Published in final edited form as:

Proteomics Clin Appl. 2008 December ; 2(12): 1658–1669. doi:10.1002/prca.200780146.

A unified sample preparation protocol for proteomic and genomic profiling of cervical swabs to identify biomarkers for cervical cancer screening

Janet S. Rader^{1,2}, James P Malone³, Julia Gross³, Petra Gilmore³, Rebecca A. Brooks¹, Loan Nguyen¹, Dan L. Crimmins⁴, Sheng Feng⁵, Jason D. Wright¹, Nicolas Taylor¹, Israel Zigelboim¹, Margo C Funk¹, Phyllis C. Huettner⁴, Jack H. Ladenson⁴, David Gius⁷, and R. Reid Townsend^{3,6}

¹Division of Gynecologic Oncology, Department of Obstetrics and Gynecology, Washington University School of Medicine and Alvin J. Siteman Cancer Center, St. Louis, MO 63110

²Department of Genetics, Washington University School of Medicine and Alvin J. Siteman Cancer Center, St. Louis, MO 63110

³Department of Internal Medicine, Washington University School of Medicine and Alvin J. Siteman Cancer Center, St. Louis, MO 63110

⁴Pathology and Immunology, Washington University School of Medicine and Alvin J. Siteman Cancer Center, St. Louis, MO 63110

⁵Division of Biostatistics, Washington University School of Medicine and Alvin J. Siteman Cancer Center, St. Louis, MO 63110

⁶Department of Cell Biology and Physiology, Washington University School of Medicine and Alvin J. Siteman Cancer Center, St. Louis, MO 63110

⁷Radiation Oncology Branch, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892

Abstract

Cervical cancer screening is ideally suited for the development of biomarkers due to the ease of tissue acquisition and the well-established histological transitions. Furthermore, cell and biologic fluid obtained from cervix samples undergo specific molecular changes that can be profiled. However, the ideal manner and techniques for preparing cervical samples remains to be determined. To address this critical issue a patient screening protein and nucleic acid collection protocol was established. *RNA later* was used to collect the samples followed by proteomic methods to identify proteins that were differentially expressed in normal cervical epithelial versus cervical cancer cells. Three hundred ninety spots were identified via two-dimensional difference gel electrophoresis (2-D DIGE) that were expressed at either higher or lower levels (>3-fold) in cervical cancer samples. These proteomic results were compared to genes in a cDNA microarray analysis of microdissected neoplastic cervical specimens to identify overlapping patterns of expression. The most frequent pathways represented by the combined dataset were: cell cycle: G2/M DNA damage checkpoint regulation; aryl hydrocarbon receptor signaling; p53 signaling; cell cycle: G1/S checkpoint regulation; and the endoplasmic reticulum stress pathway. HNRPA2B1

was identified as a biomarker candidate with increased expression in cancer compared to normal cervix and validated by Western blot.

Keywords

2-D DIGE; biomarkers; cervical cancer; cDNA microarray; *RNAlater*

1 INTRODUCTION

Human papillomavirus (HPV) infection is usually self-limiting, but it can progress to cervical cancer, which kills about 273,500 women world-wide each year [1]. Current screening techniques, which detect abnormal or pre-cancerous cells of the cervix in Pap smears, permit treatment of preinvasive disease and prevent most invasive cervical cancers if used every 1 to 3 years. However, this type of screening is labor intensive and unsuitable for underserved communities with minimal resources. Moreover, the new prophylactic vaccines against HPV are likely to lower the incidence of cervical neoplasia, necessitating more sensitive and specific screening tests to avoid a larger number of false positives and negatives.

HPV testing improves the detection of CIN3 in samples of ASCUS category Pap smears, and is approved as a primary screen for cervical disease in women aged 30 and above [2]. Most biomarker tests being developed today are using nucleic acid methods of discovery such as RNA expression microarrays or methylation screens which allow quantitation across samples[3]. Interrogating proteins in addition to nucleic acids allows more integrated assessment of the biologic processes of the cervix.

Although mass spectrometric techniques can analyze very complex mixtures of proteins, proteomic studies of samples from cervical swabs have not received significant attention since it is difficult to reproducibly extract proteins that are mixed with viscous mucus and the technology does not allow the same global assessments as nucleic acid platforms. Nevertheless, cells and secretions are easily obtained from the cervix, come directly from neoplastic sites, and are not contaminated with serum proteins in the absence of bleeding lesions. Therefore, we have developed a protocol for extracting both high-quality mRNA and protein from cervical swabs in an aqueous RNA-preservation agent called *RNAlater*, which has been shown to increase the amount and quality of RNA that can be extracted from minute tissue samples and to preserve viral antigen for ELISA assays [4,5]. Using this method, we have characterized proteins that are present in cervical specimens and have compared with previous cDNA microarray studies to identify overlapping candidate biomarkers for the disease.

2 MATERIALS AND METHODS

2.1 Clinical samples

Washington University Human Studies Committee approved this study, and all the patients gave informed consent. Specimens were obtained from the squamocolumnar junction of the cervix, using a plastic spatula and cytobrush. We took great care not to contaminate the samples with blood when we swabbed the cervix. Cervical samples were always collected after the clinical Pap test except from a few patients who had two repeat cervical swabs obtained for proteomics only. The specimens were placed in a tube containing 1.5 ml *RNAlater* (Ambion, Austin, TX), kept on ice during the clinical session, stored overnight at 4°C, and then aliquoted into two 750 µL samples that were stored at -80°C until protein and/or nucleic acid extraction. Specimens for this study were collected between October

2005 and July 2006 (Table 1). After testing several preparation and extraction methods, we devised a protocol involving *RNAlater*, an aqueous solution that rapidly permeates and preserves fresh tissue and did not require special processing in the clinic.

2.2 Protein and RNA extraction

For protein extraction samples in *RNAlater* were transferred to an Ultrafree-MC 0.45 μ filter unit (Millipore, Billerica, MA) and centrifuged at 14,000 x g in an Eppendorf micro-centrifuge for 30 s. The liquid from the lower part of the unit was discarded. The precipitated sample was washed with ~ 500 μ L of cold (4°C) acetonitrile/water (80:20) and centrifuged. After discarding the biphasic liquid, the wash cycle was repeated until there was no evidence of crystals in the upper chamber and there was no longer a biphasic liquid visible in the collection tube. The filter was then extracted with 100 μ L of 2D lysis buffer (30mM Tris-Cl pH 8.5, 7 M urea, 2 M thiourea, 4% CHAPS) containing a cocktail of protease inhibitors (Roche, Indianapolis, IN) and phosphatase inhibitor cocktails I and II (Calbiochem, San Diego, CA), both at 1X concentration as recommended by the manufacturer. The main mechanism for improvement in protein preparation was extensive washing in ice cold 80% acetonitrile to remove *RNAlater*, which formed crystals in the pellet and were toxic to the protein gels. The membrane allowed efficient and complete capture of the entire soft protein pellet from the contaminating salt crystals.

RNA was extracted with Trizol reagent (Invitrogen, Carlsband, CA) according to the manufacturer. The ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE) was used to measure A260/A280 nm ratio. Little degradation was identified on agarose gel electrophoresis (Figure 1).

2.3 Two-dimensional difference gel electrophoresis (2-D DIGE)

Protein samples were labeled with charge-matched cyanine dyes as previously described [6-8]. The cervical samples prepared from *RNAlater* using the above-described filtration-wash technique were classified as normal, low-grade viral change or CIN 1 (referred to as CIN 1), or cervical cancer, and labeled with blue, green, or red NHS cyanine dye. The specific dye labeling was varied across samples to prevent experimental bias due to different intensities from gene-dye effects. Equal amounts of protein (50 μ g) from each sample in 2D lysis buffer (Tris HCL pH 8.5 (30mM), 7 M urea, 2 M thiourea and 4% CHAPS) were labeled with 400 pmol of 3-[(4-carboxymethyl) phenylmethyl]-3'-ethyloxacarbocyanine halide N-hydroxysuccinimidyl ester (Cy2), 1-(5-carboxypentyl)-1'-propylindocarbocyanine halide N-hydroxysuccinimidyl ester (Cy3) or 1-(5-carboxypentyl)-1'-methylindocarbocyanine halide N-hydroxysuccinimidyl ester (Cy5). All labeling reactions were carried out for 45 min at 4°C and protected from light after which the reaction was quenched with 10 nmol of lysine for 10 min. Protein samples were combined as indicated in the experimental design and the resulting mixture was diluted by addition of 300 μ L of rehydration buffer (7 M urea/2 M thiourea, 4% CHAPS, 0.5% v/v ampholytes (pH 3-10)). Each combined labeled sample (450 μ L) was rehydrated into 24 cm, 3-10 NL immobilized pH gradient (IPG) strips (GE Healthcare, Uppsala, Sweden) under low voltage (100 V) for 12 h, followed by isoelectric focusing using an IPGPhor (GE Healthcare) for a total of 65.5 kVh (using a three step voltage protocol: 500 V and held for 500 Vh, 1000 V and held for 1000 Vh, 8000 V and held for 64,000 Vh). After focusing, proteins were reduced by placing the IPG strips in 10 ml of equilibration buffer (10 ml, 50 mM Tris (pH 8.8), 6 M urea, 30% glycerol, 2% SDS, bromphenol blue) containing freshly prepared DTT (50 mg) for 15 min at room temperature. The proteins were then alkylated by adding iodoacetamide (600 mg in 10 ml of equilibration buffer). IPG strips were then rinsed with 1X SDS running buffer (25 mM Tris, 192 mM glycine, 0.1% SDS) and layered on 10% polyacrylamide gels and sealed with agarose (1% w/v in 1X running buffer). Pre-cast gels on low-fluorescence glass plates were

obtained from Jule Biotechnologies Inc. (Milford, CT). The second-dimension SDS-PAGE separation was carried out using 5 W/gel for the first 15 min and was followed by 1 W/gel for 17 h with circulating buffer (20°C) in the lower chamber. The three individually-labeled samples within each gel were imaged with Typhoon Imager (GE Healthcare) using specific excitation/emission wavelengths for each dye (488/520 nm for Cy2, 520/580 nm for Cy3, 620/670 nm for Cy5). After imaging, the non-silanized glass plate was removed and the gels were placed in fixing solution (33% ethanol, 7.5% acetic acid) for 2 h. Gels were then rinsed with deionized water (18 mΩ) for 15 min and stored in water-filled, sealed bags at 4 °C.

2.4 Gel Image and multivariate analyses

ImageQuant (Molecular Dynamics, Sunnyvale, CA) software was used to crop the gel images to remove those portions of the image that corresponded to the IPG strip, the spacers between the gel plates and any remaining dye front from the electrophoresis. The DeCyder (GE Healthcare, Piscataway, NJ) software tools (version 6.5) were used for two-dimensional gel image analysis. The DIA (differential in-gel analysis) module was used to align and normalize the three fluorometrically distinct images within each gel. The DIA analysis was performed by setting the estimated number of spots to 5000 per gel image. Once gel features were identified, the 'area of interest' tool was used to exclude artifacts at the gel edges. Gel features were further globally filtered by excluding spots with slopes > 1.1, areas < 100, volumes < 10,000 and peak heights < 100. The DIA module calculated abundance ratios for feature pairs using a previously-described normalization algorithm [9].

2.5 Protein identification of gel features

The coordinates of the gel features that were selected in the DeCyder software were saved in a file for robotic spot excision. Translation from the coordinates of the gel image to the coordinates used by the picking robot (ProPic, Genomic Solutions, Ann Arbor, MI) was accomplished using a triangulation algorithm implemented with 'in-house' software. The central cores (1.8 mm) of the selected gel features were excised and transferred to a 96-well source plate. The gel pieces were digested *in situ* with trypsin as previously described [10]. The gel cores were washed sequentially with 200 µL of 50% methanol/water (2 times), 50% acetonitrile/0.5 mM NH₄OH (2 times), and once with acetonitrile. The gel cores were then dried in a Speed Vac, covered with 5 µL of trypsin (300 ng in 1 mM NH₄Cl, pH 8.3), and incubated for 1.5 h at 56°C. Ten µL of formic acid/acetonitrile (1%/1%) was added and after rocking for 30 min, the entire volume was transferred to an autosampler vial. The peptide pools were analyzed using capillary reversed-phase HPLC-MS/MS interfaced to an electrospray-quadrupole-time-of-flight mass spectrometer (Q-STAR XL, Applied Biosystems, Foster City CA). Chromatography was performed at 200 nL/min using a liquid chromatograph (Eksigent nano-LC, Eksigent, Livermore, CA) coupled to a nanocapillary source (PicoView, New Objective, Woburn, MA). Sample injection was performed with an autosampler (Endurance, Spark, Plainsboro, NJ). Peptides were separated on a C-18 column (PicoFrit, New Objective) with an inner diameter of 75 µm, pore size of 5 µm and a length of 10 cm. The solvents were 1% formic acid in HPLC grade water (solvent A, Fisher Scientific, Pittsburgh), and acetonitrile (Honeywell Burdick and Jackson American Inc, Muskegon, MI) with 1% formic acid (solvent B). Column equilibration and sample injection were performed at 5% A for 10 min each, followed by an increase of solvent B by 2%/min. The spray voltage was set to 2.5 kV with a curtain gas setting of 10, a nebulizer gas setting of 15, and a pressure in the collision cell of 3.5×10^{-7} torr. The declustering potential was 90 and the focusing potential was 280. Collision energies were calculated by the ANALYST QS software according to the following function with dependence on the m/z value of the parent ion: Collision Energy = $0.058 \times (m/z)^{-2}$. The instrument was operated in a data-dependent mode with selection of the 5 most abundant signals from each MS scan. The tandem spectra were searched against the National Center for Biotechnology Information

non-redundant protein database NR (downloaded 02-18-2007) using MASCOT, version 2.0 (Matrix Sciences, London). The database searches were constrained by allowing for trypsin cleavage (with up to 2 missed cleavage sites), fixed modifications (carbamidomethylation of Cys residues) and variable modifications (oxidation of Met residues and N-terminal pyroglutamate formation).

2.6 cDNA microarrays

Gene expression was profiled using NCI ROSP 8k human cDNA arrays [11]. Hybridizations were carried out in triplicate and scanned using a GenePix 4000A scanner (Axon Instruments, Inc., Foster City, CA). Details of the cDNA array procedures has previously been reported [11]. A total of 26 matched neoplastic and normal epithelial samples (one pair from each patient) were analyzed. Protein results were cross-referenced to genes that we found to be differentially expressed in CIN3 and invasive cervical cancer compared to normal cervix and CIN1 when we expanded the analysis of the cDNA microarrays [11].

2.7 Statistical analysis

The cDNA microarrays used for expression profiling of cervical epithelium were analyzed to identify genes that were upregulated or downregulated in CIN3 and ICC samples compared to normal or CIN1. Patients are categorized as “non-severe” (normal, VE, CIN1) and “severe” (CIN3 and Invasive Cancer). The goal of the analysis was to identify a group of genes with expression levels that change significantly between the two categories and compare to the proteomic results. A linear mixed model was generated to fit the experimental design structure and accommodate multiple covariates. The inter- and intra-patient variations are estimated by variance components. The false discovery rate was calculated using the multitest procedure in SAS by Benjamini and Hochberg [12].

2.8 Functional analysis and Interaction networks

To analyze this proteomic dataset and cDNA microarray, we used several systems-biology data-mining tools, including GENCODIS[13] which uses annotations from Gene Ontology and Ingenuity Pathway Analysis software (Ingenuity Systems, www.ingenuity.com). We used the latter to display known biological interactions among the genes and gene products on our lists.

2.9 Western Blot

Expression of HNPRA2B1 was determined in cervical swabs and tissue samples by Western blot. Protein from 4 cervical swab specimens (2 normal and 2 cancer) was extracted from RNA^{later} samples as described above. Protein was also extracted from a normal and squamous cell cancer (SCCA) tissue specimen. A 5mm³ piece of normal or SCCA tissue stored at -80 °C was pulverized and resuspended in XT sample buffer with reducing agent (Bio-Rad Laboratories, Hercules, CA) and heated for 10 minutes at 90 °C to denature the protein. Proteins were separated on a 4-12% Bis-Tris criterion XT precast gel (Bio-Rad Laboratories, Hercules, CA) using 36µg of protein in each lane, with precision plus protein dual color standard (Bio-Rad Laboratories, Hercules, CA). The gel was transferred to a PVDF membrane.

Mouse anti-HNPRA2B1 antibody (H-00003181-M01, Abnova Corporation, Taiwan) was first evaluated using a purchased blot (TB106, GBiosciences, St. Louis, MO) containing 50 µg of protein from SCCA tumor and normal tissue run on a 4-20% SDS page gel then transferred to a PVDF membrane. Protein concentration was determined using Advanced Protein Assay (Cytoskeleton, Denver CO) with serial dilutions of bovine serum albumin standard (Pierce, Rockford, IL).

Blots were blocked with ChemiBlocker (Chemicon International, Temecula, CA) for 1 hour at room temperature and then incubated overnight with 4 ug/ml of primary antibody solution [14]. The blot was washed three times with Tris-Buffered Saline Azide (TBSA), incubated with at a 1/1,000 dilution of goat anti-mouse alkaline phosphatase (115-055-072, Jackson ImmunoResearch Laboratories Inc., West Grove, PA) for 1 hour, washed 4 more times with TBSA and then developed with substrate [14]. The immuno-stained membrane was scanned with an Epson 3200 scanner (Seiko Epson Corporation, Nagano, Japan) and densitometry performed with UN-SCAN-IT 6.1 software (Silk Scientific Corporation, Orem, UT).

3 RESULTS

3.1 Gel quality and reproducibility

Initial experiments were done to determine a suitable method for acquiring nucleic acid and protein samples from cervical swabs, since the hydrogel structure of cervical mucus hampers sample preparation. *RNAlater*, an effective reagent for the preservation of mRNA, has been reported to be incompatible with proteomics analysis using 2D-DIGE [15]. This publication found the severity of gel distortion and poor spot resolution correlated with the concentration of *RNAlater*. Therefore, we tested several methods to remove *RNAlater* from precipitated protein. These included adjusting pH of the lysis buffer, 2-D clean up kit, buffer exchanges and centrifugal concentration filters. We tested a precipitation approach with extensive organic solvent washing and a filtration method for the recovery of proteins for 2D-DIGE analyses of cervical samples. Figure 2A shows a 2D-DIGE gel analysis of cervical swab proteins. As previously shown *RNAlater* detrimentally affects proteomic analysis. There was poor first and second dimensional resolution leading to large pH gaps, streaking and changes in spot position. Few individual features were observed. We reasoned that very small quantities of *RNAlater* interfered with the focusing step and developed a method to wash the precipitated protein on a filter membrane as described in 'Materials and Methods'. Figure 2B shows the 2D-DIGE image of three samples from a normal, CIN 1 and cancer cervical specimen that was prepared using this filtration-wash method. In addition to the absence of the defects observed with the precipitation-wash strategy, we were able to reproducibly produce these high-quality images from cohorts of cervical samples as described below. Extracting protein from a 50% aliquot of the cervical specimen yielded a mean of 81 μ g (37–140 μ g).

We next quantified the recovery of proteins using the filtration-wash method using cell line samples. Mouse lymphoma tissue culture cells (YAC-1) were pelleted and processed with and without *RNAlater* and filtration-wash. The individual samples were labeled with either Cy3 (control) or with Cy5 (*RNAlater* treatment) and analyzed by 2D-DIGE. Figure 2C shows the combined image from the YAC-1 cell samples. The appearance of only yellow features indicates that there are only small changes in abundance of all extracted proteins. Quantitative image analysis showed that 9 (.3%) out of 2647 gel features differed by ≥ 2 -fold in abundance between identical samples that were prepared with and without *RNAlater* and filtration-wash (Figure 2D). Preparations made from serum or protein extracts stored in *RNAlater* for up to 24 h at room temperature gave similar results. Two consecutive cervical swabs obtained from the same patient were labeled with Cy3 and Cy5 and analyzed by 2-D DIGE (Figure 2E). Quantitative image analysis showed 374 (10.78%) out of 3468 gel features differed by ≥ 2 -fold in abundance (Figure 2D).

After establishing the method for protein preparation, we set up five gels, using individual and pooled samples. Each gel was loaded with normal, CIN1, and cancer samples that were labeled with blue, green, and red NHS cyanine dyes (Figure 3, Table 1). The five gels have been analyzed, and we have identified many of the differentially expressed proteins with a >3 fold change using tandem mass spectrometry and compared to cDNA microarray results.

3.2 Gene identification

Three hundred and ninety protein spots were transferred to 96-well plates, digested, and the resulting tryptic peptide pools analyzed by tandem mass spectrometry. Twelve spots contained no identifiable peptides. The remaining 378 gel features contained peptides representing 153 different gene products (Table S1). We identified 161 genes from the cDNA array whose differential expression was deemed significant ($p < .003$) in cervical cancer epithelium compared to normal or CIN1. These results were compared to proteomic data using their respective official gene symbols [11]. Differential expression ranged from 0.4- to 2.2-fold in the microarray analysis and over 3-fold change in the protein gels. The proteins showing the largest differential expression in the gels were KRT6A, KRT13, KRT4, KRT14, HSPD1, and FABP5 (Figure 4A). Figure 4B shows the location of the 5 proteins.

3.3 Functional annotation

Function annotation was subsequently used to obtain a more comprehensive assessment of the global similarities and differences between the transcriptome and proteomic datasets. The functional annotation of the differentially expressed genes and their affiliations with specific genetic pathways was interrogated by GENCODIS [13]. The analysis summarized here involved the 161 genes and 153 gene products, and it provides a global view of the cellular components, molecular functions, and biologic processes, found in the cDNA microarray and 2-D DIGE experiments (Figure 5-7).

This analysis demonstrated that nuclear genes accounted for 26% of the genes identified in the cDNA dataset but only 14% of those in the protein dataset (Figure 5). The corresponding values for cytoplasmic genes were 10% and 18%. Genes encoding the plasma membrane and associated structures were observed more frequently on the cDNA microarray list than in the protein list. Conversely, genes encoding components of the extracellular space, the cytoskeleton, and intermediate filament protein were seen more frequently on the protein list (Figure 5). The major molecular processes that we identified—protein binding, nucleotide binding, and ATP binding—were represented to the same extent on both lists (Figure 6). Genes involved in the cell cycle, signal transduction, and transcription were more heavily represented in the cDNA list, whereas genes involved in protein folding and epidermis development were seen more often in the protein list (Figure 7).

3.4 Biologic pathway analysis

The Ingenuity Pathways Analysis software (Ingenuity Systems, www.ingenuity.com) was used to map the proteins and genes to networks and pathways. For the network analysis, we restricted our list to the 28 proteins with over 3-fold differential expression in cancer to normal on at least 2 of the 5 gels and to the top 50 genes ($p < .00045$) in the cDNA analysis. HNRPA2B1 (heterogeneous nuclear ribonucleoprotein A2B1) appeared on both lists. The top canonical pathways represented by this dataset were: cell cycle: G2/M DNA damage checkpoint regulation; aryl hydrocarbon receptor signaling; p53 signaling; cell cycle: G1/S checkpoint regulation; and endoplasmic reticulum stress pathway. To the 78 gene list we added five genes whose significance was high but did not exceed our strict cutoff: CASP3, RB1, and NF1 from the microarray results and YWHAZ and ACTB from the protein list. These additions provided important connectivity to the genes and gene products, and the final pathway displayed 52 nodes (Figure 8). The microarray data allowed better connectivity between nodes than if proteomics was used only in analysis.

3.5 Western blot HNRPA2B1

The differential expression of HNRPA2B1 was examined in the cervix from normal and SCCA samples (Figure 9). In tissue samples the expression level was 2.2 –2.7 fold higher in cancer than normal cervix. The antibody was also visualized in the cervical swabs obtained from SCCA but not normal cervix. Further work is ongoing on an expanded sample set and additional antibodies.

4 DISCUSSION

This study is the first report in cervical cancer to comprehensively integrate proteomic profiling with genomic analysis to identify potential biomarkers for cervical cancer screening. This was made possible by the development of a protocol for obtaining high-quality protein from the cervix. One of the primary strengths of this method is that these samples are easily accessible from routine gynecologic exams. In this analysis we use *RNAlater*, which solubilizes cervical mucus and precipitates proteins from cervical swabs. The proteins can be used on 2-D DIGE gels with excellent resolution and no significant alteration in isoelectric point or molecular weight, and they produce clear spots on 2-D DIGE. Furthermore, samples yielded high-quality DNA, RNA and protein preparations despite the lag between obtaining and freezing the samples.

In this study we used combined analysis of global microarray data and proteomics to identify candidate protein biomarkers. In addition, microarray incorporates control tissue from the same individual, environmental and genotypic differences between samples should be much less marked than in the proteomic experiments. When thousands of genes are simultaneously analyzed on cDNA microarrays, it is challenging to prioritize the data for protein biomarkers. As this study demonstrates, however, a proteomic approach may provide additional guidance for selecting genes from microarray experiments, and knowledge of gene expression patterns can support pathway development (Figure 8).

Using tissue rather than serum or plasma for biomarker discovery improves diagnosis by focusing on abundant proteins. The best signal gradient should occur in proximal tissue and fluids, which are close to the source and should be enriched in candidate markers. For example, concentrations of CA125 are significantly higher at the site of origin than in blood. Thus, patients with ovarian carcinoma have markedly higher levels of CA125 in both the cyst and ascites than in their serum ($p < 0.001$) [16]. In addition, cancer biomarkers are likely to be obscured in plasma because albumin and other major proteins account for about 90% of its protein content. Furthermore, nonspecific cancer biomarkers may be more specific for cervical cancer when they are obtained directly from the cervix. For example, overexpression of several S100 proteins[17] and HNRPA2B1 has been reported in various types of cancer.

We also characterized the representative Gene Ontology (GO) annotations in the RNA and protein results, using GENCODIS. The GO cellular components (Figure 5) revealed that about 10% of the top RNA and protein discovery candidates are related to the extracellular region. The top RNA candidates are also related to the nucleus, cytoplasm, or membranes structures. Not surprisingly, a large proportion of the proteins were components of the cytoplasmic framework, and were not represented by RNA. These included the cytoskeleton and intermediate filaments composed mostly of keratins. The most common GO molecular functions were equally represented on the cDNA arrays and gels, with the exception of cytoskeletal components, isomerase activity, and Zn binding, which were detected only among the protein candidates (Figure 6). The cell cycle and DNA repair and replication were better represented in RNA (18%) than in protein (2%) in GO biologic processes (Figure 7).

Several gene products that showed the greatest differential expression in the 2-D DIGE were members of the keratin family, and the expression decreased as cervical cancer progressed. Keratins are intermediate filament proteins that are important constituents of the cytoskeleton and therefore help stabilize the cell interior. KRT13, a type I cytokeratin, is paired with KRT4, and, in 2-D DIGE, both were expressed at lower levels in the cervical cancer specimens than in the normal and CIN1 specimens. KRT13 and KRT4 are also involved in epidermal development, and cDNA microarray studies show them to be downregulated in esophageal squamous cell carcinoma[18]. Keratins are used routinely for histopathologic diagnosis in clinical practice, and knowledge of various keratin profiles often helps determine where metastatic tumors have originated.

Our discovery platform also found several markers that previously have been identified as potential biomarkers for cervical cancer screening. They include CDKN2A, MCM7, and MCM2 [3]. An enzyme-linked immunosorbant assay for p16^{INK4a} (CDKN2A) detected CIN3 90% to 62.5% of the time and was positive in CIN1 53% to 34% of the time [19]. Therefore, improved antibody specificity or multiple panel biomarkers will be needed.

In this initial proof-of-principal study, HNRPA2B1 was identified as a biomarker candidate for cervical cancer. Our microarray analysis detected a 1.48-fold increase in HNRPA2B1 ($p=.0003$) in cancer/CIN3 samples compared with normal samples, and 2-D DIGE detected an average of 3.98 fold increase over 7 gel features from 2/5 gels. We identified an increase in cancer over normal in tissue and cytology samples by Western blot. HNRPA2B1 belongs to the A/B subfamily of ubiquitously expressed heterogeneous nuclear ribonucleoproteins (hnRNPs), which bind to RNA to influence pre-mRNA processing, metabolism, and transport. The gene generates two alternatively spliced transcripts that encode isoforms A2 and B1. Sueoko *et al* found that a polyclonal antibody specific to hnRNP B1 stained nuclei of cancer cells, particularly in squamous carcinomas [20]. Plasma levels of hnRNP B1 mRNA were higher in lung cancer patients than in healthy volunteers [21]. The monoclonal antibody 703D4, generated from a human large cell lung cancer line and used to evaluate sputum in clinical epidemiology studies, was identified as hnRNP A2/B1 [22]. Two prospective studies accurately predicted that 67% and 69% of subjects with upregulation of hnRNP A2/B1 (703D4) in their sputum would develop lung cancer in the first year of follow-up, compared with background lung cancer rates of 2.2% and 0.9% [23]. Thus, HNRPA2B1 should be further tested as a candidate biomarker for cervical cancer.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

The work was supported by grants from Barnes-Jewish Hospital Foundation and NIH grants CA094141 and CA95713. This work was supported, in part, by the Alvin J. Siteman Cancer Center at Barnes-Jewish Hospital and Washington University School of Medicine and by institutional resources provided to the Proteomics Center at Washington University. The Siteman Cancer Center is supported in part by a NCI Cancer Center Support Grant #P30 CA91842. This research was supported, in part, by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, and the Center for Cancer Research.

Abbreviations

CIN	cervical intraepithelial neoplasia
2-D DIGE	two-dimensional difference gel electrophoresis
HPV	human papillomavirus

REFERENCES

1. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. *CA Cancer J Clin*. 2005; 55:74–108. [PubMed: 15761078]
2. Cuzick J, Clavel C, Petry KU, Meijer CJ, et al. Overview of the European and North American studies on HPV testing in primary cervical cancer screening. *Int J Cancer*. 2006; 119:1095–1101. [PubMed: 16586444]
3. Malinowski DP. Multiple biomarkers in molecular oncology. I. Molecular diagnostics applications in cervical cancer detection. *Expert Rev Mol Diagn*. 2007; 7:117–131. [PubMed: 17331061]
4. Blacksell SD, Khouny S, Westbury HA. The effect of sample degradation and RNA stabilization on classical fever virus RT-PCR and ELISA methods. *J Virol Methods*. 2004; 118:33–37. [PubMed: 15158066]
5. Dunmire V, Wu C, Symmans WF, Zhang W. Increased yield of total RNA from fine-needle aspirates for use in expression microarray analysis. *Biotechniques*. 2002; 33:890–892. [PubMed: 12398198]
6. McDunn JE, Townsend RR, Cobb J. P. The murine plasma response to polymicrobial intra-abdominal sepsis. *Clinical Proteomics*. 2007; 1:373–386.
7. Tonge R, Shaw J, Middleton B, Rowlinson R, et al. Validation and development of fluorescence two-dimensional differential gel electrophoresis proteomics technology. *Proteomics*. 2001; 1:377–396. [PubMed: 11680884]
8. Unlu M, Morgan ME, Minden JS. Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis*. 1997; 18:2071–2077. [PubMed: 9420172]
9. Alban A, David SO, Bjorkestén L, Andersson C, et al. A novel experimental design for comparative two-dimensional gel analysis: two-dimensional difference gel electrophoresis incorporating a pooled internal standard. *Proteomics*. 2003; 3:36–44. [PubMed: 12548632]
10. Havliš J, Thomas H, Šebela M, Shevchenko A. Fast-response proteomics by accelerated in-gel digestion of proteins. *Anal Chem*. 2003; 75:1300–1306. [PubMed: 12659189]
11. Gius D, Funk MC, Chuang EY, Feng S, et al. Profiling Microdissected Epithelium and Stroma to Model Genomic Signatures for Cervical Carcinogenesis Accommodating for Covariates. *Cancer Res*. 2007; 67:7113–7123. [PubMed: 17671178]
12. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *J R Stat Soc Ser B*. 1995; B, 57:289–300.
13. Carmona-Saez P, Chagoyen M, Tirado F, Carazo JM, Pascual-Montano A. GENECODIS: a web-based tool for finding significant concurrent annotations in gene lists. *Genome Biology*. 2007; 8:R3. [PubMed: 17204154]
14. Li A, Crimmins DL, Luo Q, Hartupée J, et al. Expression of a novel regenerating gene product, Reg IV, by high density fermentation in *Pichia pastoris*: production, purification, and characterization. *Protein Expr Purif*. 2003; 31:197–206. [PubMed: 14550637]
15. Butt RH, Pfeifer TA, Delaney A, Grigliatti TA, et al. Enabling coupled quantitative genomics and proteomics analyses from rat spinal cord samples. *Mol Cell Proteomics*. 2007; 6:1574–1588. [PubMed: 17510048]
16. Sedlacek P, Frydecka I, Gabrys M, Van Dalen A, et al. Comparative analysis of CA125, tissue polypeptide specific antigen, and soluble interleukin-2 receptor alpha levels in sera, cyst, and ascitic fluids from patients with ovarian carcinoma. *Cancer*. 2002; 95:1886–1893. [PubMed: 12404282]
17. Emberley ED, Murphy LC, Watson PH. S100 proteins and their influence on pro-survival pathways in cancer. *Biochem Cell Biol*. 2004; 82:508–515. [PubMed: 15284904]
18. Luo A, Kong J, Hu G, Liew CC, et al. Discovery of Ca²⁺-relevant and differentiation-associated genes downregulated in esophageal squamous cell carcinoma using cDNA microarray. *Oncogene*. 2004; 23:1291–1299. [PubMed: 14647409]
19. Mao C, Balasubramanian A, Yu M, Kiviat N, et al. Evaluation of a new p16(INK4A) ELISA test and a high-risk HPV DNA test for cervical cancer screening: results from proof-of-concept study. *Int J Cancer*. 2007; 120:2435–2438. [PubMed: 17294450]

20. Sueoka E, Sueoka N, Goto Y, Matsuyama S, et al. Heterogeneous nuclear ribonucleoprotein B1 as early cancer biomarker for occult cancer of human lungs and bronchial dysplasia. *Cancer Res.* 2001; 61:1896–1902. [PubMed: 11280744]
21. Sueoka E, Sueoka N, Iwanaga K, Sato A, et al. Detection of plasma hnRNP B1 mRNA, a new cancer biomarker, in lung cancer patients by quantitative real-time polymerase chain reaction. *Lung Cancer.* 2005; 48:77–83. [PubMed: 15777973]
22. Zhou J, Mulshine JL, Unsworth EJ, Scott FM, et al. Purification and characterization of a protein that permits early detection of lung cancer. Identification of heterogeneous nuclear ribonucleoprotein-A2/B1 as the antigen for monoclonal antibody 703D4. *J Biol Chem.* 1996; 271:10760–10766. [PubMed: 8631886]
23. Tockman MS, Mulshine JL, Piantadosi S, Erozan YS, et al. Prospective detection of preclinical lung cancer: results from two studies of heterogeneous nuclear ribonucleoprotein A2/B1 overexpression. *Clin Cancer Res.* 1997; 3:2237–2246. [PubMed: 9815620]

Figure 1. Agarose gel electrophoresis of RNA extracted from pap smears
100 ng total RNA was loaded in each lane. Cytologic diagnosis from concurrent pap smear:
Normal lanes 4-6, CIN 1 lanes 1 and 2, squamous cancer lanes 7 and 8.

Figure 2. Gel Quality and Reproducibility

A. 2D-DIGE gel analysis of cervical swab proteins. *RNAlater* contamination caused poor first and second dimension resolution leading to large pH gaps, streaking and changes in spot positions. Few individual features were observed. Very small quantities of *RNAlater* interfere with the focusing step. **B.** The 2D-DIGE image of cervical swab proteins prepared using the filtration-wash method described in Materials and Methods. Washing *RNAlater* from samples allowed these high-quality images from cohorts of cervical samples. **C.** Control and *RNAlater*-treated filtration-washed cell lysates from the YAC-1 cell line were labeled with Cy3 (green) or Cy5 (red) and analyzed by 2-D DIGE. The yellow color(equal proportions of red- and green-labeled proteins) indicates no significant alterations in isoelectric point or molecular weight for the ~2000. **D.** The 2D-DIGE image of two consecutive cervical swabs obtained from the same patient were labeled with Cy3 or Cy5. **E.** Graph showing fold change difference between *RNAlater* treated filtration-washed YAC-1 and untreated YAC-1 (gray) and two swabs from the same patient (black).

Figure 3. Comparison of protein expression patterns in cervical swabs by 2-D DIGE

A. Gels 1 to 3 contained samples from individual patients, whereas gels 4 and 5 contained pooled samples obtained by mixing equal portions of cell lysates (see Table 1 for details). Each gel was loaded with normal, CIN1, and cancer samples that were labeled with blue, green, and red NHS cyanine dyes. The specific dye labeling was varied across samples to prevent experimental bias due to different intensities from gene-dye effects. Three registered images corresponding to the cervical samples were recovered from the gel for analysis. Proteins whose levels were higher or lower in the cancer specimens than in the normal and CIN1 specimens were identified by mass spectrometry. Supplemental figures S1 - S5 contain details of gels. **B.** Frequency of fold changes across the 5 gels for cancer to normal and the YAC-1 reference.

Figure 4. A. Magnitude of fold change for the five proteins with the greatest degree of differential expression

Open bars represent the fold change between normal and CIN1 samples, checked bars represent LSIL vs cancer, and black bars represent normal versus cancer. The results represent averages across the gels. **B.** Color gel image showing the location of peptides from the 5 proteins.

Figure 5. Global view of the cellular components involving the differentially expressed genes and proteins, analyzed by GENCODES

The black bar indicates cDNA, and the patterned bar indicates protein. Genes may be represented by more than one Gene Ontology term. About 10% of the top RNA and protein discovery candidates related to the extracellular region. The top RNA candidates are also related to the nucleus, cytoplasm, or membranes structures. Not surprisingly, a large proportion of the proteins were components of the cytoplasmic framework, and were not represented by RNA. These included the cytoskeleton and intermediate filaments composed mostly of keratins.

Figure 6. Global view of the molecular functions involving the differentially expressed genes and proteins, as analyzed by GENCODIS

The black bar indicates cDNA, and the patterned bar indicates protein. Genes may be represented by more than one Gene Ontology term. The most common GO molecular functions were equally represented on the cDNA arrays and gels, with the exception of cytoskeletal components, isomerase activity, and Zn binding, which were detected only among the protein candidates.

Figure 7. Global view of the biological processes involving the differentially expressed genes and proteins, as analyzed by GENCODIS

The black bar indicates cDNA, and the patterned bar indicates protein. Genes may be represented by more than one Gene Ontology term. The cell cycle and DNA repair and replication were better represented in RNA (18%) than in protein (2%) in GO biologic processes.

Figure 8. For the network analysis for gene expression changes in cervical lesions

For this model our gene list was restricted to the 28 most significant proteins, as determined by their presence on at least two of the five gels, and to the top 50 genes ($p < .00045$) from the cDNA microarray analysis. These 78 genes were interrogated, using Ingenuity Pathways Analysis software (Ingenuity Systems, www.ingenuity.com). The final pathway displayed 52 nodes after we added five more genes (whose significance was high but fell below the cutoff) to improve connectivity and include most of the candidate genes. CASP3, RB1, and NF1 were added from the microarray results, and YWHA2 and ACTB came from the protein list. Genes from the protein list are underlined. HNRPA2B1 is underlined twice, because it appeared on both lists.

Figure 9. Western Blot showing expression of HNPRA2B1 in cervical samples

A. TB106 blot of normal cervix and SCCA B. Cervical swabs and tissue samples. T/N – Tumor to Normal ratio for tissue.

Table 1

Patient Characteristics

Dye labeling	Screening Pap	Biopsy	Age
Gel 05510			
Green	Normal ¹		49
Blue	LGSIL ²	CIN 1 ³	20
Red		Adenocarcinoma	36
Gel 05497			
Blue	Normal		48
Red	LGSIL	Negative	38
Green		Squamous carcinoma	41
Gel 60813			
Blue	Normal		49
Green	LGSIL	Squamous metaplasia	38
Cancer		Squamous carcinoma	45
Gel 60824			
Blue	Normal		31
Blue	Normal		31
Blue	Normal		22
Green	LGSIL	CIN 1	22
Green	LGSIL	Negative	21
Green	LGSIL	Negative	38
Green	LGSIL	CIN 1	22
Red		Squamous carcinoma	46
Red		Squamous carcinoma	33
Gel 60825			
Red	Normal		58
Red	Normal		60
Blue	LGSIL	CIN 1	41
Blue	LGSIL	Inflammation	45
Green		Squamous carcinoma	62
Green		Squamous carcinoma	71

¹ Normal – no present or previous intraepithelial neoplasia or cancer

² LSIL – low-grade squamous intraepithelial neoplasia

³ CIN 1 – cervical intraepithelial neoplasia 1