka AJ, Goldenring JR. American Journal of Physiology - Gastrointestinal and Liver Physiology. 2007; 292:G1249–G1262. [PubMed: 17255364]
9. Ivanov AR, Zang L, Karger BL. Anal Chem. 2003; 75:5306–5316. [PubMed: 14710807]
10. Luo QZ, Tang KQ, Yang F, Elias A, Shen YF, Moore RJ, Zhao R, Hixson KK, Rossie SS, Smith RD. J Proteome Res. 2006; 5:1091–1097. [PubMed: 16674098]
11. Shen YF, Moore RJ, Zhao R, Blonder J, Auberry DL, Masselon C, Pasa-Tolic L, Hixson KK, Auberry KJ, Smith RD. Anal Chem. 2003; 75:3596–3605. [PubMed: 14570215]
12. Smith RD, Shen YF, Tang KQ. Accounts of Chemical Research. 2004; 37:269–278. [PubMed: 15096064]
13. Tang L, Kebarle P. Anal Chem. 1993; 65:3654–3668.
14. Yue GH, Luo QZ, Zhang J, Wu SL, Karger BL. Analytical Chemistry. 2007; 79:938–946. [PubMed: 17263319]
15. Patel KD, Jerkovich AD, Link JC, Jorgenson JW. Anal Chem. 2004; 76:5777–5786. [PubMed: 15456298]
16. Luo Q, Page JS, Tang K, Smith RD. Anal Chem. 2006; 79:540–545. [PubMed: 17222018]
17. Sgroi DC, Teng S, Robinson G, LeVangie R, Hudson JR, Elkahoul AG. Cancer Res. 1999; 59:5656–5661. [PubMed: 10582678]
18. Ma XJ, Salunga R, Tuggle JT, Gaudet J, Enright E, McQuary P, Payette T, Pistone M, Stecker K, Zhang BM, Zhou YX, Varnholt H, Smith B, Gadd M, Chatfield E, Kessler J, Baer TM, Erlander MG, Sgroi DC. Proc Natl Acad Sci U S A. 2003; 100:5974–5979. [PubMed: 12714683]
19. Gu Y, Wu SL, Meyer JL, Hancock WS, Burg LJ, Linder J, Hanlon DW, Karger BL. J Proteome Res. 2007; 6:4256–4268. [PubMed: 17902640]
20. Luo Q, Yue G, Valaskovic GA, Gu Y, Wu SL, Karger BL. Anal Chem. 2007; 79:6174–6181. [PubMed: 17625912]
21. Rauch A, Bellew M, Eng J, Fitzgibbon M, Holzman T, Hussey P, Igra M, Maclean B, Lin CW, Detter A, Fang RH, Faca V, Gafken P, Zhang HD, Whitaker J, States D, Hanash S, Paulovich A, McIntosh MW. J Proteome Res. 2006; 5:112–121. [PubMed: 16396501]
22. Elias JE, Haas W, Faherty BK, Gygi SP. Nature Methods. 2005; 2:667–675. [PubMed: 16118637]
23. Elias JE, Gygi SP. Nature Methods. 2007; 4:207–214. [PubMed: 17327847]
24. Higdon R, Hogan JM, Van Belle G, Kolker E. Omics-a Journal of Integrative Biology. 2005; 9:364–379. [PubMed: 16402894]
25. Liu HB, Sadygov RG, Yates JR. Anal Chem. 2004; 76:4193–4201. [PubMed: 15253663]
26. Carvalho PC, Fischer JS, Chen EI, Yates JR, Barbosa VC. BMC Bioinformatics. 2008; 9

27. Dennis G, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. *Genome Biology*. 2003; 4
28. Huang DW, Sherman BT, Lempicki RA. *Nature Protocols*. 2009; 4:44–57.
29. Martin SE, Shabanowitz J, Hunt DF, Marto JA. *Anal Chem*. 2000; 72:4266–4274. [PubMed: 11008759]
30. Røgeberg M, Wilson SR, Greibrokk T, Lundanes E. *J Chromatogr A*. 2010; 1217:2782–2786. [PubMed: 20299025]
31. Hendrix A, Westbroek W, Bracke M, De Wever O. *Cancer Res*. 2010; 70:9533–9537. [PubMed: 21098711]
32. Emery LA, Tripathi A, King C, Kavanah M, Mendez J, Stone MD, de las Morenas A, Sebastiani P, Rosenberg CL. *Am J Pathol*. 2009; 175:1292–1302. [PubMed: 19700746]

Highlights

- A microproteomic workflow for sample sizes of only 10,000 cells that integrates selective sample procurement *via* laser capture microdissection (LCM), sample clean up and protein level fractionation using short-range SDS-PAGE, followed by ultrasensitive LC-MS/MS analysis using a 10 μ m i.d. porous layer open tubular (PLOT) column is described.
- An in-gel digest, equivalent to only ~1000 cells, was needed per LC-MS/MS analysis.
- Application to differential proteomic analysis of 10,000 LCM collected primary and metastatic breast cancer cells from the same patient is described. More than 1100 proteins were identified from each injection with >1700 proteins identified from three LCM samples (1123 with at least two unique peptides).
- Informatics analysis of the resulting data indicated that vesicular transport and extracellular remodeling processes were significantly altered between the two cell types.

Figure 1. Selection of gel type and SDS-PAGE separation distance for proteomic analysis of small sample amounts. **(A)** NuPAGE Bis-Tris gel (4–12%T) was used to separate 1 µg and 5 µg of hepatocyte lysate over a length of 10 cm. Lane 1 corresponds to the 1 µg protein load. For improved visualization, the second lane represents 5 µg of protein loading. The third lane contains the molecular weight markers. **(B)** NuPAGE Tricine gel (16%T) was used to separate 1 µg (lane 1) and 5 µg (lane 2) of hepatocyte lysate over a gel length of 2.5 cm. Note that molecular weight markers (third lane) are separated with the 2.5 cm separation distance.

Figure 2. Optimization of LC-MS parameters. **(A)** Optimization of sample loading amount. Shown is a plot of the number of proteins identified *versus* the estimated percentage of corresponding in-gel digest (peptides) injected onto the PLOT column. **(B)** Optimization of chromatographic gradient time, the number of proteins identified *versus* the LC gradient time used.

Figure 3.

Assessment of the variability in proteomic profiles associated with three analyses of the primary and metastatic breast cancer samples (three LCM samples of 10,000 cells each from the same patient). Box plots of relative standard deviation are shown for the primary (PRI) and metastatic (MET) samples. The relative standard deviation of the peptide counts obtained from the three analyses is shown on the y-axis. The upper and lower side of the box represents the 25th and 75th percentile values, respectively. The horizontal line inside the box indicates the median value. The lines extending from the box show the spread (10–90th percentile) of the data.

Figure 4. Overlaid base peak chromatograms from each of the digested gel sections as indicated from the LCM collected primary tumor cell lysate.

Table 1

Total number of peptides and proteins identified from LCM cells of murine hepatocytes using three SDS-PAGE separation conditions.

Gel Type	Gradient Gel 4 – 12% Bis-Tris	Gradient Gel 4 – 12% Bis-Tris	Constant Gel 16% Tricine
Separation length	10 cm	2.5 cm	2.5 cm
Unique peptides	146	109	789
Total peptides	219	172	1315
Total proteins	48	41	181

Table 2

The number of proteins identified per gel section in each replicate analysis of 10,000 murine hepatocytes. Numbers in brackets correspond to proteins identified with at least two unique peptides.

Gel Section	Run 1	Run 2	Run 3	Total
1	536	522	509	
2	528	721	694	
3	442	392	345	
Total Unique Proteins	1035 (641)	1187 (739)	1138 (692)	1490 (1103)

Table 3

The number of proteins identified per gel section in each sample of 10,000 breast cancer cells LCM collected from the primary tumor site. Numbers in brackets correspond to proteins identified with at least two unique peptides.

Gel Section	Sample 1	Sample 2	Sample 3	Total
1	305	189	233	
2	550	563	596	
3	479	464	435	
Total Unique Proteins	1132 (558)	1050 (537)	1098 (523)	1708 (1123)

Metastatic enriched (relative to primary) Gene-Ontology (GO) terms with FDR less than 5% and p value less than 0.05 are shown in bold. The differentially abundant proteins associated with each cluster are presented with their Swiss-Prot accession number.

Table 4

Cluster Description and Associated Proteins		Enrichment Score	p-value	FDR (%)
<u>Cluster 1: Cytoplasmic vesicle</u>				
P30040	Endoplasmic reticulum protein ERp29	4.45	1.24E-05	0.02
P61106	Ras-related protein Rab-14			
P07339	Cathepsin D			
P10909	Clusterin			
P53621	Coatomer subunit alpha			
P53618	Coatomer subunit beta			
O14579	Coatomer subunit epsilon			
Q9Y678	Coatomer subunit gamma	2.19	2.25E-04	0.03, 0.15, 0.24 ^a
P11142	Heat shock cognate 71 kDa protein			
<u>Cluster 2: Vesicle targeting, to, from or within Golgi, Golgi vesicle transport, vesicle localization</u>				
P55735	Protein SEC13 homolog	P05783	Keratin, type I cytoskeletal 18	
Q92538	Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1	P54920	Alpha-soluble NSF attachment protein	
<u>Cluster 3: Signal peptide</u>				
Q9Y3A6	Transmembrane emp24 domain-containing protein 5	1.84	2.52E-03	3.53
Q9Y3B3	Transmembrane emp24 domain-containing protein 7			
P12111	Collagen alpha-3(VI) chain			
P12110	Collagen alpha-2(VI) chain			
P12109	Collagen alpha-1(VI) chain			
<u>Cluster 4: ECM-receptor interaction, extracellular matrix</u>				
P12111	Collagen alpha-3(VI) chain	1.59	3.78E-03	1.69, 4.73 ^a
P12110	Collagen alpha-2(VI) chain			
<u>Cluster 5: Glycolysis pathway</u>				
P60174	Triose phosphate isomerase	1.40	6.20E-03	3.34
P00558	Phosphoglycerate kinase 1			

^a Multiple FDRs resulting from multiple units within the cluster.