

RESEARCH

Open Access



Blood-based lung cancer biomarkers identified through proteomic discovery in cancer tissues, cell lines and conditioned medium

Charles E. Birse^{1*}, Robert J. Lagier¹, William FitzHugh¹, Harvey I. Pass², William N. Rom³, Eric S. Edell⁴, Aaron O. Bungum⁴, Fabien Maldonado⁴, James R. Jett⁵, Mehdi Mesri¹, Erin Sult¹, Elizabeth Joseloff¹, Aiqun Li¹, Jenny Heidbrink¹, Gulshan Dhariwal¹, Chad Danis¹, Jennifer L. Tomic¹, Robert J. Bruce¹, Paul A. Moore¹, Tao He¹, Marcia E. Lewis¹ and Steve M. Ruben¹

Abstract

Background: Support for early detection of lung cancer has emerged from the National Lung Screening Trial (NLST), in which low-dose computed tomography (LDCT) screening reduced lung cancer mortality by 20 % relative to chest x-ray. The US Preventive Services Task Force (USPSTF) recently recommended annual screening for the high-risk population, concluding that the benefits (life years gained) outweighed harms (false positive findings, abortive biopsy/surgery, radiation exposure). In making their recommendation, the USPSTF noted that the moderate net benefit of screening was dependent on the resolution of most false-positive results without invasive procedures. Circulating biomarkers may serve as a valuable adjunctive tool to imaging.

Results: We developed a broad-based proteomics discovery program, integrating liquid chromatography/mass spectrometry (LC/MS) analyses of freshly resected lung tumor specimens ($n = 13$), lung cancer cell lines ($n = 17$), and conditioned media collected from tumor cell lines ($n = 7$). To enrich for biomarkers likely to be found at elevated levels in the peripheral circulation of lung cancer patients, proteins were prioritized based on predicted subcellular localization (secreted, cell-membrane associated) and differential expression in disease samples. 179 candidate biomarkers were identified. Several markers selected for further validation showed elevated levels in serum collected from subjects with stage I NSCLC ($n = 94$), relative to healthy smoker controls ($n = 189$). An 8-marker model was developed (TFPI, MDK, OPN, MMP2, TIMP1, CEA, CYFRA 21–1, SCC) which accurately distinguished subjects with lung cancer ($n = 50$) from high risk smokers ($n = 50$) in an independent validation study (AUC = 0.775).

Conclusions: Integrating biomarker discovery from multiple sample types (fresh tissue, cell lines and conditioned medium) has resulted in a diverse repertoire of candidate biomarkers. This unique collection of biomarkers may have clinical utility in lung cancer detection and diagnoses.

Keywords: Lung cancer, Early detection, Biomarker, Mass spectrometry, Proteomics, Discovery

* Correspondence: charles.birse@celera.com

¹Celera employees during the course of these studies, Celera, 1311 Harbor Bay Parkway, Alameda, CA 94502, USA

Full list of author information is available at the end of the article

Background

Lung cancer is the leading cause of cancer mortality in the United States. Estimates for 2014 indicate that 224,210 individuals will be diagnosed with lung cancer and 159,260 will die from the disease [1]. The average 5-year survival is about 17 %, with 79 % of cases being diagnosed as regional or distant disease. If lung cancer is detected when localized, survival increases to over 50 % [1].

Support for early lung cancer detection has emerged from the landmark NLST, where LDCT screening was shown to confer a 20 % reduction in lung cancer mortality in a high risk population [2]. Despite concerns associated with the low specificity (73.4 %) of CT screening [3] and the resulting large number of false-positive findings for lung cancer (96.4 %), the USPSTF recently recommended annual LDCT-screening for lung cancer in high-risk individuals [4, 5]. In their recommendation statement, the USPSTF stressed the need for more research into the use of biomarkers to complement LDCT screening. Two key clinical opportunities exist. First, the use of biomarkers for early detection of lung cancer could define a new high-risk population or refine the screening criteria recommended by USPSTF (age: 55 to 80 years, smoking history: >30 pack-year). Such biomarkers would serve as a pre-imaging filter, reducing the overall cost of screening and lowering the number of false-positive findings and unnecessary follow-up procedures. The second opportunity lies in improving the accuracy of lung cancer diagnosis. Given the high frequency of positive findings (pulmonary nodules) with CT screening [2], new means of accurately determining malignant risk are urgently required. In the NLST, 24 % of surgically resected nodules were found to be benign [2]. By improving the accuracy with which malignant risk is determined, biomarkers could potentially enhance diagnostic management by reducing unnecessary surgical intervention, minimizing the use of costly PET-CT and lowering radiation exposure associated with CT monitoring, while enabling detection of lung cancer at an early, more curable, stage.

A wide variety of approaches have been utilized to discover new blood-based lung cancer protein biomarkers [6]. These range from splice variant analysis and the isolation of tumor-enriched transcripts [7], to the development of novel proteomic platforms with the capacity to resolve candidate markers in a highly multiplexed fashion [8]. Advances in mass spectrometry (MS)-based technologies have also enabled discovery of new lung cancer biomarker candidates directly in serum or plasma [9–13]. While the identification of biomarkers directly in blood-based matrices can be problematic due to their complexity and the presence of multiple highly abundant factors [14], some of these challenges can be minimized through extensive fractionation [15]. Differentially expressed candidate markers have also been successfully identified through comparison

of blood draining from the tumor vascular bed matched with systemic arterial blood from the same patient [16].

Alternative, “indirect” MS approaches have also been successfully employed, wherein candidate markers initially identified in lung cancer tissue specimens, cell lines or conditioned medium, have subsequently been shown to be differentially expressed in the peripheral circulation using immunoassay-based methodologies. Pioneering discovery studies employed conditioned medium derived from the lung cancer cell line A549 or cell and organ cultures, followed by confirmation of expression profiles in serum and plasma [17, 18]. Thereafter, detailed analysis of conditioned medium collected from multiple lung cancer lines revealed a novel collection of candidate biomarkers [19]. More recently, subcellular fractionation and organelle isolation from freshly collected tissue specimens has enabled further candidate discovery, with verification achieved through multiple reaction monitoring (MRM) [20].

We have expanded on these approaches, broadening the scope of biomarker discovery by performing proteomic analyses across multiple types of specimens: freshly resected lung tissues, cancer cell lines and conditioned medium, enabling the discovery of a diverse collection of candidate markers. We have confirmed disease-enriched profiles for several of these candidates in sera collected from patients with early-stage disease. Moreover, a multi-marker model has been assembled which accurately distinguishing patients with NSCLC from smokers with no known malignancies. These studies suggest that the integration of multiple indirect discovery approaches may serve as a valuable means of identifying novel blood-based biomarkers that may be employed in the early detection and diagnosis of lung cancer.

Results

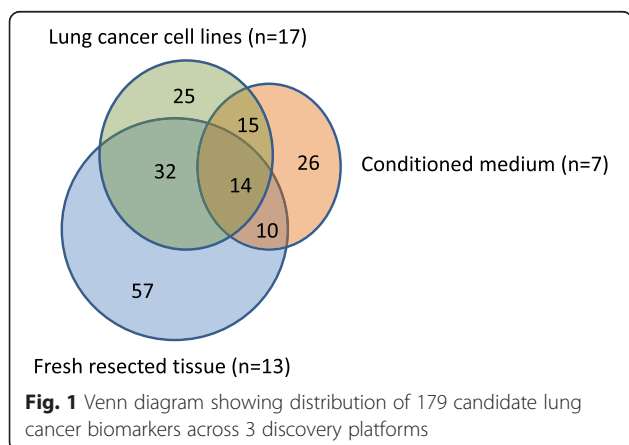
Tissue/cell-line based biomarker discovery

Three distinct LC/MS-based discovery programs were established to identify a diverse spectrum of candidate biomarkers which could serve as the basis for a blood-based immunoassay for detection or diagnosis of lung cancer. To enrich for markers destined to be found in the peripheral circulation of lung cancer patients, discovery focused on glycoproteins predicted to be located either at the cell membrane or secreted/shed from lung cancer cells. Cell-membrane discovery was performed in two distinct sample types: freshly resected tissue specimens ($n = 13$), and a collection of lung cancer cell lines ($n = 17$). The clinical specimens and cell lines studied provide broad coverage of tumor stage and prevalent histological cell types (Additional file 1: Table S1 and Additional file 2: Table S2). To focus discovery on differentially expressed candidate markers, peptide levels measured in surgically resected malignant samples were compared with adjacent

normal tissue. Expression in lung cancer cell lines was analyzed relative to the non-cancerous immortalized lung epithelial cell line Beas-2B [21]. A third discovery program, which served to complement the cell-membrane analyses, resolved proteins secreted or shed from lung cancer cells into conditioned medium, in a subset of lines amenable to growth in serum-free conditions: A549, H1299, H358, H522, H2291, H520 and Calu-1.

Candidate lung cancer markers were prioritized based on MS data and predicted subcellular localization. To be selected, proteins had to: (i) be represented by multiple differentially expressed peptides ($n > 1$); (ii) be identified in multiple malignant samples ($n > 1$) and (iii) exhibit elevated expression in lung cancer specimens, with a cancer: control expression ratio of > 4.0 . Candidate biomarkers were also prioritized based on secondary structure, with proteins predicted to be associated with either the cell membrane or secreted from the cell being selected; these two compartments are enriched with markers destined for the peripheral circulation. 179 candidate markers were identified which met these criteria (Additional file 3: Table S3).

Each of the cellular systems employed in these studies yielded a large number of candidate biomarkers: fresh resected tissues ($n = 113$), cell lines ($n = 86$) and conditioned medium ($n = 65$). While a small proportion of these biomarkers were identified in all three sample types ($n = 14/179$, 8 %), the majority were uniquely resolved in only one of the three cellular systems ($n = 108/179$, 60 %), highlighting the value of the multi-faceted approach (Fig. 1). 29 markers were discovered in both conditioned medium and cell-membrane preparations derived from lung cancer cell lines. Interestingly, a subset of these markers ($n = 9/29$, 31 %), was resolved in the same cell lines used for both membrane-bound and extracellular protein discovery. The overlapping detection of 9 markers in membrane-bound and secreted/shed preparations suggests multiple forms of these proteins may be expressed at elevated levels in NSCLC.



The Panther based classification system was used to categorize markers based on Protein Class and Pathway [22, 23]. Protein classes were defined for 141/179 (79 %) of the candidate markers evaluated. The most common classes reported were: receptors (14 %), cell adhesion molecules (14 %), hydrolases (13 %), defense/immunity proteins (10 %), proteases (9 %), enzyme modulators (8 %) and signaling molecules (8 %; Additional file 4: Figure S1). Further protein class analysis revealed similar profiles for biomarkers identified in the two cell-surface based discovery programs, resected tissue and cultured lung cancer cell lines (Additional file 5: Figure S2). Panther classification resolved protein categories for 91/113 (81 %) of the markers identified in tissues and 70/86 (81 %) of those found in cell lines. While some differences clearly exist, the most abundant protein classes (cell adhesion, defense/immunity, enzyme modulator, extracellular matrix, hydrolase, protease, receptor, signaling, transfer/carrier and transporter) were resolved in both tissues and cell lines. Panther-based pathway analysis also revealed many similarities between the two discovery platforms. Pathways commonly identified in resected tissues (integrin signaling, inflammation, gonadotropin releasing hormone receptor, Alzheimer disease-presenilin and plasminogen activating cascade) were also frequently found in the cell lines studied. Some differences were resolved between the two sources, including enrichment of blood-coagulation related proteins in the tissue based discovery system (22 %) relative to cell line studies (9 %; Additional file 6: Figure S3).

Serum-based biomarker verification

ELISA analysis was undertaken to investigate whether the differential expression profiles observed in lung cancer tissues, cell lines and conditioned medium, would also be detected in the bloodstream of subjects with lung cancer. A small number of candidates were selected for serological characterization: CEA, MDK, MMP2, SLPI, TFPI and TIMP1 (Table 1). These biomarkers were selected in part due to the reagent availability, but also, with the exception of CEA, because they represented some of the more novel lung cancer markers identified, with few studies indicating elevated expression in the circulation of patients with early stage disease [24]. While all six markers had been shown to be present in plasma [25], they had not been resolved in other proteomic studies aimed at identifying differentially expressed lung cancer markers using alternative biological fluids: bronchial lavage [26, 27] sputum [28] or pleural fluid [29, 30], or in profiling experiments aimed at identifying markers associated with other common lung disorders: COPD [27], asthma [31] or tuberculosis [32].

With the goal of identifying markers to be used to screen for early-stage disease, or to guide diagnosis following CT-based detection, expression levels were

Table 1 Candidate lung cancer biomarkers identified through MS discovery that were selected for serological characterization

| Protein | Differentially expressed peptides | Number of samples where differentially expressed peptides observed | | | Differential expression (Lung cancer/control) | Predicted subcellular localization |
|---------------|-----------------------------------|--|---------------------|----------------------------|---|------------------------------------|
| | | Tissues (n = 13) | Cell lines (n = 17) | Conditioned medium (n = 7) | | |
| CEA (CEACAM5) | CETQNPVSAR | 9 | 1 | | 20.4 | Cell Membrane |
| | TLTLFNVTRNDTASYK | | | | | |
| | YNAQCQETIR | | | | | |
| | EGTCGAQTQR | | | | | |
| MDK | FENWGACDGGTGTK | | | 7 | 4.9 | Secreted |
| | VTKPCTPK | | | | | |
| | YNAQCQETIRVTKPCTPK | | | | | |
| | ESCNLFVLK | | | | | |
| MMP2 | TDKELAVQYLNTFYGCPK | | | | 21.8 | Secreted |
| | ELAVQYLNTFYGCPK | | | | | |
| | CGNPDVANYNFFPR | | | | | |
| | YGFCPHEALFTMGGNAEQPCK | | | | | |
| SLPI | SCVSPVKA | 3 | | | 5.5 | Secreted |
| | AGVCPPK | | | | | |
| | AGVCPPKK | | | | | |
| | ADDGPCK | | | | | |
| TFPI | QCEEFYGGCEGNQNR | | 1 | 3 | 10.0 | Secreted |
| | TTLQKEKPDFCFLEEDPGICR | | | | | |
| | YFYNNQTK | | | | | |
| | FVYTPAMESVCGYFHR | | | | | |
| TIMP1 | HLACLPR | | | | | |
| | LQSGTHCLWTDQLLQGSEK | | | | | |
| | LQDGLLHITTCFVAPWNSLSLAQR | | | | | |
| | FVGTPEVNQTTLYQRYEIK | | | | | |
| TIMP1 | AKFVGTPEVNQTTLYQR | 11 | 1 | 4 | 34.6 | Secreted |
| | FVGTPEVNQTTLYQR | | | | | |
| | SHNRSEEFLIAGK | | | | | |
| | EPGLCTWQSLR | | | | | |

Differentially expressed peptides resolved for each marker are listed together with the number of samples (tissues, cell lines or conditioned medium) where peptides with elevated expression in lung cancer were identified. Median ratio represents the overall level of elevated expression combining levels observed for each differentially expressed peptide in disease samples relative to appropriate controls

determined in subjects with stage I NSCLC ($n = 94$), relative to normal smoker controls ($n = 189$; Table 2). In an effort to minimize selection of markers associated with pre-analytical variability, where differential expression profiles may be derived from serum sample collection procedures specific to any single clinical study site, subjects from two independent clinical studies were combined into a single testing set. The first study collected at CRCCC (Clinical Research Center of Cape Cod; West Yarmouth, MA), comprised patients with stage I NSCLC ($n = 30$) and healthy smoker controls ($n = 99$). The second cohort, collected at New

York University (NYU) School of Medicine/Langone Medical Center, was selected from a high-risk population with a history of heavy tobacco usage. Serum samples were collected from patients with stage I NSCLC ($n = 64$) and healthy controls ($n = 90$).

Levels of five of the six candidate biomarkers tested (CEA, MDK, MMP2, SLPI, TIMP1, TFPI) were significantly higher in serum from subjects with NSCLC than in controls (Table 3, Additional file 7: Figure S4), serving to support this indirect discovery approach. Three extensively characterized markers: CYFRA 21-1, SCC and OPN were also evaluated. These markers served as a

Table 2 Demographic and clinical profiles of subjects tested with lung cancer biomarker candidates

| | Serum verification/model training | | Model testing | |
|--------------------|-----------------------------------|-------------------|----------------------|-------------------|
| | Controls (N = 189) | Cases (N = 94) | Controls (N = 50) | Cases (N = 50) |
| Gender | | | | |
| Male | 109 | 33 | 28 | 28 |
| Female | 80 | 61 | 22 | 22 |
| Age | | | | |
| Mean (SD) | 62.1 (11.8) | 66.6 (9.6) | 63.0 (6.4) | 65.6 (6.7) |
| Smoking Pack Years | | | | |
| Mean (SD) | 37.6 (21.7) | 43.9 (20.6) | 54.3 (22.4) | 54.2 (22.2) |
| Benign Nodules (n) | | | 22 | |
| Lesion Size (cm) | | | | |
| Mean (range) | | | 0.5 (0.2-1.2) | 3.3 (0.8-12) |
| Stage | | | | |
| I (%) | | 94 (100) | | 18 (36) |
| II (%) | | | | 6 (12) |
| III (%) | | | | 16 (32) |
| IV (%) | | | | 5 (10) |
| NA (%) | | | | 5 (10) |
| Histology | | | | |
| Adenocarcinoma (%) | | 63 (67.0) | | 23 (46) |
| BAC (%) | | 4 (4.3) | | |
| Large Cell (%) | | 6 (6.4) | | |
| NSCLC (%) | | 4 (4.3) | | 6 (12) |
| Neuroendocrine (%) | | 2 (2.1) | | |
| Small Cell (%) | | | | 4 (8) |
| Squamous Cell (%) | | 15 (16.0) | | 17 (34) |

Table 3 Expression levels of biomarker candidates in serum collected from patients with NSCLC (*n* = 94) and healthy volunteer controls (*n* = 189)

| | Control Median: ng/mL (Interquartile range) | Case Median: ng/mL (Interquartile range) | KS test (p-val) | AUC |
|---------------|---|--|--------------------|-------|
| CEA (CEACAM5) | 1.65 (0.85-2.92) | 2.68 (1.85-4.90) | <0.0001 | 0.706 |
| MDK | 0.15 (0.04-0.35) | 0.43 (0.20-0.66) | <0.0001 | 0.714 |
| MMP2 | 207 (184-234) | 207 (171-254) | 0.1472 | 0.492 |
| SLPI | 39.6 (34.8-46.2) | 43.3 (35.5-54.5) | 0.0036 | 0.595 |
| TFPI | 39.7 (25.6-55.1) | 54.1 (29.1-70.3) | <0.0001 | 0.617 |
| TIMP1 | 302 (269-346) | 361 (306-440) | <0.0001 | 0.692 |
| CYFRA 21-1 | 0.58 (0.00-1.05) | 1.60 (0.91-3.00) | <0.0001 | 0.816 |
| OPN | 19.3 (10.0-31.0) | 31.4 (16.7-52.4) | <0.0001 | 0.666 |
| SCC | 0.58 (0.34-0.93) | 1.21 (0.55-1.70) | <0.0001 | 0.696 |

Median levels (ng/mL) and interquartile range are shown. The Kolmogorov–Smirnov test (KS test) was used to compare patient groups. Area under the curve (AUC) calculated from receiver operator characteristic (ROC) curve analysis. Markers in the upper section of the table (*n* = 6) represent proteins resolved through MS analysis. The lower section (*n* = 3) represents well-characterized lung cancer biomarkers

reference in evaluating clinical accuracy of the MS-identified markers.

Multi-marker model development and testing

The identification of multiple differentially expressed markers prompted the development of a multi-marker panel. Elastic net modeling [33] started with all 9 candidate markers (Table 3). The optimal value of the regularization parameter, as determined by bootstrap resampling, reduced the parameter estimate for SLPI to zero, while the remaining 8 markers: TFPI, MDK, OPN, MMP2, TIMP1, CEA, CYFRA 21-1 and SCC, which retained non-zero coefficients, were selected in the final model. In the training dataset (Table 2), this 8-marker model resolved lung cancer patients from smoker controls with 75 % sensitivity at 90 % specificity (AUC = 0.913). A bootstrap validation procedure confirmed clinical performance of the model, AUC = 0.903.

The accuracy of the 8-marker model was tested in an independent study (Mayo Clinic). Controls ($n = 50$) were selected from the high risk control population evaluated in the Mayo CT-Screening Trial [34] and included subjects with pulmonary nodules ($n = 22$). Lung cancer cases were pre-operative surgical referrals ($n = 50$). Malignant lesions were significantly larger than screen detected benign nodules. Cases and controls were matched on age, gender and smoking history (Table 2). EDTA plasma samples were utilized in this study. Levels of all markers included in the model had been shown to be highly correlated in serum and EDTA plasma (Additional file 8: Table S4). The 8-marker model distinguished patients with malignant lesions from all smoker controls with an AUC = 0.775 (Fig. 2), accurately classifying control subjects with (AUC = 0.745) or without pulmonary nodules (AUC = 0.799).

While the 8-marker model was found to be substantially correlated with nodule size ($r = 0.739$; $p < 0.0001$), it was not associated with any of the other clinicopathological variables tested: age, sex, smoking history (unpublished data). Elevated expression of the multi-marker model was observed in tumors with a squamous cell histology, relative to adenocarcinoma cases ($p = 0.019$), driven in part by higher levels of CYFRA 21-1 ($p < 0.0001$) and OPN ($p = 0.013$) in squamous cell carcinomas (unpublished data).

Discussion

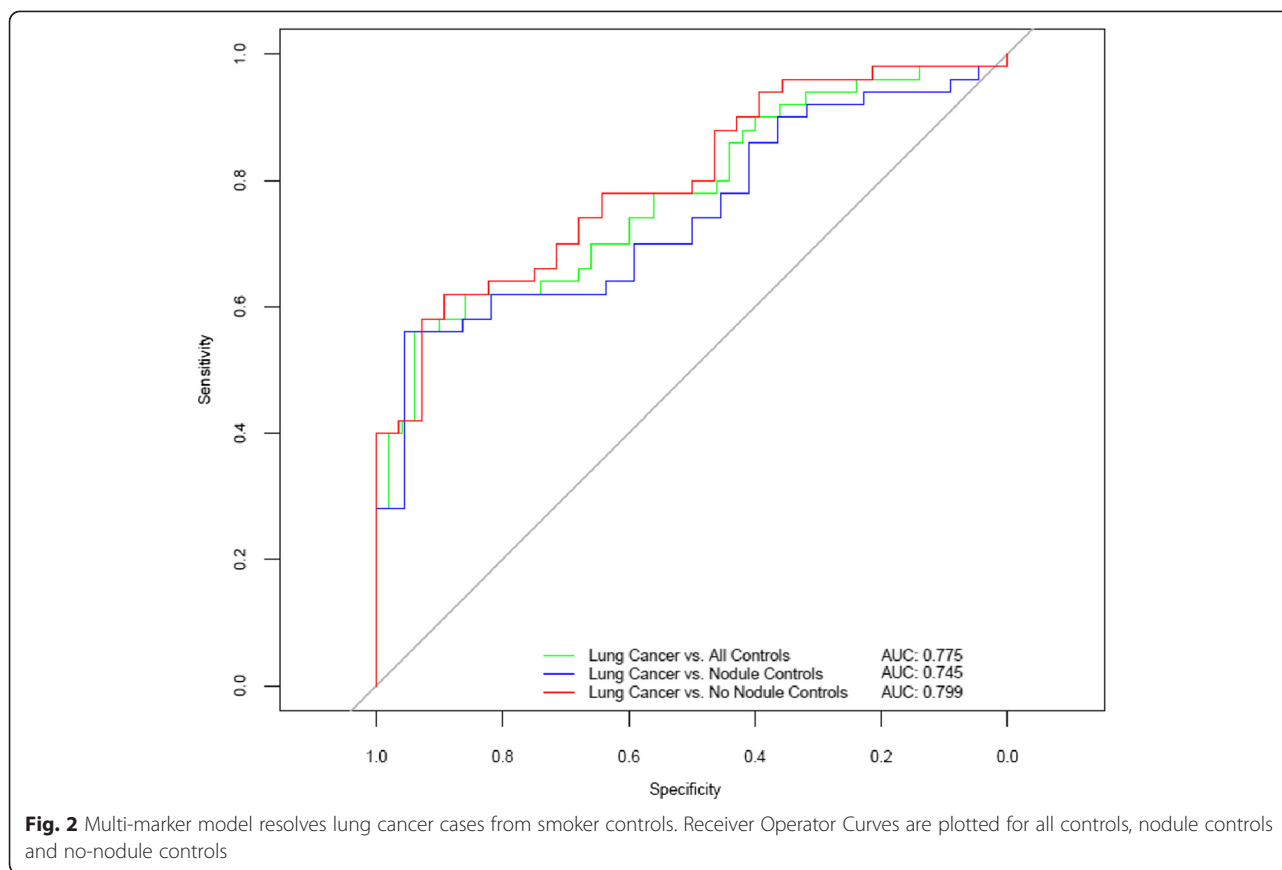
LDCT screening of high-risk smokers has been shown to reduce lung cancer mortality by 20 %, relative to chest radiography. However, of the 24.2 % of participants with an abnormal screening test, the vast majority (96.4 %) were false positives for lung cancer. The low positive predictive value of LDCT results in (i) higher screening costs and (ii) unnecessary invasive procedures for benign disease. Non-invasive biomarkers are urgently needed to

improve LDCT-based screening. Biomarkers could be used to refine the high-risk population, thus limiting the number of individuals being screened by LDCT. Alternatively, biomarkers could be employed following screening, to distinguish relatively rare malignant nodules from commonly found benign nodules. A number of novel blood-based markers have recently been characterized [7, 35], some of which have been evaluated in the form of multi-marker panels [15, 36–38]. However, to date, very few have been shown to add value to clinical variables already being employed in evaluating malignant risk [39].

Marker discovery in blood-based systems (serum and plasma) has been hampered by the complexity of these matrices and presence of multiple highly abundant proteins. Alternative “indirect” approaches have successfully been applied to: freshly resected clinical specimens, primary cultures, cell lines cultured *in vitro* and *in vivo* and conditioned medium, with collections of candidate biomarkers identified in each. However, as each of these studies has been performed in isolation, it has been difficult to evaluate the relative merits of each of these approaches. We report, for the first time, a discovery approach that combines multiple cellular systems: resected tissues, cultured cell lines and conditioned medium. In so doing, we have identified a number of markers commonly resolved across the platforms (Fig. 1). It is noteworthy that while significant overlap across the systems clearly exists, with similar signaling pathways apparently activated across the different discovery systems (Additional file 6: Figure S3), the majority of candidate markers were identified in only one of the three programs. Integrating discoveries from the three systems has not only served as a starting point to understand the relative merits of these distinct approaches, but has also produced a diverse pool of candidate markers for future validation.

All 179 candidate markers selected exhibited at least a four-fold increased level of expression in lung cancer samples relative to appropriate controls. While this 4.0X cut point provided a simple means of identifying the most differentially expressed biomarkers, additional approaches using different cut-points and possibly integrating key clinical variables such as histology and stage, will likely reveal a more extensive collection of candidates for future studies.

Analyses of the glycoproteins residing at the cell surface, in both tissues and cell lines, enabled discovery of cell-membrane markers that may be shed and released into the peripheral circulation. CEA (CEACAM5) provides an example of this type of cell-surface marker, as it is shed into the bloodstream and detected at elevated levels in a wide variety of malignant disorders [40]. While the molecular mechanism responsible for shedding remains unclear [41], CEA is a widely employed serum biomarker used in prognosis, staging and monitoring of colorectal



cancer. In addition to CEA, several other cell-surface markers known to be shed into the circulation were identified in these studies including: MET (c-Met proto-oncogene product, hepatocyte growth factor receptor) [42], mesothelin [43], EPCAM [44], and ICAM-1 [45]. It is noteworthy that many additional cell-membrane markers, with similar secondary structures, were also resolved (Additional file 3: Table S3) and may serve as a valuable pool of candidate markers for future studies.

While CEA (CEACAM5) represents a well-characterized tumor biomarker, the association of other biomarkers with lung cancer varies considerably, with limited evidence of differential expression in early-stage disease. Increased activity of TFPI, a Kunitz-type serine protease inhibitor, has been reported in the circulation of patients with late-stage NSCLC [46]. MDK appears to play a role in both angiogenesis [47] and lung cancer metastasis [48]. Elevated levels of MDK, a heparin-binding growth factor, have been observed in serum collected from patients with a broad range of solid tumors, including lung cancer [49]. A number of tumor-stimulating functions have been demonstrated for TIMP1 [50, 51]. Elevated levels of this metalloproteinase have been observed in serum collected from subjects with late-stage NSCLC [52], with the highest levels reported in squamous cell carcinoma [53]. SLPI, a

member of the Kazal superfamily of serine-proteinases, appears to play a role in tumor growth and metastasis [54–57]. Elevated levels of SLPI protein have been observed in the bloodstream of patients with NSCLC [24, 58]. While the matrix metalloproteinase MMP2 appears to play a role in lung cancer growth and migration [59–61], studies investigating levels of MMP2 in the bloodstream have reported inconsistent findings [62–64]. The diverse range of biological functions observed for markers identified in these studies are summarized in Table 4.

Our ELISA-based serological studies evaluated six candidate markers identified in these LC/MS analyses, along with three additional markers (CYFRA 21-1, SCC and OPN) that served as a benchmark of clinical accuracy. Two of these markers, CYFRA 21-1 and SCC, would not have been predicted to be resolved in these LC/MS studies as they lack an N-linked glycosylation site required for selection in the glycoprotein enrichment procedure. In contrast, two N-linked glycosylation sites are present in the mature form of the secreted protein OPN; as such, we would have expected peptides derived from this marker to be identified in these studies. It is unclear why OPN was not resolved; however it is possible that this marker may not have been differentially expressed in the samples analyzed, or that OPN-derived peptides may have been masked in the LC/MS separation.

Table 4 Biological function of markers identified through mass spectroscopy that were selected for further validation

| Gene name | Protein name | Protein ID (UniProtKB) | Function |
|------------------|--|------------------------|--|
| CEA (CEACAM5) | Carcinoembryonic antigen-related cell adhesion molecule 5 | P06731 | Oncofetal glycoprotein not typically detected in adults. Plays role in cell adhesion and intracellular signaling [70, 71]. |
| MDK | Midkine | P21741 | Heparin binding cytokine promotes cellular transformation, angiogenesis and metastasis [47, 48]. |
| MMP2 | 72 kDa type IV collagenase (Matrix metalloproteinase-2) | P08253 | Degrades extracellular matrix, associated with tissue invasion, cell-induced angiogenesis and tumor growth and metastasis [59–61]. |
| SLPI | Antileukoproteinase (Secretory leukocyte protease inhibitor) | P03973 | Protects epithelial cells from serine proteases, promotes tumorigenic and metastatic pathways [54–57]. |
| TFPI | Tissue factor pathway inhibitor | P10646 | Serine protease inhibitor involved in clotting homeostasis [72]. |
| TIMP1 | Tissue inhibitor of metalloproteinase 1 | P01033 | Inactivates metalloproteinases by binding to zinc cofactor. Promotes proliferation and inhibits apoptosis [50, 51]. |

The ability of the multi-marker model to distinguish lung cancer cases from control subjects with or without nodules indicates potential roles for the test, either as an adjunct to CT- screening, in determining risk of malignancy of pulmonary nodules, or in early lung cancer screening. Clearly additional studies are required to better characterize clinical performance of the current model and to evaluate of larger numbers of candidate biomarkers revealed in this study.

Conclusions

Given the low PPV (4 %) of LDCT in screening the high risk population, there is a pressing need to discover non-invasive biomarkers to complement radiographic imaging in lung cancer screening and diagnosis.

We describe a broad-based discovery platform that has enabled the identification of a large, diverse collection of candidate lung cancer biomarkers. A subset of these markers identified “indirectly” in freshly resected tissue, cell lines and conditioned medium retained elevated cancer-associated expression profiles in the circulation of patients with early-stage disease. A multi-marker model was developed which accurately distinguished lung cancer cases from high risk smokers. This unique collection of markers should serve as a valuable resource for future clinical validation studies.

Methods

Tissue specimens

Freshly resected lung specimens (malignant lesions and normal adjacent tissue) were collected from 4 clinical sites using IRB-approved protocols: 1. Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA; 2. Division of Thoracic Surgery, University of Maryland Medical Center, Baltimore, MD; 3. Department of Cardiothoracic Surgery, George Washington University, Washington DC; 4. Asterand, Detroit, MI. To enrich for samples likely to produce strong MS signal, tissue specimens with a mass of at least 1 g were selected for this study. Single cell suspensions were prepared from each

resected sample using a standard methodology before removal of red blood cells through addition of ACK lysis buffer [65]. Epithelial (EpCAM), leukocyte (CD45) content and cellular viability (PI exclusion) were determined by flow cytometry analysis (LSR I, BD Biosciences, San Jose, CA). Epithelial enrichment was undertaken using flow cytometry based cell sorting (EpCAM, Clone EBA-1, BD Biosciences). Samples yielding a minimum of 1×10^6 viable epithelial cells were submitted for MS analysis.

Cell lines and tissue culture

Lung cancer cell lines obtained from American Type Culture Collection (ATCC, Manassas, VA) or European Collection of Cell Cultures (ECACC, Salisbury, UK) were cultured in the appropriate media as recommended.

Conditioned medium

Cell-lines were cultured to 70 % confluence, transferred to protein free media (293 SFM II, Invitrogen, Carlsbad, CA) for 72 h, after which cell debris was removed by centrifugation.

Blood-based studies

Serum: A total of 283 subjects were evaluated in the verification/model training study, healthy smoker controls $n = 189$ and early stage NSCLC cases $n = 94$ (Table 2). Samples were collected during the period 2003–2008. Histological classification followed WHO guidelines recommended at the time of diagnosis.

Plasma: EDTA plasma was collected from 100 subjects in the model testing study, controls $n = 50$ and cases $n = 50$ (Table 2).

Serum: plasma correlations: Blood was drawn from subjects ($n = 10$) and collected into serum (red-cap) and EDTA plasma tubes on the same visit to the clinic (CRCCC). Concentrations levels were determined for all candidate biomarkers ($n = 9$). Marker levels were highly correlated (Additional file 8: Table S4).

For all studies, written informed consent was obtained from each subject. Samples were obtained prior to any treatment and were stored at -80°C until use.

Mass spectrometry

The discovery approach combined the enrichment of cell surface glycoproteins and secreted proteins with a decoupled (label-free), quantitative proteomics method. These programs focused discovery on markers containing short, tryptically-cleaved 5–25 amino acid peptides encoding either a cysteine residue or an N-linked glycosylation site, providing broad coverage of the proteome. A quantitative liquid LC/MS analysis of normal and tumor samples was used to identify peptide ions that were expressed at $>4x$ levels in the cancer cells relative to the adjacent normal tissue. In cell line studies tumor lines were compared to Beas-2B. Subsequent MS/MS identification focused exclusively on peptides that had a relative change in abundance. To ensure data quality, manual inspection of each differentially expressed peptide ion was performed.

Cell surface protein enrichment

Viable cells were incubated with 1 mM sodium periodate for 10 min to oxidize glycoproteins [66]. Oxidized glycoproteins were conjugated to hydrazide resin (Bio-Rad, Hercules, CA) at 4°C overnight [67]. After washing sequentially with: 2 M NaCl, 2 % SDS, 200 mM propanolamine (0.1 M NaAcetate, pH 5.5), 40 % ethanol and 80 % ethanol; bound proteins were reduced with dithiothreitol, and alkylated with ICAT™ reagent (Life Technologies/Thermo Fisher Scientific, Applied Biosystems, Framingham, MA). Alkylated proteins were digested with trypsin and cysteine-containing peptides were captured using an avidin column (Life Technologies/Thermo Fisher Scientific). In addition to the cysteine-containing peptide fraction, peptides bound to the resin were also collected and analyzed. Release of peptides was achieved through PNGase-F digestion (New England BioLabs, Ipswich, MA.). While we found some overlap between the proteins identified in the two fractions, analysis of both the cysteine-containing fraction and the resin-bound fraction resulted in complementary coverage of the cell surface protein population.

Conditioned medium preparation

Samples were lyophilized, reconstituted with deionized H_2O , and dialyzed against 0.6 M Guanidine HCl, 10 mM Tris buffer, pH 8. Proteins were reduced with Tris (2-carboxyethyl) phosphine and alkylated with ICAT™ reagent (Life Technologies/Thermo Fisher Scientific). Following dialysis (0.1 M $\text{NH}_4\text{Acetate}$), alkylated proteins were digested with trypsin. Cysteine-containing peptides were

purified using an avidin column (Life Technologies/Thermo Fisher Scientific).

LC/MS analysis

Peptides, including standards used for mass calibration and retention time normalization, were separated and analyzed using methods of Kim et al. [68].

Data alignment and expression analysis

Peptide ion peaks of LC/MS maps were aligned based on mass to charge ratio (m/z), retention time (Rt), and charge state (z). Retention time normalization was accomplished in two steps: a primary alignment using the internal standard peptides and a secondary fine tuning using all of the common features. Ion intensities were normalized across normal and tumor samples by minimizing the sum of the differences between the intensities of each of the ions and the mean intensity for that ion across all maps. Differentially expressed peptide ions were manually verified before LC-MS/MS-based peptide sequencing. Subcellular predictions determined by UniProt, release 2014_07 [69].

Serum/plasma analyses

Enzyme-linked immunosorbent assay (ELISA) kits were obtained from a variety of commercial sources: Bio-Techne/R&D Systems, Minneapolis, MN (MMP2, OPN, SLPI, TFPI); Siemens Healthcare Diagnostics, Cambridge, MA (TIMP1); and IBL-America, Minneapolis, MN (CEA, CYFRA 21–1, MDK, SCC). Assays were performed following the manufacturers' instructions. Plates were read on a Spectra Max M2 Microplate Reader (Molecular Devices, Sunnyvale, CA.).

Model development

Logistic regression of lung cancer status on the 9 candidate markers (ng/mL) via elastic net regularization was employed to select a final set of markers and their associated parameter estimates. Elastic net regularization penalizes the parameter estimates (shrinks them toward zero) and performs variable subset selection by allowing sufficiently small parameter estimates to be reduced entirely to zero. Regularization of the parameter estimates tends to produce stable regression models with smaller prediction error than those that are not regularized. Bootstrap resampling (10,000 iterations), and maximization of the mean area under the ROC curve (AUC) for the out-of-bag samples, was used to select the optimal weight of the shrinkage penalty.

Additional files

Additional file 1: Table S1. Demographics and clinical profiles for subjects selected for fresh tissue discovery study.

Additional file 2: Table S2. Histology of lung cancer cell lines used for MS discovery study.

Additional file 3: Table S3. Candidate lung cancer biomarkers ($n = 179$) identified through LC/MS analysis.

Additional file 4: Figure S1. Panther-based classification of protein class for candidate lung cancer markers ($n = 179$).

Additional file 5: Figure S2. Panther-based classification of protein class comparing lung cancer markers identified in tissues and cell lines.

Additional file 6: Figure S3. Panther-based classification of protein pathways comparing lung cancer markers identified in tissues and cell lines.

Additional file 7: Figure S4. Expression levels of biomarker candidates in serum collected from patients with NSCLC ($n = 94$) and healthy volunteer controls ($n = 189$).

Additional file 8: Table S4. Correlation of marker levels in serum and plasma.

Abbreviations

ACK: Ammonium-Chloride-Potassium; AUC: Area under curve; CEA: Carcinoembryonic antigen-related cell adhesion molecule 5; CYFRA 21-1: Cytokeratin-19 fragment; EDTA: Ethylenediaminetetraacetic acid; ELISA: Enzyme-linked immunosorbent assay; EpCAM: Epithelial cell adhesion molecule; ICAM: Intercellular Adhesion Molecule; IRB: Institutional Review Board; LC/MS: Liquid chromatography/mass spectrometry; LDCT: Low-dose computed tomography; MDK: Midkine; MET: c-Met proto-oncogene product; MRM: Multiple reaction monitoring; MMP2: Matrix metalloproteinase-2; NLST: National Lung Screening Trial; NSCLC: Non-small cell lung cancer; PI: Propidium Iodide; OPN: Osteopontin; PET-CT: Positron emission tomography-computed tomography; PPV: Positive predictive value; ROC: Receiver Operating Characteristic; SCC: Squamous cell carcinoma antigen; SLPI: Secretory Leukocyte Peptidase Inhibitor; TFPI: Tissue factor pathway inhibitor; TIMP1: Tissue inhibitor of metalloproteinase 1; USPSTF: US Preventive Services Task Force.

Competing interests

CEB and RJL are currently employed by Celera, a wholly-owned subsidiary of Quest Diagnostics. CEB, MM, MEL and SMR are inventors of lung cancer markers patents issued to Celera. The other authors have nothing to declare.

Authors' contributions

CEB, RJL, WF, MM, PAM, TH, MEL, SMR designed the study. HIP, WNR, ESE, AOB, FM, JRJ provided clinical samples and contributed in study design. RJL and WF provided statistical analyses. MM, ES, EJ, AL, JH, GD, CD, JLT, RJB performed mass spectrometry analyses and provided immunoassay data. CEB wrote the manuscript with assistance from others. All authors reviewed the manuscript and provided their approval for publication.

Acknowledgements

We thank Sam Broder for his enthusiastic support during the course of these studies and for assistance in reviewing manuscript. We also thank Carl H. June and Steven M. Albelda, Department of Pathology and Laboratory Medicine and Department of Medicine, University of Pennsylvania, Philadelphia, PA; King Kwong, Division of Thoracic Surgery, University of Maryland Medical Center, Baltimore, MD; The Department of Cardiothoracic Surgery, George Washington University, Washington DC; and colleagues at Asterand, Detroit, MI, and Clinical Research Center of Cape Cod (West Yarmouth, MA) for help in providing clinical specimens employed in these studies. Support for specimen collection at New York University School of Medicine was provided by grants from the NCI Early Detection Research Network: U01CA086137 (WNR) and 2U01CA111295-04 (HIP).

Author details

¹Celera employees during the course of these studies, Celera, 1311 Harbor Bay Parkway, Alameda, CA 94502, USA. ²Department of Cardiothoracic Surgery, NYU Langone Medical Center, 530 First Avenue, New York, NY, USA. ³Division of Pulmonary, Critical Care, and Sleep Medicine, NYU School of Medicine, New York, NY, USA. ⁴Division of Pulmonary and Critical Care

Medicine, Mayo Clinic, Rochester, MN, USA. ⁵Division of Oncology, National Jewish Health, Denver, CO, USA.

Received: 1 December 2014 Accepted: 7 July 2015

Published online: 16 July 2015

References

- Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64:9–29.
- Aberle DR, Adams AM, Berg CD, Black WC, Clapp JD, Fagerstrom RM, et al. Reduced lung-cancer mortality with low-dose computed tomographic screening. *N Engl J Med*. 2011;365:395–409.
- Church TR, Black WC, Aberle DR, Berg CD, Clingan KL, Duan F, et al. Results of initial low-dose computed tomographic screening for lung cancer. *N Engl J Med*. 2013;368:1980–91.
- Humphrey LL, Deffebach M, Pappas M, Baumann C, Artis K, Mitchell JP, et al. Screening for lung cancer with low-dose computed tomography: a systematic review to update the US Preventive services task force recommendation. *Ann Intern Med*. 2013;159:411–20.
- Moyer VA, USPSTF. Screening for lung cancer: U.S. Preventive Services Task Force recommendation statement. *Ann Intern Med*. 2014;160:330–8.
- Hassanein M, Callison JC, Callaway-Lane C, Aldrich MC, Grogan EL, Massion PP. The state of molecular biomarkers for the early detection of lung cancer. *Cancer Prev Res (Phila)*. 2012;5:992–1006.
- Higgins G, Roper KM, Watson IJ, Blackhall FH, Rom WN, Pass HI, et al. Variant Ciz1 is a circulating biomarker for early-stage lung cancer. *Proc Natl Acad Sci U S A*. 2012;109:E3128–35.
- Ostroff RM, Bigbee WL, Franklin W, Gold L, Mehan M, Miller YE, et al. Unlocking biomarker discovery: large scale application of aptamer proteomic technology for early detection of lung cancer. *PLoS One*. 2010;5:e15003.
- Howard BA, Wang MZ, Campa MJ, Corro C, Fitzgerald MC, Patz Jr EF. Identification and validation of a potential lung cancer serum biomarker detected by matrix-assisted laser desorption/ionization-time of flight spectra analysis. *Proteomics*. 2003;3:1720–4.
- Patz Jr EF, Campa MJ, Gottlin EB, Kusmartseva I, Guan XR, Herndon 2nd JE. Panel of serum biomarkers for the diagnosis of lung cancer. *J Clin Oncol*. 2007;25:5578–83.
- Yildiz PB, Shyr Y, Rahman JS, Wardwell NR, Zimmerman LJ, Shakhtour B, et al. Diagnostic accuracy of MALDI mass spectrometric analysis of unfractionated serum in lung cancer. *Food Funct*. 2007;2:893–901.
- Zeng X, Hood BL, Sun M, Conrads TP, Day RS, Weissfeld JL, et al. Lung cancer serum biomarker discovery using glycoprotein capture and liquid chromatography mass spectrometry. *J Proteome Res*. 2010;9:6440–9.
- Zeng X, Hood BL, Zhao T, Conrads TP, Sun M, Gopalakrishnan V, et al. Lung cancer serum biomarker discovery using label-free liquid chromatography-tandem mass spectrometry. *J Thorac Oncol*. 2011;6:725–34.
- Anderson NL, Anderson NG. The human plasma proteome: history, character, and diagnostic prospects. *Mol Cell Proteomics*. 2002;1:845–67.
- Taguchi A, Politi K, Pitteri SJ, Lockwood WW, Faca VM, Kelly-Spratt K, et al. Lung cancer signatures in plasma based on proteome profiling of mouse tumor models. *Cancer Cell*. 2011;20:289–99.
- Yee J, Sadar MD, Sin DD, Kuzyk M, Xing L, Kondra J, et al. Connective tissue-activating peptide III: a novel blood biomarker for early lung cancer detection. *J Clin Oncol*. 2009;27:2787–92.
- Huang LJ, Chen SX, Huang Y, Luo WJ, Jiang HH, Hu QH, et al. Proteomics-based identification of secreted protein dihydrodiol dehydrogenase as a novel serum markers of non-small cell lung cancer. *Lung Cancer*. 2006;54:87–94.
- Xiao T, Ying W, Li L, Hu Z, Ma Y, Jiao L, et al. An approach to studying lung cancer-related proteins in human blood. *Mol Cell Proteomics*. 2005;4:1480–6.
- Planque C, Kulasingam V, Smith CR, Reckamp K, Goodlick L, Diamandis EP. Identification of five candidate lung cancer biomarkers by proteomics analysis of conditioned media of four lung cancer cell lines. *Mol Cell Proteomics*. 2009;8:2746–58.
- Li XJ, Hayward C, Fong PY, Dominguez M, Hunsucker SW, Lee LW, et al. A blood-based proteomic classifier for the molecular characterization of pulmonary nodules. *Sci Transl Med*. 2013;5:207ra142.
- Reddel RR, Ke Y, Gerwin BI, McMenamin MG, Lechner JF, Su RT, et al. Transformation of human bronchial epithelial cells by infection with SV40 or adenovirus-12 SV40 hybrid virus, or transfection via strontium phosphate

- coprecipitation with a plasmid containing SV40 early region genes. *Cancer Res.* 1988;48:1904–9.
22. Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res.* 2013;41:D377–86.
 23. Mi H, Muruganujan A, Casagrande JT, Thomas PD. Large-scale gene function analysis with the PANTHER classification system. *Nat Protoc.* 2013;8:1551–66.
 24. Ameshima S, Ishizaki T, Demura Y, Imamura Y, Miyamori I, Mitsuhashi H. Increased secretory leukoprotease inhibitor in patients with non-small cell lung carcinoma. *Cancer.* 2000;89:1448–56.
 25. Nanjappa V, Thomas JK, Marimuthu A, Muthusamy B, Radhakrishnan A, Sharma R, et al. Plasma Proteome Database as a resource for proteomics research: 2014 update. *Nucleic Acids Res.* 2014;42:D959–65.
 26. Almatroodi SA, McDonald CF, Collins AL, Darby IA, Pouniotis DS. Quantitative proteomics of bronchoalveolar lavage fluid in lung adenocarcinoma. *Cancer Genomics Proteomics.* 2015;12:39–48.
 27. Pastor MD, Nogal A, Molina-Pinelo S, Melendez R, Salinas A, Gonzalez De la Pena M, et al. Identification of proteomic signatures associated with lung cancer and COPD. *J Proteomics.* 2013;89:227–37.
 28. Yu L, Shen J, Mannoor K, Guarnera M, Jiang F. Identification of ENO1 as a potential sputum biomarker for early-stage lung cancer by shotgun proteomics. *Clin Lung Cancer.* 2014;15:372–8. e371.
 29. Li Y, Lian H, Jia Q, Wan Y. Proteome screening of pleural effusions identifies IL1A as a diagnostic biomarker for non-small cell lung cancer. *Biochem Biophys Res Commun.* 2015;457:177–82.
 30. Tyan YC, Wu HY, Lai WW, Su WC, Liao PC. Proteomic profiling of human pleural effusion using two-dimensional nano liquid chromatography tandem mass spectrometry. *J Proteome Res.* 2005;4:1274–86.
 31. Verrills NM, Irwin JA, He XY, Wood LG, Powell H, Simpson JL, et al. Identification of novel diagnostic biomarkers for asthma and chronic obstructive pulmonary disease. *Am J Respir Crit Care Med.* 2011;183:1633–43.
 32. Xu D, Li Y, Li X, Wei LL, Pan Z, Jiang TT, et al. Serum protein S100A9, SOD3, and MMP9 as new diagnostic biomarkers for pulmonary tuberculosis by iTRAQ-coupled two-dimensional LC-MS/MS. *Proteomics.* 2015;15:58–67.
 33. Hastie T, Zou H. Regularization and variable selection via the elastic net. *JR Statist Soc.* 2005;67:301–20.
 34. Swensen SJ, Jett JR, Hartman TE, Midthun DE, Sloan JA, Sykes AM, et al. Lung cancer screening with CT: Mayo Clinic experience. *Radiology.* 2003;226:756–61.
 35. Ajona D, Pajares MJ, Corrales L, Perez-Gracia JL, Agorreta J, Lozano MD, et al. Investigation of complement activation product c4d as a diagnostic and prognostic biomarker for lung cancer. *J Natl Cancer Inst.* 2013;105:1385–93.
 36. Bigbee WL, Gopalakrishnan V, Weissfeld JL, Wilson DO, Dacic S, Lokshin AE, et al. A multiplexed serum biomarker immunoassay panel discriminates clinical lung cancer patients from high-risk individuals found to be cancer-free by CT screening. *J Thorac Oncol.* 2012;7:698–708.
 37. Daly S, Rinewalt D, Fhied C, Basu S, Mahon B, Liptay MJ, et al. Development and validation of a plasma biomarker panel for discerning clinical significance of indeterminate pulmonary nodules. *J Thorac Oncol.* 2013;8:31–6.
 38. Mehan MR, Williams SA, Siegfried JM, Bigbee WL, Weissfeld JL, Wilson DO, et al. Validation of a blood protein signature for non-small cell lung cancer. *Clinical proteomics.* 2014;11:32.
 39. Pecot CV, Li M, Zhang XJ, Rajanbabu R, Calitri C, Bungum A, et al. Added value of a serum proteomic signature in the diagnostic evaluation of lung nodules. *Cancer Epidemiol Biomarkers Prev.* 2012;21:786–92.
 40. Zamcheck N. Carcinoembryonic antigen. Quantitative variations in circulating levels in benign and malignant digestive tract diseases. *Adv Intern Med.* 1974;19:413–33.
 41. Hakim AA, Siraki CM, Joseph CE. Carcinoembryonic antigen from human malignant melanoma cells. I. Production and shedding characteristics. *Ann Immunol.* 1983;134D:319–31.
 42. Galvani AP, Cristiani C, Carpinelli P, Landonio A, Bertolero F. Suramin modulates cellular levels of hepatocyte growth factor receptor by inducing shedding of a soluble form. *Biochem Pharmacol.* 1995;50:959–66.
 43. Ho M, Onda M, Wang QC, Hassan R, Pastan I, Lively MO. Mesothelin is shed from tumor cells. *Cancer Epidemiol Biomarkers Prev.* 2006;15:1751.
 44. Abe H, Kuroki M, Imakiire T, Yamauchi Y, Yamada H, Arakawa F, et al. Preparation of recombinant MK-1/Ep-CAM and establishment of an ELISA system for determining soluble MK-1/Ep-CAM levels in sera of cancer patients. *J Immunol Methods.* 2002;270:227–33.
 45. Harning R, Mainolfi E, Bystryn JC, Henn M, Merluzzi VJ, Rothlein R. Serum levels of circulating intercellular adhesion molecule 1 in human malignant melanoma. *Cancer Res.* 1991;51:5003–5.
 46. Koldas M, Gummus M, Seker M, Seval H, Hulya K, Dane F, et al. Thrombin-activatable fibrinolysis inhibitor levels in patients with non-small-cell lung cancer. *Clin Lung Cancer.* 2008;9:112–5.
 47. Choudhuri R, Zhang HT, Donnini S, Ziche M, Bicknell R. An angiogenic role for the neurokinins midkine and pleiotrophin in tumorigenesis. *Cancer Res.* 1997;57:1814–9.
 48. Salama RH, Muramatsu H, Zou P, Okayama M, Muramatsu T. Midkine, a heparin-binding growth factor, produced by the host enhances metastasis of Lewis lung carcinoma cells. *Cancer Lett.* 2006;233:16–20.
 49. Ikematsu S, Yano A, Aridome K, Kikuchi M, Kumai H, Nagano H, et al. Serum midkine levels are increased in patients with various types of carcinomas. *Br J Cancer.* 2000;83:701–6.
 50. Hayakawa T, Yamashita K, Tanzawa K, Uchijima E, Iwata K. Growth-promoting activity of tissue inhibitor of metalloproteinases-1 (TIMP-1) for a wide range of cells. A possible new growth factor in serum. *FEBS Lett.* 1992;298:29–32.
 51. Liu XW, Bernardo MM, Fridman R, Kim HR. Tissue inhibitor of metalloproteinase-1 protects human breast epithelial cells against intrinsic apoptotic cell death via the focal adhesion kinase/phosphatidylinositol 3-kinase and MAPK signaling pathway. *J Biol Chem.* 2003;278:40364–72.
 52. Jumper C, Cobos E, Lox C. Determination of the serum matrix metalloproteinase-9 (MMP-9) and tissue inhibitor of matrix metalloproteinase-1 (TIMP-1) in patients with either advanced small-cell lung cancer or non-small-cell lung cancer prior to treatment. *Respir Med.* 2004;98:173–7.
 53. Pesta M, Kulda V, Kucera R, Pesek M, Vrzalova J, Liska V, et al. Prognostic significance of TIMP-1 in non-small cell lung cancer. *Anticancer Res.* 2011;31:4031–8.
 54. Devoogdt N, Hassanzadeh Ghassabeh G, Zhang J, Brys L, De Baetselier P, Revets H. Secretory leukocyte protease inhibitor promotes the tumorigenic and metastatic potential of cancer cells. *Proc Natl Acad Sci U S A.* 2003;100:5778–82.
 55. Devoogdt N, Revets H, Ghassabeh GH, De Baetselier P. Secretory leukocyte protease inhibitor in cancer development. *Ann N Y Acad Sci.* 2004;1028:380–9.
 56. Nukiwa T, Suzuki T, Fukuhara T, Kikuchi T. Secretory leukocyte peptidase inhibitor and lung cancer. *Cancer Sci.* 2008;99:849–55.
 57. Sugino T, Yamaguchi T, Ogura G, Kusakabe T, Goodison S, Homma Y, et al. The secretory leukocyte protease inhibitor (SLPI) suppresses cancer cell invasion but promotes blood-borne metastasis via an invasion-independent pathway. *J Pathol.* 2007;212:152–60.
 58. Zelvyte I, Wallmark A, Piitulainen E, Westin U, Janciauskiene S. Increased plasma levels of serine proteinase inhibitors in lung cancer patients. *Anticancer Res.* 2004;24:241–7.
 59. Ku MJ, Park JW, Ryu BJ, Son YJ, Kim SH, Lee SY. CK2 inhibitor CX4945 induces sequential inactivation of proteins in the signaling pathways related with cell migration and suppresses metastasis of A549 human lung cancer cells. *Bioorg Med Chem Lett.* 2013;23:5609–13.
 60. Sun Q, Yao X, Ning Y, Zhang W, Zhou G, Dong Y. Overexpression of response gene to complement 32 (RGC32) promotes cell invasion and induces epithelial-mesenchymal transition in lung cancer cells via the NF-kappaB signaling pathway. *Tumour Biol.* 2013;34:2995–3002.
 61. Zucker S, Cao J, Chen WT. Critical appraisal of the use of matrix metalloproteinase inhibitors in cancer treatment. *Oncogene.* 2000;19:6642–50.
 62. Ali-Labib R, Louka ML, Galal IH, Tarek M. Evaluation of matrix metalloproteinase-2 in lung cancer. *Proteomics Clin Appl.* 2014;8:251–7.
 63. Hida Y, Hamada J. Differential expressions of matrix metalloproteinases, a disintegrin and metalloproteinases, and a disintegrin and metalloproteinases with thrombospondin motifs and their endogenous inhibitors among histologic subtypes of lung cancers. *Anticancer Agents Med Chem.* 2012;12:744–52.
 64. Kanoh Y, Abe T, Masuda N, Akahoshi T. Progression of non-small cell lung cancer: diagnostic and prognostic utility of matrix metalloproteinase-2, C-reactive protein and serum amyloid A. *Oncol Rep.* 2013;29:469–73.
 65. Mesri M, Birse C, Heidbrink J, McKinnon K, Brand E, Bermingham CL, et al. Identification and characterization of angiogenesis targets through proteomic profiling of endothelial cells in human cancer tissues. *PLoS One.* 2013;8, e78885.

66. Bobbitt JM. Periodate oxidation of carbohydrates. *Adv Carbohydr Chem.* 1956;48:1–41.
67. Zhang H, Li XJ, Martin DB, Aebersold R. Identification and quantification of N-linked glycoproteins using hydrazide chemistry, stable isotope labeling and mass spectrometry. *Nat Biotechnol.* 2003;21:660–6.
68. Kim YJ, Zhan P, Feild B, Ruben SM, He T. Reproducibility assessment of relative quantitation strategies for LC-MS based proteomics. *Anal Chem.* 2007;79:5651–8.
69. UniProt Consortium. Activities at the Universal Protein Resource (UniProt). *Nucleic Acids Res.* 2014;42:D191–8.
70. Benchimol S, Fuks A, Jothy S, Beauchemin N, Shirota K, Stanners CP. Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. *Cell.* 1989;57:327–34.
71. Nittka S, Bohm C, Zentgraf H, Neumaier M. The CEACAM1-mediated apoptosis pathway is activated by CEA and triggers dual cleavage of CEACAM1. *Oncogene.* 2008;27:3721–8.
72. Lwaleed BA, Bass PS. Tissue factor pathway inhibitor: structure, biology and involvement in disease. *J Pathol.* 2006;208:327–39.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

