

RESEARCH ARTICLE

Use of narrow-range peptide IEF to improve detection of lung adenocarcinoma markers in plasma and pleural effusion

Maria Pernemalm^{1,2}, Luigi De Petris^{1,2}, Hanna Eriksson¹, Eva Brandén³, Hirsh Koyi³, Lena Kanter⁴, Rolf Lewensohn^{1,2} and Janne Lehtiö^{1,2}

¹Karolinska Biomics Center, Stockholm, Sweden

²Department of Oncology-Pathology, Karolinska Institutet, Stockholm, Sweden

³Department of Respiratory Medicine and Allergy, Karolinska University Hospital Solna, Stockholm, Sweden

⁴Division of Pathology, Department of Oncology and Pathology, Karolinska University Hospital Solna, Stockholm, Sweden

In this study we applied narrow-range peptide IEF to plasma or pleural effusion prior to LC/MS/MS. Two methods for narrow-range IEF were run; IPG strips and free-flow electrophoresis. Data from this study was compared with cell line data to evaluate the method performance in body fluids. To test the methods potential in quantitative biomarker discovery studies, plasma and pleural effusion from patients with lung adenocarcinoma ($n = 3$) were compared with inflammatory pleuritis ($n = 3$) using iTRAQ quantification. Using narrow-range IEF on the peptide level we were able to identify and quantify 282 proteins in plasma and 300 proteins in pleural effusion. These body fluid proteomes demonstrated high degree of overlap; however, more proteins significantly differently altered levels related to adenocarcinoma were found in pleural effusion compared with plasma, suggesting enrichment of lung tissue-related proteins in pleural effusion. Nine proteins were chosen for initial validation with Western blot, and one protein (NPC2) was chosen for further validation using immunohistochemistry. Overall, the quantitative results from IEF/LC/MS/MS showed good correlation with the results from Western blot and immunohistochemistry, showing the potential of this methodology in quantitative biomarker discovery studies.

Received: October 17, 2008

Revised: February 27, 2009

Accepted: March 17, 2009

**Keywords:**

Free-flow electrophoresis / Narrow-range peptide IEF / Plasma / Pleural effusion

1 Introduction

To increase the analytical depth in proteomics studies, a vast number of pre-fractionation strategies have been described

Correspondence: Dr. Janne Lehtiö, Karolinska Institute – Oncology/Pathology, Karolinska Biomics center, Z5:02, Karolinska University Hospital, Solna, Stockholm 17176, Sweden

E-mail: janne.lehtio@ki.se

Fax: +46-85-1776099

Abbreviations: FFE, free-flow electrophoresis; FPR, false-positive rate; GO, gene ontology; IHC, immunohistochemistry

in the literature [1, 2]. In human body fluids, the analytical challenges mainly arise from the high dynamic range of concentrations, where some high-abundant proteins completely dominate the protein content, making transient low-abundant biomarkers extremely hard to detect, particularly in a non-targeted discovery phase [3]. The pre-fractionation approach applied in this study is a combination of high-abundant protein depletion and narrow-range IEF at the peptide level, followed by reversed phase liquid chromatography coupled off-line to MALDI TOF/TOF MS. In this approach simple fractionation at the protein level is followed by two orthogonal separation steps at the peptide

level. The benefit of high-abundant protein depletion prior to proteomics analyses has been shown in a previous publication [4].

The rationale behind using narrow-range IEF is that the increase in complexity caused by digestion is reduced. It has previously been shown that approximately 80% of proteins have at least one tryptic peptide in the pH-range 3.5–4.5; hence by analyzing only this *pI* range of the tryptic peptide mixture, the complexity can be reduced without significant loss of proteome coverage [5, 6].

We have previously shown excellent performance of this method, identifying approximately 3700 proteins from a lung cancer cell line [6].

To evaluate the feasibility of the narrow-range peptide IEF/LC/MS/MS approach on body fluids we selected 12 highly complex samples, namely human plasma and pleural effusion samples from patients suffering from lung adenocarcinoma and pleural inflammation. In addition, we compared the data from this study with the cell line study of membrane proteins previously performed in our lab [6].

One way to circumvent the above-described analytical challenge to detect biomarkers in plasma sample is to analyze sample closer to origin of the disease. Non-traumatic pleural effusion is a pathological fluid derived from either inflammation in the chest cavity (high protein content) or passive osmotic diffusion of fluid into the pleural cavity due to left heart failure (low protein content). Inflammatory pleural effusion can be present due to a variety of etiologies ranging from pneumonia to breast cancer metastasis or, more important to this study, direct pleural infiltration from lung cancer tumor cells.

Wide-range IEF of plasma (pH 3–10) on the peptide level has previously been applied using both the Off-Gel™ system [7] and free-flow electrophoresis (FFE) [8].

However, the narrow-range IEF approach described above has, to our knowledge, not previously been applied to either plasma or pleural effusion. In this study, two different narrow-range IEF methods have been applied, IPG strips and FFE. For a recent review on the FFE technology, see Nissum and Foucher [9].

Several proteomics techniques have been applied previously to detect markers of lung cancer using plasma and pleural effusion as starting material [10–15]. However, as of today, no biomarkers have reached the clinic.

The overall aim of this study was to evaluate narrow-range peptide IEF/LC/MS/MS approach on body fluid proteomics. To assess the performance of these methods in quantitative biomarker discovery studies we selected 12 highly complex samples, namely human plasma and pleural effusion samples from patients suffering from lung adenocarcinoma and pleural inflammation. The secondary aim was to compare plasma and pleural effusion proteomes to assess the potential of pleural effusion as source of lung disease biomarkers. For this aim our strategy

was to use plasma and pleural effusion samples from the same patient.

2 Materials and methods

2.1 Human plasma and pleural effusion samples

Patients with pleural effusion, regardless of diagnosis, were asked to participate in the study. All donors signed informed consent and the study was approved by the ethics committee at the Karolinska University Hospital, Stockholm (2003-413 amendment 2005-71-32). Sample collection and storage was performed in line with guidelines presented in [16]. Peripheral venous blood and pleural effusion was collected in EDTA tubes (BD Vacutainer K2E 7.2 mg, BD Diagnostics) and kept at 4°C until preparation to prevent coagulation and minimize protein degradation. From each patient, blood was drawn at the time of pleural effusion. EDTA tubes were centrifuged at 1500 × *g* at 4°C for 10 min. Supernatant was transferred to a new tube and centrifuged at 3000 × *g* at 4°C for 10 min. Supernatant was aliquoted and kept in –80°C until analysis. All samples used in this study were prepared within 1 hour of sample collection and showed no signs of hemolysis.

2.2 Patient selection

Patients included in this study had to fulfill the following primary inclusion criteria to be considered eligible: availability of plasma and pleural effusion samples from the same patient; a diagnosis of inflammatory pleuritis or of primary lung adenocarcinoma that had not previously been treated with specific anticancer drugs; in the lung cancer group, the presence of tumor cells in pleural effusion had to be confirmed by cytological analysis of the cellular pellet; both plasma and pleural effusion samples had to be free from sampling problems, such as hemolysis, coagulation, or sample degradation deriving from improper storage (*e.g.* >12 h at 4°C after sampling). Three patients with primary lung adenocarcinoma (patients 1–3) and three patients with inflammatory pleuritis (patients 4–6) were selected from the biobank. Diagnosis, age, smoking status, and gender of the patients can be found in Supporting Information Table 1.

2.3 Multiple affinity removal system column

Agilent Plasma 7 Multiple Removal System 4.6 × 100 (Agilent Technologies) was set up on an ÄKTA system (GE Healthcare) and run according to manufacturer's instructions. Both plasma and pleural effusion were applied to the column. Flow through was concentrated on 5 kDa molecular weight cut-off filter (Agilent Technologies) followed by buffer exchange >100 times with MilliQ grade water.

2.4 Digestion and iTRAQ labeling

From each patient 100 µg of depleted plasma and 100 µg depleted pleural effusion were taken out and freeze dried in SpeedVac. Freeze dried samples were dissolved in iTRAQ dissolution buffer and digested according to manufacturer's instructions. To ensure complete digestion 2 µg of each digest were taken off and run on a 12% BisTris gel and stained with Silver QuestTM silver staining kit (Invitrogen) (data not shown). Samples were then pooled and labeled with iTRAQ reagent as shown in Supporting Information Table 2. Digestion and labeling was performed twice, once for IPG strips and once for FFE. For IPG strips, 50 µg of individual sample were used for each of the labels and 40 µg were used for FFE.

2.5 Strong cation exchange cleanup

Digests were applied to 1 mL Strata X-C 33 µm polymeric strong cation exchange microcolumns (Phenomenex). The microcolumns were initially washed with 1 mL 100% methanol followed by 1 mL MilliQ grade water. The sample was adjusted to 500 µL 0.1% formic acid and then applied to the columns. After washing with 1 mL 30% methanol and 0.1% formic acid the samples were eluted with 30% methanol and 5% ammonium hydroxide. Samples were then dried in a SpeedVac system.

2.6 Narrow-range IPG strips

The pooled iTRAQ labeled samples (200 µg) were dissolved in 150 µL rehydration solution containing 8 M urea 1% IPG buffer, pH 3.5–5.0 (GE Healthcare). Twenty-four-centimeter 3.5–4.5 linear gradient strips (GE Healthcare) were incubated overnight according to the manufacturer's instructions. Samples were applied to IPG strips by cup-loading at the cathode end and run as described in [17]. After focusing, the strips were cut in 24 pieces starting at the acidic end. Peptides were then eluted in two steps. First, 240 µL 0.1% TFA was added to each piece from the strip and incubated for 2 h on a shaking board. The liquid was collected and 240 µL 0.1% and TFA 50% ACN were added for a second elution step. After 2 h incubation on the shaking board the passive elution solution was then collected and pooled with the extracted peptides from the first elution. Samples were then freeze dried in SpeedVac and kept at –20°C until analysis.

2.7 FFE

The pooled iTRAQ labeled samples (160 µg) were loaded into the FFE and run essentially according to a previously described procedure [18]. A non-linear gradient between 3

and 7.5 was applied and confirmed by measurement of pH (data not shown). Fractions were collected in 96-well plates and kept frozen until analysis. Using the same analytical setup, a tryptic digest of BSA was loaded in triplicate and separated in three separate analyses to assess the reproducibility of the FFE separation.

2.8 LC/MS/MS analysis

Freeze-dried IPG fractions were dissolved in 50 µL 3% ACN and 0.05% TFA. TFA was added to the FFE fractions to a final concentration of 0.05%.

Aliquots of 6.4 and 12.4 µL of IPG fraction 10–19 and FFE fraction B4–C5, respectively, were applied to an Ultimate 3000 HPLC system (LC-Packings) using µL-pickup. An aliquot of 0.05% heptafluorobutyric acid was used as loading solvent as well as transport liquid.

Monolithic trap cartridge, 200 µm × 5 mm PS-DVB (LC-Packings), was used for desalting and concentration and followed by an analytical monolithic column, PS-DVB 200 µm (LC-Packings). The flow rate was set to 1.5 µL/min. Solvent A was 3% ACN and 0.05% v/v TFA and solvent B was 50% ACN 0.04% v/v TFA. Peptides were separated using the following gradient: 0–8 min 0% v/v B, 8–9 min 0–15% v/v B, 9–39 min 15–95% v/v B, 39–45 min 95% v/v B, 45.1–55 min 0% v/v B. The Probot fraction collector was set to collect fractions every 6 s between 13 and 28 min onto a blank MALDI target plate (Applied Biosystems). The eluent was mixed 1:1 v/v post column with 7 mg/mL CHCA (Bio-Rad) in 70% ACN before being spotted onto the MALDI target. Eight external calibration spots with 0.5 µL of standard mix diluted 1:600 (Applied Biosystems) were spotted manually on the plate. All samples were spotted in duplicates.

An AB4800 MALDI TOF-TOF (Applied Biosystems) instrument was used to analyze the samples. Before analysis the plate was externally calibrated in MS and MS/MS mode. Using the 4000 Series Explorer TM Software v.3.5.28193 (Applied Biosystems) a maximum of 15 precursors with s/n over 100 was set to be picked from each spot and 1000 shots in the range 700–4000 m/z were collected for each MS spectrum. MS/MS was performed averaging 3000 shots.

2.9 Protein identification

All MS/MS data obtained were submitted to ProteinPilot [19] (Applied Biosystems) for database searching and iTRAQ reporter ion quantification. The searches were performed against IPI human 3.36 (20071113) protein sequence database, containing 69 012 sequences and 29 002 682 residues, using a 95% confidence cut-off limit and methyl methanethiosulfonate (MMTS) and trypsin specified as digestion enzyme. Biological modifications as well as amino acid substitutions were allowed for in the search.

False-positive rate (FPR) was calculated using an IPI human 3.36 (20071113) target-decoy search strategy as previously described [20], all peptides $\geq 95\%$ confidence were included in the calculations.

2.10 pI calculations

Theoretical pI calculations of the identified peptides were performed using the 3_4pI cal 1.0.0.0 software (Research triangle institute RTP, NC, USA) based on the pI algorithm described in [21], kindly provided by J. L. Stephenson. All non-modified peptides with $\geq 99\%$ confidence were included in the calculations.

2.11 Selection of candidate markers

Average iTRAQ intensities from each reporter ion were normalized against the pooled internal standard. *p*-Value calculations were then performed using two-sided Student's *t*-test, assuming equal variances. Proteins levels, as indicated by iTRAQ reporter ion intensities, between the cancer and non-cancer patients with a *p*-value ≤ 0.05 , were considered as candidate markers.

2.12 Pathway analysis

Localization, pathway analysis, and network interactions for the identified proteins ($\geq 95\%$ confidence) were evaluated using Ingenuity Pathways Analysis. Right-tailed Fischer's exact test was used to calculate a *p*-value for the probability that the pathways assigned to the identified proteins are not due to chance alone (application version 5.5.1, content version 1002, Ingenuity Systems Inc.).

Gene Ontology (GO) annotation comparison was performed for the identified proteins ($\geq 95\%$ confidence) using ProteinCenter (version 1.5.1.1, Proxeon Bioinformatics A/S). ProteinCenter uses corrected *p*-values based on the false discovery rate [22] to calculate overrepresentation of GO terms assigned to any of the compared datasets.

2.13 Western blot analysis

Crude plasma and pleural effusion (40 μ g) were separated on a 12% SDS-PAGE gel (NuPAGE Bis-Tris, Invitrogen) and transferred to a nitrocellulose membrane. After blocking with non-fat milk, the membranes were probed with the following antibodies; Alpha-2-macroglobulin (Abcam ab36995), CLEC3B (Abcam ab51883), Cystatin-C precursor (Upstate 06–458), EFEMP1 (Genetex gtx14926), Gelsolin (Santa Cruz sc57509), Keratin, type II cytoskeletal 8 (Abcam ab53280), NPC2 (Santa Cruz sc30347), SERPINA1 (Abcam ab9400), and VCAM1 (Santa Cruz sc13160) followed by

appropriate HRP-conjugated secondary antibodies (GE Healthcare, Santa Cruz). Ponceau S staining of the membranes (Sigma-Aldrich) was used as loading control and showed equal total protein loading in all Western blot experiments.

2.14 Immunohistochemistry

Formalin-fixed and paraffin embedded tissue samples were processed for immunohistochemistry (IHC). The specimens consisted of tumor samples from 48 patients with lung adenocarcinoma and normal bronchial and lung parenchyma tissue from 14 patients with benign lung diseases.

Antigen retrieval was performed with sodium citrate buffer at pH 6 in microwave for 20 min and the endogenous peroxidase activity was blocked by hydrogen peroxide, 0.5% for 30 min. Sections were incubated overnight with a primary rabbit polyclonal antibody anti human NPC-2 (Atlas Antibodies AB). A biotinylated goat anti-rabbit IgG was used as secondary antibody. The signal from the resulting immune complexes was amplified by applying an avidin–biotin–peroxidase complex. The peroxidase reaction was developed using 3,3-diaminobenzidine for 6 min. IHC stainings were evaluated by a pathologist.

3 Results

In this study, high-abundant protein depletion of plasma and pleural effusion followed by peptide fractionation using narrow-range peptide IEF and quantitative LC/MS/MS analysis was performed. The effect of the narrow-range IEF/LC/MS/MS workflow in plasma and pleural effusion was evaluated in a comparative analysis with cell line data from a previous study [6]. Further, the narrow-range IEF/LC/MS/MS workflow was applied to study plasma and pleural effusion proteome and further to detect potential markers of lung adenocarcinoma. Significantly altered proteins were validated using Western blot and IHC.

3.1 Peptide IEF

Two different peptide IEF approaches, IPG strips and FFE, were used in this study. To evaluate the performance of the pI gradient in the two methods, theoretical pI, for all non-modified peptides with $\geq 99\%$ confidence, was calculated. A plot of the calculated gradients for the two methods, including ± 1 SD is shown in Fig. 1. The average SDs for the two methods are 0.14 pH units for the IPG strip and 0.1 pH units for the FFE separation, showing good performance of the gradient.

To evaluate focusing efficacy the spread of all non-modified peptides with $\geq 99\%$ confidence over the ten fractions was evaluated. As shown in Fig. 2 approximately 80% of the

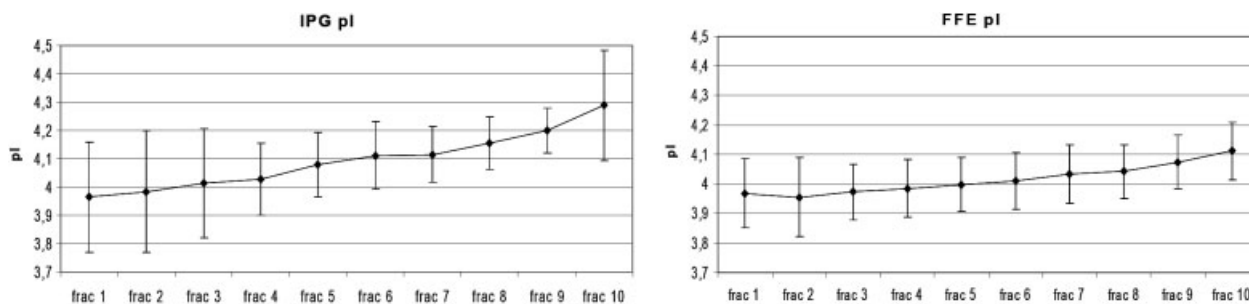


Figure 1. Peptide *pI* distribution on IPG and FFE IEF gradients. Calculated *pI* of all non-redundant peptide sequences with $\geq 99\%$ confidence ± 1 SD is plotted by sample fraction.

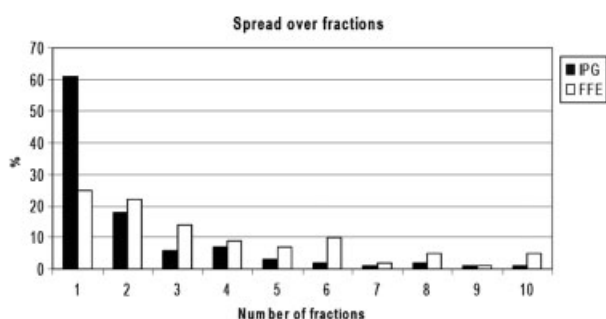


Figure 2. Performance of the focusing of the IPG and FFE systems. Spread of non-redundant peptide sequences $\geq 99\%$ confidence over the IPG (black) and FFE (white) fractions.

peptides from the IPG strip could be found in one or two fractions whereas over 50% of the peptides in the FFE experiment were found in three or more fractions. To evaluate the technical reproducibility of FFE, three repeated BSA digest runs were performed. On average 57% of the peptides were found in one or two fractions. Comparing the focusing efficacy of the three individual runs it was evident that the *pI* spread was larger between the runs than within the runs. When combining the peptides detected in all three individual runs there was a total spread of ± 0.5 pH units *per* fraction (data not shown). Reproducibility and compatibility of iTRAQ-based quantification in narrow-range IPG strips has previously been published [23].

3.2 LC/MS/MS

A summary of the number of identified proteins ($\geq 95\%$ confidence) as well as false positive rate for each search is presented in Table 1. FPRs in these experiments are all below 1%, indicating high-quality data. Number of identified proteins is higher from the IPG strip than from FFE both in the plasma (266 *versus* 101) and in the pleural effusion samples (277 *versus* 130). Merging the data sets from the two different focusing methods 282 proteins were identified from plasma and 300 from pleural effusion. The data demonstrate a high overlap of the proteins identified from the IPG strip and FFE. In general more proteins were identified from pleural effusion, despite the

Table 1. Total number of identified proteins ($\geq 95\%$ confidence)

	IPG	FPR	FFE	FPR	IPG+	FPR
	IPG		FFE		IPG+	IPG+ FFE
	(%)		(%)		(%)	
Plasma	266	0.74	101	0.33	282	0.75
Pleural effusion	277	0.52	130	0.78	300	0.64

False positive rate (FPR) was calculated using target-decoy search strategy including all peptides $\geq 95\%$ confidence.

fact that exactly the same amount of plasma and pleural effusion was used as starting material. A detailed list of all proteins identified in this experiment can be found as Supporting Information data.

3.3 Applicability of IEF/LC/MS/MS workflow for plasma proteomics

We have previously shown the benefit of using narrow-range IEF for detection of soluble protein from a colon cancer cell line (HCT116) [23] and membrane proteins from a lung cancer cell line (H69) [6]. Comparing the number of identified proteins from the present study (in total approximately 300) with the number of identified proteins from the membrane protein analysis (in total approximately 3700) it is obvious that there is a large difference in performance of the IEF/LC/MS/MS workflow. To evaluate what caused this large difference in numbers of protein identifications we carried out a comparison of the performance of the IEF/LC/MS/MS workflow in the two experiments. The comparison was based on the plasma proteomics data from one of the replicates of sample one in this study and the proteomics data from one of the H69 lung cancer cell line triplicates from the replicate experiment previously published by Eriksson *et al.* [6].

First we compared the narrow-range IEF step. In the present study a 24 cm IPG strip with a 3.5–4.5 linear gradient was cut into 24 pieces. The strip used by Eriksson *et al.* was a 24 cm strip with a 3.7–4.9 linear gradient cut into 48 pieces. Ten fractions were applied to LC/MS/MS analysis in both studies.

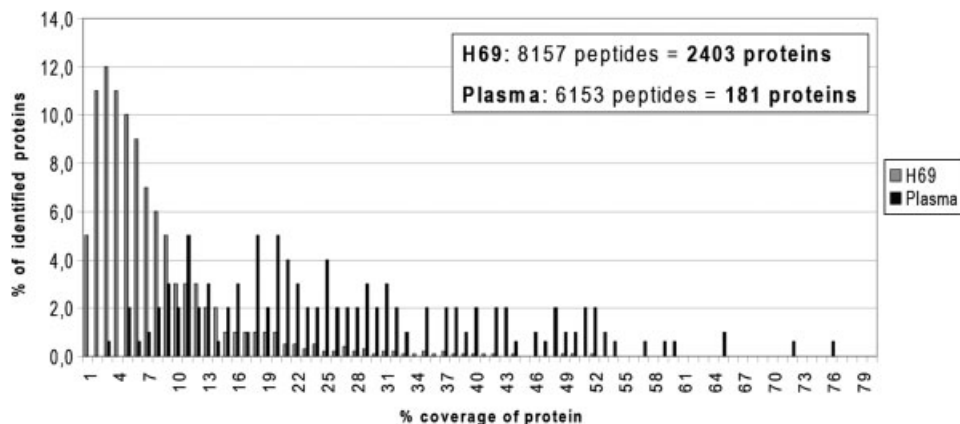


Figure 3. Sequence coverage of proteins identified in H69 cell line (gray) and plasma experiment (black).

In the present study the *pI* spread in the IPG strip was in average SD 0.14 *pI* units, and, in the study by Eriksson *et al.*, in average SD 0.09 *pI* units. To compare the performance of the gradient in the chromatography step in the two experiments, LC/MS heat maps were created for the two experiments (Supporting Information Fig. 1a). Visualized by the LC-MALDI-MS heat maps the two chromatography runs demonstrate equal performance. Comparing the number of peptides and proteins significantly identified in the two experiments 6153 peptides identified 181 proteins in the plasma sample and 8157 peptides identified 2403 proteins in the membrane proteomics experiment. Plotting the sequence coverage of the identified proteins in the two experiments, (Fig. 3) shows that 13% of the proteins in the plasma experiment have less than 10% sequence coverage, whereas in the membrane protein experiment 76% of the proteins have sequence coverage below 10%.

3.4 Comparison plasma and pleural effusion

The sample preparation of the pleural effusion used in this study, *i.e.* collection in EDTA tubes followed by high-abundant protein depletion, has not previously been described. The rationale for using EDTA tubes was to prevent coagulation and protease activity. To evaluate the coagulatory potential of pleural effusion the number of proteins assigned to the coagulation pathway by Ingenuity pathway analysis was investigated. In summary, 21 proteins related to coagulation could be detected in pleural effusion, compared to 22 proteins related to coagulation in plasma (Supporting Information list of all identified proteins).

To further evaluate the overlap in protein content between plasma and pleural effusion the identified proteins ($\geq 95\%$ confidence, merged IPG and FFE data) in the two samples were compared (Fig. 4). Indeed, about two-thirds of the proteins did overlap (189) and 93 and 111 proteins were unique for plasma and pleural effusion, respectively. To explore the characteristics of the proteins in the two sample types Ingenuity pathway analysis (Ingenuity Systems) and ProteinCenter (Proxeon) were used.

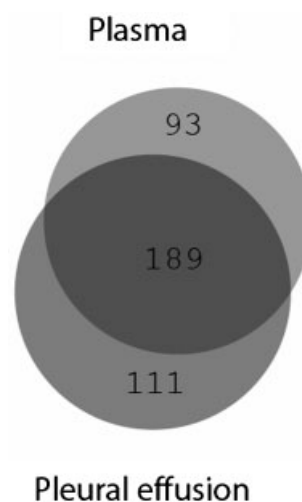


Figure 4. Number of overlapping identified proteins ($\geq 95\%$ confidence) between plasma and pleural effusion. All identified proteins from both FFE and IPG are included in the calculation.

As shown in Fig. 5A the protein classes found in the two sample types, based on annotated localization in Ingenuity pathway analysis, are almost identical. The majority of proteins, about 60% are of extracellular origin. Second most common are cytoplasmic proteins, followed by proteins from the plasma membrane and the nucleus annotations.

Comparing the GO terms assigned to the identified proteins by ProteinCenter no significant differences ($p \geq 0.05$) in molecular functions, cellular components or biological process were found between plasma and pleural effusion demonstrating the similarity of these proteomes.

Since roughly a third of the identified proteins in plasma and pleural effusion are unique for that sample type we decided to further evaluate this particular set of proteins. Analysis of the one-third of proteins that was unique for each sample type show a clear shift in the distribution of annotated localizations, from extracellular to cellular proteins. Together, the cellular proteins make up 60%

of the proteins specific for plasma or pleural effusion (Fig. 5B).

One hypothesis could be that there is a different distribution of concentrations between the proteins shared between plasma and pleural effusion. The analysis of the results from sample 1 (Supporting Information Table 2), where plasma and pleural effusion are pooled and run in the same iTRAQ experiment, demonstrates several large concentration differences between the pooled plasma sample (reporter ion 116) and the pooled pleural effusion sample (reporter ion 117). The top five proteins with increased abundance in plasma and pleural effusion respectively are found in Table 2.

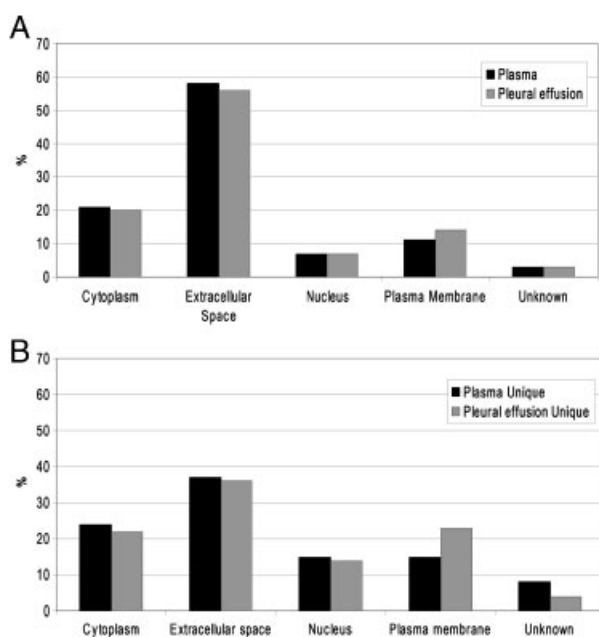


Figure 5. Histograms showing protein localization, of proteins identified from plasma and pleural effusion, as classified by Ingenuity software. (A) Annotated localization of proteins in plasma and pleural effusion given in percent ($\geq 95\%$ confidence, merged IPG and FFE data). (B) Annotated localization of proteins unique for plasma and pleural effusion respectively ($\geq 95\%$ confidence, merged IPG and FFE data).

3.5 Pleuritis versus lung adenocarcinoma

With the aim of evaluating the feasibility of this proteomics workflow for biomarker discovery in lung adenocarcinoma, plasma, and pleural effusion from three patients diagnosed with pleuritis were compared with patients diagnosed with late stage adenocarcinoma. To facilitate the comparison of three *versus* three samples, the iTRAQ labeling approach was applied, using a pooled internal standard (Supporting Information Table 2). After normalizing the reporter ion intensities in the patient samples against the pooled internal standard, the normalized reporter ion intensities were used to find proteins with significantly different expression levels between the two sample groups.

A two-component principal component analysis model based on the data including both plasma and pleural effusion protein levels were performed for patient 1–6. Principal components 1–2 were plotted (Supporting Information Fig. 2). Based on the measured data, there were no outliers in the data set. The patients are spread within a 95% confidence region. Furthermore, the unsupervised principal component analysis showed that there were no grouping of the patients based on the clinical parameters; gender, age, or smoking history. This suggests that the observed differences in protein expression were dependent upon the disease status, namely lung adenocarcinoma or inflammatory pleuritis, and excludes that such results may be biased by the disparity in terms of clinical characteristics observed between the cancer and the control groups.

In total 20 proteins were found to be significantly different between the lung adenocarcinoma and the pleuritis group (Supporting Information Table 3). Out of them, 15 proteins were detected in pleural effusion, four in both plasma and pleural effusion and only one protein was detected exclusively in plasma. Comparing the two focusing techniques, 12 and 4 proteins were only found in either the IPG strips or the FFE, respectively, while three proteins were detected in both IPG and FFE and only one in the merged data set from the two techniques.

3.6 Western blot and IHC

To confirm the identifications of the differently expressed proteins, Western blot analysis was performed. Based on

Table 2. Quantitative differences between proteins in plasma and pleural effusion

Top 5 plasmas	Gene	Fold change	Top 5 pleural effusions	Gene	Fold change
Haptoglobin	HP	4.488	TIMP metalloproteinase inhibitor 1	TIMP1	9.346
Hemoglobin, beta	HBB	4.442	Secreted phosphoprotein 1	SPP1	8.78
Coagulation factor X	F10	2.491	Prostaglandin D2 synthase 21 kDa	PTGDS	8.354
Coagulation factor IX	F9	2.38	Matrix metalloproteinase 2	MMP2	4.912
Serum amyloid p-component	APCS	2.114	Actin, beta	ACTB	4.861

The top five largest quantitative differences among proteins identified from plasma and pleural effusion; based on iTRAQ ratios between pooled plasma (reporter ion 116) and pooled (reporter ion 117) pleural effusion in sample 1 (see Supporting Information Table 2).

literature search, pathway analysis and antibody availability nine proteins (Alpha-2-macroglobulin, CLEC3B, Cystatin-C precursor, EFEMP1, Gelsolin, Keratin, type II cytoskeletal 8, NPC2, SERPINA1, and VCAM1) were chosen. All nine proteins could be detected by Western blot from plasma and pleural effusion, which strengthens the protein identification by ms/ms (Fig. 6). An overview of the success rate of the Western blot analysis can be seen in Fig. 7A. In summary 80% of the quantitative results from the IEF/LC/MS/MS experiment could be confirmed using Western blot. For two of the proteins (Cytokeratin 8 and Cystatin C) the quantitative differences found in the shotgun experiment

could not be confirmed in plasma or pleural effusion and for three proteins (Gelsolin, NPC2, and VCAM1) the differences could only be confirmed in pleural effusion.

One of the proteins that were up-regulated in cancer (NPC2) was chosen for validation using IHC. In normal lung NPC2 was found to be up-regulated in the bronchial cells, showing a cytoplasmic staining of medium intensity with in addition a granular pattern polarized toward the surface of the bronchus. Conversely, the cytoplasm of alveolar cells were negative (Fig. 6B). In the cancer samples NPC2 showed a cytoplasmic staining, with a medium-strong intensity in 75% of cases. In adenocarcinomas with a

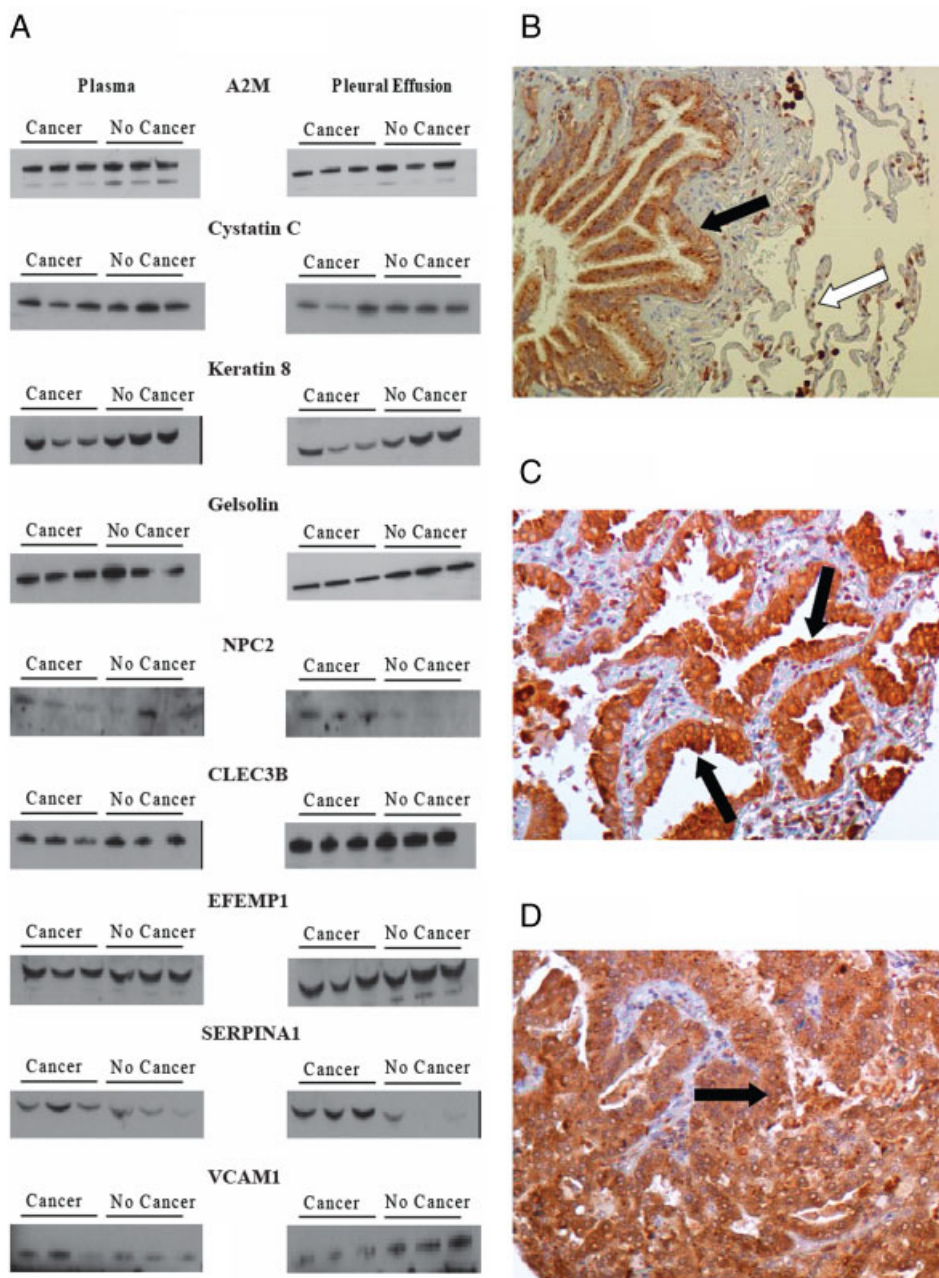


Figure 6. (A) Western blot results from plasma and pleural effusion from same six patients used in the discovery experiment. Lane 1: patient 1, lane 2: patient 2, lane 3: patient 3, lane 4: patient 4, lane 5: patient 5, lane 6: patient 6. (B) Representative IHC staining of NPC2 in normal lung (200 times magnification). In bronchial cells NPC2 show a moderate cytoplasmic expression, and an apical granular pattern (black arrow). Alveolar cells are negative (white arrow). (C and D) A strong cytoplasmic expression of NPC2 is detected in lung adenocarcinoma, with a more prevalent granular pattern in cases with papillary differentiation (arrow in (C)) than in cases with a prevalent glandular differentiation (arrow in (D)) (200 times magnification).

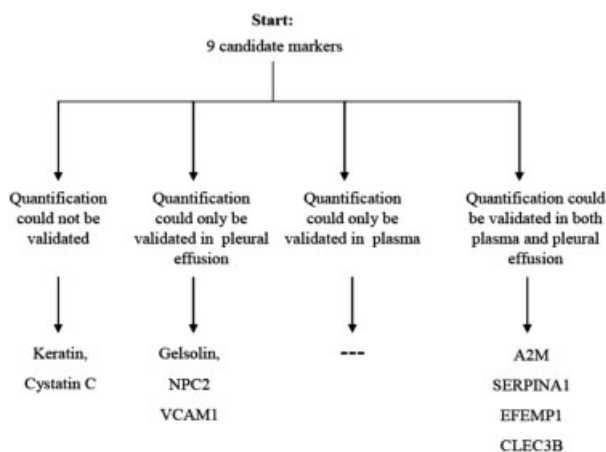


Figure 7. Schematic summary of Western blot validation step.

prevalent papillary differentiation, the staining pattern resembled what observed in bronchial cells, with a much stronger granular expression polarized toward the surface of the papillae (Fig. 6C). In adenocarcinomas with a prevalent glandular differentiation, the granular pattern was less represented, and lacking specific polarization (Fig. 6D).

4 Discussion

In this study we applied a narrow-range peptide IEF/LC/MS/MS approach to assess the feasibility of this method for body fluids proteome analysis. The developed workflow was evaluated using plasma and pleural effusion in relation to lung adenocarcinoma.

In the scope of this study two different methods for narrow-range IEF was used; IPG strips and FFE. One large difference in performance between the FFE and the IPG experiments was the spread over fractions, which was considerably larger in the FFE experiment where over 50% of peptides could be detected in three or more fractions as compared with about 20% using the IPG strip in the current experiment, despite the fact that the IPG strip was manually cut and the fraction collection in the FFE is automated. This phenomenon is probably due to the shallower gradient used in the FFE experiment, where the calculated *pI* averages in the ten fractions spans over approximately 0.1 pH units as compared with 0.3 pH units in the IPG strip, as the *pI* precision in the focusing is approximately equal (average SD 0.1–0.14).

The difference in analytical set up was also reflected in the number of identified proteins from the two experiments. Approximately twice as many proteins were identified from the IPG fractions compared with the FFE fractions, again highlighting the benefit of the wider *pI* range (0.3) at this level of *pI* precision (SD 0.1–0.14).

In addition, the number of identified proteins from the FFE fractions could probably be improved by applying a concentration and desalting step prior to LC-MALDI-TOF/

TOF analysis as the peptides elute in quite large volumes (up to 300 μ L *per* fraction).

To evaluate the benefits of IPG narrow-range peptide IEF/LC/MS/MS, for plasma and pleural effusion, we compared the results from this study with a previous study from our lab applying narrow-range peptide IEF/LC/MS/MS for detection of membrane proteins from the H69 lung cell line [6]. Comparing the IPG *pI* gradients by calculating the theoretical *pI* of the peptides detected in each fraction the average SD was ± 0.14 *pI* units in the present study and ± 0.09 *pI* units in the membrane protein experiment. The slightly smaller spread in the latter study can most likely be assigned to the differences in cutting approach in the two studies; 24 pieces in this study and 48 pieces in the membrane protein study, resulting in a theoretical *pI* range of 0.042 *per* fraction in this study and 0.025 *per* fraction in the membrane protein study. The fact that the *pI* range of the IPG strips, as well as the cutting approach, differed in the two studies could possibly also explain the difference in number of peptides, 6153 in the current experiment compared to 8157 in the previous study. The most pronounced difference between the two methods, however, is the number of identified proteins; 181 from one depleted plasma sample compared with 2403 from one membrane protein enriched cell line sample. In contrast to the results in the cell line sample, the peptides from the plasma sample all map to the same high/mid-abundant proteins in plasma, and not to unique proteins. This is further supported by the difference in protein sequence coverage distribution between the two samples, where 13% of the proteins in the plasma experiments has less than 10% sequence coverage, whereas in the membrane protein experiment 76% of the proteins have sequence coverage below 10%.

This study shows that even after narrow-range IEF body fluids comprise a tough challenge for currently available MS technologies. Similar results have been reported with other plasma fractionation methods, as highlighted in a recent review comparing different affinity fractionation approaches [2].

To our knowledge, this is the first time plasma and pleural effusion from the same patient has been compared. Previous proteomics studies [14, 15] indicate a high level of similarity between plasma and pleural effusion, hence the same sample collection, sample preparation and analytical methods that are applied to plasma should be applicable when analyzing pleural effusion and, in addition, hold the same advantages. The high presence of coagulatory proteins identified in pleural effusion demonstrated in this study further supports this. To compare plasma and pleural effusion we, in this study, collected pleural effusion in EDTA tubes, commonly used for preparation of plasma and prepared pleural effusion following the same protocol. We also applied high-abundant protein depletion to the pleural effusion. Neither of these strategies has previously been described. Analysing the identified proteins in depleted plasma and pleural effusion it is obvious that plasma and pleural effusion indeed are very similar, with an overlap of

60–70% among the identified proteins. Based on these results we suggest the use of EDTA tubes for collection of pleural effusion for proteomics studies.

Somewhat disappointing was the discovery that no higher content of tissue leakage (cellular) proteins could be seen in pleural effusion as compared with plasma. In addition no significant difference in GO terms related to molecular functions, cellular components or biological processes could be found. However, the non-overlapping subset of proteins between the two sample types represents a larger proportion of cellular proteins compared with the entire data set, indicating that the cellular proteins detected in the plasma and pleural effusion samples are indeed different and that they are likely to be derived from different tissues.

One of the benefits with pleural effusion might be that it would contain a higher concentration of tissue-derived proteins compared with plasma. This means that the same proteins could be detected in plasma and pleural effusions but in different concentrations. To evaluate this we looked at the top five fold differences when comparing plasma and pleural effusion. The comparison clearly shows a higher level of classical plasma proteins in plasma as compared with pleural effusion and an enrichment of tissue derived proteins in pleural effusions. This was further supported by the fact that more potential markers linked to the clinical question could be detected in pleural effusion than in plasma. These results are in agreement with previous studies where the potential of using proximal fluids for discovery of biomarkers has been highlighted [24, 25].

In this feasibility study, a total of 20 proteins were found to be significantly altered between the patients diagnosed with pleuritis and those with lung adenocarcinoma. As control group for this study, we selected patients with an acute inflammatory disease of the chest, namely pleuritis. This was intended to avoid the identification of acute phase proteins as false-positive markers of lung cancer, which would very likely have been the case if we had not chosen an inflammatory syndrome as a control.

Out of the 20 differently expressed proteins nine were chosen for initial validation using Western blot on the same sample. The detection of all nine markers in plasma and pleural effusion with Western blot confirms the identification by MS/MS, and indicates a good performance of the MS identification analysis. Further, the quantitative analysis in the IEF/LC/MS/MS showed a high agreement with the Western blot analysis, showing same pattern in 80% of the proteins.

One of the proteins (NPC2) that were found to be up-regulated in patients with adenocarcinoma was chosen for further validation using IHC. The immunostainings show that NPC2 is strongly expressed in lung adenocarcinoma. It is also present in normal bronchus, but its anatomical localization justifies its low expression in pleural effusions from patients with inflammatory disease, because bronchial cells are not at all involved in the inflammatory pleural process. NPC2 is known to be expressed in the liver and to

be involved in cholesterol transport, which could explain why the tumor specific up-regulation seen in pleural effusions and tumor tissue is not seen in plasma [26]. Although it is not clear yet if NPC2 could play a role in determine the tumor phenotype, our observations suggest that the presence of this protein in pleural effusion of patients with lung adenocarcinoma directly derives from tumor cells, further supporting the potential of the proteomics method implemented in this study to effectively explore pleural effusion as a source of cancer biomarkers.

In summary, we show that narrow-range IEF on the peptide level can be used for body fluids proteomics. However, additional protein level fractionation prior to peptide IEF could be tested to exploit full potential of narrow IEF using plasma samples. Further, we show high level of similarity between plasma and pleural effusion, justifying the use of same sample preparation approaches for the two sample types. Applying IEF/LC/MS/MS to plasma and pleural effusion 20 significantly differently expressed proteins could be detected, highlighting the applicability of this methodology in quantitative biomarker discovery studies.

The authors thank Dr. Mikkel Nisum at BD Diagnostics for valuable comments and Christian Obermaier at BD Diagnostics in Martinsried, Germany for the FFE separation in this study. The research leading to these results has received funding from the European Community's Seventh Framework Programme FP7/2007-2011 under grant agreement no. 200767. This work is additionally financially supported by the European Community's FP6/2007-2012 under grant agreement no. 037665, CHEMORES, The Swedish Cancer Society, Stockholm's cancer society (project 061202), Karolinska Institutet (KIB funding) and the Swedish National board of health.

The authors have declared no conflict of interest.

5 References

- [1] Canas, B., Pineiro, C., Calvo, E., Lopez-Ferrer, D., Gallardo, J. M., Trends in sample preparation for classical and second generation proteomics. *J. Chromatogr A* 2007, 1153, 235–258.
- [2] Pernemalm, M., Lewensohn, R., Lehtiö, J., Affinity prefractionation for MS-based plasma proteomics. *Proteomics* 2009, 9, 1420–1427.
- [3] Anderson, N. L., Anderson, N. G., The human plasma proteome: history, character, and diagnostic prospects. *Mol. Cell. Proteomics* 2002, 1, 845–867.
- [4] Pernemalm, M., Orre, L. M., Lengqvist, J., Wikstrom, P. *et al.*, Evaluation of three principally different intact protein prefractionation methods for plasma biomarker discovery. *J. Proteome Res.* 2008, 7, 2712–2722.
- [5] Cargile, B. J., Sevinsky, J. R., Essader, A. S., Stephenson J. L., Jr., Bundy, J. L., Immobilized pH gradient isoelectric focusing as a first-dimension separation in shotgun proteomics. *J. Biomol. Tech.* 2005, 16, 181–189.

- [6] Eriksson, H., Lenggqvist, J., Hedlund, J., Uhlen, K. *et al.*, Quantitative membrane proteomics applying narrow range peptide isoelectric focusing for studies of small cell lung cancer resistance mechanisms. *Proteomics* 2008, 8, 3008–3018.
- [7] Heller, M., Michel, P. E., Morier, P., Crettaz, D. *et al.*, Two-stage off-gel isoelectric focusing: protein followed by peptide fractionation and application to proteome analysis of human plasma. *Electrophoresis* 2005, 26, 1174–1188.
- [8] Nissum, M., Kuhfuss, S., Hauptmann, M., Obermaier, C. *et al.*, Two-dimensional separation of human plasma proteins using iterative free-flow electrophoresis. *Proteomics* 2007, 7, 4218–4227.
- [9] Nissum, M., Foucher, A. L., Analysis of human plasma proteins: a focus on sample collection and separation using free-flow electrophoresis. *Expert Rev. Proteomics* 2008, 5, 571–587.
- [10] Patz E. F., Jr., Campa, M. J., Gottlin, E. B., Kusmartseva, I. *et al.*, Panel of serum biomarkers for the diagnosis of lung cancer. *J. Clin. Oncol.* 2007, 25, 5578–5583.
- [11] Yildiz, P. B., Shyr, Y., Rahman, J. S., Wardwell, N. R. *et al.*, Diagnostic accuracy of MALDI mass spectrometric analysis of unfractionated serum in lung cancer. *J. Thorac. Oncol.* 2007, 2, 893–901.
- [12] Ueda, K., Katagiri, T., Shimada, T., Irie, S. *et al.*, Comparative profiling of serum glycoproteome by sequential purification of glycoproteins and 2-nitrobenzensulfonyl (NBS) stable isotope labeling: a new approach for the novel biomarker discovery for cancer. *J. Proteome Res.* 2007, 6, 3475–3483.
- [13] Soltermann, A., Ossola, R., Kilgus-Hawelski, S., von Eckardstein, A. *et al.*, N-glycoprotein profiling of lung adenocarcinoma pleural effusions by shotgun proteomics. *Cancer* 2008, 114, 124–133.
- [14] Tyan, Y. C., Wu, H. Y., Lai, W. W., Su, W. C., Liao, P. C., Proteomic profiling of human pleural effusion using two-dimensional nano liquid chromatography tandem mass spectrometry. *J. Proteome Res.* 2005, 4, 1274–1286.
- [15] Tyan, Y. C., Wu, H. Y., Su, W. C., Chen, P. W., Liao, P. C., Proteomic analysis of human pleural effusion. *Proteomics* 2005, 5, 1062–1074.
- [16] Schroh, A. S., Wurtz, S., Kohn, E., Banks, R. E. *et al.*, Banking of biological fluids for studies of disease-associated protein biomarkers. *Mol. Cell. Proteomics* 2008, 7, 2061–2066.
- [17] Essader, A. S., Cargile, B. J., Bundy, J. L., Stephenson J. L., Jr., A comparison of immobilized pH gradient isoelectric focusing and strong-cation-exchange chromatography as a first dimension in shotgun proteomics. *Proteomics* 2005, 5, 24–34.
- [18] Malmstrom, J., Lee, H., Nesvizhskii, A. I., Shteynberg, D. *et al.*, Optimized peptide separation and identification for mass spectrometry based proteomics via free-flow electrophoresis. *J. Proteome Res.* 2006, 5, 2241–2249.
- [19] Shilov, I. V., Seymour, S. L., Patel, A. A., Loboda, A. *et al.*, The Paragon Algorithm, a next generation search engine that uses sequence temperature values and feature probabilities to identify peptides from tandem mass spectra. *Mol. Cell. Proteomics* 2007, 6, 1638–1655.
- [20] Elias, J. E., Gygi, S. P., Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat. Methods* 2007, 4, 207–214.
- [21] Stephenson J. L., Jr., Bunger, M. K., Cargile, B. J., Sevensky, J. R., *7th Siena meeting From genome to proteome: Back to the future*, Siena, Italy 2006, p. 65.
- [22] Benjamini, Y., Hochberg, Y., Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser. B* 1995, 57, 289–300.
- [23] Lenggqvist, J., Uhlen, K., Lehtio, J., iTRAQ compatibility of peptide immobilized pH gradient isoelectric focusing. *Proteomics* 2007, 7, 1746–1752.
- [24] Lescuyer, P., Hochstrasser, D., Rabilloud, T., How shall we use the proteomics toolbox for biomarker discovery? *J. Proteome Res.* 2007, 6, 3371–3376.
- [25] Rifai, N., Gillette, M. A., Carr, S. A., Protein biomarker discovery and validation: the long and uncertain path to clinical utility. *Nat. Biotechnol.* 2006, 24, 971–983.
- [26] Infante, R. E., Wang, M. L., Radhakrishnan, A., Kwon, H. J. *et al.*, NPC2 facilitates bidirectional transfer of cholesterol between NPC1 and lipid bilayers, a step in cholesterol egress from lysosomes. *Proc. Natl. Acad. Sci. USA* 2008, 105, 15287–15292.