Lan Dai[1]
Chen Li[2]
Kerby A. Shedden[3]
David E. Misek[4]
David M. Lubman[1,2,4,5]

[1]Bioinformatics Program,
 University of Michigan Medical
 Center, Ann Arbor, MI, USA
[2]Department of Chemistry,
 University of Michigan, Ann
 Arbor, MI, USA
[3]Department of Statistics,
 University of Michigan, Ann
 Arbor, MI, USA
[4]Department of Surgery,
 University of Michigan Medical
 Center, Ann Arbor, MI, USA
[5]Department of Pathology,
 University of Michigan Medical
 Center, Ann Arbor, MI, USA

Research Article

Comparative proteomic study of two closely related ovarian endometrioid adenocarcinoma cell lines using cIEF fractionation and pathway analysis

The proteomic profiles from two distinct ovarian endometrioid tumor-derived cell lines, (MDAH-2774 and TOV-112D) each with different morphological characteristics and genetic mutations, have been studied. Characterization of the differential global protein expression between these two cell lines has important implications for the understanding of the pathogenesis of ovarian endometrioid carcinoma. In this comparative proteomic study, extensive fractionation of peptides generated from whole-cell trypsin digestion was achieved by coupling cIEF in the first-dimensional separation with capillary LC (RP-HPLC) in the second dimensional separation. Online analysis was performed using tandem mass spectra acquired by a linear ion trap mass spectrometer from triplicate runs. A total of 1749 and 1955 proteins with protein probability above 0.95 were identified from MDAH-2774 and TOV-112D after filtering through Peptide Prophet/Protein Prophet software. Differentially expressed proteins were further investigated by ingenuity pathway analysis (IPA) to reveal the association with important biological functions. Canonical pathway analysis using IPA demonstrates that important signaling pathways are highly associated with one of these two cell lines *versus* the other, such as the PI3K/AKT pathway, which is found to be significantly predominant in MDAH-2774 but not in TOV-112D. Also, protein network analysis using IPA highlights p53 as a central hub relating to other proteins from the connectivity map. These results illustrate the utility of high throughput proteomics methods using large-scale proteome profiling combined with bioinformatics tools to identify differential signaling pathways, thus contributing to the understanding of mechanisms of deregulation in neoplastic cells.

1 Introduction

Ovarian cancer is the fifth leading cause of cancer-related death in the Western world and causes more deaths of women in the United States than any other gynecological

Correspondence: Professor David M. Lubman, Department of Surgery, The University of Michigan Medical Center, 1150 West Medical Center Dr., Building MSRB1 Rm A510B, Ann Arbor, MI 48109-0656, USA
E-mail: dmlubman@umich.edu
Fax: +1-734-615-2088

Abbreviations: ER, estrogen receptor; **FA,** formic acid; **FDR,** false discovery rate; **IPA,** ingenuity pathway analysis; **IPKB,** ingenuity pathway knowledge base; **LTQ,** linear ion trap mass spectrometer; **OEA,** ovarian endometrioid adenocarcinoma; **TPP,** transproteomic pipeline

malignancy [1]. The five-year survival rate can be as high as 90% with early detection; however, early detection of ovarian cancer is rare and known markers have limited utility for general population screening. The most common form of ovarian cancer is epithelial ovarian cancer, which can be further divided into four major histological subtypes: serous, clear cell, mucinous and endometrioid [2]. Ovarian endometrioid adenocarcinoma (OEA) represents approximately 20% of common epithelial tumors.

In the present comparative study, we have employed two closely related OEA cell lines, MDAH-2774 and TOV-112D [3, 4]. Both these cell lines were derived from female Caucasian patients with OEA. In particular, the TOV-112D cell line originates from an aggressive ovarian endometrioid tumor (stage 3, grade 3). The growth characteristics and tumorigenic potential of this cell line parallels the clinical behavior of aggressive OEAs. Categorization of the tumor grade/stage of MDAH-2774 is not available. Differential

global gene expression analyses have been performed, and different genetic defects have been previously detected between these two cell lines, possibly leading to different levels of deregulation of important signaling pathways [5]. It has been shown that both the MDAH-2774 and TOV-112D cell lines have elevated constitutive Wnt signaling deregulation. A missense AXIN1 sequence alteration was identified in MDAH-2774 and mutant beta-catenin was identified in TOV-112D. A mutated K-ras gene, involved in the PI3K/AKT signaling pathway, was detected in MDAH-2774 but not in TOV-112D. Both the MDAH-2774 and the TOV-112D OEA cell lines have a mutant p53 gene (unpublished data, R. Wa *et al.* Medical School, Unversity of Michigan, USA).

Protein expression and gene expression data, while being mutually exclusive, are complementary to each other. A lack of direct correlation between protein expression and gene expression has been reported [6]. Protein over-/under-expression is expected to relate to deregulated tumor cell behavior more directly than would gene expression. The proteomic profiles of these two cell lines have been generated in previous work using different methods. In one study, Rotofor IEF and nonporous silica RP separation was coupled with ESI-TOF-MS and MALDI-TOF-MS to analyze the proteome of MDAH-2774 *via* intact protein fractionation [7]. In a second study, 2-D PAGE coupled to MALDI-TOF-MS and SDS-PAGE coupled to LC-MS/MS were both used to obtain protein profiles from TOV-112D [8]. Alternatively, a shotgun [9] proteomics strategy of a whole-cell digest can be used to compare the global proteome profile of MDAH-2774 and TOV-112D (both qualitatively and quantitatively) in order to analyze protein expression differences in neoplastic dedifferentiation. Within, we have utilized cIEF to separate peptides based on pH [10, 11], followed by capillary RP separation with online nanoESI-ion trap mass spectrometer analysis. This method is capable of identifying large numbers of proteins over an extended pH range where 1749 and 1955 proteins from triplicate runs of MDAH-2774 and TOV-112D, respectively, have been identified in this work.

Quantitation is always an important issue in pathway analysis using either isotopic labeling or label-free methods. Label-free quantitation has gained increasing popularity in recent years and has been successfully applied in large quantitative studies [12, 13] due to the development of computational and statistical methods and advances in LC-MS/MS systems. Extraction of peptide ion intensities and spectral counting (defined as the number of MS/MS spectra identified *per* protein) are two widely adopted methods for performing comparative quantitative analysis of LC-MS proteomics experiments. It has been shown that spectral counting is highly reproducible and is sensitive to protein abundance changes [14]. Further, in controlled experiments it was found that the correlation of protein abundance with spectral count is superior to that of protein sequence coverage or peptide count [13]. Thus, we have utilized spectral counting to measure protein abundance. The ratio of the spectral count of the same protein repre-

sents the relative expression level between two samples. Spectral sampling can enable protein ratios larger than ∼twofold to be determined with high confidence.

The large number of identified proteins between these two cell lines provides a means for qualitative and quantitative bioinformatics pathway analysis. Differentially expressed proteins can be further investigated to reveal the associated biological pathways using bioinformatic tools such as ingenuity pathway analysis (IPA). The IPA program uses a knowledge base derived from the literature containing information on interactions between genes, proteins and other biological molecules. After uploading differentially expressed protein lists to the IPA server, IPA uses these focused proteins to extract connectivity networks that relate candidate proteins to each other based on their interactions and generates global canonical pathways that are shown to be significantly associated with these candidates [15]. As illustrated in Fig. 1, we have used a strategy of shotgun proteomics with subsequent bioinformatics analysis to study pathways in the TOV-112D and MDAH-2774 cell lines in order to understand the different interactions in these two OEA cell lines.

2 Materials and methods

2.1 Sample preparation

MDAH-2774 and TOV-112D cells were gently washed three times with PBS (pH 7.4) by repetitive pipetting, followed each time by centrifugation at 1500*g* for 5 min at 4°C. The cell pellets were resuspended with 1 mL lysis-denaturing buffer (7.5 M urea, 2.5 M thiourea, 12.5 v/v glycerol, 62.5 M Tris-HCl, 2.5% w/v *n*-octylglucoside and 1% v/v protease inhibitor cocktail). All chemicals were purchased from Sigma (St. Louis, MO) unless otherwise noted. The lysates were vortexed and then centrifuged at 35 000*g* for 1 h at 4°C. The supernatant was collected and dialyzed against 50 mM ammonia bicarbonate overnight using Slide-A-Lyzer dialysis cassettes with a 3500 Da molecular cutoff from Pierce (Rockford, IL). The proteins were quantified with the micro-BCA assay kit from Pierce, and then lyophilized to 100 μL using a SpeedVac concentrator (Labconco, Kansas City, MO) operating at 45°C.

2.2 Trypsin digestion

An aliquot of 5 mM of DTT was added and the mixture was incubated at 60°C for 30 min. After cooling, 5 mM iodoacetamide was added and the mixture was placed in the dark at room temperature for 30 min in order to carboxamidomethylate the Cysteine residues. Then, 1:50 w/v L-1-tosylamido-2-phenylethyl chloromethylketone modified sequencing-grade porcine trypsin from Promega (Madison, WI) was added. Following vortexing, the mixture was incubated overnight at 37°C in a water bath with agitation, followed by addition of 2% formic acid (FA) to terminate the reaction.
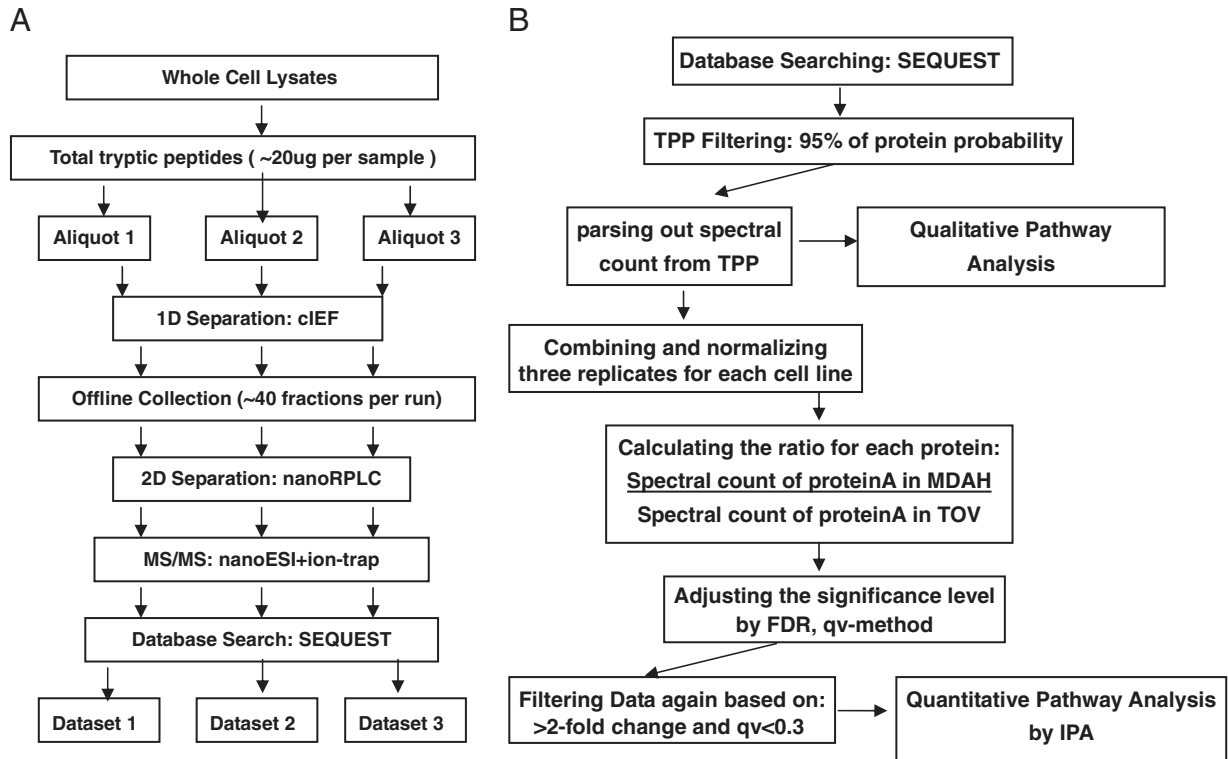
Figure 1. (A) Experimental flow chart; and (B) Data Processing Strategy.

2.3 First dimension separation: cIEF

Peptides were sequentially resolved based on their different p*I*s and hydrophobicity. cIEF was performed on a Beckman CE instrument with sample collection as shown in Fig. 2. A 70 cm cIEF (100 μm id, 365 μm od) capillary was coated with hydropropyl cellulose for eliminating electroosmotic flow and absorption of peptides onto the capillary wall. The capillary was initially filled with sample gel buffer containing 2% ampholyte 3–10 and 1 μg tryptic peptides. Sodium hydroxide solution at pH 10.8 and 0.1 M phosphate acid solution were employed as catholyte and anolyte, respectively. One end of the capillary was emerged in the anolyte, while the other end was kept in coaxial metal tubing with a sheath flow composed of catholyte eluting flush with the exit of the capillary. The flow rate was controlled by a syringe pump at 2 μL/min, and was adjusted to ensure that a proper droplet formed at the exit to carry the peptides fractionated into individual wells in the sample plate. IEF was performed at 21 kV (300 V/cm) over the entire capillary. The current decreased continuously as the peptides were focused and the process was considered complete after the current no longer changed. The focused bands of peptides were sequentially mobilized slowly under pressure towards the cathode and delivered as droplets with catholyte sheath flow into individual wells on a sample plate, where the fractions were collected with a modified Beckman HPLC sample collector. Each cIEF separation runs approximately 90 min. One-third of the run time is spent focusing the peptides in the capillary while the remaining time is used to deposit the offline fractions.
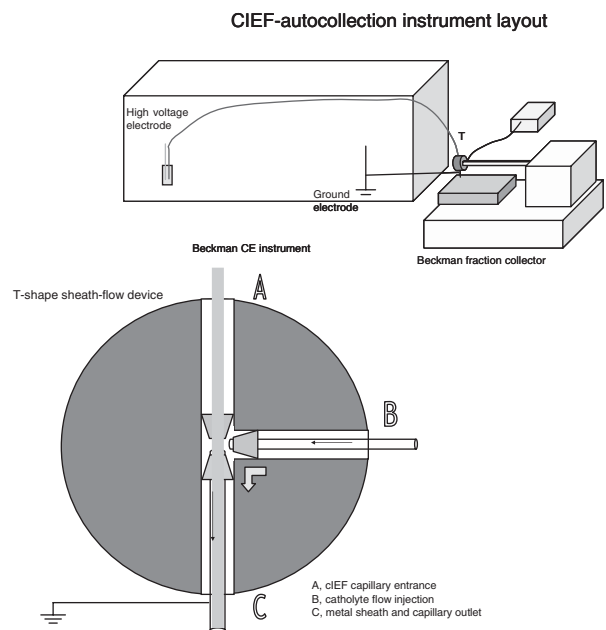

Figure 2. cIEF-autocollection instrument layout.

2.4 Second dimensional separation: nanoRPLC+ nanoESI-MS/MS

When cIEF separation was completed, each p*I* fraction of tryptic-digested sample was injected *via* Paradigm autosam-

pler (Michrom Biosciences, Auburn, CA) and loaded onto a desalting nano-trap (150 μm × 50 mm) (Michrom) connected to a nano-RP column (C18AQ, 5 μm 200 Å 0.1 × 150 mm) (Michrom) by a Paradigm AS1 micropump (Michrom). The mobile phases A and B were composed of 0.3% FA in water and 0.3% FA in ACN, respectively. Peptides were first desalted and enriched starting at 100%A with a flow rate of 50 μL/min for 5 min. The sample was subsequently separated by a nano-RP column with a flow rate of 0.3 μL/min after splitting. The linear gradient for separation was as follows: from 3% ACN to 12% ACN in 5 min, from 12% ACN to 40% ACN in 30 min, from 40% ACN to 80% ACN in 15 min and finally decreased from 80% ACN to 3% ACN in 10 min. The resolved peptides were then introduced into a ThermoFinnigan linear ion trap mass spectrometer (LTQ) (Thermo Electron, San Jose, CA) equipped with a nanospray ion source (Thermo Electron). The LTQ was operated in data-dependent mode in which one cycle of experiments consisting of one full MS scan was followed by five pairs of zoom scans and MS/MS scans with dynamic exclusion set to 30 s. The capillary temperature was set at 175°C, spray voltage was 2.8 kV, capillary voltage was 30 V and the normalized collision energy was 35% for the fragmentation.

2.5 Database search and protein identification

MS/MS spectra were then searched against the human UniProt FASTA database (updated in Dec. 2007) by TurboSEQUEST provided by Bioworks ver3.1 SR1 (Thermo-Finnigan). The search was performed using the following parameters: (i) enzyme: trypsin; (ii) one missed cleavage allowed; (iii) peptide ion mass tolerance: 1.5 Da; (iv) fragment ion mass tolerance: 0.0 Da; (v) mass tolerance for precursor ions: 1.4 Da; (vi) peptide charges: $+1, +2, +3$; (vii) possible modifications: 15.99 Da shift for oxidized Met residues; 79.97 Da for phosphorylated Ser, Thr, Tyr residues, respectively; 58.1 Da shift for carboxymethylated Cys residues. The identified peptides were subsequently processed through Peptide Prophet and Protein Prophet incorporated in the trans-proteomic pipeline (TPP) [16]. In TPP, the search results were first validated by Peptide Prophet, which converts various SEQUEST parameters to a discriminant score and uses Bayesian statistics to compute the probability that each identified peptide is correct. Protein Prophet reads in peptides and assigned probabilities to compute the probabilities of proteins that are present in the original sample (http://proteinprophet.sourceforge.net/prot-software.html). In this study, we use a protein probability score of ≥0.95 as the threshold for protein identification, to ensure that the minimized overall error rate is below 0.05.

2.6 Label-free quantitation and normalization

After processing the SEQUEST data through TPP, the spectral counts were parsed out of TPP protXML files using a perl script (see Fig. 1B). Three datasets of identified proteins with 0.95 protein probability and their associated spectral count have been generated for each sample. We divided the data into two groups. Qualitative data consists of proteins that are only identified in one of these two cell lines, whereas quantitative data consists of proteins that are identified in both these cell lines with their expression values. In the first group one cannot compare the same protein expression level between two cell lines. In the second group the relative protein abundance fold change of the same protein can be calculated by the ratio of their spectral count in two samples as explained later in this work. The data was processed in two different ways. For the qualitative analysis, only the qualitative data was used (*i.e.* only different protein names from two samples were included). We also performed a quantitative analysis in which we combined the qualitative data and quantitative data by replacing any missing value with 0. For example, protein "A" is only detected in MDAH-2774 with its assigned spectral count. In order to make the comparison of the differential expression of this protein plausible, we assume protein "A" is also present in TOV-112D but at a very low level that is not detectable and assign a spectral count of 0 to protein "A" in TOV-112D.

Subsequent normalization was used to reduce technical bias when acquiring spectral count data from different runs across the two different cell lines. The bias may come from instrument error or the inherent random sampling nature of the LTQ. In order to normalize the data, we first calculated the ratio of the total spectral count of three runs between MDAH-2774 and TOV-112D and then multiplied the spectral count of each protein in the numerator by this ratio. Statistical significance levels of the pairwise comparison were then adjusted for multiple testing using the false discovery rate (FDR), *q*-value method. Differentially expressed genes used to learn network structure were declared at an FDR *q*-value threshold of 0.3. The FDR *q*-values were calculated using the R package (http://cran.r-project.org/web/packages/qvalue/).

2.7 Ingenuity pathway analysis

To infer global network functions between all differentially expressed proteins from MDAH-2774 and TOV-112D, we conducted two types of analysis using IPA software. For the qualitative analysis, we uploaded into the IPA database two sets of proteins with corresponding primary accession number, which were only identified from one of these two cell lines. Out of 652 and 838 proteins uploaded from MDAH-2774 and TOV-112D, the IPA software identified 515 and 665 "focus genes" that were eligible for generating connectivity networks and 443 and 582 "focus genes" that were eligible for generating biological functions/disease and associated pathways.

In order to gain further insight into the dynamic changes of the cell states between these two cell lines at the molecular level, we performed a quantitative analysis by incorporating quantitative data in addition to qualitative

data. In this analysis, we uploaded a list of 828 differentially expressed proteins with fold change larger than 2 based on normalized spectral count data. The relevant proteins with their fold change, *qv*-value and corresponding primary accession number were uploaded as an Excel spreadsheet file. A total of 609 proteins were eligible for generating networks and 532 proteins were used to retrieve functions/pathways after applying a threshold of *qv*-value of <0.3.

The significance values for analyses of network and pathway generation were calculated using the right-tailed Fisher's Exact Test by comparing the number of proteins that participate in a given function or pathway relative to the total number of occurrences of these proteins in all functional/pathway annotations stored in the ingenuity pathway knowledge base (IPKB).

2.8 Western blot analysis

MDAH-2774 and TOV-112D cell lines were lysed in lysis buffer as described above. An aliquot of 100 µg of total protein from each of the cell lysates was separated by 10% SDS-PAGE in parallel. The resolved proteins were transferred to PVDF membranes (Immobilon-P, Millipore) by conventional procedures using a TE70 semi-dry transfer unit (Amersham Biosciences, Princeton, NJ). Beta-actin protein expression levels were used as an internal control to ensure equal loading between lanes. After transfer, membranes were incubated with a blocking buffer consisting of PBS and 0.1% Tween 20 containing 5% nonfat dry milk overnight. The membranes were incubated for 1 h at room temperature with primary antibodies against UCHL1 (rabbit polyclonal antibody, Biogenesis, NH), SFN (mouse monoclonal antibody; Abcam, Cambridge, MA), MARKS (mouse monoclonal antibody, Abcam) and beta-actin (mouse monoclonal antibody, Sigma) for 1 h at 1:5000 dilution in Tris-buffered saline. Membranes were simultaneously incubated with the mouse anti-beta actin antibody and either the rabbit anti-UCHL1 antibody, the mouse anti-SFN antibody or the mouse anti-MARKS antibody. After three washes with washing buffer (PBS containing 0.1% Tween 20), the membranes were incubated with the appropriate secondary antibody (highly cross absorbed HRP-conjugated goat anti mouse and/or highly cross-absorbed HRP-conjugated goat anti-rabbit; Abcam) for 1 h at 1:2000 dilution. Immunodetection was accomplished by enhanced chemiluminescence (Amersham Biosciences) followed by autoradiography on Hyperfilm MP (Amersham Biosciences).

3 Results and discussion

3.1 cIEF performance

An aliquot of 20 ug of whole cell lysate was extracted from MDAH-2774 and TOV-112D followed by trypsin digestion. Each aliquot of tryptic peptides (~5 µg) was then loaded to
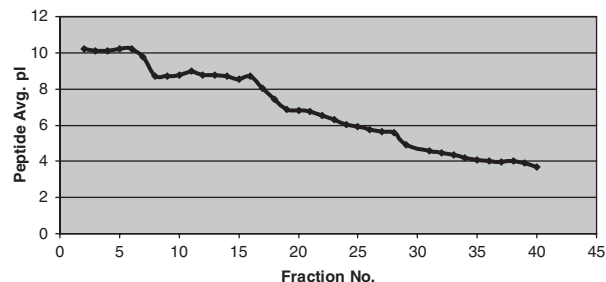

Figure 3. Theoretical p*I* distribution plot of the second run of MDAH-2774. Fraction number shown in the *X*-axis is plotted against the average of peptides' p*I* value within each fraction shown in the *Y*-axis.

cIEF separation. About 40 fractions were collected *per* run and each fraction was further subjected to the second-dimensional separation coupled with nanoESI-ion trap. We have repeated this process three times for each cell line sample.

The theoretical p*I* value for each identified peptide within each fraction was calculated after database searching. The p*I* distribution plot from the second run of MDAH-2774 is shown in Fig. 3. As expected, the p*I* trend follows a non-perfect linear velocity. Peptides with p*I* in the region 3.5 to 8 tend to show improved separation performance compared with basic peptides with p*I* from 8 to 10. Peptides with p*I* above 10 or below 3 are not expected to be resolved as they fall outside the pH range of the ampholytes (pH 3–10) used in these experiments. Overall, cIEF exhibits high separation resolution with little overlap of the same peptides identified between adjacent fractions.

An important issue here is the use of offline collection of cIEF fractions coupled to nano-RPLC. With the use of online cIEF coupled to nano-RPLC one can directly load each fraction to a nano-RP column by sacrificing the separation resolution due to the transfer of cIEF fractions to the second dimension from the increased back-pressure and dead volume. In the offline collection method, the sheath-flow eluting from the coaxial tubing was adjusted flush with the exit of the capillary in order to eliminate back-pressure and dead volume as shown in Fig. 2B. Compared with the online integration of cIEF/nano-RPLC, the offline collection mode does not degrade the cIEF separation, significantly reducing the mixing of separated peptides during the transfer process. This is also central for precise quantification by spectral counting in this work.

3.2 Proteomic profiling

3.2.1 Number of proteins identified

MS/MS spectra were searched against the UniProt database using SEQUEST software and search results were then validated using the Peptide Prophet program. Peptide Prophet provides an empirical statistical model that

estimates the accuracy of peptide identifications made by SEQUEST. For each tandem mass spectrum, Peptide Prophet determines the probability that the spectrum is correctly assigned to a peptide based upon its SEQUEST scores. A second program, Protein Prophet, was subsequently used to group peptides by their corresponding proteins to compute probabilities that those proteins were present in the original sample. A stringent cutoff of 0.95 was used to filter all the SEQUEST results based on Protein Prophet's estimate of error rate. For each cell line, we have repeated the same experimental procedure and combined the results from all three runs instead of selecting only the overlapping proteins. This is done since some proteins can only be identified in a single run of a sample due to the random sampling nature of tandem MS. The Venn diagram in Fig. 4A summarizes the intersection of proteins

identified from all three runs of MDAH-2774. In the first run of MDAH-2774, 656 distinct proteins were identified from 25 fractions. A total of 1181 and 1095 distinct proteins were identified in the second run and the third run of MDAH-2774, respectively, when we increased the fractionation number to approximately 40. The total number of proteins from the combined list is 1749 for MDAH-2774 and 1955 for TOV-112D with an overlap of 1092 as shown in Fig. 4B. More detailed peptide information for each proteome analysis can be found in Supporting Information, together with the number of distinct peptides, the number of spectral counts and other related information.

3.2.2 Cellular localization

Each identified protein was assigned a cellular localization based on information from the Swiss-Prot, entrez gene and genome ontology databases. Figure 5 shows the cellular distribution of 1749 identified proteins from MDAH-2774 and 1092 identified proteins from TOV-112D. The majority are cytoplasmic and nuclear proteins for both these cell lines. Membrane proteins only occupy 6% of each total proteome, which is not surprising since the protein extraction method used in this study is not optimized for hydrophobic proteins.

3.3 Label-free quantitation

Detecting protein quantity and the changes in this quantity between various stages or different samples is central to understanding the molecular process of the cell. We used the spectral count as the measurement of relative protein abundance because it has been shown to accurately reflect
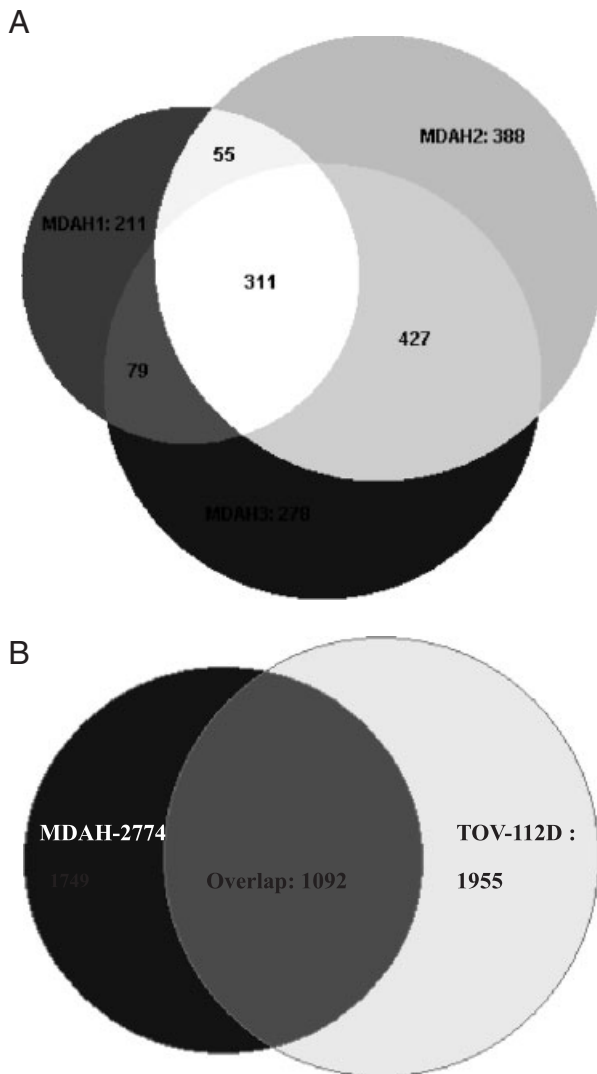

Figure 4. Venn diagram of the number of proteins identified from: all three runs of MDAH-2774 (A); MDAH-2774 and TOV-112D (B) with a minimum protein probability of 0.95 as given by ProteinProphet™.
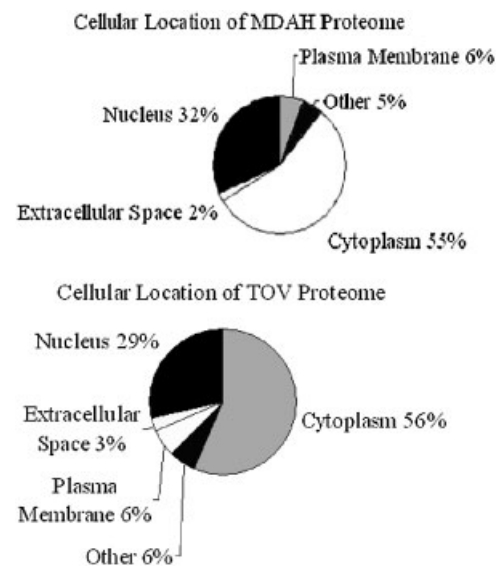

Figure 5. Cellular distribution of identified proteins from MDAH-2774 and TOV-112D

relative protein abundance with a linear correlation of over two orders of magnitude of dynamic range [14]. Spectral count was assigned to each identified protein followed by normalization and log transformation. The signal distribution in Fig. 6A shows that the ratio of protein expression level between MDAH-2774 and TOV-112D follows a symmetric distribution. These two cell lines have approximately an equal number of proteins that are up-regulated or down-regulated when compared with each other. Only proteins with fold change larger than 2 between MDAH-2774 and TOV-112D, which is equal to 1 in $\log_2$ scale, are shown in Fig. 6A and were used for further comparative analysis. About two-thirds of the identified proteins fall into the column with fold change range between 2 and 4. The rest of proteins fall into a fold change from 4 to 32, with a few exceptions over 2 orders of magnitude. The distribution
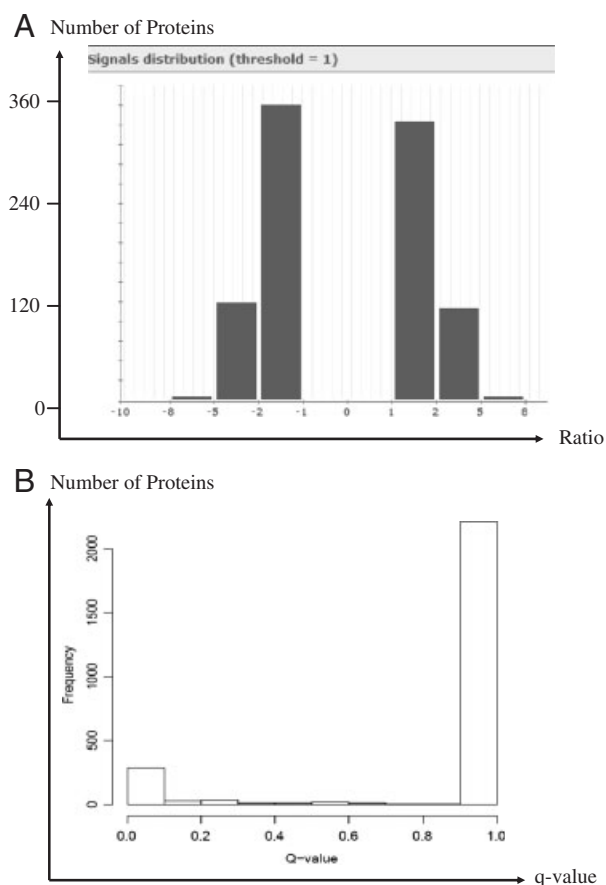
Table 1. Pearson correlation analysis among replicates within the same cell line and between two different cell line samples

Test	R	Confidence interval (95%)
MDAH2 *versus* MDAH3	0.75	[0.72, 0.78]
TOV2 *versus* TOV3	0.85	[0.83, 0.87]
MDAH1 *versus* TOV1	0.5	[0.41, 0.56]
MDAH1 *versus* TOV2	0.5	[0.46, 0.62]
MDAH1 *versus* TOV3	0.6	[0.5, 0.64]
MDAH2 *versus* TOV1	0.6	[0.51, 0.63]
MDAH2 *versus* TOV2	0.4	[0.36, 0.51]
MDAH2 *versus* TOV3	0.4	[0.34, 0.49]
MDAH3 *versus* TOV1	0.6	[0.51, 0.62]
MDAH3 *versus* TOV2	0.5	[0.38, 0.53]
MDAH3 *versus* TOV3	0.4	[0.34, 0.5]

of q-values for all proteins regardless of the expression fold change is also plotted in Fig. 6B.

Pearson correlation analyses have also been applied to assess the reproducibility and quality of the quantitative data. The first run of MDAH termed MDAH1 and the first run of TOV termed TOV1 are slightly experimentally different than the rest of runs in terms of the number of fractionations in the cIEF separation step. It would therefore make more sense to evaluate the reproducibility between (MDAH2 and MDAH3), (TOV2 and TOV3), which results in a moderate-to-high Pearson correlation coefficient of 0.75 and 0.85. Pearson correlation analyses have also been applied to any of the two runs from two distinct cell line samples. Results are summarized in Table 1.

Table 2 lists the top 10 most abundant proteins in MDAH-2774 and in TOV-112D and the top 10 most differentially expressed proteins based on the ratio of their spectral count from MDAH-2774 over TOV-112D. From the quantitation list, we observed that the most abundant proteins from both these cell lines are proteins related to structural elements like vimentin, actin and tubulin, as well as chaperone proteins and members of the heat shock protein family. Proteins that are most differentially expressed between these two cell lines cover a wide range of molecular functions and associations with different diseases. For example, collagen 3 alpha 1 is a structural constituent of intracellular matrix. Tubulin beta 4 is the major constituent of the microtubules. They have both been shown to be associated with epithelial ovarian cancer. Stratifin, a protein kinase C inhibitor, is involved in regulation of progression through the cell cycle and has been shown to be associated with breast cancer and prostate cancer. Eukaryotic translation elongation factor 1 alpha 2 has been shown to be associated with breast cancer. Myristoylated alanine-rich protein kinase C substrate has been shown to be associated with endometriosis.

Other differentially expressed proteins that are not shown in this table also have important implications on the

A Number of Proteins

B Number of Proteins

Figure 6. (A) Distribution of the protein abundance ratio between MDAH-2774 and TOV-112D on $\log_2$ scale. Eight hundred and twenty eight differentially expressed proteins with fold changes larger than 2 based on normalized spectral count data were used to generate this histogram. Horizontal axis shows the ratio of the relative abundance between filtered proteins from MDAH and TOV on $\log_2$ scale. Vertical axis shows the number of proteins within each column. (B) shows the distribution of the q-values for all proteins regardless of the expression fold change. A clear separation can be seen between significant range (q-value < 0.1) and insignificant range (q-value > 0.1).

Table 2. Top 10 expression molecules and top 10 differential expression molecules[a]

	Protein accession number	Gene name	Exp. value (log)	*qv*-value	Description
Top 10 expression molecules in MDAH-2774	P60709	ACTB	8.1		Actin, beta
	P08670	VIM	8.04		Vimentin
	P38646	HSPA9	7.86		Heat shock 70 kDa protein 9 (mortalin)
	P68032	ACTC1	7.84		Actin, alpha, cardiac muscle 1
	P11142	HSPA8	7.75		Heat shock 70 kDa protein 8
	P62736	ACTA2	7.74		Actin, alpha 2, smooth muscle
	P10809	HSPD1	7.67		Heat shock 60 kDa protein1 (chaperonin)
	P043350	TUBB4	7.44		Tublin, beta 4
	P11021	HSPA5	7.43		Heat shock 70 kDa protein 5 (glucose-regulated protein, 78 kDa)
	P68104	EEF1A1	7.39		Eukaryotic translation Elongation factor 1 alpha 1
Top 10 expression molecules in TOV-112D	P08670	VIM	10.67		Vimentin
	P07737	PFN1	8.52		Profilin1
	P15531	NME1	8.3		Non-metastatic cells 1
	P60709	ACTB	8.2		Actin, beta
	P22392	NME2	8.15		Non-metastatic cells 2
	P11142	HSPA8	8.03		Heat shock 70 kDa protein 8
	P68104	EEF1A1	7.9		Eukaryotic translation Elongation factor 1 alpha 1
	P04083	ANXA1	7.81		Annexin A1
	P02461	COL3A1	7.78		Collagen, type 3, alpha 1
	P61978	HNRPK	7.65		Heterogeneous nuclear ribonucleoprotein K
Top 10 differentially expressed molecules between MDAH-2774 and TOV-112D	P02461	COL3A1	−7.78	0	Collagen, type 3, alpha 1
	P04350	TUBB4	7.44	0	Tublin, beta 4
	P09936	UCHL1	−6.77	0	Ubiquitin carboxyl terminal Esterase L1
	P31947	SFN	6.56	0	Stratifin
	P20671	HIST1H2AD	6.15	0.26	Histone cluster1, H2ad
	Q05639	EEF1A2	−5.38	0.26	Eukaryotic translation elongation factor 1 alpha 2)
	P29966	MARCKS	−5.24	0.01	Myristoylated alanine-rich protein kinase C substrate
	P31949	S100A11	5.19	0	S100 calcium-binding protein A11
	P37840	SNCA	−4.94	0.26	Synuclein, alpha
	Q9BQE3	TUBA1C	4.55	0	Tublin, alpha 1c

a) The expression values for top 10 expression proteins in MDAH-2774 and TOV-112D are normalized mean spectral counts across different runs and are shown in $\log_2$ scale. The expression values for top 10 differentially expressed proteins are normalized spectral count ratio in $\log_2$ scale. *qv*-value indicates the significance of difference from multiple test corrections.

mechanisms of OEA. For example, beta-catenin (CTNNB1), a critical component of the Wnt signaling pathway, was found to be over-expressed 4.2-fold in TOV-112D as compared with MDAH-2774 based on our spectral count data. This compares favorably with previously reported data by Wu *et al.* in which CTNNB1 was expressed 4.4-fold in TOV-112D over MDAH-2774 from the CTNNB1/TCF transcription reporter assay [5]. Although CTNNB1/TCF transcriptional activity in MDAH-2774 is modest compared with TOV-112D, it is known to be present at elevated levels in both these cell lines compared with other ovarian cell lines leading to constitutive activation of the Wnt signaling

pathway. Notably, the CTNNB1 missense mutation was detected in TOV-112D by PCR sequencing [5]. It has a mutation in its NH_2-terminal regulatory domain, thereby rendering the mutant protein resistant to degradation thus resulting in a higher CTNNB1 level in TOV-112D than in MDAH-2774.

The most significant drawback of spectral counting is that it is more likely to be influenced by the acquisition program of the mass spectrometer compared with other label-free comparative quantitation methods such as peptide ion intensity-based quantification. High-abundance peptides can mask low-abundant peptides if the

data-dependent MS/MS acquisition exclusion list is too small. If the exclusion list is too large, the spectral count can become rapidly saturated, resulting in reduced sensitivity. We have optimized the conditions in this case by extensive fractionation and setting the exclusion list time to 30 s.

3.4 Comparison with previously reported proteins

The proteome of both MDAH-2774 and TOV-112D cell lines have been previously analyzed by different methods. In the first study, a 2-D all liquid phase (Rotofor IEF nonporous silica RP-HPLC) separation method was used combined with ESI-TOF-MS and MALDI-MS/MS to compare the proteome profile of cultured ovarian cancer cell lines [8]. In this study, 161 unique proteins from MDAH-2774 were identified from five fractions with pH range from 5.8 to 8.3 by using PMF and peptide sequencing analysis after applying a 0.95 probability from the Mascot Search Engine. Around 70% of the proteins identified in the first study were also observed in the current study, including some important cancer-associated proteins such as the oncoprotein 18/stathmin, ezrin and p53 protein. Oncoprotein 18/stathmin, a conserved cytosolic phosphoprotein that regulates microtubule dynamics, was identified in two of the three runs of the MDAH-2774 cell line. It was previously reported that over-expression of oncoprotein 18 is associated with a variety of human cancers, including breast cancer and lung cancer [17, 18]. Ezrin is a member of the ezrin/radixin/moesin family of membrane-axin cross-linking proteins, which also transduces signals from growth factors. Previous studies have shown frequent ezrin over-expression in ovarian carcinomas, particularly in metastatic lesions [20]. Mutant P53 is also known to be over-expressed in MDAH-2774 [20].

In another publication [8], the proteome profiling of TOV-112D has been examined by two complementary proteomic approaches, two-dimensional gel electrophoresis (2-D PAGE) protein separation coupled to MALDI-TOF/MS and SDS-PAGE coupled to LC-MS/MS. A total of 172 proteins were identified from 2D-PAGE and a total of 589 proteins were identified from SDS-PAGE LC-MS/MS after applying a 0.9 probability cutoff by Protein Prophet, of which 436 proteins are also found in the current study. Relatively high expression of stress proteins like HSP90 and HSP71 were observed when compared with other proteins in both studies, as well as in numerous malignant tumors [21]. Two forms of aldehyde dehydrogenase1, which have previously been shown to be over-expressed in aggressive EOC *versus* non-aggressive EOC or normal ovarian surface epithelia cells at the RNA level, were also observed in both studies [22]. Proteins that have been previously proposed as biomarkers or targets for diagnostic studies of invasive ovarian cancer because of their over-expression in invasive carcinomas as compared with benign tumors have been identified in previous studies [23] including FK506-binding

protein 4 and several reported differentially expressed proteins such as proliferating cell nuclear antigen; leukemia-associated phosphoprotein (stathmin); glutathione S-transferase π; triose-phosphate isomerase and tumor metastatic process-associated protein (Nm23), which have been the subject of extensive investigation in ovarian cancer. In addition, Cytokeratin 18 and Cytokeratin 8 reported as biomarkers by Alaiya *et al.* [24] were also identified in our study, but not in the work of reference [8].

Overall, more than 70% of the proteins identified in previous work [7, 8] were also found in our study when comparing our proteomic profiling results with previously reported data. Coupling of offline cIEF with online nano-RPLC and nano-ESI-LTQ in our study has enabled a more comprehensive proteomic profiling of differentially expressed proteins between MDAH-2774 and TOV-112D.

3.5 IPA

3.5.1 Qualitative analysis

The 15 most variant canonical signaling pathways between these two cell lines were generated by IPA and are shown in Fig. 7 with a threshold *p*-value < 0.1 indicated. The length of the bar only indicates that the differentially expressed proteins are related to this pathway, but is by no means indicative of the pathway being either up-regulated or down-
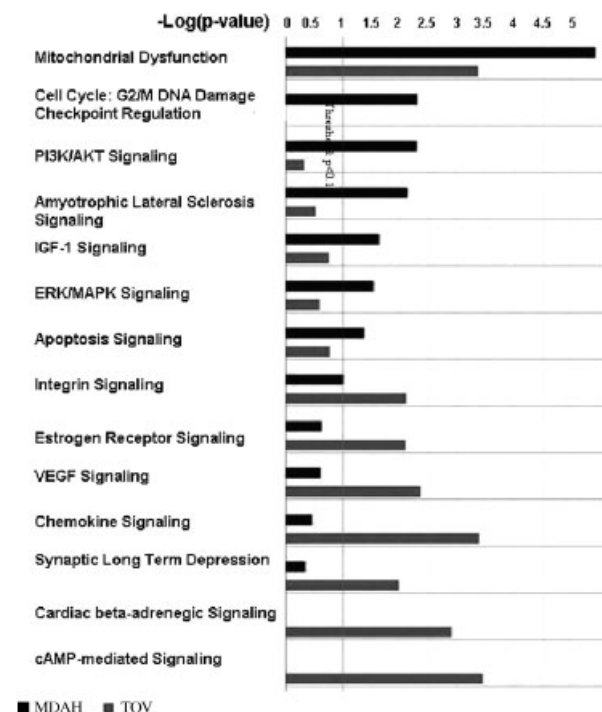

Figure 7. Comparison of canonical signaling pathways between MDAH-2774 and TOV-112D. Only the 14 most different pathways are shown in the figure, as ranked by the significance in MDAH-2774. The vertical line indicates a threshold of *p*<0.1.

regulated. It is possible that the overall activity of a pathway is up-regulated or down-regulated, but it is not sufficient to draw a conclusion of the direction of change based on the data forming network alone. It is shown that MDAH-2774 and TOV-112D have different levels of association with different signaling pathways. For example, PI3K/AKT signaling was found to be more significant in MDAH-2774 than in TOV-112D from this figure. Previous studies have shown that frequent activation and over-expression of PI3 K are associated with ovarian carcinoma [25]. Specifically, amplification of the catalytic subunit alpha of PI3K (PIK3CA) is detected in most ovarian cancer cell lines and primary tumors, as well as the somatic mutations in the gene encoding the p85α regulatory subunit of PI3K (PIK3R1), which leads to constitutive activation of PI3K. PIK3R1 was identified from MDAH-2774 but not in TOV-

112D in this study, implying that PI3K/AKT signaling up-regulation is potentially more activated in MDAH-2774 than in TOV-112D.

It has also been shown that estrogen signaling was found to have a stronger connection with TOV-112D than MDAH-2774 from our IPA. The estrogen receptor (ER) was found to be over-expressed in most ovarian cancers and anti-estrogen drugs have been used to inhibit the growth of ER-positive epithelia ovarian cancer cells, implying a strong connection between ER signaling and the tumor, but little is known about the detailed mechanism [26]. The stronger connection with ER signaling in TOV-112D is probably due to the activation of K-ras, which has been detected in TOV-112D but not in MDAH-2774 according to our analysis. K-ras is known to be present as the wild type in TOV-112D and mutated in MDAH-2774. Active K-ras can activate the ER through ERK-mediated ER phosphorylation and enhance the steady level of ER. Therefore, ER signaling may turn out to be more pronounced in TOV-112D than MDAH-2774. Also, VEGF and chemokine signaling, which are both related to metastasis, were both shown to be more significant in TOV-112D than MDAH-2774. Other pathways such as insulin-like growth factor-1 signaling, ERK/MAPK signaling, integrin signaling, cAMP-mediated signaling have all been previously reported to be involved in ovarian cancer [26].

Signature proteins were further characterized based on protein–protein interaction. The *de novo* network constructed by IPA shown in Fig. 8 produced a network comprising network eligible molecules that have been combined to maximize their specific connectivity. Additional molecules are imported from the IPKB to connect two or more smaller sub-networks by merging them into a larger one. The proteins in the network of MDAH-2774 and TOV-112D and their major functions are listed in Table 3. The network for MDAH-2774 is a compact one, centering on P53 while the network for TOV-112D is more scattered, composed of small sub-networks with ATM, Jnk and GLI1 in the center (Table 3).

3.5.2 Quantitative analysis

In order to gain further insight into the dynamic changes of the cell state between these two cell lines at the molecular level, we performed another analysis by incorporating quantitative data along with qualitative data. In this analysis, we uploaded a list of 828 differentially expressed proteins with fold changes larger than 2 based on normalized spectral count data. A total of 609 proteins were eligible for generating networks and 532 proteins were used to retrieve functions/pathways after applying a threshold qv-value of <0.3. The canonical signaling pathways enriched with differentially expressed proteins were constructed and ranked by significance. A total of 14 pathways have been calculated to be significant (Fig. 9), where 8 of these pathways are also shown in Fig. 7. In addition, some new pathways were also uncovered, such as protein ubiquitination signaling and the hypoxia signaling pathway. We
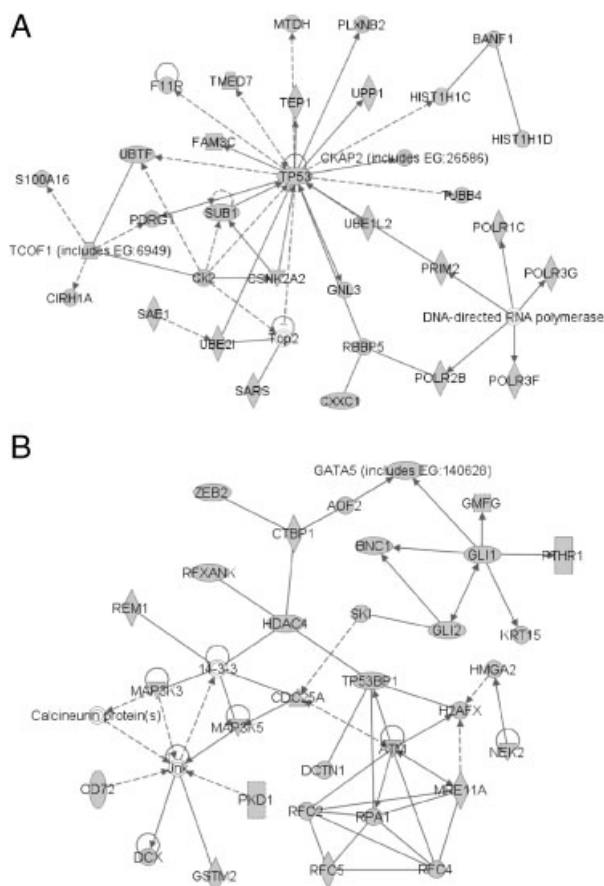

Figure 8. Top connectivity map for MDAH-2774 with *p* values of <10^{-49} (A) and TOV-112D (B) with *p* values of <10^{-45}. Nodes with gray background are network-eligible molecules and nodes with plain background (TOP2, DAN-directed RNA polymerase for MDAH-2774 and 14-3-3, Jnk for TOV-112D) are imported from IPKB. A line indicates interactions, with the arrowhead indicating directionality. The absence of arrowheads refers to a binding interaction. The dotted line indicates an inferred or indirect interaction. The score is based on a *p*-value calculation, which calculates the likelihood that the network-eligible molecules that are part of a network are found therein by random chance alone.

Table 3. Molecules in the top1 network of MDAH-2774 and TOV-112D

Analysis	Molecules in the network	Score	Top functions
MDAH-2774	BANF1, CIRH1A, Ck2, CKAP2 (includes EG26586), CSNK2A2, CXXC1, DNA-directed RNA polymerase, F11R, FAM3C, GNL3, HIST1H1C, HIST1H1D, MTDH, PDRG1, PLXNB2, POLR1C, POLR2B, POLR3F, POLR3G, PRIM2, RBBP5, S100A16, SAE1, SARS, SUB1, TCOF1 (includes EG 6949), TEP1, TMED7, Top2, TP53, TUBB4, UBE1L2, UBE2I, UBTF, UPP1	−49	Cell cycle, cellular assembly and organization, DNA replication, recombination and repair
TOV-112D	14-3-3, AOF2, ATM, BNC1, calcineurin protein(s), CD72, CDC25A, CTBP1, DCTN1, DCX, GATA5 (includes EG 140628), GLI1, GLI2, GMFG, GSTM2, H2AFX, HDAC4, HMGA2, Jnk, KRT15, MAP3K3, MAP3K5, MRE11A, NEK2, PKD1, PTHR1, REM1, RFC2, RFC4, RFC5, RFXANK, RPA1, SKI, TP53BP1, ZEB2	−45	Cancer, cell cycle, DNA replication, recombination and repair
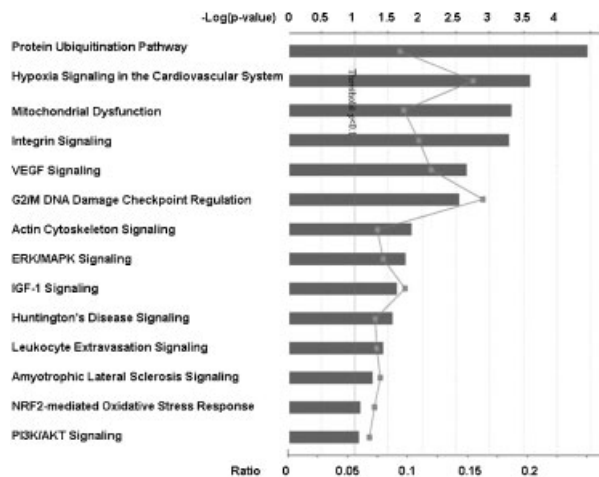
Figure 9. Canonical signaling pathways enriched with differential expressed proteins from quantitative analysis. A threshold of *p*-value <0.1 is applied. Bar indicates the significance and box indicates the ratio of the number of proteins found in our study over the number of proteins in IPA knowledge base under the same signaling pathway.

have also sought to examine the differentially expressed components of these pathways in depth. For example, the detailed signaling cascade of PI3K/AKT is depicted in Fig. 10. By incorporating the normalized spectral count results, we have been able to calculate the relative expression level of identified proteins under this pathway in addition to detecting their presence or absence.

The top connectivity network from integrative analysis for MDAH and TOV depicting protein–protein interactions is shown in Fig. 11. The major network, which comprises 34 identified differentially expressed proteins and two imported from IPKB, is displayed in Fig. 10 with a *p*-value of $<10^{-49}$. The major functions extracted from this network are related to cancer, reproductive system disease, and skeletal and muscular disease. P53 is the hub of this network, implying that the differential expression level of P53 in these two cell lines is one of the major driving forces for their differentiation in tumor growth.

Pathway analyses of the qualitative data and the quantitative data partially coincide with each other by using IPA. Qualitative data represents a group of proteins with enri-
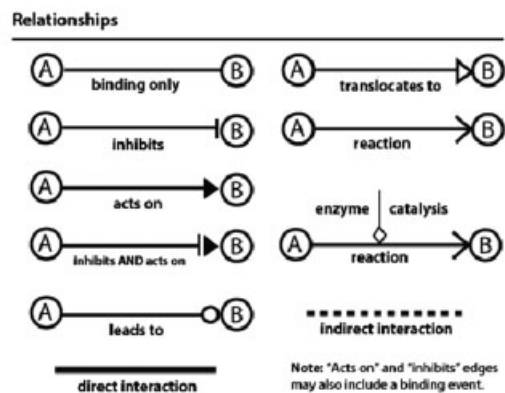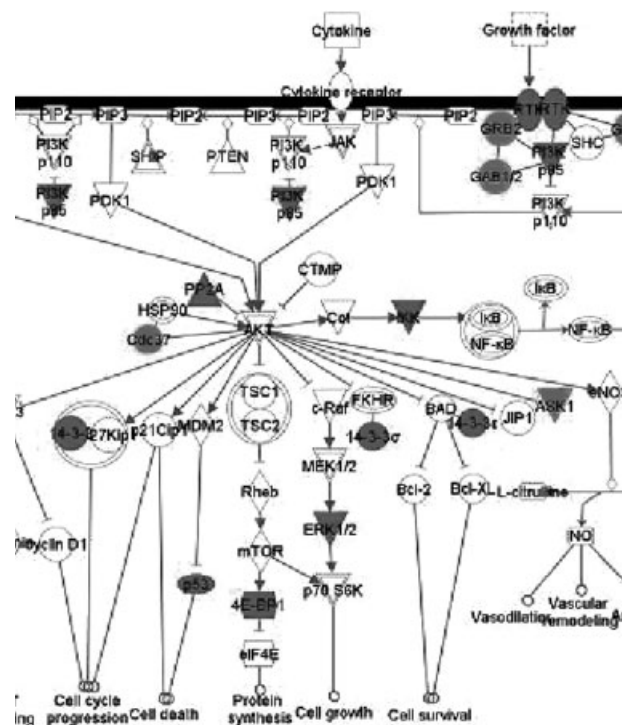

Figure 10. Signaling cascade of PI3K/AKT pathway. Green nodes represent over-expression in MDAH-2774 and red nodes represent over-expression in TOV-112D. Plain nodes are imported from IPKB. This figure has been manually modified from integrative analysis by adding some proteins that were identified from only one cell line but did not meet the threshold of integrative analysis.
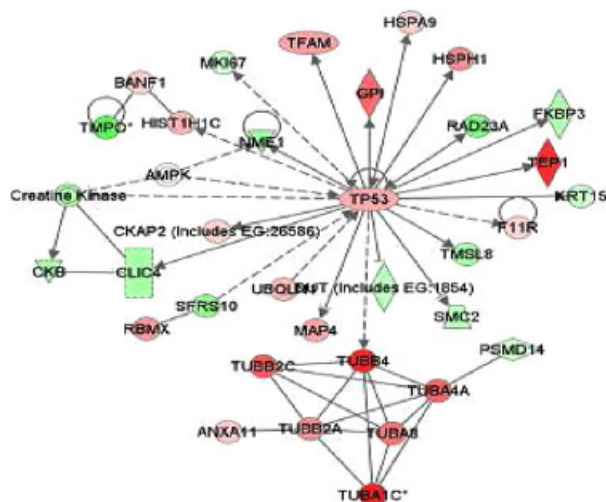
Figure 11. Top connectivity network from integrative analysis. Red and green nodes represent proteins that are identified to be over-expressed in MDAH-2774 and TOV-112D, respectively. Darker color indicates larger fold change. The detailed description of these molecules and their relative expression values can be found in the Supporting Information.
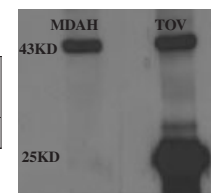

Protein Accession Number	Gene Name	Exp.Value(log)	qv-value
P09936	UCHL1	-6.77	0

Protein Accession Number	Gene Name	Exp.Value(log)	qv-value
P31947	SFN	6.56	0

Protein Accession Number	Gene Name	Exp.Value(log)	qv-value
P29966	MARKS	-5.24	0.01

Figure 12. Western results of UCHL1 (A), Stratifin(1433-sigma) (B) and MARKS (C). Expression values are normalized spectral count ratio in $\log_2$ scale. Positive value indicates over-expression in MDAH-2774 and negative value indicates over-expression in TOV-112D. Band shown on 37 KD is Beta-actin, which was used as a control. (A) Western results of UCHL1. (B) Western results of Stratifin. (C) Western results of MARKS.

ched difference between MDAH-2774 and TOV-112D, as they are only detectable in either of these two cell lines. Quantitative data consists of proteins that are detected in both these cell lines with a fold change larger than 2 in addition to those detected in only one of these two cell lines. Analysis based on qualitative data alone is simple to handle; meanwhile it is biased as it excludes the information containing dynamic change in protein abundance whereas quantitative analysis is more comprehensive. The combination of qualitative data and quantitative data is based on the assumption that the spectral count of the protein detected in only one sample is assigned to 0 in the other one. However, we have observed a decrease of sensitivity induced by replacing any missing values with 0. After multiple testing corrections, fold changes of some proteins between MDAH and TOV are decreased and the *q*-values are increased, which suggests global signal suppression by this replacement method. One of the explanations could be these missing values are not truly 0, simply because we cannot detect them by the current technique. This is especially the case for low-abundance peptides, which could be masked by their co-eluting high-abundance peptides. In the future, target proteomics method (*e.g.* multiple reaction monitoring methods) will be adopted to verify important proteins.

3.6 Western blot validation

It is becoming increasingly important to validate the proteome profiling results using alternative technologies. In this study, we used one-dimensional Western blot analyses to confirm some of the differential expression results inferred by spectral counting. Three proteins were
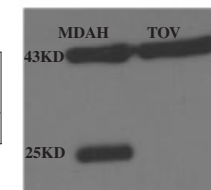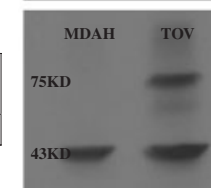
selected from Table 1: UCHL1, Stratifin and MARKS. As can be seen from Fig. 12, the intensities from these three proteins correlate well with the spectral counting results shown in the left panel.

4 Concluding remarks

Proteomic profiles from two ovarian endometrioid-derived cell lines with different genetic mutations have been studied using a shotgun proteomic approach. This involved whole lysate digestion by trypsin with extensive fractionation in the first dimension using cIEF based upon a pH-based separation followed by capillary RP-HPLC. Online analysis was performed using tandem MS acquired by an LTQ. A large number of proteins were identified after filtering through the Peptide Prophet/Protein Prophet TPP. Differentially expressed proteins were quantitated using label-free methods and studied by IPA to reveal the association with important biological functions. It was shown that some important signaling pathways may be highly associated with one of the two cell lines. The PI3K/AKT pathway was found to be significantly predominant in MDAH-2774 but not in TOV-112D. The p53 pathway is shown by network analysis to be important in both cell lines but the network in MDAH-2774 is a more compact one centered on p53 while the network for TOV-112D is more scattered and composed of small networks with ATM, Jnk and GLI1 in the center. The fact that p53 is an important hub of this network

implies that this pathway is a major driving force for differentiation and growth. Other pathways such as estrogen signaling were found to have a stronger connection to TOV-112D than MDAH-2774 and activation of K-ras has been detected in TOV-112D but not in MDAH-2774. Thus, the method described can define the important pathways involved in cancer development and how it may differ between samples. This strategy may be important for biomarker discovery and may lead to development of candidates for drug treatment of disease.

The authors have declared no conflict of interest.

5 References

[1] McCluskey, L. L., Dubeau, L., *Curr. Opin. Oncol.* 1997, *9*, 465–470.

[2] Scully, R. E., Young, R. H., Clement, P. B., *Atlas of Tumor Pathology*, Third Series, Fascicle 23, AFIP, Washington, DC 1998.

[3] Freedman, R. S., Pihl, E., Kusyk, C., Gallager, H. S., Rutledge, F., *Cancer* 1978, *42*, 2352–2359.

[4] Provencher, D. M., Lounis, H., Champoux, L., Tetrault, M., Manderson, E. N., Wang, J. C., Eydoux, P., *In Vitro Cell Dev. Biol. Anim.* 2000, *36*, 357–361.

[5] Wu, R., Zhai, Y., Fearon, E. R., Cho, K. R. *Cancer Res.* 2001, *61*, 8247–8255

[6] Wang, J. H., Hewick, R. M. *Drug Discov. Today* 1999, *4*, 129–133.

[7] Wang, H., Kachman, M. T., Schwartz, D. R., Cho, K. R., Lubman, D. M. *Proteomics* 2004, *4*, 2476–2495.

[8] Gagne, J. P., Gagne, P., Hunter, J. M., Bonicalzi, M. E., *Mol. Cell. Biochemistry* 2005, *275*, 25–55.

[9] Washburn, M. P., Wolters, D., Yates, J. R., *Nat. Biotechnol.* 2001, *19*: 242–247.

[10] Wang, Y., Balgley, B. M., Rudnick, P. A., Evans, E. L., *J. Proteome Res.* 2005. *4*, 36–42.

[11] Zhou, F., Johnston, M. V., *Electrophoresis* 2005, *26*, 1383–1388.

[12] Andreev, V. P., Li, L., Cao, L., Gu, Y., *J. Proteome Res.* 2007, *6*, 2186–2194.

[13] Wong, J. W. H., Sullivan, M. J., Cagney, G., *Brief. Bioinform.* 2008, *9*, 156–165.

[14] Liu, H ., Sadygov, R. G ., Yates, J. R ., *Anal. Chem.* 2004, *76*, 4193–4201.

[15] Silvia, M., Uriarte, D. W., Powell, G. C., Luerman, M. L., *J. Immunol.* 2008. *180*, 5575–5581.

[16] Keller, A., Nesvizhskii, A. I., Kolker, E., Aebersold, R., *Anal. Chem.* 2002, *74*, 5383–5392.

[17] Curmi, P. A., Nogues, C., Lachkar, S., Carelle, N., *Br. J. Cancer* 2000, *82*, 142–150.

[18] Nishio, K., Nakamura, T., Koh, Y., Kanzawa, F., *Cancer* 2001, *91*, 1494–1499.

[19] Chen, Z., *Cancer* 2001, *92*, 3068–3075.

[20] Jacobberger, J. W., Sramkoski, R. M., Zhang, D. S., Zumstein, L.A., *Cytometry* 1999, *38*, 201–213.

[21] Conroy, S. E., Latchman, D. S., *Br. J. Cancer* 1996, *74*, 717–721

[22] Tonin, P. N., Hudson, T. J., Rodier, F., Bossolasco, M., *Oncogene* 2001, *20*, 6617–6626.

[23] Jones, M. B., Krutzsch, H., Shu, H., Zhao, Y., *Proteomics* 2002, *2*, 76–84.

[24] Alaiya, A. A., Franzen, B., Fujioka, K., Moberger, B. *Int. J. Cancer* 1997, *73*, 678–683.

[25] Campbell, I. G., Russell, S. E., Choong, D. Y., Montgomery, K. G. *Cancer Res.* 2004, *64*, 7678–7681.

[26] Nicosia, S. V., Bai, W., Cengu, J. Q., Coppola, D. *Hematol. Oncol. Clin. North Am.* 2003, *17*, 927–943.