Discovery of KPNA2 as a NSCLC potential biomarker by integration of cancer cell secretome and tissue transcriptome

Chun-I Wang†, Chih-Liang Wang‡,#, Chih-Wei Wang§, Chi-De Chen¶, Chih-Ching Wu∥, Ying-Huang Tsai∥, Yu-Sun Chang¶,¥, Jau-Song Yu¶,¥,¶, and Chia-Jung Yu†,¶,¥,*

†Graduate Institute of Biomedical Sciences, ‡Graduate Institute of Clinical Medical Sciences, §Chang Gung Molecular Medicine Research Center, ¶Department of Cell and Molecular Biology, Chang Gung University, Tao-Yuan, Taiwan. #Division of Pulmonary Oncology and Interventional Bronchoscopy, Department of Thoracic Medicine, ¥Department of Pathology, Chang Gung Memorial Hospital, Tao-Yuan, Taiwan.

* Correspondence author.

To whom should be addressed: Department of Cell and Molecular Biology, Chang Gung University, 259 Wen-Hwa 1st Road, Kwei-Shan, Tao-Yuan, Taiwan; Tel: 886-3-2118800 ext. 3424; Fax: 886-3-2118042; E-mail: yucj1124@mail.cgu.edu.tw
Abstract

The cancer cell secretome may contain potentially useful biomarkers. In this study, we integrated the profiles of secreted proteins in lung cancer cell lines with mRNA expression levels from pulmonary adenocarcinoma tissue, with a view to identifying effective biomarkers for non-small cell lung cancer (NSCLC). Among the novel candidates isolated, importin subunit alpha-2 (also known as karyopherin subunit alpha-2, KPNA2), was selected for further validation. Immunohistochemical staining revealed overexpression and altered subcellular localization of KPNA2 in tumor cells, compared with adjacent normal cells. A sandwich ELISA assay developed to detect KPNA2 levels in serum samples showed significantly higher serum KPNA2 in NSCLC patients than in healthy controls. A combination of serum KPNA2 and CEA displayed higher diagnostic capacity than either marker alone. Importantly, protein levels of KPNA2 in pleural effusion from NSCLC patients were also significantly higher than those from non-lung cancer. Moreover, knockdown of KPNA2 inhibited the migration ability and viability of lung cancer cells. Our results collectively suggest that integration of the cancer cell secretome and transcriptome datasets provides an efficient means of identifying novel body fluid accessible biomarkers for NSCLC, such as KPNA2.

Key words: Secretome, Transcriptome, Non-small cell lung cancer, KPNA2, Biomarker
Introduction

Lung cancer is the most common type of cancer worldwide\(^1\) accounting for 12.4% of all newly diagnosed cases,\(^2\) and represents the leading cause of cancer-related deaths in Taiwan. Based on biology, therapy and prognosis, lung cancer is broadly divided into two classes, specifically, small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC).\(^3\) NSCLC is the most common lung cancer type, comprising ~80% of all lung cancers.\(^4\) Despite major advances in cancer therapy over the past two decades, the prognosis of patients with NSCLC has improved only minimally, and the 5-year survival rate remains less than 15%.\(^5\) One reason for this poor outcome is that most patients are diagnosed at the late stages as they initially present at outpatient clinics.\(^6\)

Patient history, physical examination and radiological imaging studies, including chest X-ray and computerized axial tomography scans, are currently the gold standard for lung cancer screening. However, these tests are not sufficiently accurate for effective cancer diagnosis and disease staging.\(^7\)\(^8\) Serum proteins, such as carcinoembryonic antigen (CEA), CYFRA 21-1 (cytokeratin 19 fragment), CA125 (cancer antigen 125), squamous cell carcinoma antigen, neuron-specific enolase, progastrin-releasing peptide, tumor M2-pyruvate kinase and C-reactive protein are potential markers for lung cancer, and their levels may signify the presence of tumors, facilitate histological analysis, and allow prediction of cancer progression.\(^9\)\(^10\) Due to
limited sensitivity and specificity, these potential markers are not currently recommended or encouraged in routine clinical practice.\textsuperscript{11}

Proteomic approaches have been widely applied to investigate malignant diseases, particularly in the field of plasma/serum tumor marker identification.\textsuperscript{12, 13} For practical usage in tumor screening, biomarkers should be measurable in body fluid samples. Thus, proteins secreted by or shed from tumor cells are of particular interest.\textsuperscript{14, 15} Proteins present in the conditioned media of cultured cell lines derived from specific cancer types present attractive potential tumor biomarker candidates, as they are more likely to be detected in body fluids, such as serum or plasma. Several potential biomarkers have been identified from cancer cell secretome analysis,\textsuperscript{16-23} including collapsin response mediator protein-2 as a colon cancer marker,\textsuperscript{19} Mac-2 binding protein as a oral cancer marker,\textsuperscript{20} cathepsin D, syntenin, and gp100 as uveal melanoma biomarkers,\textsuperscript{21} and protein gene product 9.5, translationally controlled tumor protein, tissue inhibitors of metalloproteinase-2 and triosephosphate isomerase as lung cancer biomarkers.\textsuperscript{23} Thus, identification of proteins secreted from cancer cells is an efficient strategy for discovering markers measurable in the body fluid of suspected cancer patients. Importantly, secretome datasets offer a good opportunity to identify promising candidates for early diagnosis as well as prognosis of cancer.\textsuperscript{19-20, 23}

The cDNA microarray-based transcriptome technology has allowed high-throughput and simultaneous comparison of the expression patterns of thousands
of genes in tissue. The National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) is a public array-based database containing 2041 microarray (including 726 human-related) datasets. Taking advantage of the available cancer cell secretome datasets and public-accessible gene expression microarray databases of several human cancer types, we have proposed a simple strategy to facilitate the discovery of potential body fluid-accessible cancer biomarkers overexpressed in tumor tissue and secreted into body fluids. This theory is supported by our recent finding that Mac-2 binding protein is an effective colorectal carcinoma (CRC) plasma biomarker, based on its apparent secretion by CRC cell lines and elevated transcriptional level in public array-based analysis of CRC tissues.

Here, we use the above strategy to detect potential lung cancer biomarkers. We integrated two lung adenocarcinoma cell line secretome datasets (CL1-0 and CL1-5) with one pulmonary adenocarcinoma microarray dataset to identify targets that are significantly up-regulated in lung cancer tissues and secreted/released from lung cancer cells. Importin subunit alpha-2 (also known as karyopherin subunit alpha-2, KPNA2), one of the 19 novel candidates, was selected for further validation using clinical specimens from lung cancer patients. We detected overexpression and altered subcellular localization of KPNA2 in lung cancer tissues, compared with adjacent non-tumor counterparts using immunohistochemical analyses, and observed elevated serum levels of KPNA2 in lung cancer patients versus healthy controls with in-house
sandwich ELISA. Importantly, a combination of serum KPNA2 and CEA serum levels displayed higher diagnostic capacity in predicting cancer in healthy individuals than either marker alone. Protein levels of KPNA2 in pleural effusion from NSCLC patients were also significantly higher than those from non-lung cancer. Furthermore, data from RNA interference-based knockdown of KPNA2 expression in lung cancer cells suggest that the protein is involved in regulating the growth and motility of lung cancer cells.
Materials and Methods

Cell Culture. Human lung adenocarcinoma cancer cell lines, CL1-0 and CL1-5, were kindly provided by Professor P.C. Yang (Department of Internal Medicine, National Taiwan University Hospital, Taipei, Taiwan, Republic of China).\textsuperscript{27} A549, a representative human lung adenocarcinoma cell line, was obtained from American Type Culture Collection (ATCC, VA, USA). CL1-0, and CL1-5 cells were maintained in RPMI 1640, while A549 was maintained in Dulbecco's modified Eagle's medium (DMEM) with 10% FBS plus antibiotics at 37\textdegree C at a humidified atmosphere of 95% air/5% CO\textsubscript{2}.

Transcriptome Dataset of Pulmonary Adenocarcinoma Tissue. The NCBI GEO database (http://www.ncbi.nlm.nih.gov/projects/geo/) was searched for mRNA expression profiles of pulmonary adenocarcinoma tissues. Currently, three human lung cancer microarray datasets are deposited in the GEO database, including two squamous cell carcinoma (SCC) datasets (GDS2373 and GDS1312) and one human pulmonary adenocarcinoma dataset (GDS1650). GDS1650 contains differential gene profiles generated from 20 pulmonary adenocarcinoma tissues and 19 paired adjacent normal tissues, and was selected as our target transcriptome.

Patient Populations and Clinical Specimens. The 66 paired surgically resected lung cancer and adjacent normal tissue samples (42 males and 24 females; 46 adenocarcinoma, 20 squamous cell carcinoma; stages I to IV) were obtained from
patients subjected to surgery at Chang Gung Memorial Hospital. Written informed consent was received from all patients before collection, and the study approved by the Institutional Review Board. Patients’ medical records were reviewed, and we ensured that identities were protected. All tissue specimens were formalin-fixed and paraffin-embedded. Prior to treatment, serum samples were obtained from 126 patients (83 males and 43 females; range 27-79 years) with lung cancer and 64 healthy control volunteers (42 males and 22 females; range 25-81 years). Control subjects did not display any type of cancer at least six months after serum sample collection. Pleural effusion samples were obtained from 82 patients (51 males and 31 females; range 38-85 years) with lung cancer, 23 with breast cancer (23 females; range 34-83 years), 12 with gastric cancer (6 males and 6 females; range 40-72 years), and 13 (6 males and 7 females; range 53-83 years) with colon cancer. Venous blood (10 ml) was collected and centrifuged at 4°C for serum collection. The serum and pleural effusion samples collected were stored at -80°C until further analysis.

**Immunohistochemistry and Scoring.** Immunohistochemistry was performed as described previously.28 Consecutive sections (5 μm thickness) of formalin-fixed paraffin-embedded specimens from 66 NSCLC patients were subjected to H&E staining and immunohistochemical analysis (IHC) with the anti-KPNA2 antibody (Proteintech, Chicago, USA). Tissue sections were deparaffinized, treated with 3% hydrogen peroxide for 10 min at room temperature, and microwaved in 0.01 M citrate
buffer (pH 6.0) for retrieval of antigenicity. Sections were incubated with blocking solution (1% bovine serum albumin in PBS) for 20 min at room temperature. Samples were incubated with anti-KPNA2 overnight at 4°C. Secondary anti-rabbit antibody-coated polymer peroxidase complexes (DAKO Corp., Carpinteria, CA) were applied for 30 min at room temperature, followed by treatment with substrate/chromogen (DAKO Corp., Carpinteria, CA) and further incubation for 5-10 min at room temperature. Slides were counterstained with hematoxylin. Scoring of KPNA2 staining was performed according to a previous study, which was reproducible. Scoring was based on the overall percentage of stained cells (0, 0-9%; 1, 10%-49%; 2, 50%-89% and 3, >90%) and median intensity of staining (0, no staining; 1, weak; 2, moderate and 3, strong staining). The ranks of percentage and intensity of staining described above were multiplied by each other, divided by 3 and rounded up to the nearest whole number. Immunoreactivity was interpreted as negative (score 0) or positive (scores 1 to 3).

**Western Blot Analysis.** Western blotting was carried out as described earlier. Protein extracts were separated by SDS-PAGE, and transferred to a PVDF membrane. The membrane was gently shaken for 1 h at room temperature in 5% (wt/vol) non-fat dried milk in Tris-buffered saline (TBS), washed three times with TBS, and incubated for 1 h with the indicated antibody. After further washing, the membrane was incubated for 1 h with horseradish peroxidase-conjugated antibodies against rabbit or
mouse IgG (GE Healthcare UK). Immune complexes were visualized by chemiluminescence using the ECL detection kit (PerkinElmer Life Sciences, Ontario, Canada), according to the manufacturer’s instructions.

**Fluorometric Sandwich ELISA.** KPNA2 protein levels in human serum and pleural effusion were determined using a sandwich ELISA assay developed in-house. Briefly, white polystyrene microtiter plates (Corning, Canton, NY, USA) were coated with an anti-KPNA2 mouse monoclonal antibody (Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA) and blocked with blocking buffer (1% bovine serum albumin /PBS). Serum or pleural effusion samples diluted 1:10 in blocking buffer and various amounts of KPNA2 recombinant protein (Abnova, Taipei, Taiwan) were added to the wells. Each well was incubated with anti-KPNA2 rabbit polyclonal antibody (Proteintech Inc, Chicago, USA). After washing, alkaline phosphatase-conjugated rabbit anti-rabbit IgG (Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA) was added to the individual wells and plates. Following six further washes, 4-methylumbelliferyl phosphate (Molecular Probes, Eugene, OR, USA) was added as the substrate, and the fluorescence intensity (excitation: 355 nm, emission: 460 nm) measured with a SpectraMax M5 microplate reader (Molecular Devices, Sunnyvale, CA).

**Gene Knockdown of KPNA2 with Small Interfering RNA.** For gene knockdown, 25-nucleotide RNA duplexes targeting human KPNA2 were designed
and purchased from Invitrogen (Invitrogen, Grand Island, NY, USA). Briefly, CL1-0 and CL1-5 cells were transfected with control siRNA (Invitrogen, Grand Island, NY, USA) or KPNA2 siRNA (UUUCUGGCAGCUUGAGUAGCUUG) using Lipofectamine RNAiMAX reagents (Invitrogen, Grand Island, NY, USA), according to the protocol provided by the manufacturer. At 48 h after transfection, cell lysates were prepared for Western blotting to determine gene knockdown efficacy.

**Cell Viability Assay.** Cell viability was evaluated with the 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) colorimetric growth assay. Briefly, cells (1 × 10^3 cells/well) were plated in 96-well plates, and cultured for the indicated time intervals. Following culture, MTT solution (5 mg/mL) was added and the cells incubated at 37°C for 1 h. The supernatant was aspirated, cells treated with DMSO (100 μL), and absorbance measured at 540 nm using an ELISA reader (Molecular Devices, SpectraMax M2).

**Transwell Migration Assay.** Cells transfected with control siRNA or KPNA2 siRNA for 48 h were harvested by trypsinization, and suspended in serum-free culture medium. Cell migration was assayed using a 24-well format transwell chamber (8.0 μm pore size filter, Corning, Canton, NY, USA). The cell suspension (300 μL; 3×10^4 cells) was added to each insert of the upper chamber, and each lower chamber was filled with 600 μL culture medium containing 10 μL/ml fibronectin. After 6 h (CL1-0 and CL1-5) or 24 h (A549) incubation at 37°C, chambers were gently washed twice
with PBS and fixed with methanol, followed by Giemsa staining. Cells that had traversed the filter to the lower chamber were counted microscopically (200X) in 6 different fields per filter.

Statistical Analysis. The statistical package, SPSS 13.0 (SPSS Inc, Chicago, IL), was used for all analyses. All continuous variables were expressed as means ± SD. The nonparametric Mann-Whitney U test was employed to analyze variations in ELISA results for different clinical characteristics. The Chi-square test was used to determine the proportional differences in IHC intensity between clinicopathologic factors. Two-tailed p values of 0.05 or less were considered significant. The receiver operating characteristic (ROC) curve was constructed by plotting sensitivity versus (1-specificity), considering each observed value as a possible cutoff value. The area under the ROC curve (AUC) was calculated as a single measurement to establish the discriminative efficacy of each marker.31
Results and Discussion

Generation of a Potential Adenocarcinoma Biomarker Dataset by Combined Analysis of Lung Cancer Cell Secretome and Transcriptome of Pulmonary Adenocarcinoma Tissues. Previously, we analyzed the secretomes of two adenocarcinoma cell lines, CL1-0 and CL1-5, by one-dimensional SDS-PAGE in conjunction with the nano LC tandem mass spectrometry (GeLC-MS/MS) approach. Our analysis led to the identification of 1096 and 1830 proteins from serum-free conditioned media of CL1-0 and CL1-5 cells, respectively (J. Proteome Res., revision, manuscript ID: pr-2009-00160h.R1). Here, we integrate the two secretome datasets with the GDS1650 pulmonary adenocarcinoma tissue transcriptome to identify potential serum biomarkers for NSCLC. A schematic representation of the strategy is shown in Fig. 1. We detected 285 genes displaying >4-fold upregulation in pulmonary adenocarcinoma tissues, compared with adjacent normal tissues. Upon combination of these 285 upregulated genes with CL1-0 and CL1-5 cancer cell secretome datasets, 35 potential biomarkers present in single or both secretome datasets were isolated (Fig. 1 and Table 1). Among these potential biomarkers, 16 were previously reported as dysregulated proteins/genes in lung cancer and 19 were novel candidates (Table 1). Five novel candidates were identified in both cancer cell secretome datasets, including DNA replication complex GINS protein PSF1 (GINS1), importin subunit alpha-2 (KPNA2), platelet-activating factor acetylhydrolase IB subunit gamma (PAFAH1B3),
6-phosphofructokinase type C (PFKP), and sorbitol dehydrogenase (SORD).

**Overexpression of KPNA2 in NSCLC tissues.** KPNA2, a nucleocytoplasmic shuttle protein that recognizes the nuclear localization signal on cargo proteins and mediates the nuclear import of macromolecules,\textsuperscript{32,33} was selected for further analysis. To our knowledge, KPNA2 is a reported potential biomarker for breast cancer, but has not been investigated in other types of human cancer to date. Dahl et al.\textsuperscript{34} initially identified KPNA2 as a potential novel prognostic marker in breast cancer following comprehensive transcriptome analysis. Subsequently, Dankof and colleagues examined KPNA2 protein expression in invasive breast carcinoma, which matched peritumoral ductal carcinoma *in situ*.\textsuperscript{35} Recently, Gluz and co-workers reported that nuclear KPNA2 expression is a predictor of poor survival in patients with advanced breast cancer, irrespective of treatment intensity.\textsuperscript{36} However, the clinical and biological significance of KPNA2 in other cancers remains to be established. To ascertain the clinical significance of KPNA2 in NSCLC, we examined protein expression in 66 NSCLC tissue specimens containing tumors and their adjacent normal tissues by immunohistochemical staining with the anti-KPNA2 antibody. The results showed that 71.2% (47/66) of lung tumor tissues, but none of the adjacent normal tissues (0%; 0/66) stained positive for KPNA2 in the nucleus (Table 2). In addition, 27.3% (18/66) of lung tumor tissues displayed KPNA2-positive staining in the cytoplasm, whereas all adjacent normal tissues (100%; 66/66) stained positive for
KPNA2 in the cytoplasm. The results clearly indicate that KPNA2 is located exclusively in the cytoplasmic fractions of normal lung epithelial cells, but mainly expressed in the nuclei of tumor cells. One representative case of positive staining of KPNA2 in tumor cell nuclei was shown in Fig 2A. The precise mechanism(s) responsible for the distinct subcellular distribution of KPNA2 between tumor and adjacent normal tissues remain unclear at present. KPNA2 belongs to the karyopherin family and delivers numerous cargo proteins to the nucleus, followed by translocation from nuclear to cytoplasmic compartments in a Ran-GTP dependent manner. In view of the finding that KPNA2 is predominantly localized in the nucleus in lung tumor cells, in contrast to its cytoplasmic location in most normal lung epithelial cells, we speculate that the function of KPNA2 in cargo protein transport between nucleus and cytoplasm is altered. This speculation is supported by increasing cytosolic fraction of NBS1, one of cargo protein of KPNA2, in NSCLC. Analysis of the intracellular distribution of NBS1 and KPNA2 in the same NSCLC tissue specimens revealed that compared with adjacent normal cells, the nuclear NBS1 signal was dramatically decreased while the cytoplasmic NBS1 signal was prevalent in 60 % (3/5) of KPNA2-positive tumor cells (Supplementary Fig. 1). The exact molecular mechanism underlying dysregulation of expression/location of KPNA2 and its cargo proteins in tumorigenesis of lung cancer remains to be established.

It is notable that positive staining of KPNA2 was more frequently observed in
poorly differentiated than well differentiated tumor cells (82% vs. 37.5%, \( p < 0.001 \), Supplementary Table S1). Two representative cases displaying poorly and well differentiated adenocarcinoma are presented in Fig. 2B. Microarray and quantitative real-time PCR analysis of the mRNA expression profiles in normal human epidermal keratinocytes with either overexpression or knockdown of KPNA2 clearly indicate that KPNA2 is involved in various signal transduction pathways that regulate epidermal proliferation and differentiation.\(^{40}\) Our findings imply that KPNA2 could modulate cell differentiation in lung cancer cells, which is similar to the biological function reported in epidermal cells. Interestingly, 95% (19/20) of squamous cell carcinoma tissues exhibited KPNA2-positive nuclear staining (Supplementary Table S1), indicating that dysregulation of KPNA2 is not limited to adenocarcinomas, but is a common phenomenon in NSCLC. Using univariate linear regression analysis followed by a multiple stepwise linear regression model, no statistical association was observed between KPNA2 expression levels in lung tumor tissue specimens and patient gender, age, smoking, tumor cell mitosis, angiolymphatic invasion and tumor stage of NSCLC (Supplementary Table S1).

To ascertain whether KPNA2 expression is a prognostic factor in NSCLC, we examined the probabilities of disease-free and overall survival in 42 NSCLC patients with surgically resectable stage I and II disease. An inverse correlation between KPNA2 expression level and survival was observed. Specifically, early-stage NSCLC
patients with high KNPA2 levels tend to have a shorter survival time, although the probabilities of disease-free and overall survival were not significantly associated with KPNA2 expression ($p = 0.204$ and $0.306$, respectively) (Supplementary Fig. 2). Further investigations using a larger cohort with an increased number of patients are required to clarify this issue.

**Elevation of KPNA2 Serum Levels in NSCLC Patients.** As KPNA2 is overexpressed in NSCLC tissues and present in the secretome of lung cancer cell lines, we speculated that it may be detectable in serum samples of NSCLC patients. To determine the serum levels of KPNA2, we developed an in-house fluorometric sandwich ELISA technique, as described in Materials and Methods. The linear dynamic range for detection of soluble KPNA2 ranged from 7.8~500 ng/ml. We examined the KPNA2 levels in serum samples collected from NSCLC patients ($n = 126$; $n = 37$ for SCC and $n = 89$ for adenocarcinoma) and healthy controls ($n = 64$) using this ELISA system. As shown in Fig. 3A, the serum levels of KPNA2 were significantly higher in NSCLC patients versus those in healthy controls (mean ± SD, $646.8 ± 392.3$ ng/mL vs. $485.1 ± 168.5$ ng/mL, $p<0.001$, Supplementary Table S2).

We additionally determined the CEA levels in the same serum samples. Serum levels of CEA were higher in NSCLC patients versus healthy controls ($10 ± 16.9$ ng/mL vs. $0.1 ± 0.5$ ng/mL, $p <0.001$; Supplementary Table S2). At 5.0 ng/mL CEA, a cutoff value currently used for detection of lung cancer, sensitivity and specificity values
were 34.9% and 100%, respectively. Notably, upon selection of a cutoff value of 548.89 ng/mL for KPNA2, 38 of 82 lung cancer patients with CEA levels lower than 5.0 ng/mL were further distinguished from healthy individuals. Additionally, ROC curves were constructed for KPNA2, CEA and both proteins together (Fig. 3B). The area under the ROC curve (AUC), a commonly used indicator for evaluating the diagnostic efficacy of a potential biomarker, was determined as 0.63 (95% CI, 0.55–0.71) for KPNA2 and 0.78 (95% CI, 0.72–0.85) for CEA, respectively. Importantly, a combination of KPNA2 and CEA displayed higher diagnostic capacity than either marker alone (AUC = 0.89; 95% CI, 0.85–0.94; Fig. 3B). These results collectively indicate that KPNA2 is a potentially useful serum biomarker for NSCLC, particularly in conjunction with CEA.

Measurement of KPNA2 Levels in Pleural Effusion from Cancer Patients.

Here, we determine the KPNA2 protein levels in human serum specimens for the first time. KPNA2 levels in pleural effusion, a type of easily accessible body fluid frequently observed in patients suffering from specific cancer types, including lung cancer, breast cancer, gastric cancer and colon cancer, were evaluated. We assessed KPNA2 levels in pleural effusion from 82 NSCLC, 23 breast cancer, 12 gastric cancer and 13 colon cancer patients using ELISA, as described above. In NSCLC patients, the KPNA2 levels in pleural effusion were comparable to those in serum samples (699.5 ± 684.6 ng/mL vs 646.8 ± 392.3 ng/mL). Moreover, protein levels of KPNA2
in pleural effusion from NSCLC patients were significantly higher than those from non-lung cancer (691 ± 684.6ng/mL vs 361± 240.7 ng/mL) (Fig. 4A). The AUC was determined as 0.78 (95% CI, 0.70-0.87) for KPNA2 (Fig. 4B). Our data show for the first time that the KPNA2 level in pleural effusion of NSCLC patients is similar to that in NSCLC serum samples, further supporting the theory that KPNA2 is an ideal candidate pleural effusion marker to differentiate NSCLC from non-lung cancer.

**A Knockdown of KPNA2 Inhibits Migration Ability and Reduces Survival of Lung Cancer Cells.** KPNA2 is possibly involved in the regulation of cell proliferation, differentiation, DNA repair and tumorigenesis. To examine the possible role(s) of KPNA2 in tumor progression of lung cancer, we applied the siRNA approach to suppress the expression of endogenous KPNA2 in three lung cancer cell lines, CL1-0, CL1-5 and A549, and assessed the effects on cell migration and survival. A Western blot assay showed that KPNA2 protein levels were reduced significantly in cells transfected with KPNA2 siRNA, compared with control siRNA (Fig. 5A). Data obtained from the transwell migration assay indicate that the migration capability of KPNA2 knockdown cells is severely impaired, compared with that of control cells. The migration capacity was reduced to ~40% in KPNA2-depleted CL1-5 cells, and ~70 % in KPNA2-depleted CL1-0 or A549 cells (Figs. 5B and 5C). The MTT assay revealed decreased survival of cells transfected with KPNA2 siRNA, compared to
those with control siRNA. Cell viability was reduced to ~70 % in CL1-5 cells and ~90 % in CL1-0 cells, but A549 cells displayed marginal or no changes (Fig. 5D). Notably, cell survival was significantly decreased in KPNA2-suppressed CL1-5 cells. Examination of cleaved PARP, the biochemical marker for apoptosis, revealed upregulated levels in KPNA2-depleted CL1-5 cell lysates, compared to control cell lysates (Fig. 5A, middle panel). Accordingly, we propose that apoptosis contributes to the reduced survival of CL1-5 cells treated with KPNA2 siRNA. Our results collectively indicate that KPNA2 is involved in cell migration and survival during tumorigenesis of lung cancer.
Conclusions

In this study, we integrate cancer cell secretome and tissue transcriptome dataset to identify potential lung cancer serum biomarkers. In total, we identified 35 potential biomarkers (16 well known and 19 novel candidates) for lung cancer. The novel candidate, KPNA2, displays higher mRNA expression in pulmonary adenocarcinoma tissues and is secreted/released from both the lung cancer cell lines analyzed. In our experiments, KPNA2 was overexpressed in NSCLC tissues, and subcellular localization of KPNA2 was altered in tumor cells. Furthermore, KPNA2 levels were successfully determined in serum and pleural effusion samples with in-house ELISA. Higher serum levels of KPNA2 were detected in NSCLC patients, compared to healthy controls. Higher pleural effusion levels of KPNA2 were also detected in NSCLC patients, compared to non-lung cancer patients. Our results additionally show that KPNA2 is involved in the regulation of lung cancer cell migration, viability and apoptosis. These findings support the feasibility of this strategy in effectively identifying potential lung cancer serum biomarkers.

Acknowledgements

This work was supported by grants from the Chang Gung Medical Research Fund (CMRPG33099, CMRPD160097 and CMRPD 150172) and National Science Council, R.O.C. (NSC 96-2320-B-182-035 and 97-2320-B-182-026-MY3) as well as the
Supporting Information Available

Supplementary Table S1, relationship between tissue KPNA2 expression and clinicopathologic characteristics of NSCLC patients; Supplementary Table S2, serum levels of KPNA2 and CEA in NSCLC patients with different clinical characteristics; Supplementary reference; Supplementary Figure S1, altered distribution of NBS1 in KPNA2-positive NSCLC tissues; Supplementary Figure S2, Kaplan-Meier estimates of disease-free and overall survival in NSCLC: relationship to KPNA2 protein expression. This information is available free of charge via the Internet at http://pubs.acs.org.
Figure Legends

Figure 1. Schematic illustration of the strategy used to identify potential serum biomarkers for NSCLC. The diagram represents integration of CL1-0 and CL1-5 secretome datasets with the GDS1650 pulmonary adenocarcinoma tissue transcriptome dataset. The CL1-0 and CL1-5 secretome datasets contain 1096 and 1830 proteins, while the GDS1650 transcriptome dataset contains 285 unique genes with tumor/normal ratio > 4-fold. In total, 35 potential biomarkers are presented in single or both secretome datasets.

Figure 2. Detection of KPNA2 in NSCLC tissues. (A) Immunohistochemical analysis of KPNA2 in paired tumor and adjacent normal tissues from one representative NSCLC case. Brown signals indicate the distribution of KPNA2-positive staining in nuclei of tumor cells. (B) Two representative cases of well and poorly differentiated NSCLC are shown.

Figure 3. Significant increase in serum KPNA2 levels of NSCLC patients. (A) Serum levels of KPNA2 in healthy controls (Control) and NSCLC patients were determined using ELISA, as described in Materials and Methods. Data are presented as upper and lower quartiles and range (box), median value (horizontal line), and middle 90% distribution (dashed line). (B) ROC curve analysis of the diagnostic
efficacy of KPNA2, CEA, and a combination of both proteins.

Figure 4. Significant increase in pleural effusion KPNA2 levels of NSCLC. (A) The protein levels of KPNA2 in pleural effusion from patients with NSCLC and non-lung cancer were determined by fluorometric sandwich ELISA, as described in Materials and Methods. Data are presented as upper and lower quartile and range (box), median value (horizontal line), and the middle 90% distribution (dashed line). A $p$ value of less than 0.05 indicates statistical significance using the nonparametric Mann-Whitney U test. (B) ROC curve analysis of the diagnostic efficacy of KPNA2.

Figure 5. KPNA2 involves in the migration ability and survival of lung cancer cells. (A) Lung cancer cell lines were transfected with control siRNA or KPNA2-specific siRNA. After transfection for 2 days, cell lysates were prepared and the extracted proteins (20 μg) analyzed by Western blotting. Simultaneously, cells were subjected to migration and MTT assays, as described in Materials and Methods. (B) Representative microphotographs of filters obtained from the migration assay. Original magnification: 100×. Quantitative analysis of the migration assay (C) and MTT assay (D). Data are presented as mean values obtained from three independent experiments. Error bars indicate standard deviations. A $p$ value of less than 0.05 indicates statistical significance using the unpaired Student's t-test.
References


Yu, C. J.; Lee, I. N.; Chang, Y. J.; Lee, S. Y.; Yeh, Y. M.; Chang, Y. S.; Chien, K. Y.;
Yu, J. S., Enhanced interferon signaling pathway in oral cancer revealed by
quantitative proteome analysis of microdissected specimens using 16O/18O labeling
and integrated two-dimensional LC-ESI-MALDI tandem MS. *Mol. Cell Proteomics*

29. Ravn, V.; Rasmussen, B. B.; Hojholt, L.; Barfoed, M.; Heiberg, I.; Thorpe, S. M.,
Reproducibility of subjective immunohistochemical estrogen- and progesterone
1015-22.


31. Zweig, M. H.; Campbell, G., Receiver-operating characteristic (ROC) plots: a


33. Conti, E.; Muller, C. W.; Stewart, M., Karyopherin flexibility in


Table 1. Protein list of 35 potential serum biomarkers for NSCLC.

<table>
<thead>
<tr>
<th>Gene name</th>
<th>Accession Number</th>
<th>CL₁₀</th>
<th>CL₁-₅</th>
<th>Lung cancer markers&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABP1</td>
<td>P19801</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C5</td>
<td>P01031</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CIT</td>
<td>O14578</td>
<td>●</td>
<td>●</td>
<td>1, 2</td>
</tr>
<tr>
<td>COL3A1</td>
<td>P02461</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>COL5A2</td>
<td>P05997</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ECM1</td>
<td>Q16610</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EEF1A2</td>
<td>Q05639</td>
<td>●</td>
<td>●</td>
<td>3</td>
</tr>
<tr>
<td>GINS1</td>
<td>Q14691</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>GMDS</td>
<td>O60547</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>HMG1A1</td>
<td>P17096</td>
<td>●</td>
<td>●</td>
<td>4, 5</td>
</tr>
<tr>
<td>HMG2A</td>
<td>P52926</td>
<td>●</td>
<td></td>
<td>6, 7</td>
</tr>
<tr>
<td>HMG3B</td>
<td>O15347</td>
<td>●</td>
<td>●</td>
<td>8</td>
</tr>
<tr>
<td>ITGA2</td>
<td>P17301</td>
<td>●</td>
<td></td>
<td>9</td>
</tr>
<tr>
<td>KIF23</td>
<td>Q02241</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>KPNA2</td>
<td>P52292</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>MAD2L1</td>
<td>Q13257</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MDK</td>
<td>P21741</td>
<td>●</td>
<td>●</td>
<td>10, 11</td>
</tr>
<tr>
<td>MMP1</td>
<td>P03956</td>
<td>●</td>
<td></td>
<td>12</td>
</tr>
<tr>
<td>MMP9</td>
<td>P14780</td>
<td>●</td>
<td></td>
<td>13</td>
</tr>
<tr>
<td>NME1</td>
<td>P15531</td>
<td>●</td>
<td>●</td>
<td>14</td>
</tr>
<tr>
<td>NQO1</td>
<td>P15559</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAEP</td>
<td>P09466</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PAFH1B3</td>
<td>Q15102</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>PAICS</td>
<td>P22234</td>
<td>●</td>
<td>●</td>
<td>15</td>
</tr>
<tr>
<td>PFKP</td>
<td>Q01813</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>PPAT</td>
<td>Q06203</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PYCR1</td>
<td>A6NFM2</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RRM2</td>
<td>P31350</td>
<td>●</td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>SERPINB5</td>
<td>P36952</td>
<td>●</td>
<td>●</td>
<td>17</td>
</tr>
<tr>
<td>SLC7A5</td>
<td>Q01650</td>
<td>●</td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>SORD</td>
<td>Q00796</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>SPP1</td>
<td>P10451</td>
<td>●</td>
<td></td>
<td>19, 20</td>
</tr>
<tr>
<td>TOP2A</td>
<td>A6NDFM0</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRIP13</td>
<td>Q15645</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UCHL1</td>
<td>P09936</td>
<td>●</td>
<td>●</td>
<td>21</td>
</tr>
</tbody>
</table>

<sup>a</sup>Protein detected in secretome dataset.

<sup>b</sup>References are denoted in Supporting Information.
Table 2. Immunohistochemical analysis of KPNA2 expression in 66 NSCLC tissue specimens containing both tumor and adjacent-tumor cells.

<table>
<thead>
<tr>
<th>KPNA2 localization</th>
<th>Stain score(^a)</th>
<th>Patient number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0(^b)</td>
<td>1(^c)</td>
</tr>
<tr>
<td>Normal nuclei</td>
<td>66 (100)</td>
<td>0 (0)</td>
</tr>
<tr>
<td>Tumor nuclei</td>
<td>19 (28.8)</td>
<td>39 (59.1)</td>
</tr>
<tr>
<td>Normal cytoplasm</td>
<td>0 (0)</td>
<td>66 (100)</td>
</tr>
<tr>
<td>Tumor cytoplasm</td>
<td>48 (72.7)</td>
<td>18 (27.3)</td>
</tr>
</tbody>
</table>

\(^a\) Stain score was calculated according to overall percentage of stained cell combined with the median intensity of staining, as described in Materials and Methods.

\(^b\) Score 0: negative staining.

\(^c\) Score 1 to 3: positive staining.
**Table of Contents Synopsis**

To discover serum biomarkers for improved non-small cell lung cancer (NSCLC) diagnosis and prognosis, the NSCLC secretome and transcriptome datasets were analyzed. KPNA2, a protein secreted by NSCLC cells and over-expressed in NSCLC patients, was selected and verified as a potential NSCLC serum biomarker using clinical specimens by immunohistochemistry and ELISA. To integrated cancer cell secretome and transcriptome datasets provides a feasible strategy for efficient identification of novel NSCLC serum biomarkers.
Figure 1

CL_{1-0} secretome dataset (1096)

CL_{1-5} secretome dataset (1830)

Transcriptome dataset (285)

161

917

884

12

6

17

250
Figure 2

(A) Tumor

200X

Adjacent normal

400X

(B) Well differentiated

Poor differentiated
Figure 3

(A) Boxplot showing the distribution of KPNA2 levels (ng/mL) in Control (n=64) and NSCLC (n=126) groups. The boxplot indicates a statistically significant difference between the two groups, with a p-value of less than 0.001.

(B) ROC curves for CEA (AUC=0.78), KPNA2 (AUC=0.63), and the combined marker (AUC=0.83). The combined marker shows the highest sensitivity and specificity.
Figure 4

(A)

<table>
<thead>
<tr>
<th>KPNA2 (+)</th>
<th>Tumor</th>
<th>Adjacent normal</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Image" /></td>
<td><img src="image2.png" alt="Image" /></td>
<td><img src="image3.png" alt="Image" /></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>NBS1</th>
<th>Tumor</th>
<th>Adjacent normal</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image4.png" alt="Image" /></td>
<td><img src="image5.png" alt="Image" /></td>
<td><img src="image6.png" alt="Image" /></td>
</tr>
</tbody>
</table>
Figure 5

(A) Table showing migration and relative viability percentages for different conditions.

<table>
<thead>
<tr>
<th>Sample</th>
<th>CL_{1.0}</th>
<th>CL_{1.5}</th>
<th>A549</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control siRNA</td>
<td>siRNA</td>
<td>siRNA</td>
<td>siRNA</td>
</tr>
<tr>
<td>KPNA2 siRNA</td>
<td>siRNA</td>
<td>siRNA</td>
<td>siRNA</td>
</tr>
</tbody>
</table>

(B) Images showing cell morphology under different conditions.

(C) Bar graphs showing migration and relative viability percentages for different conditions.

- CL_{1.0} p=0.054
- CL_{1.5} p<0.05
- A549 p=0.87
Supplement Figure. 1. Kaplan-Meier estimates of disease-free and overall survival in patients with surgically resectable stage I and II NSCLC: relationship to KPNA2 protein expression. Probabilities of (A) disease-free survival and (B) overall survival in 42 NSCLC patients. KPNA2 protein expressions of NSCLC specimens were examined by immunohistochemical analysis, and the experiment procedure was described in materials and methods. Cases were considered high KPNA2 expression if the scoring of tumor nuclear stain intensity plus percentage scores (intensity scores: 0-3, percentage scores: 0-3; sum of both: 0-6) was more than 3; low PKNA2 expression if stain intensity plus percentage scores was equal or less than 3. Blue line, patients with low KPNA2 staining; green line, patients with high KPNA2 staining. P-value was determined using a two-sided log-rank test.
Supplement Figure. 2. The pleural effusion levels of KPNA2 are significantly higher than those from other cancers. The pleural effusion levels of KPNA2 with lung cancer, breast cancer and colon cancer patients were determined by ELISA, and the experiment procedure was described in materials and methods. Data are presented as upper and lower quartile and range (box), median value (horizontal line), and the middle 90% distribution (dashed line). A $p$ value of less than 0.05 indicates statistical significance using the nonparametric Mann-Whitney U test.
Lung cancer (n=82)
Non-Lung cancer (n=48)

KPNA2 (ng/mL)

$AUC=0.78$
p < 0.001