

# Label-Free Quantitative Proteomics and N-terminal Analysis of Human Metastatic Lung Cancer Cells

Hophil Min<sup>1</sup>, Dohyun Han<sup>1,2</sup>, Yikwon Kim<sup>1</sup>, Jee Yeon Cho<sup>3</sup>, Jonghwa Jin<sup>1</sup>, and Youngsoo Kim<sup>1,2,\*</sup>

**Proteomic analysis is helpful in identifying cancer-associated proteins that are differentially expressed and fragmented that can be annotated as dysregulated networks and pathways during metastasis. To examine metastatic process in lung cancer, we performed a proteomics study by label-free quantitative analysis and N-terminal analysis in 2 human non-small-cell lung cancer cell lines with disparate metastatic potentials—NCI-H1703 (primary cell, stage I) and NCI-H1755 (metastatic cell, stage IV). We identified 2130 proteins, 1355 of which were common to both cell lines. In the label-free quantitative analysis, we used the NSAF normalization method, resulting in 242 differential expressed proteins. For the N-terminal proteome analysis, 325 N-terminal peptides, including 45 novel fragments, were identified in the 2 cell lines. Based on two proteomic analysis, 11 quantitatively expressed proteins and 8 N-terminal peptides were enriched for the focal adhesion pathway. Most proteins from the quantitative analysis were upregulated in metastatic cancer cells, whereas novel fragment of CRKL was detected only in primary cancer cells. This study increases our understanding of the NSCLC metastasis proteome.**

## INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths worldwide (30%) but constitutes only 15% of new cancer diagnoses (Parkin and Fernandez, 2006). Despite of the advances in cancer research, the 5-year survival rate of lung cancer remains low at 16%, compared with 65% for colon cancer, 89% for breast cancer, and 100% for prostate cancer (Jemal et al., 2010). Lung cancer is divided into 2 major histological types: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC) (Hoffman et al., 2000). SCLC is commonly treated

with chemotherapy and radiotherapy, and NSCLC is usually treated with surgery. Yet, surgery for NSCLC is effective only in those who are diagnosed at an early stage. More than 70% of NSCLC patients are diagnosed at the late stage with metastasis, resulting in a loss of opportunity for effective surgery and, ultimately, a poor prognosis (Tan et al., 2012).

Metastasis is a major cause of death from lung cancer that accompanies several processes, including the detachment of cancer cells, invasion of cancer cells into the surrounding tissue, and colonization of and proliferation in distant organs (Hwang et al., 2012; Tian et al., 2007). During metastasis, irreversible protein fragmentation occurs (Lopez-Otin and Bond, 2008). Dysregulation of protein fragment reactions in organs can cause pathological developmental disorders, such as cancer, inflammation, infection, and Alzheimer disease (Dawson and Dawson, 2003; Opferman and Korsmeyer, 2003; Rao, 2003).

In lung cancer, serum cytokeratin 19 fragments (CYFRA 21-1) are generated by protein fragmentation reaction and have recently been implicated as a biomarker for the diagnosis and prognosis of NSCLC (Nisman et al., 2008). Pro1708/Pro2044 (the C-terminal fragment of albumin) (Kawakami et al., 2005) and HER2 rb2 (the ectodomain of human epithelial growth factor receptor-2) (Streckfus et al., 1999) are also cancer biomarkers that are generated by protein fragmentation. The identification of natural protease substrates and their cleavage sites is essential information with which we can understand the regulation of metastatic pathways. Thus, the pathways that culminate in protein fragment events must be examined to develop novel and more effective molecular markers and therapeutic targets.

Proteomic analysis for global protein identification is a powerful tool that can be used to identify novel biomarkers in various diseases. Of such methods, label-free quantification determines the expression levels of nontarget proteins (Fanayan et al., 2013). Many global quantitative proteomics studies have examined metastasis in various cancers, such as colorectal cancer (Xue et al., 2010), breast cancer (Xie et al., 2010), and hepatocellular carcinoma (Wang et al., 2011). However, there are few reports on the proteomic profile in metastatic lung cancer. For instance, Tian et al. identified metastasis-related proteins in NSCLC cell lines (nonmetastatic CL1-0 and the highly metastatic CL1-5) by 2-DE analysis (Tian et al., 2007).

The recent development of N-terminal peptide analysis, based on mass spectrometry, has enabled us to generate data on the protein targets and fragment sites (Brown and Hartley, 1966). To this end, several groups have established a method

<sup>1</sup>Department of Biomedical Sciences, <sup>2</sup>Institute of Medical and Biological Engineering, Medical Research Center, Seoul National University College of Medicine, Seoul 110-799, Korea, <sup>3</sup>Division of Life Sciences and Biotechnology, Korea University, Seoul 136-701, Korea  
\*Correspondence: biolab@snu.ac.kr

Received 20 February, 2014; revised 7 April, 2014; accepted 8 April, 2014; published online 8 May, 2014

**Keywords:** label-free quantitative analysis, metastasis, N-terminal analysis, non-small-cell lung cancer

of identifying protease-generated (neo) peptides in cellular pathways, known as N-terminomics (Enoksson et al., 2007). Combined fractional diagonal chromatography (COFRADIC) is a pioneering technique in N-terminomics. Free amines of proteins are first acetylated prior to trypsin digestion and RP-HPLC fractionation. The N-termini of neo peptides are then derivatized with a hydrophobic reagent allow the original N-terminal peptides to be purified on rechromatography (Gevaert et al., 2003). However, the COFRADIC method requires many HPLC and LC-MS/MS runs and large amounts of starting material to select N-terminal neo peptides. McDonald and Beynon (2006) developed a more rapid and simpler N-terminal peptide analysis method (positional proteomics) that is based on negative selection by chemical labeling of the  $\alpha$ -amine in proteins.

In this study, to differentiate primary cancer cells from metastatic cells, we performed 2 parallel experiments: label-free quantification and N-terminal peptide analysis (positional proteomics methods) by LC-MS/MS. Human non-small-cell lung cancer cell lines were used—NCI-H1703, a stage I primary cancer cell, and NCI-H1755, a stage IV metastatic cancer line (Anisowicz et al., 2008). Our label-free quantification identified 2130 proteins from the LC-MS/MS analysis, 242 of which were differentially expressed between NCI-H1703 and NCI-H1755 cells. Analysis of N-terminal neo peptides identified 325 N-terminal peptides, 45 of which were observed in both cell lines. This differential expression of the proteome and N-terminal neo peptides can increase our understanding of differentially regulated pathways between primary and metastatic cancer cells in human non-small-cell lung cancer.

## MATERIALS AND METHODS

### Reagents and chemicals

HPLC-grade water, HPLC-grade acetonitrile (ACN), and HPLC-grade methanol (MeOH) were obtained from FISHER (USA). Hydrochloric acid (HCl) and sodium chloride (NaCl) were purchased from DUKSAN (Korea). Urea and dithiothreitol (DTT) were purchased from AMRESCO (USA). Phenylmethanesulfonyl fluoride (PMSF), sodium dodecyl sulfate (SDS), and Tris were obtained from USB (USA). Complete protease inhibitor cocktail tablets were acquired from ROCHE (USA), and sequencing-grade modified trypsin was purchased from PRO-MEGA (USA). Sulfo-NHS acetate and NHS-Activated agarose slurry were obtained from Pierce (USA). All other reagents—iodoacetamide,  $\alpha$ -cyano-4-hydroxycinnamic acid (CHCA), and trifluoroacetic acid (TFA)—were purchased from Sigma-Aldrich (USA).

### Cell cultures and lysis

Stage 1 (NCI-H1703) and stage 4 non-small-cell lung cancer cells (NCI-H1755) were obtained from the Korean Cell Line Bank. Both lines were cultured in RPMI1640 (WELGENE, Korea) with 10% fetal bovine serum (Gibco, USA), 100 U/ml penicillin and 100  $\mu$ g/ml streptomycin (Gibco, USA) and 25 mM HEPES (Gibco, USA). The cultures were maintained in 95% humidified air and 5% CO<sub>2</sub> at 37°C.

To prepare the cell lysates, cells were grown to 80% confluence and lysed in strong SDS-based buffer, containing 4% SDS, 0.1 mM PMSF, 1 $\times$  protease inhibitor cocktail, 0.1 M DTT, and 0.1 M HEPES. Lysates were incubated at 95°C for 5 min and sonicated for 1 min. Supernatants were collected from the lysates by centrifugation at 15,000  $\times$  g for 20 min at 4°C. Protein concentrations were measured using the BCA Protein Assay Kit – reducing reagent-compatible (Pierce, USA). Finally,

each cell lysate was stored in 0.2-mg aliquot at -80°C until use.

### Filter-aided sample preparation (FASP)

Cell lysates were processed by filter-aided sample preparation (FASP) (Wisniewski et al., 2009) using a 10 K molecular weight cutoff (MWCO) filter (Millipore, USA). Briefly, 200  $\mu$ g of cell lysates in lysis buffer (4% SDS, 0.1 mM PMSF, 1 $\times$  protease inhibitor cocktail, 0.1 M DTT, and 0.1 M HEPES) was transferred to the filter and mixed with 0.2 ml 8 M urea in 0.1 M HEPES, pH 7.5 (FASP solution). Samples were centrifuged at 14,000  $\times$  g at 20°C for 20 min. The samples in the filter were diluted with 0.2 ml FASP solution and centrifuged again. The reduced cysteines remained in 0.1 ml 50 mM iodoacetamide in FASP solution, were incubated at room temperature (RT) in the dark for 30 min, and centrifuged for 20 min.

For the label-free quantification, alkylated samples were mixed with 0.2 ml 50 mM Tris solution and centrifuged at 14,000  $\times$  g at 20°C for 20 min; this step was repeated 3 times. One hundred microliters 50 mM Tris solution with trypsin (enzyme:protein ratio 1:80) was added to the resulting concentrate and incubated for 16 h at 37°C. Peptides were collected from the filter by centrifugation for 20 min to new collection tubes and acidified with 2% TFA.

### Labeling of N-terminal neo peptides

Alkylated samples were mixed with 0.1 ml 50 mM HEPES with Sulfo-NHS acetate (Sulfo-NHS acetate:protein ratio at 25:1) and incubated for 2 h at RT. The samples were centrifuged at 14,000  $\times$  g at 20°C for 20 min, mixed with 0.2 ml 1 M Tris solution, and incubated on the filter for 4 h at RT. The samples were then centrifuged at 14,000  $\times$  g at 20°C for 20 min 4 times. One hundred microliters 50 mM Tris solution with trypsin (enzyme:protein ratio of 1:80) was added to the filter and incubated for 16 h at 37°C. Digested peptides were collected by centrifugation and acidified with 2% TFA.

### Desalting of peptides

Digested samples were desalted using in-house C<sub>18</sub> StageTip desalting (STD) columns, as described (Han et al., 2012). Briefly, in-house C<sub>18</sub> STD columns were prepared by reversed-phase packing of POROS 20 R2 material into 0.2-ml yellow pipet tips that sat atop C<sub>8</sub> empore disk membranes. The STD columns were washed with 0.1 ml 100% methanol and with 0.1 ml 100% ACN 3 times and equilibrated 3 times with 0.1 ml 0.1% TFA. After the peptides were loaded, the STD columns were washed 3 times with 0.1 ml 0.1% TFA, and the peptides were eluted with 0.1 ml of a series of elution buffers, containing 0.1% TFA and 40, 60, and 80% ACN. All eluates were combined and dried in a vacuum centrifuge.

### Enrichment of labeled N-terminal peptides

Dried samples were dissolved in bupH™ PBS (Pierce, USA). One milliliter of an NHS-agarose bead slurry (50% slurry in acetone) was prepared per the manufacturer's protocol (Pierce, USA). Briefly, acetone was removed from the slurry by centrifugation, and the slurry was washed 2 times with water and equilibrated 3 times with bupH™ PBS. After mixing with the equilibrated beads, the labeled samples were incubated for 4 h at RT. Finally, the beads were centrifuged at 1,000  $\times$  g for 30 s, and the supernatant was transferred to new tubes, acidified with 2% TFA, and desalted again.

### MALDI-MS/MS analysis

Bovine serum albumin (BSA) peptides (Amresco, USA) were N-terminally labeled as described above as control. The peptides

were dissolved in 10  $\mu$ l 0.1% TFA, and 0.5  $\mu$ l of each sample was mixed with 0.5  $\mu$ l of a matrix solution that contained 5 mg/ml CHCA (Sigma, USA), 70% ACN, and 0.1% TFA. The peptides were spotted directly onto a MALDI plate (Opti-TOF™ 384-well Insert, Applied Biosystems, USA) and crystallized with the matrix. Dried peptides were analyzed on a 4800 MALDI-TOF/TOF™ Analyzer (Applied Biosystems) that was equipped with a 355-nm Nd:YAG laser. The pressure in the TOF analyzer was approximately  $7.6 \times 10^{-7}$  Torr.

The mass spectra were obtained in the reflectron mode over an m/z range of 800–3500 Da with an accelerating voltage of 20. External calibration was performed using des-Arg-Bradykinin (904,468 Da), angiotensin 1 (1,296.685 Da), Glu-Fibrinopeptide B (1,570.677 Da), adrenocorticotrophic hormone (ACTH) (1–17) (2,093.087 Da), and ACTH (18–39) (2,465.199) (4700 calibration mixture, Applied Biosystems). Raw data were reported by 4000 SERIES EXPLORER, v4.4 (Applied Biosystems).

### LC-ESI-MS/MS analysis

All peptide samples were analyzed on an LTQ-Orbitrap Velos mass spectrometer (Thermo Scientific, USA) that was coupled to an EasyLC II (Proxeon Biosystems, Denmark), equipped with a nanoelectrospray device and fitted with a 10- $\mu$ m fused silica emitter tip (New Objective, USA). Ten microliters of each sample was loaded onto a nano-LC trap column (ZORBAX 300SB-C<sub>18</sub>, 5  $\mu$ m, 0.3  $\times$  5 mm, Agilent, USA), and peptides were separated on a C<sub>18</sub> analytical column (75  $\mu$ m  $\times$  15 cm) that was packed in-house with C<sub>18</sub> resin (Magic C18-AQ 200 Å, 5- $\mu$ m particles). Solvent A was 98% water with 0.1% formic acid and 2% ACN, and Solvent B was 98% ACN with 0.1% formic acid and 2% water.

Peptides were separated using a 180-min gradient at 300 nl/min, comprising 0% to 40% B for 120 min, 40% to 60% B for 20 min, 60% to 90% B for 10 min, 90% B for 10 min, 90% to 5% B for 10 min, and 0% B for 10 min. The spray voltage was set to 1.8 kV, and the temperature of the heated capillary was 200°C. The mass spectrometer scanned a mass range of 300 to 2000. The data on the top 10 most abundant ions were analyzed in data-dependent scan mode over a minimum threshold of 1000. The normalized collision energy was adjusted to 35%, and the dynamic exclusion was set to a repeat count of 1, repeat duration of 30 s, exclusion duration of 60 s, and  $\pm$  1.5 m/z exclusion mass width. Each biological replicate was analyzed in triplicate.

### Peptide identification and label-free quantification

After the data acquisition, data searches were performed using SEQUEST Sorcerer (Sage-N Research, USA). Raw files from the LTQ-Orbitrap Velos were converted into mzXML files using Trans-Proteomics Pipeline (TPP, ISB, USA). MS/MS data were searched using a target decoy database strategy against a composite database that contained the International Protein Index (IPI) human database (v3.87, 91,464 entries), and its reverse sequences were generated using Scaffold 3 (Proteome Software Inc., USA).

For the label-free quantification dataset and N-terminal peptide data, 2 independent search parameters were used. Parameters for the label-free quantification dataset were as follows: enzyme, full-trypsin; peptide tolerance, 10 ppm; MS/MS tolerance, 1.0 Da; variable modifications, oxidation (M); and static modifications, carbamidomethylation (Cys). Identified proteins were filtered using Scaffold 3, based on a minimum of 2 unique peptides and false discovery rate (FDR) < 1%. The parameters for N-terminal peptide dataset were as follows:

enzyme, semi-arginine; peptide tolerance, 10 ppm; MS/MS tolerance, 1.0 Da; variable modifications, oxidation (Met); and static modifications, carbamidomethylation (Cys) and acetylation (N-term and Lys). Peptide-spectrum matches were filtered to have less than a 1% FDR by calculating the statistics tool in TPP.

The label-free quantitative analysis of peptides was performed by spectral counting analysis. To calculate a protein spectrum count, we exported the numbers of peptides that were assigned to each protein from Scaffold 3. Exported data were analyzed by normalized spectral abundance factor (NSAF) method to normalize run-to-run variations (Zybaylov et al., 2006). NSAF values were calculated as:

$$\text{NSAF} = (\text{SpC} / \text{Mw}) / \sum (\text{SpC} / \text{Mw}) n$$

where SpC is the spectral count, Mw is the molecular weight in kDa, and n is the total number of proteins. Because some expression ratios that are calculated from spectral counts of 0, causing certain data to be represented as '#DIV/0!' in Microsoft Office Excel 2010, we shifted the entire spectral count equally by adding 0.1 to the original values. By NSAF method, we could compare expression levels and apply independent 2-sample t-test of each protein in the cell lines.

### Bioinformatics analysis

Data were analyzed using various bioinformatics tools. To determine N-terminal peptide sites, we performed manual annotations using UniProtKB (Universal Protein Resource Knowledgebase) (<http://www.uniprot.org/>). The N-termini were categorized into 6 types, based on molecule processing part of each protein sequence annotation in UniProtKB: initial methionine depletion, initial methionine nondepletion, signal peptide depletion, propeptide depletion, mitochondrial transit peptide depletion, and novel N-terminal neo peptide. Novel N-terminal neo peptides were annotated with peptides that were not included in the other 5 categories.

The biological process and molecular function classifications of identified proteins were analyzed using PANTHER ID numbers (<http://www.pantherdb.org/>). Functional pathways were analyzed using the KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway.

## RESULTS

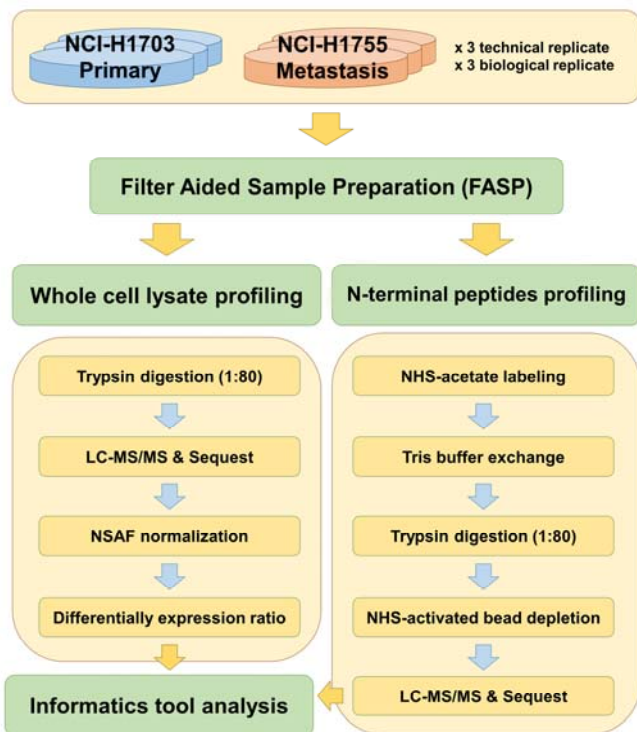
### Overall scheme

To differentiate the proteomic changes between primary and metastatic cells, whole-cell lysates of cultured human non-small-cell lung cancer cell lines (NCI-H1703 and NCI-H1755) were analyzed in parallel experiments, as depicted in Fig. 1. Each cell line was cultured as 3 independent biological replicates and prepared by FASP.

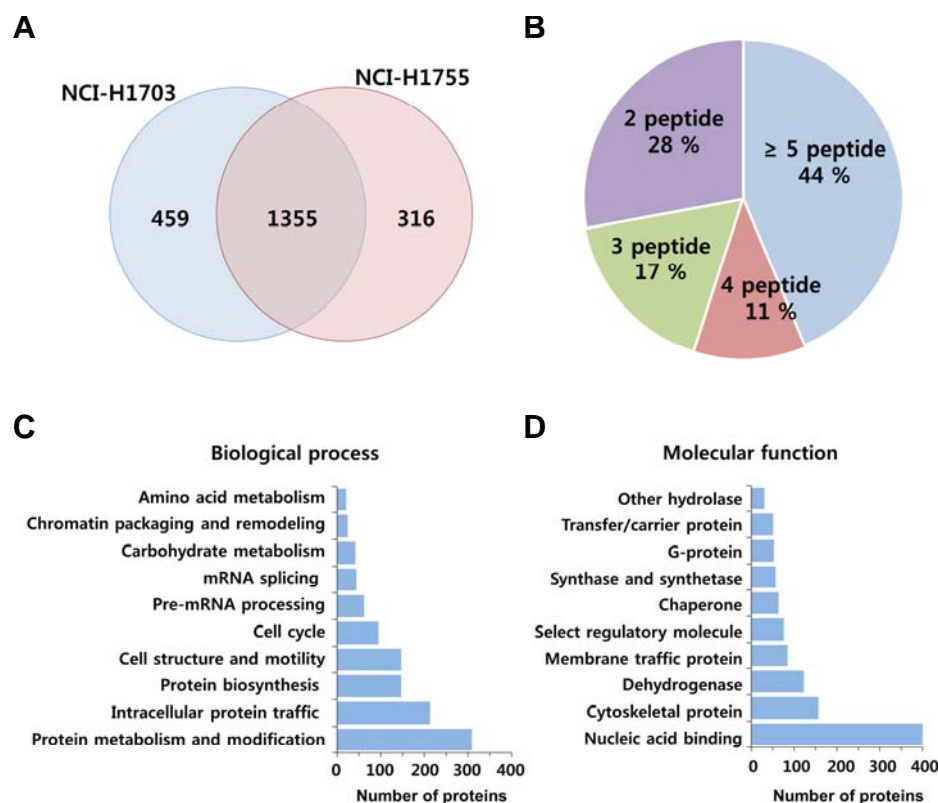
For the label-free quantitative proteomic analysis, cell lysates were digested with trypsin and desalted with a C<sub>18</sub> in-house stage tip prior to LTQ-Orbitrap Velos analysis. To ensure the reliability of the quantitative profiling, each sample was injected in triplicate (3 technical replicates) for each biological replicate. A total of 18 raw files from the LTQ-Orbitrap Velos were processed in Scaffold 3 with the SEQUEST algorithm.

To analyze the N-terminal peptide data, free amines in the cell lysates were labeled by NHS-acetate. The remaining NHS-acetate was quenched by the amine group of Tris. N-terminally labeled proteins were digested with trypsin and desalted using C<sub>18</sub> in-house stage tips and filtered by NHS-activated beads that depleted the newly generated N-termini by trypsin. The super-

nanants of the N-terminal peptide samples were desalted using



**Fig. 1.** Overall scheme. In this study, we performed comprehensive study of metastatic lung cancer using label-free quantitative analysis and N-terminal peptides analysis methods in human non-small lung cancer cell lines with different metastasis potential such as NCI-H1703 and NCI-H1755.



**Fig. 2.** Identification and proteome analysis of two different cell lines. (A) All identified proteins number were shown by Venn diagram. (B) All proteins were identified by greater 2 unique peptides. (C) Gene ontology (GO) biological process and (D) molecular function analysis with all identified proteins was performed by DAVID tool.

**Table 1.** Top 15 up- and down-regulated proteins

IPI <sup>a</sup>	MW (kDa)	Ratio <sup>b</sup>	p-value <sup>c</sup>	Gene symbol	Protein name
<i>Up-regulated protein</i>					
IPI00013744	129.3	6.60	0.0001	ITGA2	Integrin alpha-2
IPI00006663	56.4	6.57	0.0004	ALDH2	Aldehyde dehydrogenase, mitochondrial
IPI00553131	38.3	5.90	0.0003	GALE	UDP-glucose 4-epimerase
IPI00413641	35.9	3.81	0.0022	AKR1B1	Aldose reductase
IPI00216008	62.5	3.35	0.0023	G6PD	Isoform Long of Glucose-6-phosphate 1-dehydrogenase
IPI00017376	86.5	2.86	0.0015	SEC23B	Protein transport protein Sec23B
IPI00215743	152.5	2.60	0.0001	RRBP1	Isoform 3 of Ribosome-binding protein 1
IPI00001539	41.9	2.38	0.0009	ACAA2	3-ketoacyl-CoA thiolase, mitochondrial
IPI00292771	238.3	2.23	0.0018	NUMA1	Isoform 1 of Nuclear mitotic apparatus protein 1
IPI00744692	37.5	2.22	0.0000	TALDO1	Transaldolase
IPI00643920	68.8	2.04	0.0016	TKT	cDNA FLJ54957, highly similar to Transketolase
IPI00414717	134.6	2.03	0.0223	GLG1	Isoform 2 of Golgi apparatus protein 1
IPI00219525	51.9	2.01	0.0001	PGD	6-phosphogluconate dehydrogenase, decarboxylating
IPI00027223	46.6605	2.01	0.0010	IDH1	Isocitrate dehydrogenase [NADP] cytoplasmic
IPI00003479	41.3919	2.00	0.0082	MAPK1	Mitogen-activated protein kinase 1
<i>Down-regulated protein</i>					
IPI00001453	55.4	-7.47	0.0055	INA	Alpha-internexin
IPI00397526	230.8	-6.87	0.0039	MYH10	Isoform 1 of Myosin-10
IPI00607787	58.7	-6.64	0.0041	UAP1	Isoform 3 of UDP-N-acetylhexosamine pyrophosphorylase
IPI00856045	616.6	-6.59	0.0028	AHNAK2	Isoform 1 of Protein AHNAK2
IPI00333619	54.8	-6.52	0.0027	ALDH3A2	Isoform 1 of Fatty aldehyde dehydrogenase
IPI00178150	139.9	-6.36	0.0083	KIF4A	Isoform 1 of Chromosome-associated kinesin KIF4A
IPI00237884	181.0	-6.26	0.0416	AKAP12	Isoform 1 of A-kinase anchor protein 12
IPI00218775	51.2	-6.12	0.0051	FKBP5	Peptidyl-prolyl cis-trans isomerase FKBP5
IPI00023972	50.6	-6.11	0.0041	DDX47	Probable ATP-dependent RNA helicase DDX47
IPI00003505	48.6	-5.96	0.0039	TRIP13	Isoform 1 of Pachytene checkpoint protein 2 homolog
IPI00396627	92.1	-5.95	0.0118	ELAC2	Isoform 1 of Zinc phosphodiesterase ELAC protein 2
IPI00022977	42.6	-5.89	0.0085	CKB	Creatine kinase B-type
IPI00294187	75.6	-5.89	0.0008	PADI2	Protein-arginine deiminase type-2
IPI00017303	104.7	-5.89	0.0201	MSH2	DNA mismatch repair protein Msh2
IPI00218922	88.0	-5.77	0.0110	SEC63	Translocation protein SEC63 homolog

<sup>a</sup>IPI accession number of each protein<sup>b</sup>Significant difference expression log<sub>2</sub> ratio of NCI-H1755/NCI-H1703 with NSAF value<sup>c</sup>Significant difference in *t*-test (*p*-value < 0.05). See Supplementary Table S3 for the complete set of label free quantitative results.

C<sub>18</sub> in-house stage tips again. To profile the N-terminal peptides, the samples were analyzed in triplicate (3 technical replicates) for each biological replicate. A total of 18 raw data files were then processed in SEQUEST and TPP. All data from the whole-cell lysates and N-terminal peptides were classified using informatics tools.

### Proteome profiling

Samples were prepared by FASP, and LC-MS/MS analysis was performed using the LTQ-Orbitrap Velos. MS/MS data were acquired for the biological and technical triplicates for each cell line and processed to identify peptides that generated the observed spectra, and proteins were inferred, based on the identified peptides. Because the MS/MS spectral counts for peptides from shotgun proteomic approaches have recently been shown estimate protein abundance well, we performed a label-free quantitative analysis of NSCLC cell lines, based on a shotgun proteomics strategy and spectral counting techniques.

A total of 18 raw files from the 2 cell lines were combined into

a single merged output file in Scaffold 3, in which the analysis was restricted to proteins with at least 2 unique peptides and an FDR < 0.5%. Per these criteria, we reproducibly identified 2130 non redundant proteins (Fig. 2A and Supplementary Table S1), 28% of which was identified by 2 unique peptides, whereas 17% was identified by 3 unique peptides, 11% was identified by 4 unique peptides, and 44% was identified by more than 5 unique peptides (Fig. 2B).

We classified all identified proteins by gene ontology (GO) analysis as biological process and molecular function. Many proteins mapped to the GO terms "protein metabolism and modification" (309 proteins), "intracellular protein traffic" (213 proteins), "protein biosynthesis" (147 proteins), "cell structure and motility" (147 proteins), and "cell cycle in biological process" (95 proteins) (Fig. 2C). Notably, molecular functions were assigned many proteins: 493 proteins were annotated with the GO term "nucleic acid binding," 157 proteins were related to cytoskeletal protein, 123 proteins fell under "dehydrogenase," and 85 proteins were "membrane traffic proteins" (Fig. 2D) (Supplementary Table S1).

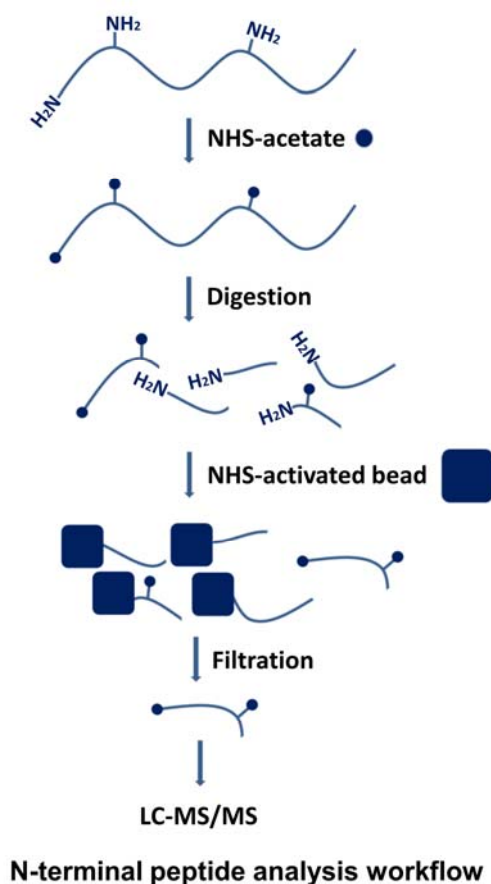
**Table 2.** Proteolytic events identified with less than 1.5 fold change

IPI	Peptide sequence <sup>a</sup>	Ratio <sup>b</sup>	N-terminal analysis <sup>c</sup>	Gene symbol	Protein name
IPI00215637	N.SSDNQSGGSTASKGR.Y	-0.48	NCI-H1703	DDX3X	ATP-dependent RNA helicase DDX3X
IPI00003918	R.SGQGAFGNMCR.G	-0.37	NCI-H1703	RPL4	60S ribosomal protein L4
IPI00219156	V.AAKKTKKSLESINSR.L	-0.15	NCI-H1703	RPL30	60S ribosomal protein L30
IPI00644712	R.SDSFENPVLQQHFR.N	0.14	NCI-H1703	XRCC6	X-ray repair cross-complementing protein 6
IPI00002520	Q.HSNAAQQTGEANR.G	0.30	NCI-H1755	SHMT2	Serine hydroxymethyltransferase, mitochondrial

<sup>a</sup>Observed peptide sequence from N-terminal peptide analysis is written by italics.

<sup>b</sup>Expression log<sub>2</sub> ratio of NCI-H1755/NCI-H1703 with NSAF value by label-free analysis

<sup>c</sup>Cell line with detected peptide sequences from N-terminal analysis



**Fig. 3.** N-terminal peptide analysis principle. Free amino groups ( $\alpha$  and  $\epsilon$ ) are acetylated prior to proteolysis, which results in a mixture of N-terminally acetylated (true N-terminal) and non-acetylated (internal) peptides. Subsequent incubation of the peptide mixture with an immobilized amine-reactive reagent creates a preparation enriched in N-terminal peptides.

#### Label-free quantitation between NCI-H1703 and NCI-H1755 cell lines

To quantify the identified proteins by spectral count, we used normalized spectral abundance factors (NSAF), with which the total number of spectra of an identified protein in each LC-MS/MS run correlates well with the abundance of the corre-

sponding protein over a wide linear dynamic range (Zybailov et al., 2006). High-confidence proteins for label-free quantitation were selected with an average spectral count  $\geq 5$  in 9 datasets (3 technical and 3 biological replicate) in either cell line. Also, missing values from each dataset were exchanged with a value of 0. Of the 2130 identified proteins, 671 satisfied our label-free quantitative protein criteria (Supplementary Table S2).

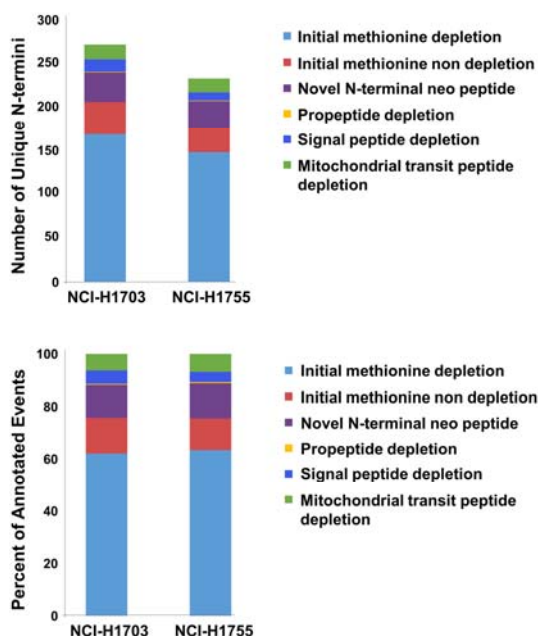
The distribution of the ratio correlation between NCI-H1703 and NCI-H1755 in the 3 biological replicates was selectively plotted, as shown in Supplementary Fig. S1A, in which 3 distributions had high similarity. To determine the fold-change in expression for each protein between the 2 cell lines, the standard deviation of the 671 quantitative proteins were calculated for the 3 biological replicates, indicating that approximately 90% fell within 0.5 standard deviation (Supplementary Fig. S1B) (Kim et al., 2012). The differential expression ratios for the 671 protein groups are shown in Supplementary Fig. S1C, in which ratios  $\geq 1.5$ -fold are shadowed. The expression of 242 proteins changed  $\geq 1.5$ -fold between NCI-H1703 and NCI-H1755 cells; 92 proteins were upregulated, and 150 proteins were downregulated. For example, integrin alpha-2 (ITGA2), aldehyde dehydrogenase, mitochondrial (ALDH2), UDP-glucose 4-epimerase (GALE), and aldose reductase (AKR1B1) were preferentially expressed in NCI-H1755 cells. Conversely, alpha-internexin (INA), isoform 1 of myosin-10 (MYH10), isoform 3 of UDP-N-acetylhexosamine pyrophosphorylase (UAP1), and isoform 1 of protein AHNAK2 (AHNAK2) were significantly downregulated in NCI-H1755 cells (Table 1 and Supplementary Table S3).

#### Identification of N-terminal peptides using BSA as control

The scheme with which N-terminal peptides were identified is shown in Fig. 3. The N-termini of proteins are characterized by an  $\alpha$ -amine, as opposed to the  $\epsilon$ -amines that are on lysine side chains. Thus,  $\epsilon$ -amines on lysine side chains had to be blocked. We blocked the  $\alpha$ -amine and  $\epsilon$ -amine groups by acetylation using NHS-acetate. After a quenching step, the unbound NHS-acetate was depleted by the amine in Tris. Next, proteins were digested with trypsin, generating N-terminal peptides with free amino groups. Then, we added NHS-activated beads, which bind free amine groups in newly generated N-terminal peptides by trypsin, whereas natural N-terminal peptides are blocked by acetylation (McDonald and Beynon, 2006).

In a control experiment, we examined whether this scheme could identify the natural N-termini of bovine serum albumin (BSA). Precursor BSA comprises 607 amino acids, whereas the mature form of BSA contains 583 amino acids, lacking residues 1-24 (Weijers, 1977). Thus, our BSA had an aspartic acid at residue 25 as its natural N-terminus.

Acetylated BSA was digested with trypsin and analyzed by



**Fig. 4.** Site annotation of N-terminal peptides. All identified peptides in N-terminal analysis were classified into six types based on their peptide site, number of unique N-termini (A) and percent of annotated events (B).

MALDI-MS (Supplementary Fig. S2A). The observed peptide masses were consistent with the expected Arg-C-specific digestion of BSA (acetylated lysine is resistant to tryptic cleavage) and included the known N-terminal peptide (Ac-DTHK(ac)SEIAHR) at 1277.6 m/z. As expected, a range of lysine-containing peptides appeared, increasing by 42.03 Da per lysine. On removal of newly generated BSA peptides by tryptic digestion by NHS-activated beads, we detected a single major peak at 1277.6 m/z by mass spectrometry. The N-terminal peptide of BSA had 1 peak that was mass-shifted by the acetylation of  $\alpha$ -amine and  $\epsilon$ -amine and confirmed with the peptide fingerprint by MS/MS analysis (Supplementary Fig. S2B).

#### Profile of N-terminal peptides in lung cancer cells

N-terminal peptides were identified in the 2 cell lines by positional proteomics analysis, as described (McDonald and Beynon, 2006). All samples were analyzed with 3 biological and technical replicates, and 307 unique proteins (272 peptides from 261 proteins in NCI-H1703 and 233 peptides from 220 proteins in NCI-H1755) were identified with more than 2 hits in the biological replicate analysis, with > 95% peptide probability and FDR < 1%. Ultimately, 92 unique N-terminal peptides were identified in NCI-H1703 cells compared to 53 in the NCI-H1755 cells (Supplementary Figs. S3A and S3B; Supplementary Table S4).

We analyzed the biological process and molecular function of the identified proteins. With regard to biological process, many proteins were enriched for the GO terms “protein metabolism and modification,” “protein biosynthesis,” and “mRNA splicing.” Many proteins mapped to the molecular function GO terms “nucleic acid binding” (62 proteins), “ribosomal protein” (30 proteins), and “chaperone in molecular function” (18 proteins)

(Supplementary Figs. S3C and S3D).

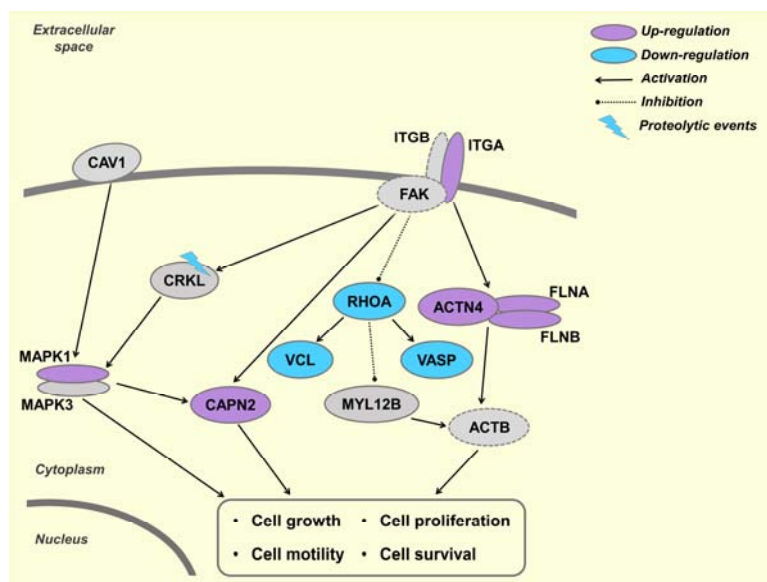
The identified N-terminal peptides were divided into natural N-terminus and novel N-terminal neo peptides. Most proteins undergo systematic depletion of their natural N-termini to function. For example, certain proteins have their signal peptides excised from the N-terminus to be secreted. Thus, natural N-termini were grouped into 5 types, based on molecule processing part of each protein sequence annotation in UniProtKB: initial methionine depletion, initial methionine nondepletion signal peptide depletion, propeptide depletion, and mitochondrial transit peptide depletion. Except for these natural N-termini, the newly identified peptides in the N-terminus analysis were annotated as novel N-terminal neo peptides that have not been assigned in the UniprotKB database.

A total of 325 unique N-terminal peptides were classified into 6 categories with regard to distributions of N-terminal peptides in NCI-H1703 and NCI-H1755 cells (Figs. 4A and 4B): (1) initial methionine depletion, NCI-H1703 (169 peptides, 62.1%) and NCI-H1755 (148 peptides, 63.5%); (2) initial methionine non-depletion, NCI-H1703 (37 peptides, 13.6%) and NCI-H1755 (25 peptides, 12.1%); (3) signal peptide depletion, NCI-H1703 (17 peptides, 5.5%) and NCI-H1755 (10 peptides, 4.3%); (4) propeptide depletion, NCI-H1703 (1 peptides, 0.4%) and NCI-H1755 (1 peptides, 0.4%); (5) mitochondrial transit peptide depletion, NCI-H1703 (17 peptides, 6.3%) and NCI-H1755 (16 peptides, 6.9%); and (6) novel N-terminal neo peptide, NCI-H1703 (33 peptides, 12.1%) and NCI-H1755 (30 peptides, 12.9%) (Supplementary Table S4).

#### Bioinformatics analysis of two parallel proteomic experiments

We performed a pathway analysis of differentially expressed proteins and identified N-terminal peptides in the 2 cell lines. To define the related pathways, all proteins in the lists were subjected to KEGG pathway analysis (Supplementary Fig. S4). Fourteen proteins were involved in the focal adhesion pathway in relation of cell invasion, growth, proliferation, and migration (Supplementary Table S5), 5 of which (FLNA, FLNB, CAV1, MYL12B, and CAPN2) were common in the two parallel experiments. Three proteins—CRKL, PPP1CB, and MAPK3—were identified only in the N-terminal peptide analysis, and 6 proteins (VASP, VCL, RHOA, ACTN4, MAPK1, and ITGA2) appeared in the label-free quantitative analysis. Thirteen of the 14 focal adhesion proteins—except FLNA, which contained a novel N-terminal neo peptide (PATEKDLAEDAPWKKIQNTFTR) in the NCI-H1703 and NCI-H1755 lines—showed differential expression in both cell lines in at least 1 experiments (Supplementary Table S5 and Fig. 5).

Six proteins (ITGA2, FLNA, FLNB, CAPN2, ACTN4, and MAPK1) were upregulated in metastatic lung cancer cells by label-free quantification analysis versus 3 downregulated proteins (RHOA, VASP, and VCL); 2 proteins (CAV1 and MY12B) were not differentially expressed. Three proteins (CRKL, PPP1CB, and MAPK3) were identified only in the N-terminal peptide analysis, in which we identified a fragment (novel N-terminal neo peptide) from CRKL in NCI-H1703 cells and methionine-depleted N-terminal peptides from PPP1CB and MAPK3 at the initial N-terminus. Protein phosphatase 1 (PPP1CB) is overexpressed in lung cancer (Liu et al., 2007) and is activated by phosphorylation. Although PPP1CB was detected by N-terminal peptide analysis only in NCI-H1755 cells, we excluded in subsequent analyses, due to the lack of phosphorylation data in this analysis.



**Fig. 5.** Deregulated focal adhesion pathway in NSCLC cell lines. Key focal adhesion proteins underwent either up-regulation (shown by violet color) or down-regulation (blue color) in NCI-H1755 cell line compared to NCI-H1703 cell line. CRKL was identified with novel N-terminal peptide in NCI-H1703 (blue lightning). Three proteins, ITGB, FAK, and ACTB, which are not identified in our data were shown by dash circle.

## DISCUSSION

Most NSCLC patients develop metastases, resulting in incurable disease at the time of diagnosis. Despite the advances in cancer research, there are few biomarkers for early-stage cancer, and our understanding of metastasis is poor (Tan et al., 2012). Also, metastasis has become the chief obstacle to the treatment of lung cancer. Thus, it will be helpful to determine the mechanisms of metastasis. To this end, our study has generated phenotypic data from primary and metastatic NSCLC using NCI-H1703 and NCI-H1755 cells, respectively.

Label-free quantitative analysis, based on MS1 peak intensities (Domon and Aebersold, 2006) and MS/MS spectral counts (Liu et al., 2004), is valuable in the large-scale analysis of proteins and peptides. General analysis of spectral counts has a limit of quantitation for low-abundance proteins ( $\leq 4$  spectrum detected) and post translational modification proteins (Freund and Prenni, 2013). However, the analysis is suitable for detection of subtle abundance changes in most proteins with high sensitivity and reproducibility (Old et al., 2005).

In this study, we identified 2130 nonredundant proteins with 218,323 spectra by cell lysate profiling at a minimum of 2 distinct peptides per protein, based on an FDR of 0.3%. We also required 5 or more spectral counts for the identifications, for which spectral counts were normalized by NSAF. Lastly, 671 proteins were used for the label-free quantification, which allowed us to identify differentially expressed proteins ( $n = 242$ ) with  $\geq 1.5$  fold-change and  $p$ -value  $< 0.05$ .

Of the 242 differentially expressed proteins, transaldolase (TALDO1) is a novel serum biomarker for a model hepatocellular carcinoma (HCC) metastasis and HCC patients (Wang et al., 2011). TALDO1 was overexpressed in NCI-H1755 versus NCI-H1703 cells. Dipanjana et al. reported global proteomic alterations in colorectal cancer cell metastasis, 8 proteins of which were consistent with our dataset; 3 upregulated proteins (ALDH2, HSP90B1, and PDIA4) and 5 downregulated proteins (EIF2S2, MCM6, MCM7, PSMC1, and PSMC2) (Ghosh et al., 2011).

Many proteins, such as isoform 2 of filamin-A (FLNA), isoform 1 of filamin-B (FLNB), isoform A of prelamin-A/C (LMNA), and vimentin (VIM), which were classified as the GO term "cell

structure and motility," were upregulated in the metastatic NCI-H1755 line (Supplementary Table S1). In particular, LMNA is a metastatic biomarker of colorectal cancer cells (Willis et al., 2008) and a marker of embryonic stem cell differentiation (Constantinescu et al., 2006), although this status not been reported in NSCLC metastasis.

Cell proliferation molecules, such as isoform 1 of protein CDV3 homolog (CDV3), isoform 1 of epidermal growth factor receptor (EGFR), and histone-binding protein RBBP7 (RBBP7), were downregulated in the NCI-H1755 cells. Conversely, isoform 1 of annexin A7 (ANXA7), 60-kDa heat shock protein mitochondrial (HSPD1), proliferating cell nuclear antigen (PCNA), and isoform 3 of thioredoxin reductase 1 cytoplasmic (TXNRD1) were upregulated in this line. ANXA7 is a biomarker of progression in prostate and breast cancer (Srivastava et al., 2001); we also noted a 1.7-fold increase in NCI-H1755 cells.

Protein fragment reaction linked to cancer metastasis. Several studies have demonstrated that potential cancer biomarkers, such as HER2 rb2 and CYFRA 21-1, are generated by protein fragmentation (Pujol et al., 1993; Streckfus et al., 2000). For example, CYFRA 21-1 that is protein fragment is known relation with lung cancer metastasis, although it is not a specific marker for lung cancer diagnosis. In searching for markers that are elicited by protein fragmentation, we identified new generated N-terminal peptides using positional proteomics methods. In brief, natural N-termini are blocked by certain labeling methods, such as acetylation (McDonald and Beynon, 2006), dimethylation (Hsu et al., 2003), iTRAQ (Prudova et al., 2010), and PITC adman (Dugaiczuk et al., 1982). In our study, N-termini were labeled by acetylation, based on its simplicity and high labeling efficiency. Ultimately, we identified 27 novel N-terminal neo peptides that were differentially generated between metastatic cells and primary cancer cells. Notably, natural cleavage of N-terminal peptides, such as initial methionine depletion, signal peptide depletion, propeptide depletion, and transit peptide depletion, were also detected and annotated using the Uniprot database (Apweiler et al., 2004). Specifically, of the initial methionine-depleted proteins, we identified 44 proteins that do not exist in the UniprotKB database.

In the N-terminal peptide analysis, 92 peptides from 87 pro-



teins were detected in NCI-H1703 cells, whereas 53 peptides from 46 proteins were identified in NCI-H1755 cells (Supplementary Fig. S3)—27 peptides were categorized as novel N-terminal neo peptides (like the fragment peptides), and 15 novel N-terminal neo peptides appeared only in NCI-H1703 cells. Notably, EPH receptor A2 (EPHA2) is a marker of NSCLC progression (Brannan et al., 2009), and a novel N-terminal neo peptide of EPHA2 was detected in primary cancer cells. However, EPHA2 was observed in both cell lines by label-free quantitative analysis (not used for quantification due to a spectral count below 5).

Five proteins were identified with fragment N-terminal peptides, whereas their expression did not differ by label-free quantification analysis (Table 2). Four of them—DDX3X, RPL4, RPL30, and XRCC6—were observed only in NCI-H1703 cells by N-terminal peptide analysis, whereas SHMT2 was detected only in NCI-H1755 cells. Further, four proteins (DDX3X, RPL4, RPL30, and XRCC6) are associated with cell proliferation and differentiation in metastasis (Bauer et al., 2012; Li et al., 2011; Yoon et al., 2006). In this study, the four proteins that were identified with novel N-terminal neo peptides were expressed in equal amounts in the cell lines, but they could not affect the metastasis of primary cancer cells (NCI-H1703).

We found 138 proteins that were common to both experiments (Supplementary Table S6). Most proteins, including natural N-terminal peptides that were differentially identified by N-terminal analysis, except for histone-binding protein RBBP7 (RBBP7), were consistent with their expression levels in the label-free quantification analysis. For example, creatine kinase B-type (CKB) was identified with initial methionine-depleted N-termini only in NCI-H1703 cells by N-terminal analysis, whereas CKB was significantly upregulated in NCI-H1703 cells by label-free quantitative analysis.

In the classification of the 138 commonly identified proteins by KEGG pathway, the proteins were primarily involved in aminoacyl-tRNA biosynthesis, the pentose phosphate pathway, the proteasome, arginine and proline metabolism, DNA replication, and focal adhesion (Supplementary Fig. S4). Focal adhesion is a major pathway of cancer metastasis, and we identified 15 proteins that were related to focal adhesion in the 2 profiling experiments (Fig. 5 and Supplementary Table S5). Of the 138 proteins, 11 proteins, identified by label-free quantification analysis, participated in focal adhesion—6 proteins were upregulated, 3 proteins were downregulated, and 2 proteins were not differentially expressed. Conversely, of the proteins that were identified by N-terminal peptide analysis, 8 were involved in focal adhesion.

Integrin alpha-2 (ITGA2) was upregulated by 2.4-fold in NCI-H1755 cells. Apparently, ITGA2 mediates metastasis to the liver by regulating the focal adhesion pathway (Yoshimura et al., 2009). Overexpression of integrin proteins (ITGA and ITGB) initiates a signaling cascade to alpha-actinin-4 (ACTN4), FLNA, FLNB, and FAK (not identified in our data) to effect cell proliferation and growth (Shibue and Weinberg, 2009) (Fig. 5). Notably, ACTN4, FLNA, and FLNB were overexpressed in NCI-H1755 cells in this study. In addition, MAPK1 (also known as ERK2), upregulated in metastatic cells, is a point at which multiple biochemical signals integrate (Wu et al., 2008) (Fig. 5).

MAP kinases mediate many processes in cancer cells, such as proliferation, migration, invasion, and metastasis (Obchoei et al., 2011). Increased expression of MAPK1 promotes the expression of CAPN2, which functions in cell movement, migration, and invasion during metastasis (Storr et al., 2011). In the N-terminal peptide analysis, v-crk sarcoma virus CT10 onco-

gene homolog (avian)-like (CRKL) was identified as a novel N-terminal neo peptide only in NCI-H1703 cells. Because CRKL activates ERK signaling to promote cell proliferation, survival, and invasion in lung cancer (Kim et al., 2010), we hypothesize that CRKL function is regulated by fragment events during metastasis.

In summary, we have identified differentially expressed proteins that distinguish primary and metastatic lung cancer. Many of these quantitative proteins and N-terminal peptides are involved in pathways in cell migration, proliferation, and metastasis. Thus, our datasets of proteins and fragment peptides in lung cells might be valuable in discovering and validating lung cancer biomarkers and metastasis markers.

*Note: Supplementary information is available on the Molecules and Cells website (www.molcells.org).*

## ACKNOWLEDGMENTS

This work was supported by the Proteogenomic Research Program through the National Research Foundation of Korea and a National Research Foundation of Korea [NRF] grant (No. 2011-0030740), funded by the Korea government [Ministry of Science, ICT and Future Planning (MSIP)]. This work was also supported by the Industrial Strategic Technology Development Program (#10045352, MKE, Korea).

## REFERENCES

- Anisowicz, A., Huang, H., Braunschweiger, K.I., Liu, Z., Giese, H., Wang, H., Mamaev, S., Olejnik, J., Massion, P.P., and Del Maestro, R.G. (2008). A high-throughput and sensitive method to measure global DNA methylation: application in lung cancer. *BMC Cancer* 8, 222.
- Apweiler, R., Bairoch, A., Wu, C.H., Barker, W.C., Boeckmann, B., Ferro, S., Gasteiger, E., Huang, H., Lopez, R., Magrane, M., et al. (2004). UniProt: the universal protein knowledgebase. *Nucleic Acids Res.* 32, D115-119.
- Bauer, K.M., Lambert, P.A., and Hummon, A.B. (2012). Comparative label-free LC-MS/MS analysis of colorectal adenocarcinoma and metastatic cells treated with 5-fluorouracil. *Proteomics* 12, 1928-1937.
- Brannan, J.M., Sen, B., Saigal, B., Prudkin, L., Behrens, C., Solis, L., Dong, W., Bekele, B.N., Wistuba, I., and Johnson, F.M. (2009). EphA2 in the early pathogenesis and progression of non-small cell lung cancer. *Cancer Prev. Res. (Phila)* 2, 1039-1049.
- Brown, J.R., and Hartley, B.S. (1966). Location of disulphide bridges by diagonal paper electrophoresis. The disulphide bridges of bovine chymotrypsinogen A. *Biochem. J.* 101, 214-228.
- Constantinescu, D., Gray, H.L., Sammak, P.J., Schatten, G.P., and Csoka, A.B. (2006). Lamin A/C expression is a marker of mouse and human embryonic stem cell differentiation. *Stem Cells* 24, 177-185.
- Dawson, T.M., and Dawson, V.L. (2003). Molecular pathways of neurodegeneration in Parkinson's disease. *Science* 302, 819-822.
- Domon, B., and Aebersold, R. (2006). Mass spectrometry and protein analysis. *Science* 312, 212-217.
- Dugaiczak, A., Law, S.W., and Dennison, O.E. (1982). Nucleotide sequence and the encoded amino acids of human serum albumin mRNA. *Proc. Natl. Acad. Sci. USA* 79, 71-75.
- Enoksson, M., Li, J., Ivancic, M.M., Timmer, J.C., Wildfang, E., Eroshkin, A., Salvesen, G.S., and Tao, W.A. (2007). Identification of proteolytic cleavage sites by quantitative proteomics. *J. Proteome Res.* 6, 2850-2858.
- Fanayan, S., Smith, J.T., Lee, L.Y., Yan, F., Snyder, M., Hancock, W.S., and Nice, E. (2013). Proteogenomic analysis of human colon carcinoma cell lines LIM1215, LIM1899, and LIM2405. *J. Proteome Res.* 12, 1732-1742.
- Freund, D.M., and Prenni, J.E. (2013). Improved detection of quantitative differences using a combination of spectral counting and MS/MS total ion current. *J. Proteome Res.* 12, 1996-2004.

- Gevaert, K., Goethals, M., Martens, L., Van Damme, J., Staes, A., Thomas, G.R., and Vandekerckhove, J. (2003). Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. *Nat. Biotechnol.* 21, 566-569.
- Ghosh, D., Yu, H., Tan, X.F., Lim, T.K., Zubaidah, R.M., Tan, H.T., Chung, M.C., and Lin, Q. (2011). Identification of key players for colorectal cancer metastasis by iTRAQ quantitative proteomics profiling of isogenic SW480 and SW620 cell lines. *J. Proteome Res.* 10, 4373-4387.
- Han, D., Moon, S., Kim, Y., Ho, W.K., Kim, K., Kang, Y., and Jun, H. (2012). Comprehensive phosphoproteome analysis of INS-1 pancreatic beta-cells using various digestion strategies coupled with liquid chromatography-tandem mass spectrometry. *J. Proteome Res.* 11, 2206-2223.
- Hoffman, P.C., Mauer, A.M., and Vokes, E.E. (2000). Lung cancer. *Lancet* 355, 479-485.
- Hsu, J.L., Huang, S.Y., Chow, N.H., and Chen, S.H. (2003). Stable-isotope dimethyl labeling for quantitative proteomics. *Anal. Chem.* 75, 6843-6852.
- Hwang, S.J., Seol, H.J., Park, Y.M., Kim, K.H., Gorospe, M., Nam, D.H., and Kim, H.H. (2012). MicroRNA-146a suppresses metastatic activity in brain metastasis. *Mol. Cells* 34, 329-334.
- Jemal, A., Siegel, R., Xu, J., and Ward, E. (2010). Cancer statistics, 2010. *CA Cancer J. Clin.* 60, 277-300.
- Kawakami, T., Hoshida, Y., Kanai, F., Tanaka, Y., Tateishi, K., Ike-noue, T., Obi, S., Sato, S., Teratani, T., Shiina, S., et al. (2005). Proteomic analysis of sera from hepatocellular carcinoma patients after radiofrequency ablation treatment. *Proteomics* 5, 4287-4295.
- Kim, Y.H., Kwei, K.A., Girard, L., Salari, K., Kao, J., Pacyna-Gengelbach, M., Wang, P., Hernandez-Boussard, T., Gazdar, A.F., Petersen, I., et al. (2010). Genomic and functional analysis identifies CRKL as an oncogene amplified in lung cancer. *Oncogene* 29, 1421-1430.
- Kim, S.J., Jin, J., Kim, Y.J., Kim, Y., and Yu, H.G. (2012). Retinal proteome analysis in a mouse model of oxygen-induced retinopathy. *J. Proteome Res.* 11, 5186-5203.
- Li, F., Glinskii, O.V., Zhou, J., Wilson, L.S., Barnes, S., Anthony, D.C., and Glinsky, V.V. (2011). Identification and analysis of signaling networks potentially involved in breast carcinoma metastasis to the brain. *PLoS One* 6, e21977.
- Liu, H., Sadygov, R.G., and Yates, J.R., 3rd (2004). A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* 76, 4193-4201.
- Liu, Y., Sun, W., Zhang, K., Zheng, H., Ma, Y., Lin, D., Zhang, X., Feng, L., Lei, W., Zhang, Z., et al. (2007). Identification of genes differentially expressed in human primary lung squamous cell carcinoma. *Lung Cancer* 56, 307-317.
- Lopez-Otin, C., and Bond, J.S. (2008). Proteases: multifunctional enzymes in life and disease. *J. Biol. Chem.* 283, 30433-30437.
- McDonald, L., and Beynon, R.J. (2006). Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. *Nat. Protoc.* 1, 1790-1798.
- Nisman, B., Biran, H., Heching, N., Barak, V., Ramu, N., Nemirovsky, I., and Peretz, T. (2008). Prognostic role of serum cytokeratin 19 fragments in advanced non-small-cell lung cancer: association of marker changes after two chemotherapy cycles with different measures of clinical response and survival. *Br. J. Cancer* 98, 77-79.
- Obchoei, S., Weakley, S.M., Wongkham, S., Wongkham, C., Sawanyawisuth, K., Yao, Q., and Chen, C. (2011). Cyclophilin A enhances cell proliferation and tumor growth of liver fluke-associated cholangiocarcinoma. *Mol. Cancer* 10, 102.
- Old, W.M., Meyer-Arendt, K., Aveline-Wolf, L., Pierce, K.G., Mendoza, A., Sevinsky, J.R., Resing, K.A., and Ahn, N.G. (2005). Comparison of label-free methods for quantifying human proteins by shotgun proteomics. *Mol. Cell. Proteomics* 4, 1487-1502.
- Opferman, J.T., and Korsmeyer, S.J. (2003). Apoptosis in the development and maintenance of the immune system. *Nat. Immunol.* 4, 410-415.
- Parkin, D.M., and Fernandez, L.M. (2006). Use of statistics to assess the global burden of breast cancer. *Breast J.* 12 Suppl 1, S70-80.
- Prudova, A., auf dem Keller, U., Butler, G.S., and Overall, C.M. (2010). Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. *Mol. Cell. Proteomics* 9, 894-911.
- Pujol, J.L., Grenier, J., Dures, J.P., Daver, A., Pujol, H., and Michel, F.B. (1993). Serum fragment of cytokeratin subunit 19 measured by CYFRA 21-1 immunoradiometric assay as a marker of lung cancer. *Cancer Res.* 53, 61-66.
- Rao, J.S. (2003). Molecular mechanisms of glioma invasiveness: the role of proteases. *Nat. Rev. Cancer* 3, 489-501.
- Shibue, T., and Weinberg, R.A. (2009). Integrin beta1-focal adhesion kinase signaling directs the proliferation of metastatic cancer cells disseminated in the lungs. *Proc. Natl. Acad. Sci. USA* 106, 10290-10295.
- Srivastava, M., Bubendorf, L., Nolan, L., Glasman, M., Leighton, X., Miller, G., Fehrle, W., Raffeld, M., Eidelman, O., Kallioniemi, O.P., et al. (2001). ANX7 as a bio-marker in prostate and breast cancer progression. *Dis. Markers* 17, 115-120.
- Storr, S.J., Carragher, N.O., Frame, M.C., Parr, T., and Martin, S.G. (2011). The calpain system and cancer. *Nat. Rev. Cancer* 11, 364-374.
- Streckfus, C., Bigler, L., Dellinger, T., Pfeifer, M., Rose, A., and Thigpen, J.T. (1999). CA 15-3 and c-erbB-2 presence in the saliva of women. *Clin. Oral Investig.* 3, 138-143.
- Streckfus, C., Bigler, L., Tucci, M., and Thigpen, J.T. (2000). A preliminary study of CA15-3, c-erbB-2, epidermal growth factor receptor, cathepsin-D, and p53 in saliva among women with breast carcinoma. *Cancer Invest.* 18, 101-109.
- Tan, F., Jiang, Y., Sun, N., Chen, Z., Lv, Y., Shao, K., Li, N., Qiu, B., Gao, Y., Li, B., et al. (2012). Identification of isocitrate dehydrogenase 1 as a potential diagnostic and prognostic biomarker for non-small cell lung cancer by proteomic analysis. *Mol. Cell. Proteomics* 11, M111 008821.
- Tian, T., Hao, J., Xu, A., Luo, C., Liu, C., Huang, L., Xiao, X., and He, D. (2007). Determination of metastasis-associated proteins in non-small cell lung cancer by comparative proteomic analysis. *Cancer Sci.* 98, 1265-1274.
- Wang, C., Guo, K., Gao, D., Kang, X., Jiang, K., Li, Y., Sun, L., Zhang, S., Sun, C., Liu, X., et al. (2011). Identification of transaldolase as a novel serum biomarker for hepatocellular carcinoma metastasis using xenografted mouse model and clinic samples. *Cancer Lett.* 313, 154-166.
- Weijers, R.N. (1977). Amino acid sequence in bovine serum albumin. *Clin. Chem.* 23, 1361-1362.
- Willis, N.D., Cox, T.R., Rahman-Casans, S.F., Smits, K., Przyborski, S.A., van den Brandt, P., van Engeland, M., Weijenberg, M., Wilson, R.G., de Bruine, A., et al. (2008). Lamin A/C is a risk biomarker in colorectal cancer. *PLoS One* 3, e2988.
- Wisniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359-362.
- Wu, W.S., Wu, J.R., and Hu, C.T. (2008). Signal cross talks for sustained MAPK activation and cell migration: the potential role of reactive oxygen species. *Cancer Metastasis Rev.* 27, 303-314.
- Xie, X., Feng, S., Vuong, H., Liu, Y., Goodison, S., and Lubman, D.M. (2010). A comparative phosphoproteomic analysis of a human tumor metastasis model using a label-free quantitative approach. *Electrophoresis* 31, 1842-1852.
- Xue, H., Lu, B., Zhang, J., Wu, M., Huang, Q., Wu, Q., Sheng, H., Wu, D., Hu, J., and Lai, M. (2010). Identification of serum biomarkers for colorectal cancer metastasis using a differential secretome approach. *J. Proteome Res.* 9, 545-555.
- Yoon, S.Y., Kim, J.M., Oh, J.H., Jeon, Y.J., Lee, D.S., Kim, J.H., Choi, J.Y., Ahn, B.M., Kim, S., Yoo, H.S., et al. (2006). Gene expression profiling of human HBV- and/or HCV-associated hepatocellular carcinoma cells using expressed sequence tags. *Int. J. Oncol.* 29, 315-327.
- Yoshimura, K., Meckel, K.F., Laird, L.S., Chia, C.Y., Park, J.J., Olino, K.L., Tsunedomi, R., Harada, T., Iizuka, N., Hazama, S., et al. (2009). Integrin alpha2 mediates selective metastasis to the liver. *Cancer Res.* 69, 7320-7328.
- Zybailov, B., Mosley, A.L., Sardi, M.E., Coleman, M.K., Florens, L., and Washburn, M.P. (2006). Statistical analysis of membrane proteome expression changes in *Saccharomyces cerevisiae*. *J. Proteome Res.* 5, 2339-2347.