이학박사 학위논문

Cancer Biomarker Discovery Using N-terminal Peptides and Multiple Reaction Monitoring-MS Techniques

N-말단단백체 및 다중반응검지 질량분석기술을 이용한 암 표지자 개발 연구

2015 년 08 월

서울대학교 대학원

의과학과 의과학전공

민 호 필

A thesis of the Degree of Doctor of Philosophy

N-말단단백체 및 다중반응검지 질량분석기술을 이용한 암 표지자 개발 연구

Cancer Biomarker Discovery Using N-terminal Peptides and Multiple Reaction Monitoring-MS Techniques

August 2015

Major in Biomedical Sciences

Department of Biomedical Sciences

Seoul National University

Graduate School

Hophil Min

N-말단단백체 및 다중반응검지 질량분석기술을 이용한 암 표지자 개발 연구

지도교수 김 영 수

이 논문을 이학박사 학위논문으로 제출함

2015 년 04 월

서울대학교 대학원

의과학과 의과학전공

민 호 필

민호필의 이학박사 학위논문을 인준함

2015 년 07 월

위 원 장 정 구 홍 (인)

부위원장 김 영 수 (인)

위 원 박 태 성 (인)

위 원 김 윤 준 (인)

위 원 장 수 환 (인)

Cancer Biomarker Discovery Using N-terminal Peptides and Multiple Reaction Monitoring-MS Techniques

by

Hophil Min

A thesis submitted to the Department of Biomedical Sciences in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Biomedical Sciences at Seoul National University Graduate School

July 2015

Approved by Thesis Committee:

Professor _____Chairman

Professor _____Vice chairman

Professor _____

Professor _____

Professor _____

ABSTRACT

Cancer Biomarker Discovery Using N-terminal Peptides and Multiple Reaction Monitoring-MS Techniques

Hophil Min

Major in Biomedical Sciences

Department of Biomedical Sciences

Seoul National University

Graduate School

Introduction: Cancer is the leading cause of death in the worldwide, and the major cause of cancer death is the difficulty for early diagnosis. To overcome this problem, the discovery of cancer biomarkers is useful for early diagnosis, outcome monitoring, or predicting recurrence. For biomarker discovery, proteomics technique is powerful tools with high-throughput and high sensitivity. Thus, proteomics study can help variable cancer biomarker discovery and understand of cancer mechanisms in body.

Methods: In chapter I, to examine metastatic events 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). In chapter II, we performed to identify new marker-candidate proteins from LiverAtlas database. And abundance of marker-candidate proteins were quantified in individual patients by multiple reaction monitoring assay.

Results: In chapter I, 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. In chapter II, we selected quantitative 104 marker candidate proteins with reference labeled peptides. Among them, we found that 17 proteins with AUC more than 0.60 were able to effectively discriminate poor responders from total patients underwent TACE. Also, we discovered powerful ensemble model panel with protein markers and clinical variables.

Conclusions: In chapter I, our datasets of proteins and fragment peptides in lung cells might be valuable in discovering and validating lung cancer biomarkers and metastasis markers. This study increases our understanding of the NSCLC metastasis proteome. In chapter II, we discovered three new marker proteins that are associated with prognosis prediction after TACE in the first time. Our study can help to identify useful biomarkers for prediction of prognosis with multi-panel modeling.

Keywords: Non-small-cell lung cancer; label-free quantitative analysis; N-terminal analysis; Metastasis; Multiple Reaction Monitoring; transcatheter arterial chemoembolization; Hepatocellular carcinoma; Prognostic factor; Proteomics; Biomarker

Student Number: 2008-21997

*This work is published in Molecules and Cells journal. (H Min, D Han, Y Kim, J Cho, J Jin, and Y Kim. Label-Free Quantitative Proteomics and N-terminal Analysis of Human Metastatic Lung Cancer Cells. Mol. Cells 2014; 37(6): 457~466)

CONTENTS

LIST OF TABLES

Chapter I

Chapter II

LIST OF FIGURES

Chapter I

Chapter II

LIST OF ABBREVIATIONS

MS, Mass spectrometry

PTM, Post-translational modification

NSCLC, Non-small-cell lung cancer

CYFRA 21-1, serum cytokeratin 19 fragments

FASP, Filter-aided sample preparation

PMSF, Phenylmethanesulfonyl fluoride

SDS, Sodium dodecyl sulfate

TFA, Trifluoroacetic acid

CHCA, α-cyano-4-hydroxycinnamic acid

STD, StageTip desalting

BSA, Bovine serum albumin

TPP, Trans-Proteomics Pipeline

NSAF, Normalized spectral abundance factor

UniProtKB, Universal Protein Resource Knowledgebase

COFRADIC, Combined fractional diagonal chromatography

CRKL, v-crk sarcoma virus CT10 oncogene homolog (avian)-like

MRM, Multiple Reaction Monitoring

HCC, Hepatocellular carcinoma

AASLD, American Association for the Study of Liver

EASL, European Association for the Study of the Liver

AFP, Alpha-fetoprotein

PIVKA-II, Prothrombin induced by the absence of vitamin K or antagonist-II

MCPs, Marker candidate proteins

mRECIST, modified Response Evaluation Criteria in Solid Tumors

TACE, Transarterial chemoembolization

BCHE, Cholinesterase

ITIH4, Inter-alpha-trypsin inhibitor heavy chain H4

SERPINF2, Alpha-2-antiplasmin

C7, Complement component C7

ROC, Receiver operating characteristic

GENERAL INTRODUCTION

Cancer is the leading cause of death in the worldwide, and developed countries take the brunt of the disease with approximately 70% of deaths (1). The major cause of cancer death is the difficulty for early diagnosis and suitable treatment without major medical devices. To overcome this problem, the discovery of cancer biomarkers is useful for early diagnosis, monitoring how well a treatment is performed, or predicting recurrence (2, 3). Thus, the utility and importance of biomarkers are growing in both academic and industrial fields (4).

Most biomarkers are molecules that are secreted by tumor organ or specific responses to the presence of cancer (5). To detect the biomarkers, proteomics technique is one of the applicable tools with high-throughput and high sensitivity (6). As the development of mass spectrometry (MS), proteomics technique could analyze the relative protein abundance, occurrence site of post translational modifications (PTMs), protein-protein interactions, and cellular functions (7-9). So, knowledge of the proteome can be useful for cancer biomarker discovery.

In chapter I, to know cellular proteome changes, we performed label-free analysis and N-terminal peptides enrichments in non-small-cell lung cancer (NSCLC) cell lines. NSCLS is usually treated with surgery, but surgery is effective only in patients who are diagnosed at an early stage. Unfortunately, more than 70% of NSCLC patients are diagnosed at the late

stage with metastasis, resulting in a loss of opportunity for effective surgery. Thus, application of our new technique for biomarker discovery can help to develop novel and more effective molecular markers and therapeutic targets.

In chapter II, for application of quantitative proteomics tools in biomarker discovery, we performed multiple reaction monitoring (MRM) assay in serum of hepatocellular carcinoma patients underwent transarterial chemoembolization (TACE). TACE is an effective treatment option for reducing systemic toxicity, increasing local antitumor effects, and improving survival for late stage patients. But, unpredictable outcomes often occur after TACE in terms of treatment response and survival. So, in this study, we identified useful biomarkers for prediction of prognosis with multi-panel modeling.

CHAPTER I

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

INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths worldwide (30%) but constitutes only 15% of new cancer diagnoses (10). 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 (11). Lung cancer is divided into 2 major histological types: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC) (12). 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 (13).

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 (14, 15). During metastasis, irreversible protein fragmentation occurs (16). Dysregulation of protein fragment reactions in organs can cause pathological developmental disorders, such as cancer, inflammation, infection, and Alzheimer disease (17-19).

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 (20). Pro1708/Pro2044 (the C-terminal fragment of albumin) (21) and HER2 rb2 (the ectodomain of human epithelial growth factor receptor-2) (22) 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 (23). Many global quantitative proteomics studies have examined metastasis in various cancers, such as colorectal cancer (24), breast cancer (25), and hepatocellular carcinoma (26). 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 (14).

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 (27). To this end, several groups have established a method of identifying protease-generated (neo) peptides in cellular pathways, known as N-terminomics (28). 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 (29). 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 et al. developed a more rapid and simpler N-terminal peptide analysis method (positional proteomics) that is based on negative selection by chemical labeling of the α-amine in proteins (30).

In this study, to differentiate primary cancer cells from metastatic cells, we performed 2 parallel experiments: label-free quantification and novel 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 (31). 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 (Waltham, MA). Hydrochloric acid (HCl) and sodium chloride (NaCl) were purchased from DUKSAN (Gyungkido, Korea). Urea and dithiothreitol (DTT) were purchased from AMRESCO (Solon, OH). Phenylmethanesulfonyl fluoride (PMSF), sodium dodecyl sulfate (SDS), and Tris were obtained from USB (Cleveland, OH). Complete protease inhibitor cocktail tablets were acquired from ROCHE (Indianapolis, IN), and sequencing-grade modified trypsin was purchased from PROMEGA (Madison, WI). Sulfo-NHS acetate and NHS-Activated agarose slurry were obtained from Pierce (Rockford, IL). All other reagents—iodoacetamide, α-cyano-4-hydroxycinnamic acid (CHCA), and trifluoroacetic acid (TFA)—were purchased from Sigma-Aldrich (St. Louis, MO).

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, Daegu, Korea) with 10% fetal bovine serum (Gibco, Grand Island, NY), 100 U/mL penicillin and 100 μg/mL streptomycin (Gibco, Grand Island, NY) and 25 mM HEPES (Gibco, Grand Island, NY). The cultures were maintained in 95% humidified air and 5% CO_2 at 37℃.

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, 1x protease inhibitor cocktail, 0.1 M DTT, and 0.1 M HEPES. Lysates were incubated at 95℃ for 5 min and sonicated for 1 min. Supernatants were collected from the lysates by centrifugation at 15,000 × g for 20 min at 4℃. Protein concentrations were measured using the BCA Protein Assay Kit – reducing reagent-compatible (Pierce, Rockford, IL). Finally, each cell lysate was stored in 0.2-mg aliquot at -80℃ until use.

Filter-Aided Sample Preparation (FASP)

Cell lysates were processed by filter-aided sample preparation (FASP) (32) using a 10K molecular weight cutoff (MWCO) filter (Millipore, Pittsburgh, PA). Briefly, 200 μg of cell lysates in lysis buffer (4% SDS, 0.1 mM PMSF, 1x 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 × g at 20℃ 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 darkn 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 × g at 20℃ 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℃. 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 × g at 20℃ 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 × g at 20℃ 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℃. Digested peptides were collected by centrifugation and acidified with 2% TFA.

Desalting of peptides

Digested samples were desalted using in-house C_{18} StageTip desalting (STD) columns, as described (33). Briefly, in-house C_{18} STD columns were prepared by reversed-phase packing of POROS 20 R2 material into 0.2-mL yellow pipet tips that sat atop C_8 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, Rockford, IL). One milliliter of an NHS-agarose bead slurry (50% slurry in acetone) was prepared per the manufacturer's protocol (Pierce Rockford, IL). 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 hours at RT. Finally, the beads were centrifuged at 1000 × *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, Solon, OH) were N-terminally labeled as described above as control. The peptides were dissolved in 10 μl 0.1% TFA, and 0.5 μl of each sample was mixed with 0.5 l of a matrix solution that contained 5 mg/ml CHCA (Sigma, St. Louis, MO), 70% ACN, and 0.1% TFA. The peptides were spotted directly onto a MALDI plate (Opti-TOF™ 384-well Insert, Applied Biosystems, Foster City, CA) 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 x e^{-07} 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), adrenocorticotropic 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, Waltham, MA) that was coupled to an EasyLC II (Proxeon Biosystems, Denmark), equipped with a nanoelectrospray device and fitted with a 10-μm fused silica emitter tip (New Objective, Woburn, MA). Ten microliters of each samples was loaded onto a nano-LC trap column (ZORBAX 300SB-C_{18}, 5 μm, 0.3 × 5 mm, Agilent, Wilmington, DE), and peptides were separated on a C_{18} analytical column (75 μm × 15 cm) that was packed in-house with C_{18} resin (Magic C18-AQ 200 Å, 5-μm 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℃. 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 ±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, Milpitas, CA). Raw files from the LTQ-Orbitrap Velos were converted into mzXML files using Trans-Proteomics Pipeline (TPP, ISB, Seattle, WA). 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., Portland, OR).

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 (34). NSAF values were calculated as:

$$NSAF = (SpC/Mw)/\Sigma(SpC/Mw)n$$

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 Figure 1-1. 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_{18} 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_{18} in-house stage tips and filtered by NHS-activated beads that depleted the newly generated N-termini by trypsin. The supernatants of the N-terminal peptide samples were desalted using C_{18} 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.

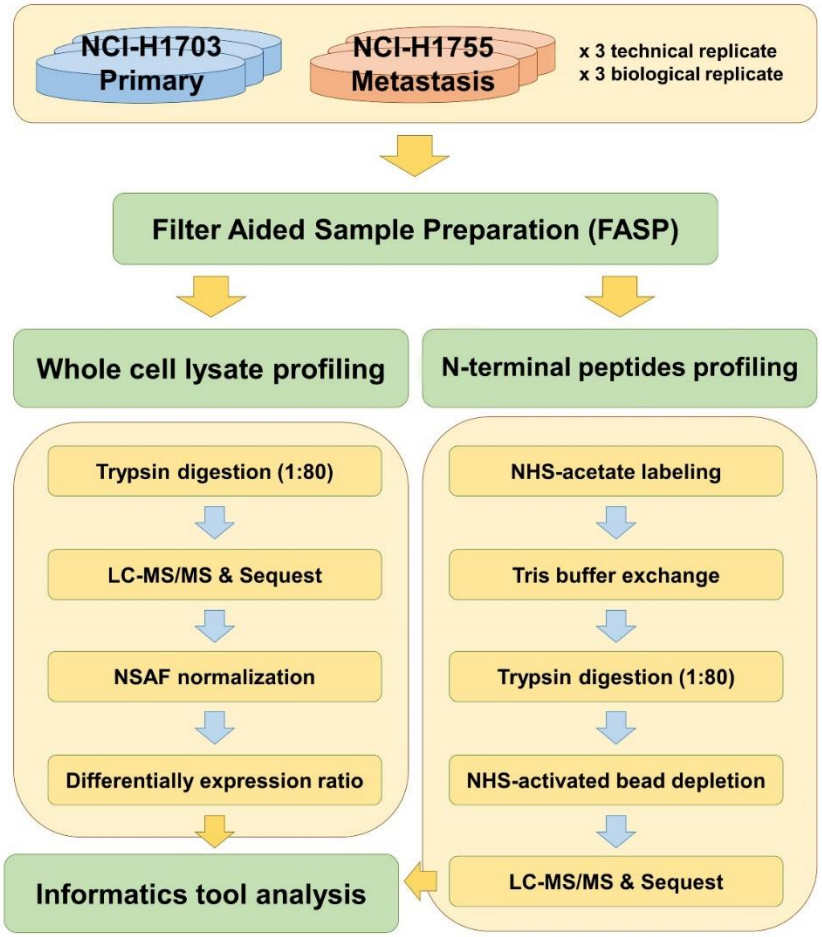


Figure 1-1. Overall scheme

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.

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 (Figure 1-2A), 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 (Figure 1-2B).

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) (Figure 1-2C). Notably, molecular functions were assigned many proteins: 493 proteins were annotate 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" (Figure 1-2D).

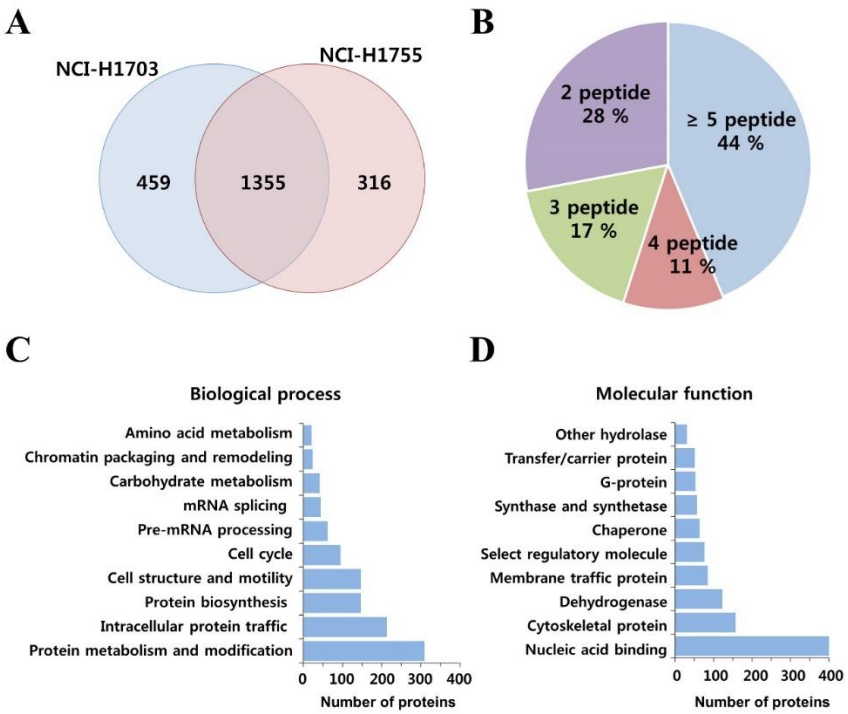


Figure 1-2. Identification and proteome analysis of two different cell lines

(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.

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 corresponding protein over a wide linear dynamic range (34). High-confidence proteins for label-free quantitation were selected with an average spectral count ≥ 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.

The distribution of the ratio correlation between NCI-H1703 and NCI-H1755 in the 3 biological replicates was selectively plotted, as shown in Figure 1-3A, 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 (Figure 1-3B) (35). The differential expression ratios for the 671 protein groups are shown in Figure 1-3C, in which ratios ≥ 1.5-fold are shadowed. The expression of 242 proteins changed ≥ 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 1-1).

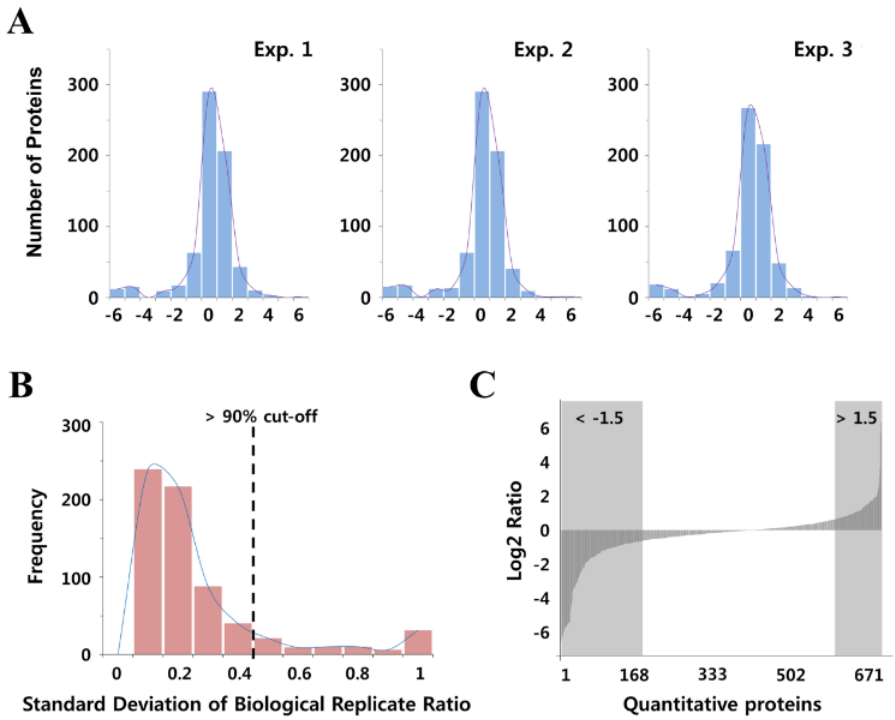


Figure 1-3. Distribution of log2 NSAF ratios and differentially expressed proteome

(A) The distributions of log2 NSAF ratios for primary cancer cells versus metastatic cancer cells were obtained by comparing 3 biological replicates from the label-free quantification experiments. (B) Fold-change cutoff of protein expression was considered the standard deviation of the 3 replicates. Ninety percent of all identified proteins were within less than 0.5 standard deviations. (C) Protein ratios are arranged in ascending order, resulting in a sigmoidal curve. The light shaded area represents unregulated protein groups with a less than 1.5-fold change in expression, and the dark shaded area represents protein groups that undergo more than a 1.5-fold change.

Table 1-1. *Up*- and *down*- regulated proteins

#	IPI	MW (kDa)	Expression ratio[a] Log_2(NCI-H1755/NCI-H1703)	*p* value[b]	Gene Symbol	Protein name
1	IPI00001453	55.3923	-7.47	0.00549	INA	Alpha-internexin
2	IPI00397526	230.7853	-6.87	0.00389	MYH10	Isoform 1 of Myosin-10
3	IPI00607787	58.6824	-6.64	0.00410	UAP1	Isoform 3 of UDP-N-acetylhexosamine pyrophosphorylase
4	IPI00856045	616.6283	-6.59	0.00284	AHNAK2	Isoform 1 of Protein AHNAK2
5	IPI00333619	54.8498	-6.52	0.00265	ALDH3A2	Isoform 1 of Fatty aldehyde dehydrogenase
6	IPI00178150	139.8838	-6.36	0.00830	KIF4A	Isoform 1 of Chromosome-associated kinesin KIF4A
7	IPI00237884	180.9821	-6.26	0.04158	AKAP12	Isoform 1 of A-kinase anchor protein 12
8	IPI00218775	51.2136	-6.12	0.00507	FKBP5	Peptidyl-prolyl cis-trans isomerase FKBP5
9	IPI00023972	50.6484	-6.11	0.00408	DDX47	Probable ATP-dependent RNA helicase DDX47
10	IPI00003505	48.5521	-5.96	0.00392	TRIP13	Isoform 1 of Pachytene checkpoint protein 2 homolog
11	IPI00396627	92.0913	-5.95	0.01184	ELAC2	Isoform 1 of Zinc phosphodiesterase ELAC protein 2
12	IPI00022977	42.6451	-5.89	0.00851	CKB	Creatine kinase B-type
13	IPI00294187	75.5654	-5.89	0.00076	PADI2	Protein-arginine deiminase type-2
14	IPI00017303	104.7458	-5.89	0.02007	MSH2	DNA mismatch repair protein Msh2
15	IPI00218922	87.9994	-5.77	0.01096	SEC63	Translocation protein SEC63 homolog
16	IPI00292894	91.8114	-5.72	0.00950	TSR1	Pre-rRNA-processing protein TSR1 homolog
17	IPI00553109	117.5145	-5.70	0.02743	PITRM1	Uncharacterized protein
18	IPI00165949	107.8444	-5.67	0.01092	ERAP1	Isoform 2 of Endoplasmic reticulum aminopeptidase 1
19	IPI00165092	53.201	-5.66	0.00146	YARS2	Tyrosyl-tRNA synthetase, mitochondrial
20	IPI00290439	72.3845	-5.63	0.00094	SRPK1	cDNA FLJ58405, highly similar to Serine/threonine-protein kinase SRPK1
21	IPI00554777	62.1702	-5.07	0.00007	ASNS	Asparagine synthetase [glutamine-hydrolyzing]
22	IPI00215893	32.8191	-4.87	0.00263	HMOX1	Heme oxygenase 1
23	IPI00294891	88.9752	-4.66	0.01774	NOP2	Isoform 1 of Putative ribosomal RNA methyltransferase NOP2
24	IPI00005024	148.8583	-4.45	0.00194	MYBBP1A	Isoform 1 of Myb-binding protein 1A
25	IPI00032158	101.2069	-4.27	0.00106	NAA15	Isoform 2 of N-alpha-acetyltransferase 15, NatA auxiliary subunit
26	IPI00550882	35.9808	-3.80	0.00734	PYCR1	Pyrroline-5-carboxylate reductase
27	IPI00033036	52.8923	-3.58	0.00087	METAP2	Methionine aminopeptidase 2
28	IPI00396203	132.6026	-3.56	0.00737	TBCD	Isoform 1 of Tubulin-specific chaperone D
29	IPI00218728	46.6374	-3.52	0.00492	PAFAH1B1	Isoform 1 of Platelet-activating factor acetylhydrolase IB subunit alpha
30	IPI00004534	144.7338	-3.47	0.00061	PFAS	Phosphoribosylformylglycinamidine synthase
31	IPI00024403	60.1316	-3.30	0.00526	CPNE3	Copine-3
32	IPI00829992	119.5254	-3.25	0.00027	MYO1C	Isoform 3 of Myosin-Ic
33	IPI00018350	82.2883	-3.25	0.00147	MCM5	DNA replication licensing factor MCM5

Continue.

34	IPI00217686	96.5605	-3.23	0.00267	FTSJ3	Putative rRNA methyltransferase 3
35	IPI00784414	88.0696	-3.15	0.00004	STAT3	Isoform 1 of Signal transducer and activator of transcription 3
36	IPI00014197	27.3347	-3.08	0.00551	CDV3	Isoform 1 of Protein CDV3 homolog
37	IPI00334907	31.5403	-2.88	0.00841	PITPNB	Isoform 1 of Phosphatidylinositol transfer protein beta isoform
38	IPI00178431	73.4589	-2.77	0.01296	RECQL	ATP-dependent DNA helicase Q1
39	IPI00384456	152.7899	-2.70	0.00265	MSH6	Isoform GTBP-N of DNA mismatch repair protein Msh6
40	IPI00001734	45.3561	-2.64	0.00001	PSAT1	Phosphoserine aminotransferase
41	IPI00015973	112.5878	-2.55	0.00056	EPB41L2	Band 4.1-like protein 2
42	IPI00016249	69.7209	-2.54	0.00811	FXR1	Isoform 1 of Fragile X mental retardation syndrome-related protein 1
43	IPI00004233	358.6286	-2.51	0.00200	MKI67	Isoform Long of Antigen KI-67
44	IPI00301263	236.0221	-2.40	0.00059	CAD	CAD protein
45	IPI01014863	41.3504	-2.38	0.01202	ACAT2	Acetyl-CoA acetyltransferase, cytosolic
46	IPI00011200	56.6506	-2.30	0.00111	PHGDH	D-3-phosphoglycerate dehydrogenase
47	IPI00306369	86.4728	-2.28	0.00412	NSUN2	tRNA (cytosine(34)-C(5))-methyltransferase
48	IPI00550385	838.3142	-2.28	0.00627	MACF1	Isoform 1 of Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5
49	IPI00744648	146.2052	-2.22	0.00719	SPAG9	Isoform 1 of C-Jun-amino-terminal kinase-interacting protein 4
50	IPI00220637	58.7787	-2.08	0.03169	SARS	Seryl-tRNA synthetase, cytoplasmic
51	IPI00333067	109.5883	-2.00	0.01436	HERC4	Isoform 1 of Probable E3 ubiquitin-protein ligase HERC4
52	IPI00376005	20.1709	-1.97	0.00290	EIF5A	Isoform 2 of Eukaryotic translation initiation factor 5A-1
53	IPI00216319	28.2196	-1.97	0.00063	YWHAH	14-3-3 protein eta
54	IPI00748303	117.6924	-1.95	0.00609	ZFR	Uncharacterized protein
55	IPI00299524	157.1863	-1.92	0.00019	NCAPD2	Condensin complex subunit 1
56	IPI00000030	66.1829	-1.88	0.00846	PPP2R5D	Isoform Delta-1 of Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform
57	IPI00025273	107.7684	-1.84	0.00853	GART	Isoform Long of Trifunctional purine biosynthetic protein adenosine-3
58	IPI00013214	95.9103	-1.80	0.00011	MCM3	cDNA FLJ55599, highly similar to DNA replication licensing factor MCM3
59	IPI00441473	72.6851	-1.75	0.00033	PRMT5	Protein arginine N-methyltransferase 5
60	IPI00218606	15.8077	-1.74	0.04921	RPS23	40S ribosomal protein S23
61	IPI00002459	75.2808	-1.70	0.02534	ANXA6	Uncharacterized protein
62	IPI00783313	93.1375	-1.69	0.01078	PYGL	Glycogen phosphorylase, liver form
63	IPI00017334	29.8046	-1.67	0.00163	PHB	Prohibitin
64	IPI00003519	105.3856	-1.66	0.00540	EFTUD2	116 kDa U5 small nuclear ribonucleoprotein component
65	IPI00644431	53.6972	-1.65	0.01897	DDX39	cDNA FLJ55484, highly similar to ATP-dependent RNA helicase DDX39
66	IPI00015897	37.4897	-1.63	0.04521	CHORDC1	Isoform 1 of Cysteine and histidine-rich domain-containing protein 1
67	IPI00783097	83.1676	-1.62	0.02221	GARS	Glycyl-tRNA synthetase
68	IPI00219616	34.8347	-1.61	0.03823	PRPS1	Ribose-phosphate pyrophosphokinase 1
69	IPI00218914	54.8628	-1.60	0.00008	ALDH1A1	Retinal dehydrogenase 1
70	IPI00007928	273.6086	-1.59	0.00073	PRPF8	Pre-mRNA-processing-splicing factor 8
71	IPI00007334	150.5571	-1.57	0.02266	ACIN1	Isoform 1 of Apoptotic chromatin condensation inducer in the nucleus

Continue.

72	IPI00026569	40.8458	-1.56	0.00017	HLA-A	HLA class I histocompatibility antigen, A-1 alpha chain
73	IPI00101186	143.7051	-1.49	0.01579	RRP12	Isoform 1 of RRP12-like protein
74	IPI00385042	73.9673	-1.46	0.00283	GTPBP4	Nucleolar GTP-binding protein 1
75	IPI00290142	66.6907	-1.46	0.00027	CTPS	CTP synthase 1
76	IPI00219217	36.6386	-1.43	0.00026	LDHB	L-lactate dehydrogenase B chain
77	IPI00001159	292.7644	-1.41	0.00070	GCN1L1	Translational activator GCN1
78	IPI00298696	152.2035	-1.40	0.00625	TCOF1	Isoform 2 of Treacle protein
79	IPI00411559	147.1879	-1.36	0.00077	SMC4	Isoform 1 of Structural maintenance of chromosomes protein 4
80	IPI00219029	46.2481	-1.29	0.02572	GOT1	Aspartate aminotransferase, cytoplasmic
81	IPI00419979	58.0772	-1.28	0.01209	LOC646214	Serine/threonine-protein kinase PAK 2
82	IPI00329633	83.4378	-1.28	0.02555	TARS	Threonyl-tRNA synthetase, cytoplasmic
83	IPI00026781	273.4271	-1.27	0.00014	FASN	Fatty acid synthase
84	IPI00218830	48.1415	-1.27	0.00822	NMT1	Isoform Short of Glycylpeptide N-tetradecanoyltransferase 1
85	IPI00008433	22.877	-1.25	0.00300	RPS5	40S ribosomal protein S5
86	IPI00029629	70.9732	-1.23	0.02829	TRIM25	E3 ubiquitin/ISG15 ligase TRIM25
87	IPI00216694	67.6019	-1.22	0.00403	PLS3	Plastin-3
88	IPI00012462	67.8525	-1.22	0.01830	EIF2A	Eukaryotic translation initiation factor 2A
89	IPI00184330	101.8981	-1.19	0.00903	MCM2	DNA replication licensing factor MCM2
90	IPI00553185	60.5354	-1.15	0.00111	CCT3	T-complex protein 1 subunit gamma
91	IPI00234252	122.8674	-1.14	0.03174	SMARCC1	SWI/SNF complex subunit SMARCC1
92	IPI00299904	81.309	-1.14	0.02542	MCM7	Isoform 1 of DNA replication licensing factor MCM7
93	IPI00029019	114.5341	-1.13	0.00104	UBAP2L	Isoform 2 of Ubiquitin-associated protein 2-like
94	IPI00013683	50.4327	-1.13	0.02167	TUBB3	Tubulin beta-3 chain
95	IPI00024664	93.3095	-1.12	0.00654	USP5	Isoform Long of Ubiquitin carboxyl-terminal hydrolase 5
96	IPI00000816	29.175	-1.10	0.00323	YWHAE	Isoform 1 of 14-3-3 protein epsilon
97	IPI00022462	84.8736	-1.09	0.00103	TFRC	Transferrin receptor protein 1
98	IPI00031801	40.0894	-1.06	0.01178	CSDA	Isoform 1 of DNA-binding protein A
99	IPI00395865	47.82	-1.06	0.00044	RBBP7	Histone-binding protein RBBP7
100	IPI00964079	57.145	-1.05	0.03078	CCT5	Uncharacterized protein
101	IPI00909703	45.7302	-1.05	0.02712	ANXA11	Uncharacterized protein
102	IPI00000873	140.4675	-1.02	0.00746	VARS	Valyl-tRNA synthetase
103	IPI00176903	43.4765	-1.01	0.01776	PTRF	Isoform 1 of Polymerase I and transcript release factor
104	IPI00002214	57.8625	-0.98	0.01622	KPNA2	Importin subunit alpha-2
105	IPI00031517	92.8906	-0.98	0.02801	MCM6	DNA replication licensing factor MCM6
106	IPI00027626	58.0253	-0.97	0.03482	CCT6A	T-complex protein 1 subunit zeta
107	IPI00414676	83.2673	-0.96	0.00102	HSP90AB1	Heat shock protein HSP 90-beta
108	IPI00294536	38.4394	-0.94	0.01691	STRAP	cDNA FLJ51909, highly similar to Serine-threonine kinase receptor-associatedprotein
109	IPI00027252	33.2976	-0.93	0.02274	PHB2	Prohibitin-2

Continue.

110	IPI00028031	70.3916	-0.93	0.01170	ACADVL	cDNA FLJ56425, highly similar to Very-long-chain specific acyl-CoAdehydrogenase, mitochondrial
111	IPI00031461	50.6655	-0.92	0.02047	GDI2	cDNA FLJ60299, highly similar to Rab GDP dissociation inhibitor beta
112	IPI00012268	100.2022	-0.92	0.00741	PSMD2	26S proteasome non-ATPase regulatory subunit 2
113	IPI00301058	39.8291	-0.92	0.02523	VASP	Vasodilator-stimulated phosphoprotein
114	IPI00172656	52.6244	-0.91	0.02005	FAF2	FAS-associated factor 2
115	IPI00003768	66.0797	-0.88	0.03541	PES1	Isoform 1 of Pescadillo homolog
116	IPI00549248	32.5755	-0.86	0.03177	NPM1	Isoform 1 of Nucleophosmin
117	IPI00013452	170.5932	-0.85	0.00338	EPRS	Bifunctional aminoacyl-tRNA synthetase
118	IPI00018931	91.71	-0.85	0.01438	VPS35	Vacuolar protein sorting-associated protein 35
119	IPI00026202	18.1104	-0.84	0.00309	RPL18A	60S ribosomal protein L18a
120	IPI00291175	123.8013	-0.84	0.00704	VCL	Isoform 1 of Vinculin
121	IPI00029079	76.7167	-0.83	0.03810	GMPS	GMP synthase [glutamine-hydrolyzing]
122	IPI00290460	35.612	-0.83	0.04217	EIF3G	Eukaryotic translation initiation factor 3 subunit G
123	IPI00179473	47.6874	-0.82	0.00591	SQSTM1	Isoform 1 of Sequestosome-1
124	IPI00290204	51.5584	-0.81	0.02268	SNRNP70	Isoform 1 of U1 small nuclear ribonucleoprotein 70 kDa
125	IPI00550689	55.2102	-0.81	0.01408	C22orf28	tRNA-splicing ligase RtcB homolog
126	IPI00021435	48.6356	-0.80	0.01208	PSMC2	26S protease regulatory subunit 7
127	IPI00007074	59.1451	-0.80	0.00705	YARS	Tyrosyl-tRNA synthetase, cytoplasmic
128	IPI00926977	45.7991	-0.78	0.02952	PSMC6	26S protease regulatory subunit 10B
129	IPI00018274	128.6815	-0.77	0.01884	EGFR	Isoform 1 of Epidermal growth factor receptor
130	IPI00304596	54.2316	-0.77	0.03193	NONO	Non-POU domain-containing octamer-binding protein
131	IPI00026625	155.203	-0.76	0.01483	NUP155	Isoform 1 of Nuclear pore complex protein Nup155
132	IPI00478231	21.7684	-0.72	0.03730	RHOA	Transforming protein RhoA
133	IPI00021728	38.3897	-0.71	0.04327	EIF2S2	Eukaryotic translation initiation factor 2 subunit 2
134	IPI00008240	101.1175	-0.70	0.01495	MARS	Methionyl-tRNA synthetase, cytoplasmic
135	IPI00301936	38.997	-0.70	0.01963	ELAVL1	cDNA FLJ60076, highly similar to ELAV-like protein 1
136	IPI01019005	53.4908	-0.70	0.02830	ATXN10	Ataxin-10
137	IPI00297779	57.4899	-0.70	0.00365	CCT2	T-complex protein 1 subunit beta
138	IPI00017617	69.1497	-0.69	0.02957	DDX5	Probable ATP-dependent RNA helicase DDX5
139	IPI00002966	94.3319	-0.69	0.00451	HSPA4	Heat shock 70 kDa protein 4
140	IPI00306960	62.9443	-0.69	0.00241	NARS	Asparaginyl-tRNA synthetase, cytoplasmic
141	IPI00783271	157.9122	-0.69	0.01934	LRPPRC	Leucine-rich PPR motif-containing protein, mitochondrial
142	IPI00021187	50.2294	-0.68	0.00712	RUVBL1	Isoform 1 of RuvB-like 1
143	IPI00218342	107.4363	-0.68	0.04579	MTHFD1	C-1-tetrahydrofolate synthase, cytoplasmic
144	IPI00000690	66.2956	-0.66	0.01517	AIFM1	Isoform 1 of Apoptosis-inducing factor 1, mitochondrial
145	IPI00479262	158.521	-0.65	0.01142	EIF4G1	eukaryotic translation initiation factor 4 gamma 1 isoform 1
146	IPI00003881	45.6719	-0.64	0.02758	HNRNPF	Heterogeneous nuclear ribonucleoprotein F
147	IPI00290566	60.3452	-0.64	0.00387	TCP1	T-complex protein 1 subunit alpha

Continue.

148	IPI00012442	52.1628	-0.63	0.01410	G3BP1	Ras GTPase-activating protein-binding protein 1
149	IPI00011126	49.1864	-0.63	0.02509	PSMC1	26S protease regulatory subunit 4
150	IPI00440493	59.7521	-0.61	0.00859	ATP5A1	ATP synthase subunit alpha, mitochondrial
151	IPI00021700	28.7693	0.59	0.01195	PCNA	Proliferating cell nuclear antigen
152	IPI00021405	74.1407	0.61	0.00303	LMNA	Isoform A of Prelamin-A/C
153	IPI00783872	76.8599	0.61	0.02125	CAPRIN1	Isoform 1 of Caprin-1
154	IPI00016801	61.3997	0.61	0.03002	GLUD1	Glutamate dehydrogenase 1, mitochondrial
155	IPI00012074	70.9439	0.62	0.00173	HNRNPR	Isoform 1 of Heterogeneous nuclear ribonucleoprotein R
156	IPI00908881	59.9918	0.63	0.04730	GPI	Glucose-6-phosphate isomerase
157	IPI00783862	22.1187	0.63	0.00048	BLVRB	Flavin reductase
158	IPI00102069	42.5039	0.64	0.00908	EIF3M	Eukaryotic translation initiation factor 3 subunit M
159	IPI00219097	24.0346	0.65	0.00358	HMGB2	High mobility group protein B2
160	IPI00418471	53.6527	0.66	0.00015	VIM	Vimentin
161	IPI00031583	109.197	0.69	0.00668	USO1	Isoform 2 of General vesicular transport factor p115
162	IPI00783641	60.373	0.69	0.02486	TXNRD1	Isoform 3 of Thioredoxin reductase 1, cytoplasmic
163	IPI00217966	36.6892	0.73	0.00859	LDHA	Isoform 1 of L-lactate dehydrogenase A chain
164	IPI00219757	23.3567	0.73	0.00432	GSTP1	Glutathione S-transferase P
165	IPI00295851	107.146	0.74	0.00523	COPB1	Coatomer subunit beta
166	IPI00220642	28.3031	0.75	0.02477	YWHAG	14-3-3 protein gamma
167	IPI00011603	60.9796	0.75	0.02893	PSMD3	26S proteasome non-ATPase regulatory subunit 3
168	IPI00219420	141.5471	0.76	0.00661	SMC3	Structural maintenance of chromosomes protein 3
169	IPI00789155	37.1077	0.76	0.00830	CALU	calumenin isoform c precurosr
170	IPI00010105	26.5982	0.77	0.01270	EIF6	Eukaryotic translation initiation factor 6
171	IPI00026182	32.9492	0.77	0.02080	CAPZA2	F-actin-capping protein subunit alpha-2
172	IPI00479186	57.9375	0.82	0.00276	PKM2	Isoform M2 of Pyruvate kinase isozymes M1/M2
173	IPI00027230	92.4717	0.82	0.00096	HSP90B1	Endoplasmin
174	IPI00002460	50.3169	0.82	0.03805	ANXA7	Isoform 1 of Annexin A7
175	IPI00010796	57.1181	0.83	0.00017	P4HB	Protein disulfide-isomerase
176	IPI00256684	105.3642	0.83	0.02573	AP2A1	Isoform B of AP-2 complex subunit alpha-1
177	IPI00784154	61.0557	0.84	0.01822	HSPD1	60 kDa heat shock protein, mitochondrial
178	IPI00016613	45.9096	0.85	0.03114	CSNK2A1	Uncharacterized protein
179	IPI00031397	80.4231	0.89	0.04289	ACSL3	Long-chain-fatty-acid--CoA ligase 3
180	IPI00014424	50.4706	0.90	0.01501	EEF1A2	Elongation factor 1-alpha 2
181	IPI00028091	47.3719	0.92	0.00307	ACTR3	Actin-related protein 3
182	IPI00465439	39.4206	0.94	0.00688	ALDOA	Fructose-bisphosphate aldolase A
183	IPI00022334	48.5362	0.95	0.02032	OAT	Ornithine aminotransferase, mitochondrial
184	IPI00418262	48.4083	0.97	0.04258	ALDOC	Fructose-bisphosphate aldolase
185	IPI00550363	22.3919	0.98	0.01963	TAGLN2	Transgelin-2

Continue.

186	IPI00011284	30.0376	1.00	0.00434	COMT	Isoform Membrane-bound of Catechol O-methyltransferase
187	IPI00017375	82.9709	1.02	0.00545	SEC23A	Protein transport protein Sec23A
188	IPI00019912	79.6885	1.05	0.00673	HSD17B4	Peroxisomal multifunctional enzyme type 2
189	IPI00302592	280.7294	1.05	0.00001	FLNA	Isoform 2 of Filamin-A
190	IPI00014898	531.7839	1.09	0.00055	PLEC	Isoform 1 of Plectin
191	IPI00140420	101.2127	1.11	0.00275	SND1	Staphylococcal nuclease domain-containing protein 1
192	IPI00008982	87.3029	1.12	0.00604	ALDH18A1	Isoform Long of Delta-1-pyrroline-5-carboxylate synthase
193	IPI00024466	174.9825	1.12	0.02067	UGGT1	Isoform 1 of UDP-glucose:glycoprotein glucosyltransferase 1
194	IPI00141318	66.0222	1.15	0.02675	CKAP4	Cytoskeleton-associated protein 4
195	IPI00009904	72.934	1.16	0.00203	PDIA4	Protein disulfide-isomerase A4
196	IPI00012069	27.2967	1.16	0.00060	NQO1	NAD(P)H dehydrogenase [quinone] 1
197	IPI00006865	24.594	1.16	0.00606	SEC22B	Vesicle-trafficking protein SEC22b
198	IPI00219301	31.5542	1.18	0.01228	MARCKS	Myristoylated alanine-rich C-kinase substrate
199	IPI00289334	276.9334	1.19	0.00010	FLNB	Isoform 1 of Filamin-B
200	IPI00030781	87.3369	1.19	0.00498	STAT1	Isoform Alpha of Signal transducer and activator of transcription 1-alpha/beta
201	IPI00410067	101.4326	1.21	0.02741	ZC3HAV1	Isoform 1 of Zinc finger CCCH-type antiviral protein 1
202	IPI00004358	96.6983	1.24	0.02456	PYGB	Glycogen phosphorylase, brain form
203	IPI00414127	23.3108	1.34	0.00868	RANBP1	Ran-specific GTPase-activating protein
204	IPI00031131	46.4815	1.35	0.02408	C20orf3	Isoform 1 of Adipocyte plasma membrane-associated protein
205	IPI00000105	99.326	1.35	0.00260	MVP	Major vault protein
206	IPI00013070	90.2927	1.39	0.00222	HNRNPUL1	Isoform 1 of Heterogeneous nuclear ribonucleoprotein U-like protein 1
207	IPI00479722	28.7239	1.44	0.00328	PSME1	Proteasome activator complex subunit 1
208	IPI00018873	55.5225	1.49	0.00106	NAMPT	Nicotinamide phosphoribosyltransferase
209	IPI00008868	270.6344	1.53	0.00110	MAP1B	Microtubule-associated protein 1B
210	IPI00013808	104.8572	1.54	0.00101	ACTN4	Alpha-actinin-4
211	IPI00025084	28.3167	1.54	0.02978	CAPNS1	Calpain small subunit 1
212	IPI00442073	20.5671	1.57	0.00283	CSRP1	Cysteine and glycine-rich protein 1
213	IPI00017726	26.9231	1.59	0.00825	HSD17B10	Isoform 1 of 3-hydroxyacyl-CoA dehydrogenase type-2
214	IPI00329801	35.9386	1.60	0.00001	ANXA5	Annexin A5
215	IPI00030009	69.9763	1.61	0.01561	PAPSS2	Isoform A of Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 2
216	IPI00294578	77.3291	1.67	0.00018	TGM2	Isoform 1 of Protein-glutamine gamma-glutamyltransferase 2
217	IPI00742682	267.2893	1.68	0.02558	TPR	Nucleoprotein TPR
218	IPI00215687	65.4611	1.74	0.00215	GLS	Isoform 3 of Glutaminase kidney isoform, mitochondrial
219	IPI00005614	274.6134	1.76	0.00146	SPTBN1	Isoform Long of Spectrin beta chain, brain 1
220	IPI00550069	49.9744	1.76	0.01042	RNH1	Ribonuclease inhibitor
221	IPI00883655	73.503	1.77	0.00294	DPYSL2	dihydropyrimidinase-related protein 2 isoform 1
222	IPI00017283	113.7941	1.80	0.02110	IARS2	Isoleucyl-tRNA synthetase, mitochondrial
223	IPI00246975	26.5611	1.80	0.01699	GSTM3	Glutathione S-transferase Mu 3

Continue.

224	IPI00016862	51.7004	1.82	0.00510	GSR	Isoform Mitochondrial of Glutathione reductase, mitochondrial
225	IPI00289758	79.9991	1.86	0.00003	CAPN2	Calpain-2 catalytic subunit
226	IPI00182757	102.9031	1.94	0.04763	KIAA1967	Isoform 1 of Protein KIAA1967
227	IPI00844215	284.5427	1.96	0.00473	SPTAN1	Isoform 1 of Spectrin alpha chain, brain
228	IPI00003479	41.3919	2.00	0.00818	MAPK1	Mitogen-activated protein kinase 1
229	IPI00027223	46.6605	2.01	0.00096	IDH1	Isocitrate dehydrogenase [NADP] cytoplasmic
230	IPI00219525	51.8742	2.01	0.00009	PGD	6-phosphogluconate dehydrogenase, decarboxylating
231	IPI00414717	134.5539	2.03	0.02230	GLG1	Isoform 2 of Golgi apparatus protein 1
232	IPI00643920	68.8155	2.04	0.00156	TKT	cDNA FLJ54957, highly similar to Transketolase
233	IPI00744692	37.5417	2.22	0.00003	TALDO1	Transaldolase
234	IPI00292771	238.2597	2.23	0.00176	NUMA1	Isoform 1 of Nuclear mitotic apparatus protein 1
235	IPI00001539	41.9242	2.38	0.00085	ACAA2	3-ketoacyl-CoA thiolase, mitochondrial
236	IPI00215743	152.4706	2.60	0.00010	RRBP1	Isoform 3 of Ribosome-binding protein 1
237	IPI00017376	86.4811	2.86	0.00153	SEC23B	Protein transport protein Sec23B
238	IPI00216008	62.4697	3.35	0.00232	G6PD	Isoform Long of Glucose-6-phosphate 1-dehydrogenase
239	IPI00413641	35.8539	3.81	0.00223	AKR1B1	Aldose reductase
240	IPI00553131	38.282	5.90	0.00032	GALE	UDP-glucose 4-epimerase
241	IPI00006663	56.3814	6.57	0.00044	ALDH2	Aldehyde dehydrogenase, mitochondrial
242	IPI00013744	129.2979	6.60	0.00007	ITGA2	Integrin alpha-2

[a)] Significant difference expression $\log_2$ ratio of NCI-H1755/NCI-H1703 with NSAF value. [b)] Significant difference in t-test (p-value < 0.05).

Identification of N-terminal peptides using BSA as control

The scheme with which N-terminal peptides were identified is shown in Figure 3. The N-termini of proteins are characterized by an α-amine, as opposed to the ε-amines that are on lysine side chains. Thus, ε-amines on lysine side chains had to be blocked. We blocked the α-amine and ε-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 (30).

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 (36). Thus, our BSA had an aspartic acid at residue 25 as its natural N-terminus.

Acetylated BSA was digested with trypsin and analyzed by MALDI-MS (Figure 1-4A). 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 α-amine and ε-amine and confirmed with the peptide fingerprint by MS/MS analysis (Figure 1-4B).

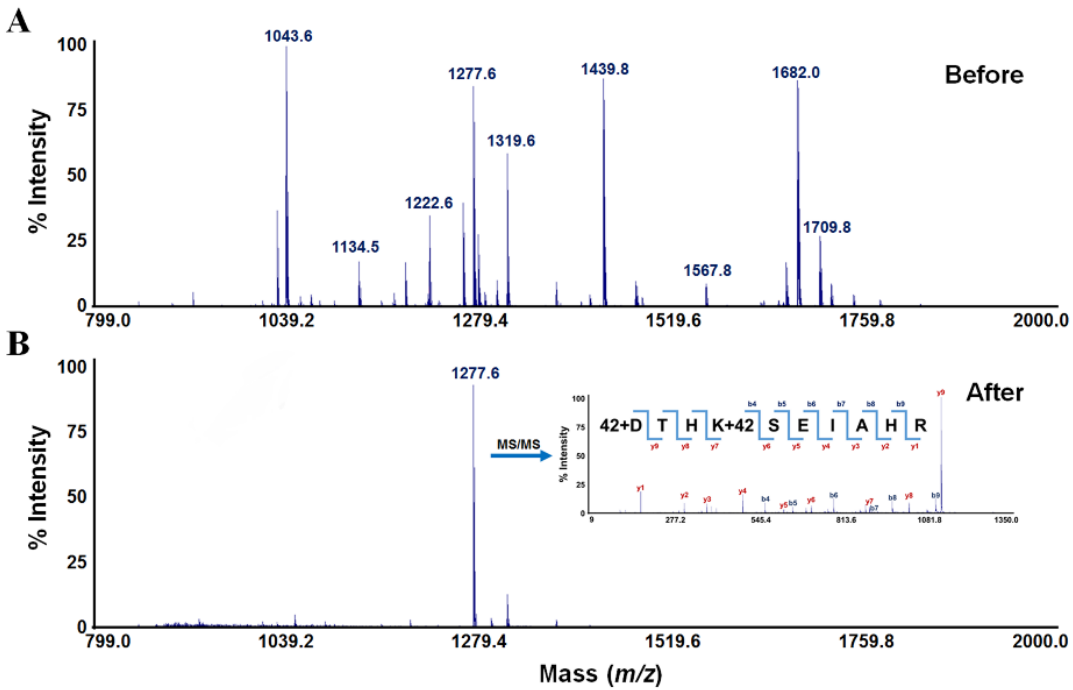


Figure 1-4. N-terminal peptide analysis of BSA control

(A) MS peaks are trypsin-digested peptides of acetylated BSA. (B) With our protocol, the labeled major ions correspond to the N-terminal peptides from BSA.

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 (30). 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 (Figure 1-5 A and B).

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) (Figure 1-5 C and D).

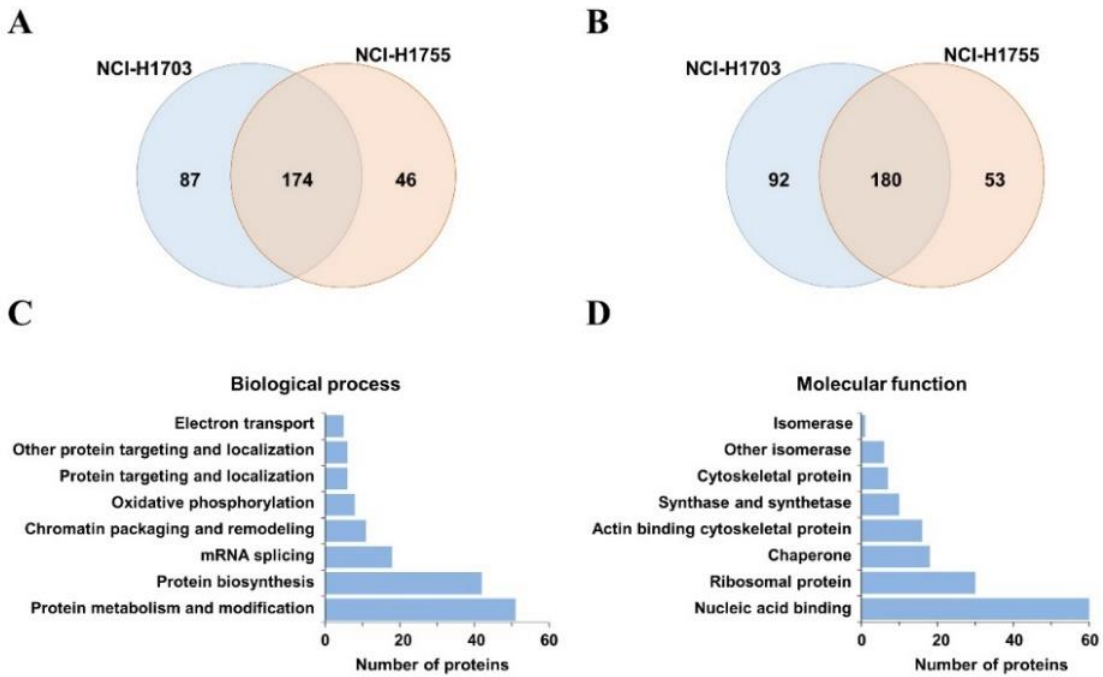


Figure 1-5. Summary of the identification of N-terminal peptides

(A) Numbers of all identified proteins and (B) peptides were shown in Venn diagrams. (C) Gene ontology (GO) for biological process and (D) molecular function of all identified proteins was performed using the DAVID bioinformatics tool.

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 (Figure 1-6 A and B): (1) initial methionine depletion, NCI-H1703 (169 peptides, 62.1%) and NCI-H1755 (148 peptides, 63.5%); (2) initial methionine nondepletion, NCI-H1703 (37 peptides, 13.6%) and NCI-H1755 (28 peptides, 12.1%); (3) signal peptide depletion, NCI-H1703 (15 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%).

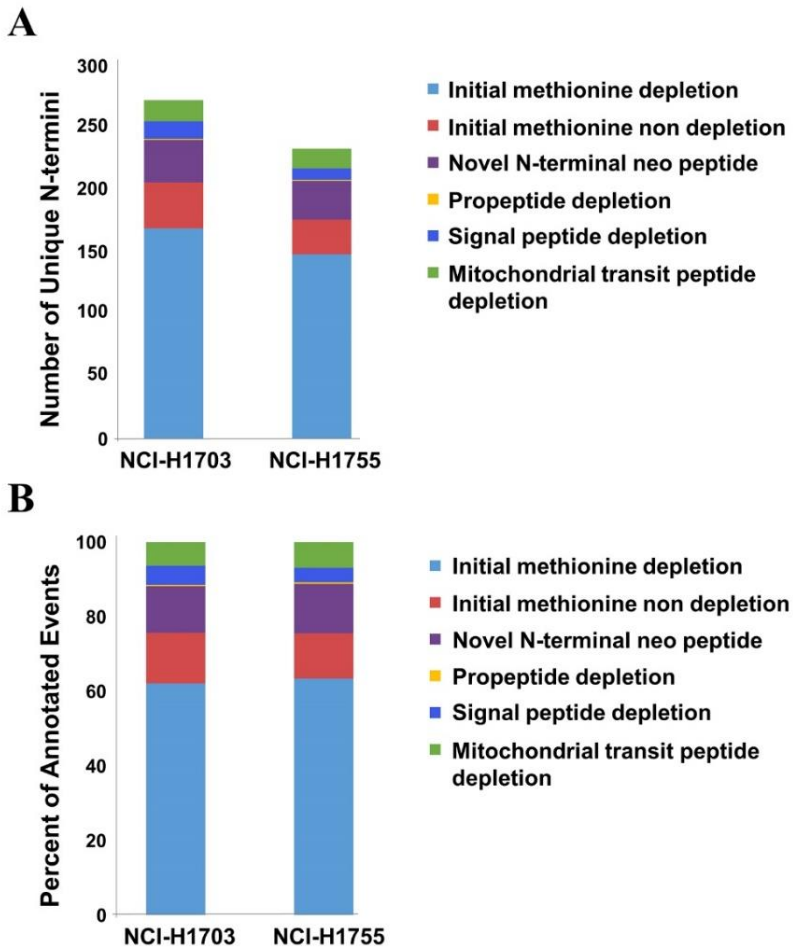


Figure 1-6. Site annotation of N-terminal peptides

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).

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 (Figure 1-7). Fourteen proteins were involved in the focal adhesion pathway in relation of cell invasion, growth, proliferation, and migration (Table 1-2), 5 of which (FLNA, FLNB, CAV1, MYL12B, and CAPN2) were common in the two parallel experiments.

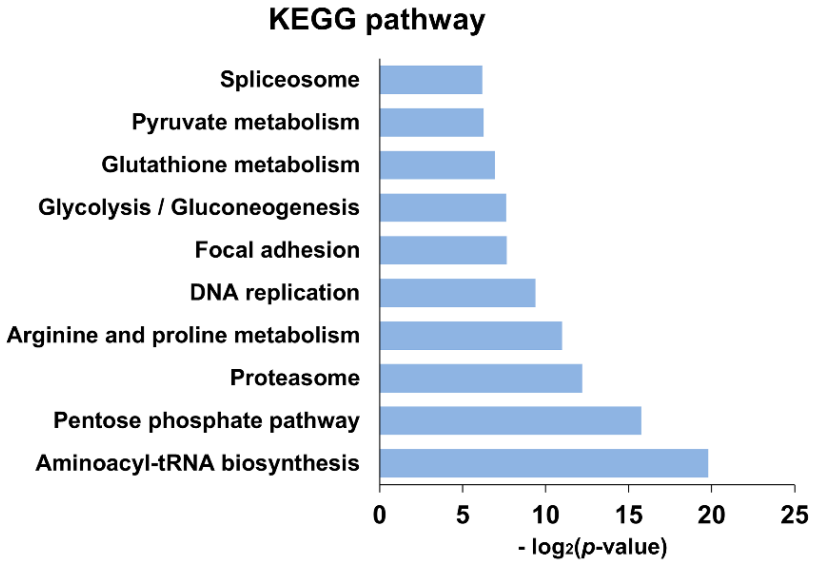


Figure 1-7. Pathways identified using differentially expressed proteins from both experiments

The numbers of significantly differentiated proteins associated with each pathway are shown in the bar graph.

Table 1-2. Focal adhesion pathway related protein list

#	IPI	peptide sequence	Peptide sequence site	Expression ratio	N-terminal peptides		Gene Symbol	Protein name
				log_2(NCI-H1755/NCI-H1703)	NCI-H1703	NCI-H1755		
1	IPI00302592	M.PATEKDLAEDAPWKKIQQNTFTR.W	Novel N-termini peptide	1.05	O	O	FLNA	Isoform 2 of Filamin-A
2	IPI00289334	M.PVTEKDLAEDAPWKKIQQNTFTR.W	Initial methionine depletion*	1.19	X	O	FLNB	Isoform 1 of Filamin-B
3	IPI00009236	M.SGGKYVDSEGHLYTVPIR.E	Initial methionine depletion	-0.04	X	O	CAV1	Isoform Alpha of Caveolin-1
4	IPI00033494	M.SSKKAKTKTTKKRPQR.A	Initial methionine depletion	0.04	X	O	MYL12B	Myosin regulatory light chain 12B
5	IPI00289758	M.AGIAAKLAKDR.E	Initial methionine depletion	1.86	X	O	CAPN2	calpain 2, (m/II) large subunit
6	IPI00004839	D.SSTCPGDYVLSVSENSR.V	Novel N-termini peptide	-	O	X	CRKL	v-crk sarcoma virus CT10 oncogene homolog (avian)-like
7	IPI00018195	M.AAAAAQGGGGGEPR.R	Initial methionine depletion	-	X	O	MAPK3	mitogen-activated protein kinase 3
8	IPI00218236	M.ADGELNVDSLITR.L	Initial methionine depletion	-	X	O	PPP1CB	protein phosphatase 1, catalytic subunit, beta isozyme
9	IPI00301058	-	-	-0.92	-	-	VASP	Vasodilator-stimulated phosphoprotein
10	IPI00291175	-	-	-0.84	-	-	VCL	Isoform 1 of Vinculin
11	IPI00478231	-	-	-0.72	-	-	RHOA	Transforming protein RhoA
12	IPI00013808	-	-	1.54	-	-	ACTN4	Alpha-actinin-4
13	IPI00003479	-	-	2.00	-	-	MAPK1	Mitogen-activated protein kinase 1
14	IPI00013744	-	-	6.60	-	-	ITGA2	Integrin alpha-2

* mark means that the peptide sequence site were not assigned in UniProtKB.

Focal adhesion pathway involved proteins were listed. List included the IPI accession number, gene symbol, peptide sequence and site from identified N-terminal peptide analysis. Expression ratio and p value were calculated by average NSAF value from label-free quantitative analysis. Information of peptide site by UniprotKB database was provided in this list.

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 (PATEKDLAEDAPWKKIQQNTFTR) in the NCI-H1703 and NCI-H1755 lines—showed differential expression in both cell lines in at least 1 experiments (Figure 1-8).

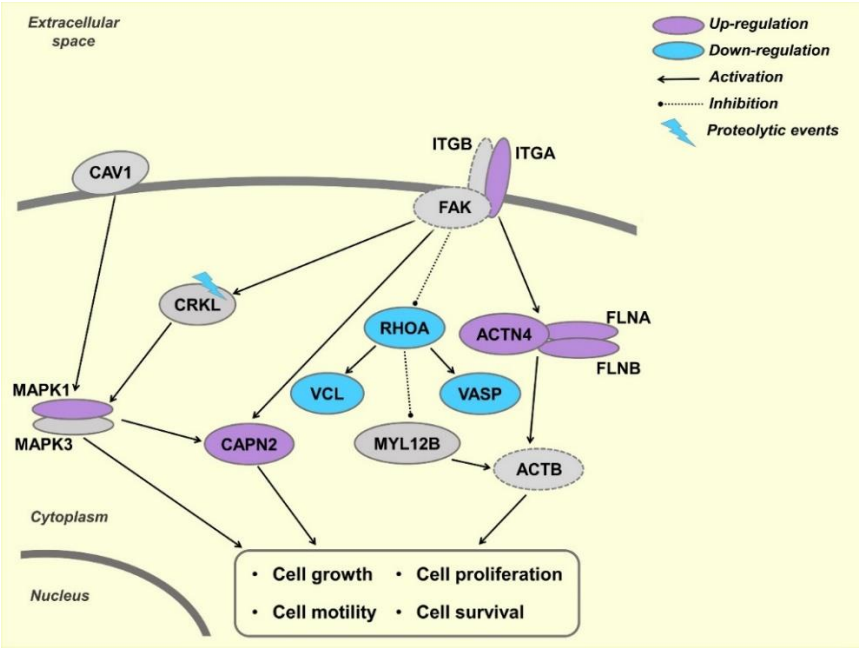


Figure 1-8. Deregulated focal adhesion pathway in NSCLC cell lines

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.

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 (37) 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.

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 (13). 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 (38) and MS/MS spectral counts (39), 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 (≤ 4 spectrum detected) and post translational modification proteins (40). However, the analysis is suitable for detection of subtle abundance changes in most proteins with high sensitivity and reproducibility (41).

In this study, we identified 2130 non-redundant 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 ≥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 (26). 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) (42).

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. In particular, LMNA is a metastatic biomarker of colorectal cancer cells (43) and a marker of embryonic stem cell differentiation (44), 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 (45); 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 (46, 47). 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 (30), dimethylation (48), iTRAQ (49), and PITC adman (50). 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 (51). 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 proteins were detected in NCI-H1703 cells, whereas 53 peptides from 46 proteins were identified in NCI-H1755 cells (Figure 1-5)—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 (52), 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 1-3). 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 (53-55). 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).

Table 1-3. Proteolytic events identified with less than 1.5 fold change

IPI	peptide sequence[a]	Ratio[b]	N-terminal analysis[c]	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.*HSNAAQTQTGEANR*.G	0.3	NCI-H1755	SHMT2	Serine hydroxymethyltransferase, mitochondrial

[a)] Observed peptide sequence from N-terminal peptide analysis is written by italics. [b)] Expression $\log_2$ ratio of NCI-H1755/NCI-H1703 with NSAF value by label-free analysis. [c)] Cell line with detected peptide sequences from N-terminal analysis.

We found 138 proteins that were common to both experiments. 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 (Figure 1-7). 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 (Figure 1-8). 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 (56). 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 (57). 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 (58).

MAP kinases mediate many processes in cancer cells, such as proliferation, migration, invasion, and metastasis (59, 60). Increased expression of MAPK1 promotes the expression of CAPN2, which functions in cell movement, migration, and invasion during metastasis (61). In the N-terminal peptide analysis, v-crk sarcoma virus CT10 oncogene 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 (62), we hypothesize that CRKL function is regulated by fragment events during metastasis.

In summary, we applied two proteome methods for biomarker discovery in lung cancer metastasis. Specially, N-terminal enrichment method was used for biomarker discovery for the first time. We can find that many of these quantitative proteins and N-terminal peptides are involved in pathways in cell migration, proliferation, and metastasis. Also, our datasets of proteins and fragment peptides in lung cells might be valuable in discovering and validating lung cancer biomarkers and metastasis markers.

CHAPTER II

Targeted proteomics predicts complete response after transarterial chemoembolization in hepatocellular carcinoma

INTRODUCTION

Hepatocellular carcinoma (HCC) is the fifth most common cancer and the third most common cause of cancer-related deaths worldwide (63). Recently, the treatment of HCC has well advanced after applications of curative therapeutic practices, such as surgical resection, liver transplantation, and local ablation (64). However, most HCC patients are diagnosed at advanced stage when curative treatment is no longer applicable. For these patients, transarterial chemoembolization (TACE) may be an effective treatment option for reducing systemic toxicity, increasing local antitumor effects, and improving survival (65). International Bridge study showed that TACE is the most widely used treatment for HCC worldwide, ahead of both surgical removal and systemic treatments (66). However, there are often shown unforeseeable outcomes after TACE in terms of treatment response and survival.

In real clinical practice, a high rate of recurrence and unsatisfactory treatment outcome after TACE remains troublesome and repeated TACE procedures are often needed, since the best response cannot always be achieved after one session of TACE, especially in large tumors (67). Georgiades et al. recommended that at least two TACE sessions should be performed before abandoning the procedure, on the basis of their observations that about half of patients who did not respond to initial TACE ultimately achieved response and that improved clinical outcomes were observed after

second course (68). Recently, Kim et al. reported that the complete response at initial TACE most strongly predicts survivals of patients with intermediate-stage HCC. However, it still remains unresolved which marker is the better for more accurate progsnostification in patients with HCC undergoing TACE.

Over the past decades, a large number of HCC diagnostic marker proteins including alpha-fetoprotein (AFP), *Lens culinaris* agglutinin A-reactive fraction of AFP (AFP-L3), and prothrombin induced by the absence of vitamin K or antagonist-II (PIVKA-II) have been discovered and , reported that dynamic change of these diagnostic marker proteins can predict outcome after TACE (69, 70). Therefore, identifying marker proteins that can help us to predict or prognosis of treatment outcomes before choosing this treatment option is an important endeavor in designing a treatment strategy.

Traditionally, the enzyme-linked immunosorbent assay (ELISA) utilizing antibodies is common quantitative assay for development of diagnostic marker proteins with high specificity and sensitivity (71). However, the immunoassay has major constraints that are the expensive and time-consuming development of specific antibodies, and the technical limitations for multiplex quantitation. In contrast, targeted proteomics approach through multiple reaction monitoring (MRM) assay suitable for multiplex quantitation of more than one hundreds of proteins with high accuracy and lower limit of quantitation (LLOQ) in efficiency cost (72). In addition, MRM assay has been shown consistent and reproducible data set across different laboratories in highly complex samples (73). More recently, Silvia *et al.* developed automated MRM data analysis workflow for validation of marker proteins in

large-scale clinical cohorts (74). In our previously study, we identified the HCC diagnostic markers using MRM and immunoassay from global data-mining. Additionally, these marker proteins showed difference level in HCC state and recovery state by treatments (75). Here, we applied marker-candidate proteins (MCPs) that have been previously reported as liver disease related proteins for TACE prognosis prediction.

Therefore, our aim was to identify pre-TACE marker proteins from the MCPs predicted to complete response after TACE, to ultimately suggest guideline for clinical decision making in future prospective studies.

MATERIALS AND METHODS

Study Population

This study was based on 180 HCC patients who were enrolled in a prospective cohort at Seoul National University Hospital (Seoul, Republic of Korea) as part of an ongoing study identifying the biomarkers associated with treatment response and prognosis in HCC. (Table 2-1). Patients with HCC who received TACE as the first-line therapy between 2008 and 2014 were considered eligible in this study. HCC was diagnosed by histological or radiological evaluation with reference to American Association for the Study of Liver (AASLD) or European Association for the Study of the Liver (EASL) guidelines.

TABLE 2-1. The clinicopathologic characteristics of the training and validation cohorts

		All cohort (*N*=180)				Training cohort (*N*=100)				Validation cohort (*N*=80)			
Variable		**Good Responders**	**(%)**	**Poor Responders**	**(%)**	**Good Responders**	**(%)**	**Poor Responders**	**(%)**	**Good Responders**	**(%)**	**Poor Responders**	**(%)**
Gender	Male	72	80.0%	82	91.1%	38	76.0%	47	94.0%	34	85.0%	35	87.5%
	Female	18	20.0%	8	8.9%	12	24.0%	3	6.0%	6	15.0%	5	12.5%
Age (years)	<60	32	35.6%	31	34.4%	19	38.0%	16	32.0%	13	32.5%	15	37.5%
	≥ 60	58	64.4%	59	65.6%	31	62.0%	34	68.0%	27	67.5%	25	62.5%
Etiology	Alcohol	4	4.4%	5	5.6%	2	4.0%	3	6.0%	2	5.0%	2	5.0%
	HBV	74	82.2%	67	74.4%	40	80.0%	39	78.0%	34	85.0%	28	70.0%
	HCV	9	10.0%	13	14.4%	6	12.0%	5	10.0%	3	7.5%	8	20.0%
	Others	3	3.3%	5	5.6%	**2**	4.0%	**3**	6.0%	1	2.5%	2	5.0%
Child-Pugh class	A	74	82.2%	63	70.0%	43	86.0%	36	72.0%	31	77.5%	27	67.5%
	B	16	17.8%	27	30.0%	7	14.0%	14	28.0%	9	22.5%	13	32.5%
MELD score	Mean ± SD	9.1 ± 2.6		9.4 ± 3.2		9.0 ± 2.4		8.8 ± 2.3		9.3 ± 2.9		10.2 ± 4.0	
Platelet (10^3/uL)	Mean ± SD	113.4 ± 53.1		116.9 ± 85.0		121.1 ± 56.3		115.0 ± 52.9		103.9 ± 47.9		120.3 ± 113.8	
ALT, IU/L	Mean ± SD	32.5 ± 19.9		39.9 ± 25.9		34.8 ± 18.3		41.8 ± 29.5		29.7 ± 21.7		36.9 ± 21.1	
Bilirubin, mg/dL	Mean ± SD	1.1 ± 0.7		1.1 ± 1.1		1.1 ± 0.5		1.2 ± 1.4		1.1 ± 0.8		1.1 ± 0.6	
Albumin	Mean ± SD	3.8 ± 0.55		3.7 ± 0.55		3.8 ± 0.58		3.7 ± 0.57		3.8 ± 0.52		3.7 ± 0.53	
Prothrombin time	Mean ± SD	1.2 ± 0.17		1.1 ± 0.13		1.2 ± 0.17		1.1 ± 0.11		1.2 ± 0.18		1.2 ± 0.15	
Creatinine	Mean ± SD	0.9 ± 0.31		1.0 ± 0.91		0.9 ± 0.29		0.9 ± 0.31		0.9 ± 0.33		1.2 ± 1.31	
No. of lesions	< 3	74	82.2%	39	43.3%	41	82.0%	20	40.0%	33	82.5%	19	47.5%
	≥ 3	16	17.8%	51	56.7%	9	18.0%	30	60.0%	7	17.5%	21	52.5%
Tumor size, cm	< 3	76	84.4%	64	71.1%	39	78.0%	34	68.0%	37	92.5%	30	75.0%
	≥ 3	14	15.6%	26	28.9%	11	22.0%	16	32.0%	3	7.5%	10	25.0%
BCLC stage	0	30	33.3%	17	18.9%	14	28.0%	9	18.0%	16	40.0%	8	20.0%
	A	43	47.8%	18	20.0%	25	50.0%	9	18.0%	18	45.0%	9	22.5%
	B	21	23.3%	40	44.4%	18	36.0%	24	48.0%	3	7.5%	16	40.0%
	C	5	5.6%	15	16.7%	2	4.0%	8	16.0%	3	7.5%	7	17.5%

Continue.

TNM stage	1	48	53.3%	23	25.6%	26	52.0%	12	24.0%	22	55.0%	11	27.5%
	2	35	38.9%	47	52.2%	20	40.0%	26	52.0%	15	37.5%	21	52.5%
	3	5	5.6%	17	18.9%	4	8.0%	11	22.0%	1	2.5%	6	15.0%
	4	2	2.2%	3	3.3%	0	0.0%	1	2.0%	2	5.0%	2	5.0%
Pre-TACE AFP, ng/mL	<20	57	63.3%	36	40.0%	28	56.0%	21	42.0%	29	72.5%	15	37.5%
	20-200	24	26.7%	38	42.2%	15	30.0%	21	42.0%	9	22.5%	17	42.5%
	>200	9	10.0%	16	17.8%	7	14.0%	8	16.0%	2	5.0%	8	20.0%
Pre-TACE PIVKA-II mAU/mL[†]	<40	65	72.2%	42	46.7%	33	68.8%	21	43.8%	32	80.0%	21	52.5%
	40-200	13	14.4%	22	24.4%	6	12.5%	14	29.2%	7	17.5%	8	20.0%
	>200	10	11.1%	24	26.7%	9	18.8%	13	27.1%	1	2.5%	11	27.5%
mRECIST[‡]	CR	90	100.0%	0	0.0%	50	100.0%	0	0.0%	40	100.0%	0	0.0%
	PR	0	0.0%	17	18.9%	0	0.0%	10	20.0%	0	0.0%	7	17.5%
	SD	0	0.0%	4	4.4%	0	0.0%	2	4.0%	0	0.0%	2	5.0%
	PD	0	0.0%	69	76.7%	0	0.0%	38	76.0%	0	0.0%	31	77.5%

AFP, alpha-fetoprotein; DCP, des gamma carboxy prothrombin; HBV, Hepatitis B virus; HCV, Hepatitis C virus; ECOG, Eastern Cooperative Oncology Group; TNM, tumor-node-metastasis, BCLC, Barcelona Clinic Liver Cancer; ALT, Alanine transaminase; mRECIST, modified response evaluation criteria in solid tumors.

[†] Missing values in training cohort, n=4 (Good responders=2, Poor responders=2)

[‡] Tumor response evaluation after 6 month with TACE

For candidate marker discovery, we adopted most recently established LiverAtlas (76) which included 19,801 genes and 50,265 proteins list related to the liver and various hepatic diseases by incorporating 53 database such as Hepatocellular carcinoma network database (HCC.net), Oncomine, Human Protein Atlas (HPA), and BiomarkerDigger. Of these databases, we selected MCPs for prognostic prediction marker discovery after TACE from pre-screening study. The training set consisted of 100 HCC patients and we collected paired samples before and 6 months after TACE. The validation set comprised 80 patients and we collected pre-TACE samples. Overall scheme of the study is summarized in Figure 1. This study protocol was in accordance with the ethical guidelines of the 1975 Declaration of Helsinki and written informed consent was obtained from each participant or responsible family member after possible complications of invasive procedures had been fully explained. This study procedure was approved by the Institutional Review Board of Seoul National University Hospital.

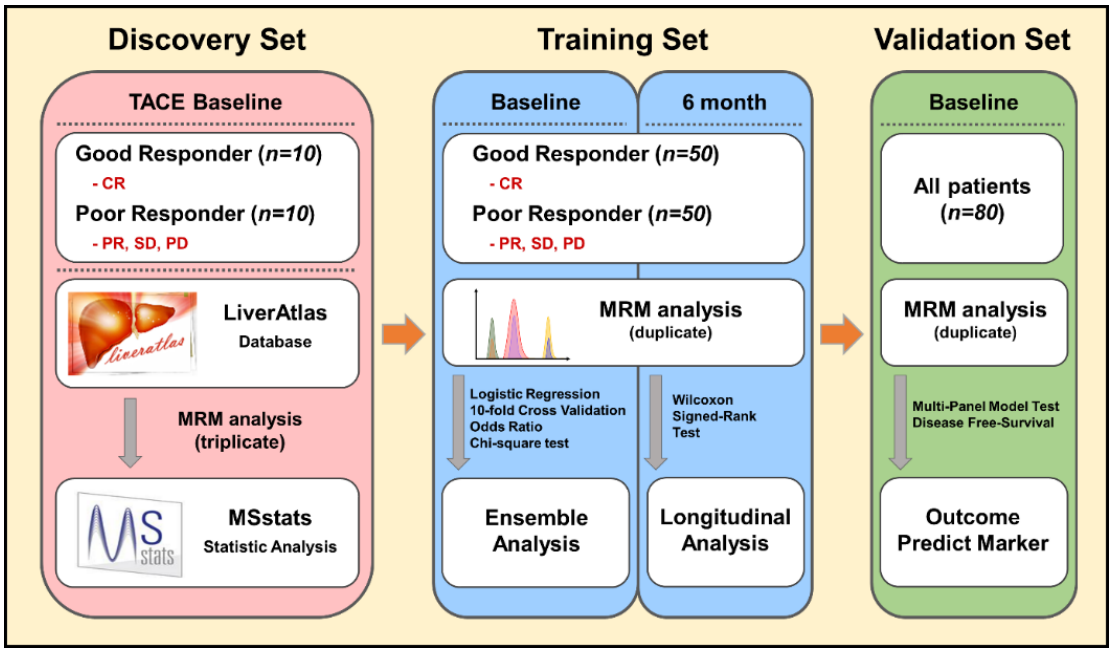


Figure 2-1. Workflow of prognostic prediction marker study

To develop prognostic prediction markers, MRM assay were performed by three strategy. First, marker candidate proteins were selected by LiverAtlas Database with MSstats statistical analysis. Next, the proteins were confirmed and combined as multi-panel model in training set by MRM assays. Finally, the model was validated in validation set.

Treatment modality

TACE was performed according to the Seoul National University Hospital protocol, as described previously. Chemoembolization was performed as selectively as possible via the lobar, segmental, or subsegmental arteries–depending on the tumor distribution and patient's hepatic functional reserve–by using a microcatheter (Microferret [Cook, Bloomington, Ind] or Progreat [Terumo, Tokyo, Japan]). The procedure was initially performed by infusing from 2 to 12 mL of iodized oil (Lipiodol; Andre Gurbet, Aulnay-sous-Bois, France) and from 10 to 60 mg of doxorubicin hydrochloride emulsion (Adriamycin RDF; Ildong Pharmaceutical, Seoul, Korea) until arterial flow stasis was achieved and/or iodized oil appeared in the portal branches. If the initial hepatic arterial blockade was insufficient because of arterioportal shunting or a large sized mass, then embolization was performed with absorbable gelatin sponge particles (1–2 mm in diameter; Gelfoam; Upjohn, Kalamazoo, MI) soaked in a mixture of from 4 to 6 mg of crystalline mitomycin (Mitomycin-C; Kyowa Hakko Kogyo, Tokyo, Japan) and 10 mL of nonionic contrast medium. The extent of chemoembolization was individually adjusted by using a superselective catheterization technique depending on the patient's hepatic functional reserve, similar to that used with surgical hepatectomy (77, 78).

Tumor Response Assessment after TACE

The tumor response evaluation for this study was assessed at CT or MRI by two expert abdominal radiologists by the modified Response Evaluation Criteria in Solid Tumors (mRECIST) for HCC (79). According to mRECIST criteria, complete response (CR) was defined as the complete disappearance of any intratumoral arterial enhancement in all recognizable tumors lesions. Partial response (PR) was defined as a decrease of at least 30 % in the sum of the longest diameter of viable (enhancement in the arterial phase) target lesions, taking as reference the baseline. Progressive disease (PD) was considered as the appearance of new lesions or as an increase of at least 20 % in the sum of the longest diameter of viable (enhancing) lesions, taking as reference the smallest sum of the longest diameters of viable (enhancing) lesions recorded since treatment started. Stable disease (SD) was defined as neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD criteria. Good responders were defined patients who maintained CR state for 6 months after TACE, but poor responders were defined as patients who did not.

Serum Protein Preparation for Selected Reaction Monitoring Assay

Serum depletion was performed using a Multiple Affinity Removal System Human-6 (MARS Hu-6, 4.6 mm x 100 mm, Agilent, CA, USA) affinity column on an HPLC system (Shimadzu, Kyoto, Japan) as described previously (80). Briefly, serum samples were centrifuged at 14,000 x g for 30 min at 4 ℃, and supernatants were transferred to fresh tubes. 40 µL of the supernatants were diluted with 160 µL MARS Buffer A (Agilent, CA, USA). The diluted sample was injected onto a MARS Hu-6 column and unbound fractions were collected into 1.5 mL tubes. Depleted serum was concentrated using 3000-MWCO centrifugal filter units (Amicon Ultra-4 3K, Millipore, MA, USA) and quantified by bicinchoninic acid (BCA) assay.

The depleted serum (0.1 mg) was denatured and reduced with 6 M urea, 20 mM dithiothreitol, 0.1 M Tris, pH 8 at 37 ℃ for 30 min, and alkylated with 50 mM iodoacetamide in the dark at room temperature for 30 min. To avoid trypsin compatibility concentration by urea, the alkylated sample was diluted 10-fold with 0.1 M Tris, pH 8 prior to incubation for 16 h at 37 ℃ with trypsin (Sequencing-grade modified, Promega, WI, USA) in a 1:50 enzyme to substrate ratio. After 16 h incubation, neat formic acid was added to 2% to quench the enzymatic reaction and desalted using Oasis® HLB 1cc (30 mg) extraction cartridges (Waters, MA, USA).

Desalting procedure was followed. Oasis cartridge was washed with 1 mL of 100% MeOH, washed with 3 mL of 100% ACN in 0.1% formic acid, and equilibrated with 3 mL of 0.1% formic acid, sequentially. After total

volume of the digested serum were loaded into the cartridge, the cartridge was washed with 3 mL of 0.1% formic acid, and eluted with 1 mL of 80% ACN in 0.1% formic acid. The eluted sample was lyophilized to vacuum centrifuged and stored at -80℃ until analysis. The sample was resolublized in 0.1% formic acid to 2 μg/μL prior to MRM analysis.

Quantification by multiple reaction monitoring assay

All samples were analyzed on an Agilent 6490 triple quadrupole (QQQ) mass spectrometry (Agilent Technologies, Santa Clara, CA) equipped with 1260 Infinity HPLC system (Agilent Technologies) and using a micro-flow (10 μL/min) gradient of 3 to 35% acetonitrile (ACN)/0.1% formic acid (FA) in 45 min. The analytical column was 150 mm × 0.5 mm id, packed with Agilent Zorbax SB-C18 (3.5-μm particle size), and maintained at 40℃. The MRM assay was conducted in the positive mode with 2500 V of the ion spray capillary and 2000 V of nozzle voltage. The drying gas and sheathe gas temperature was set to 250℃ at 15 L/min and 350℃ at 12 L/min, respectively. Delta EMV was set to 200 V, and the cell accelerator voltage and fragment voltage was 5 V and 380 V, respectively.

Statistical analysis

To analysis the MRM results, all raw files (.d format) were inputted in Skyline software. All transition signals were manually integrated with the Savitzky-Golay smoothing algorithm and exported MSstats format. And

protein significance and relative abundance were analyzed with MSstats package in R. We performed MSstats procedure as described in (81). Briefly, for data preprocessing, all transition intensities were transformed into log2 values. Then, we performed the equalizing of the median peak intensities of reference transitions between the runs. Finally, significant difference and relative abundance of the proteins were calculated by the linear mixed-effects model implemented in MSstats. Receiver operator characteristic (ROC) curves and logistic regression modeling were analyzed by panel composer web statistic tool (82) and MedCalc (Mariakerke, Belgium, ver12.2.1) with relative abundance of each proteins. Also, ROC curve was performed with 10-fold cross validation. Also, Mann-Whitney and Kuskal-Wallis with post-hoc Dunnett-T3 test were used for nonparametric group comparisons. Non-parametric statistics were done using SPSS 22.0 for Windows.

RESULTS

Selection of MCPs from LiverAtlas

From the LiverAtlas database, we focused on proteins that have reliability score more than 4 (27,410 proteins), liver specific proteins (162 proteins), or significant proteins in HCC (1,210 proteins). Out of 27,568 proteins, 948 proteins were reported as secretion proteins in UniProt Knowledgebase (UniProtKB, http://www.uniprot.org/). Then, 572 proteins were filtered with MS/MS spectrum from home-made and National Institute of Standards and Technology (NIST) MS/MS library for empirical evidence of MS detectability. From these proteins, we selected ten proteolytic peptides per protein, and 3,928 peptides were selected to represent the 572 MCPs for sorting detectable proteins in serum samples (Figure 2-2).

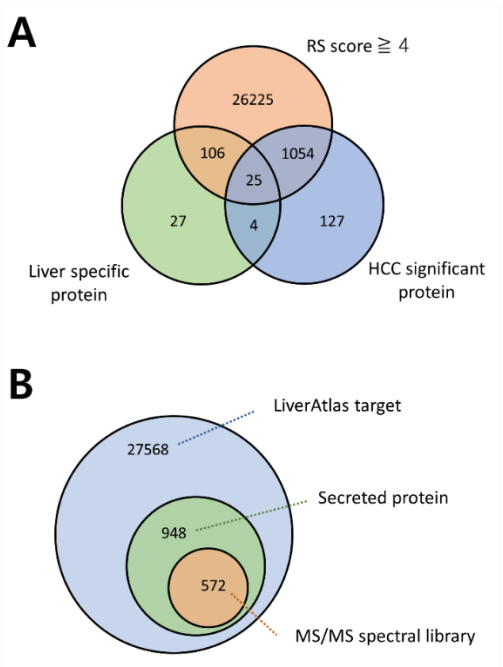


Figure 2-2. List of detectable marker candidate proteins from the LiverAtlas Database

(A) From the LiverAtlas Database, marker candidate proteins were selected by three criteria, RS score ≧ 4, liver specific proteins, and HCC significant proteins. (B) Experimental detectable proteins were selected by secretion DB and MS/MS spectral library.

Detection of MCPs in pooled serum using label-free MRM

Selection of true transition signals in complex samples is challenging due to numerous interfering (false) transition signals (83). To establish detectable proteins, we analyzed 572 MCPs with decoy peptides to pooled serum using label-free MRM. To minimize the number of MS run, we generated 393 decoy peptides (10% of total number of peptides), which is the minimal percentage acceptable for the mProphet tool in skyline software (84), by adding or subtracting a random integer to Q1 and Q3 m/z values. Total 186 MS runs were analyzed, and the results were evaluated by the mProphet tool. The mProphet tool suggests combined score of each peptide by intensity, co-elution count, library intensity dot-product, and peak shape of each peptide. As a result, 1,108 peptides corresponding to 109 proteins were detected with a false discovery rate (FDR) of 0.1%. Also, we manually selected 41 proteins that were missed from the mProphet tool but had high peak intensity and co-eluted transitions. Finally, total 175 peptides from 104 MCPs were synthesized for label MRM assays (Table 2-2).

TABLE 2-2. 104 marker candidate proteins list

N	Uniprot ID	Uniprot Accession	Gene Symbol	Protein name	HCC significant protein	RS score
1	1433S	P31947	SFN	14-3-3 protein sigma	N	4
2	A2AP	P08697	SERPINF2	Alpha-2-antiplasmin	N	5
3	A2GL	P02750	LRG1	Leucine-rich alpha-2-glycoprotein	Yes	5
4	A2MG	P01023	A2M	Alpha-2-macroglobulin	Yes	5
5	AACT	P01011	SERPINA3	Alpha-1-antichymotrypsin	Yes	4
6	ALS	P35858	IGFALS	Insulin-like growth factor-binding protein complex acid labile subunit	N	4
7	AMBP	P02760	AMBP	Protein AMBP	Yes	4
8	ANGT	P01019	AGT	Angiotensinogen	Yes	4
9	ANT3	P01008	SERPINC1	Antithrombin-III	Yes	5
10	APOA1	P02647	APOA1	Apolipoprotein A-I	Yes	5
11	APOA4	P06727	APOA4	Apolipoprotein A-IV	N	4
12	APOC1	P02654	APOC1	Apolipoprotein C-I	Yes	4
13	APOC2	P02655	APOC2	Apolipoprotein C-II	Yes	3
14	APOC3	P02656	APOC3	Apolipoprotein C-III	Yes	4
15	APOC4	P55056	APOC4	Apolipoprotein C-IV	N	4
16	APOE	P02649	APOE	Apolipoprotein E	Yes	5
17	APOF	Q13790	APOF	Apolipoprotein F	N	4
18	APOH	P02749	APOH	Beta-2-glycoprotein 1	Yes	4
19	APOL1	O14791	APOL1	Apolipoprotein L1	N	4
20	BGH3	Q15582	TGFBI	Transforming growth factor-beta-induced protein ig-h3	Yes	4
21	BTD	P43251	BTD	Biotinidase	Yes	4
22	C1QB	P02746	C1QB	Complement C1q subcomponent subunit B	N	4
23	C1QC	P02747	C1QC	Complement C1q subcomponent subunit C	N	4
24	C1RL	Q9NZP8	C1RL	Complement C1r subcomponent-like protein	Yes	3
25	C4BPA	P04003	C4BPA	C4b-binding protein alpha chain	Yes	4
26	C4BPB	P20851	C4BPB	C4b-binding protein beta chain	N	3
27	CATB	P07858	CTSB	Cathepsin B	Yes	4
28	CBPB2	Q96IY4	CPB2	Carboxypeptidase B2	N	5

Continue.

29	CD5L	O43866	CD5L	CD5 antigen-like	N	4
30	CETP	P11597	CETP	Cholesteryl ester transfer protein	Yes	5
31	CFAH	P08603	CFH	Complement factor H	Yes	4
32	CFAI	P05156	CFI	Complement factor I	Yes	4
33	CHLE	P06276	BCHE	Cholinesterase	Yes	4
34	CO2	P06681	C2	Complement C2	Yes	4
35	CO4A	P0C0L4	C4A	Complement C4-A	Yes	3
36	CO5	P01031	C5	Complement C5	N	4
37	CO6	P13671	C6	Complement component C6	Yes	4
38	CO7	P10643	C7	Complement component C7	Yes	4
39	CO8B	P07358	C8B	Complement component C8 beta chain	N	4
40	COL11	Q9BWP8	COLEC11	Collectin-11	N	4
41	CPN2	P22792	CPN2	Carboxypeptidase N subunit 2	N	5
42	CRAC1	Q9NQ79	CRTAC1	Cartilage acidic protein 1	N	4
43	CRP	P02741	CRP	C-reactive protein	Yes	5
44	CXCL7	P02775	PPBP	Platelet basic protein	N	4
45	FA10	P00742	F10	Coagulation factor X	N	5
46	FA11	P03951	F11	Coagulation factor XI	N	4
47	FA12	P00748	F12	Coagulation factor XII	N	4
48	FA9	P00740	F9	Coagulation factor IX	N	5
49	FBLN1	P23142	FBLN1	Fibulin-1	Yes	4
50	FCN3	O75636	FCN3	Ficolin-3	Yes	3
51	FETA	P02771	AFP	Alpha-fetoprotein	Yes	4
52	FETUA	P02765	AHSG	Alpha-2-HS-glycoprotein	Yes	4
53	FETUB	Q9UGM5	FETUB	Fetuin-B	N	4
54	FHR2	P36980	CFHR2	Complement factor H-related protein 2	N	4
55	FHR5	Q9BXR6	CFHR5	Complement factor H-related protein 5	N	4
56	FIBA	P02671	FGA	Fibrinogen alpha chain	Yes	4
57	FIBB	P02675	FGB	Fibrinogen beta chain	Yes	5
58	FIBG	P02679	FGG	Fibrinogen gamma chain	Yes	4
59	FINC	P02751	FN1	Fibronectin	Yes	4
60	HABP2	Q14520	HABP2	Hyaluronan-binding protein 2	N	5

Continue.

61	HEMO	P02790	HPX	Hemopexin	Yes	4
62	HGFA	Q04756	HGFAC	Hepatocyte growth factor activator	N	4
63	HPTR	P00739	HPR	Haptoglobin-related protein	Yes	4
64	IBP2	P18065	IGFBP2	Insulin-like growth factor-binding protein 2	Yes	4
65	IBP3	P17936	IGFBP3	Insulin-like growth factor-binding protein 3	Yes	3
66	IC1	P05155	SERPING1	Plasma protease C1 inhibitor	Yes	5
67	IGF2	P01344	IGF2	Insulin-like growth factor II	N	4
68	IGHG1	P01857	IGHG1	Ig gamma-1 chain C region	Yes	4
69	IGHG3	P01860	IGHG3	Ig gamma-3 chain C region	N	4
70	IGJ	P01591	IGJ	Immunoglobulin J chain	Yes	5
71	IPSP	P05154	SERPINA5	Plasma serine protease inhibitor	N	4
72	ISLR	O14498	ISLR	Immunoglobulin superfamily containing leucine-rich repeat protein	N	4
73	ITIH1	P19827	ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1	Yes	4
74	ITIH2	P19823	ITIH2	Inter-alpha-trypsin inhibitor heavy chain H2	Yes	4
75	ITIH3	Q06033	ITIH3	Inter-alpha-trypsin inhibitor heavy chain H3	N	3
76	ITIH4	Q14624	ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4	Yes	5
77	KAIN	P29622	SERPINA4	Kallistatin	Yes	4
78	KLKB1	P03952	KLKB1	Plasma kallikrein	N	5
79	LBP	P18428	LBP	Lipopolysaccharide-binding protein	N	4
80	LCAT	P04180	LCAT	Phosphatidylcholine-sterol acyltransferase	N	5
81	LG3BP	Q08380	LGALS3BP	Galectin-3-binding protein	Yes	4
82	LUM	P51884	LUM	Lumican	Yes	4
83	MBL2	P11226	MBL2	Mannose-binding protein C	N	5
84	NGAL	P80188	LCN2	Neutrophil gelatinase-associated lipocalin	Yes	4
85	PAPP1	Q13219	PAPPA	Pappalysin-1	N	5
86	PGRP2	Q96PD5	PGLYRP2	N-acetylmuramoyl-L-alanine amidase	N	4
87	PHLD	P80108	GPLD1	Phosphatidylinositol-glycan-specific phospholipase D	N	4
88	PLMN	P00747	PLG	Plasminogen	Yes	4
89	PON1	P27169	PON1	Serum paraoxonase/arylesterase 1	Yes	5
90	POSTN	Q15063	POSTN	Periostin	N	5
91	PROS	P07225	PROS1	Vitamin K-dependent protein S	N	4
92	PROZ	P22891	PROZ	Vitamin K-dependent protein Z	N	4

Continue.

93	PVR	P15151	PVR	Poliovirus receptor	N	5
94	QSOX1	O00391	QSOX1	Sulfhydryl oxidase 1	Yes	4
95	RET4	P02753	RBP4	Retinol-binding protein 4	Yes	5
96	SAMP	P02743	APCS	Serum amyloid P-component	Yes	5
97	SEPP1	P49908	SEPP1	Selenoprotein P	Yes	4
98	SODE	P08294	SOD3	Extracellular superoxide dismutase [Cu-Zn]	N	4
99	THBG	P05543	SERPINA7	Thyroxine-binding globulin	N	3
100	THRB	P00734	F2	Prothrombin	Yes	4
101	VTDB	P02774	GC	Vitamin D-binding protein	Yes	4
102	VTNC	P04004	VTN	Vitronectin	Yes	4
103	ZA2G	P25311	AZGP1	Zinc-alpha-2-glycoprotein	Yes	4
104	ZPI	Q9UK55	SERPINA10	Protein Z-dependent protease inhibitor	N	4

Selection of quantitative MCPs using MRM assay with labeled reference peptides

In MRM assay, measurement level such as limit of detection (LOD) and limit of quantitation (LOQ) is critical point (73). So, we selected quantitative MCPs by 2 steps as following; interference free transition using Automated Detection of Inaccurate and imprecise Transitions (AuDIT) algorithm and assay linearity using calibration curve.

First, in order to minimize interfering transition signals, pooled sample with 175 peptides (endogenous and reference peptide pairs) were analyzed with 5 or 6 transitions per peptide in triplicates. Of these peptides, 161 peptides were passed with more than 3 transitions having no interference signal, respectively (Figure 2-3A). On the contrary, 14 peptides that had less than 2 interference free transitions were excluded in the following step.

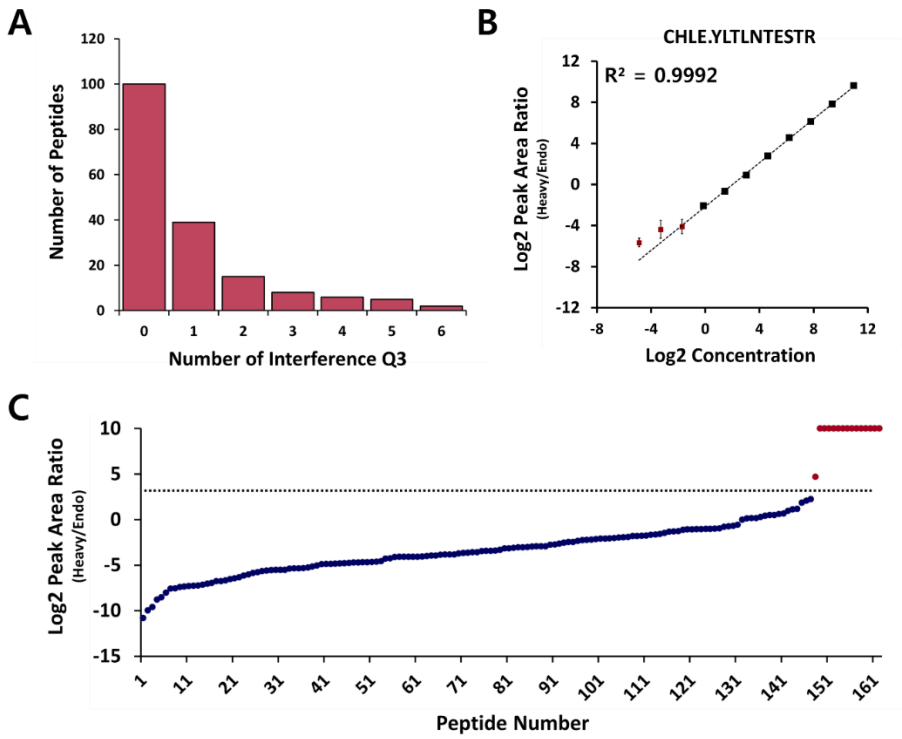


Figure 2-3. Selection of quantitative proteins/peptides by MRM assay

For selection of quantitative proteins/peptides, MRM assays were performed in pooled serum sample. (A) All peptides of the MCPs were considered with interference signal by AuDIT analysis. The peptides that have at least 3 transitions (Q3) were selected as first quantitative peptides. (B) Calibration curves were performed using each reference labeled peptide. Triplicate MRM assays were performed at 11 concentration points of each peptides. For example, calibration curve of "YLTLNTESTR" peptides of "BCHE" protein was showed. (C) Blue dots mean each protein that can be quantitate by MRM assays. Red dots mean each protein that cannot be quantitate by MRM assays.

Next, calibration curves were analyzed with series of diluted labeled reference peptide mixtures (in the range of approximate 0.01 – 2000 fmol/μL) in a pooled sample. For all peptides analysis per injection, we selected and analyzed 2 transitions that had best intensity per peptide from the AuDIT results. And then we performed MRM assay with technical triplicate for each concentration. Finally, the calibration curves were generated by linear regression analysis on the peak area ratio (reference/endogenous) versus spiked reference peptides concentration. Unfortunately, in this study, we used unpurified reference peptides. So, we determined only lower limit of quantitation (LLOQ) as minimum measurement (quantitation) level with linearity R^2 > 0.998 and 0.2 > coefficient of variation (CV). Therefore, we confirmed quantitative peptides with LLOQ less than 10 compared to endogenous peak area ratio in pooled sample. For example, "YLTLNTESTR" peptide of "BCHE" protein can be measured at about 1/5 lower level (0.23 in peak area ratio) compared to endogenous level in pooled sample (Figure 2-3B). According to the rules, we have 147 quantitative peptides from 89 MCPs (Figure 2-3C).

Feasible MCPs selection using pre-screening MRM and western blot

To confirm feasibility of 89 MCPs as prognosis prediction biomarkers after TACE, we performed pre-screening MRM using 10 patients selected blindly in each group with technical triplicate. From the analysis, we obtained the relative abundance of 89 MCPs in each samples. Statistical analysis of the relative abundance was performed using MSstats package in R. Significant differences (Fold change > 1.2 or < 0.83, and adjusted p value < 0.01) between good responders and poor responders were detected in 47 proteins; 24 proteins were highly expressed and 23 proteins were less expressed in poor responders (Table 2-3).

Table 2-3. Differentially expressed proteins from pre-screening MRM assay

N	Uniprot ID	Uniprot Accession	Gene Symbol	Log2 Fold Change	Adjust *p*-value	Standard Error
1	IPSP	P05154	SERPINA5	-1.65	< 0.005	0.019
2	CHLE	P06276	BCHE	-0.80	< 0.005	0.015
3	FCN3	O75636	FCN3	-0.53	< 0.005	0.029
4	FINC	P02751	FN1	-0.50	< 0.005	0.013
5	CPN2	P22792	CPN2	-0.49	< 0.005	0.125
6	APOA4	P06727	APOA4	-0.47	< 0.005	0.015
7	PON1	P27169	PON1	-0.46	< 0.005	0.014
8	IGHG1	P01857	IGHG1	-0.46	< 0.005	0.031
9	LCAT	P04180	LCAT	-0.45	< 0.005	0.022
10	PROZ	P22891	PROZ	-0.43	< 0.005	0.025
11	PGRP2	Q96PD5	PGLYRP2	-0.41	< 0.005	0.020
12	A2AP	P08697	SERPINF2	-0.40	< 0.005	0.016
13	CXCL7	P02775	PPBP	-0.40	< 0.005	0.017
14	KAIN	P29622	SERPINA4	-0.39	< 0.005	0.030
15	IBP3	P17936	IGFBP3	-0.36	< 0.005	0.025
16	APOC3	P02656	APOC3	-0.35	< 0.005	0.018
17	RET4	P02753	RBP4	-0.34	< 0.005	0.007
18	ALS	P35858	IGFALS	-0.33	< 0.005	0.013
19	FETUA	P02765	AHSG	-0.31	< 0.005	0.044
20	C1QB	P02746	C1QB	-0.31	< 0.005	0.029
21	KLKB1	P03952	KLKB1	-0.31	< 0.005	0.018

Continue.

22	APOA1	P02647	APOA1	-0.30	< 0.005	0.013
23	APOF	Q13790	APOF	-0.29	< 0.005	0.023
24	AACT	P01011	SERPINA3	0.28	< 0.005	0.020
25	ITIH4	Q14624	ITIH4	0.28	< 0.005	0.023
26	CO7	P10643	C7	0.32	< 0.005	0.019
27	CO5	P01031	C5	0.33	< 0.005	0.022
28	IC1	P05155	SERPING1	0.34	< 0.005	0.027
29	C4BPB	P20851	C4BPB	0.35	< 0.005	0.029
30	CO2	P06681	C2	0.36	< 0.005	0.015
31	IGJ	P01591	IGJ	0.39	< 0.005	0.023
32	APOE	P02649	APOE	0.40	< 0.005	0.015
33	LG3BP	Q08380	LGALS3BP	0.43	< 0.005	0.019
34	FETA	P02771	AFP	0.47	< 0.005	0.116
35	ITIH3	Q06033	ITIH3	0.49	< 0.005	0.019
36	C4BPA	P04003	C4BPA	0.51	< 0.005	0.019
37	CO4A	P0C0L4	C4A	0.54	< 0.005	0.033
38	SEPP1	P49908	SEPP1	0.63	< 0.005	0.023
39	FHR2	P36980	CFHR2	0.66	< 0.005	0.027
40	A2GL	P02750	LRG1	0.71	< 0.005	0.022
41	SAMP	P02743	APCS	0.77	< 0.005	0.023
42	LBP	P18428	LBP	0.83	< 0.005	0.017
43	FIBA	P02671	FGA	1.14	< 0.005	0.036
44	FIBG	P02679	FGG	1.19	< 0.005	0.041
45	FIBB	P02675	FGB	1.21	< 0.005	0.032
46	THBG	P05543	SERPINA7	1.38	< 0.005	0.138
47	CRP	P02741	CRP	1.96	< 0.005	0.034

For example, Inter-alpha-trypsin inhibitor heavy chain H4 (ITIH4), C-reactive protein (CRP), and AFP were highly expressed in poor responders. Conversely, Plasma serine protease inhibitor (SERPINA5), Cholinesterase (BCHE), and Alpha-2-antiplasmin (SERPINF2) significantly decreased in poor responders (Figure 2-4).

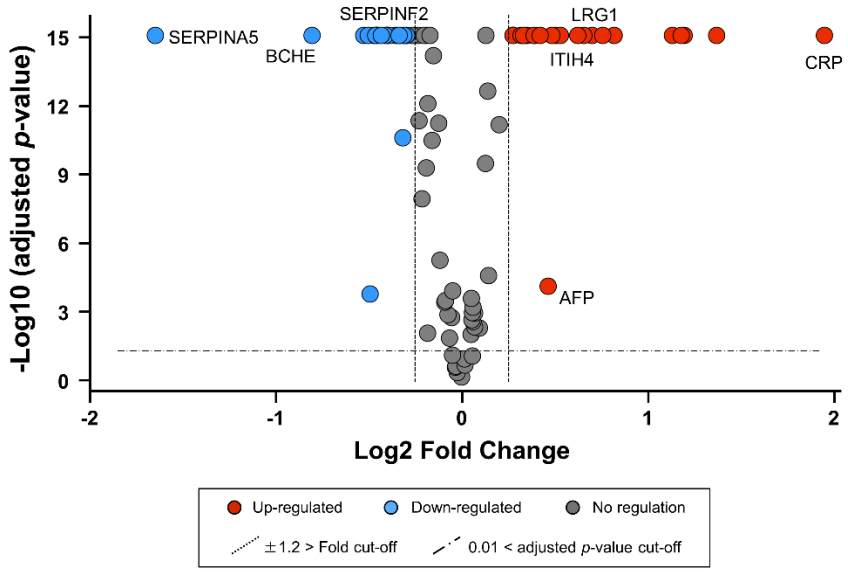


Figure 2-4. Quantification of MCPs by MSstats

Differential expression of MCPs in 20 HCC patients were calculated by MSstats. Log2 fold changes and the corresponding log10 adjusted *p*-values are summarized in a volcano plot. Significant proteins were considered by a fold change > ±1.2 and *p*-value < 0.01 and. Red dots mean up-regulation in poor responders and blue dots mean down-regulation in poor responders. Grey dots mean no regulation in both responders.

To verify our pre-screening MRM results, we performed antibody based western blot assay with 2 randomly selected proteins, ITIH4 and SERPINF2. Total 24 patients, 12 good responders and 12 poor responders, were randomly selected from training set cohorts. To normalize the variability between SDS-PAGE gels, we loaded 6 good responders and 6 poor responders per gel, and pooled sample was loaded on last lane of each gel as internal standard. As a result, ITIH4 protein showed significantly high expression in poor responders group (Figure 2-5A). In contrast, SERPINF2 protein showed low expression pattern in poor responders group (Figure 2-5B). These results were corresponded with pre-screening MRM results despite analysis using independent patients.

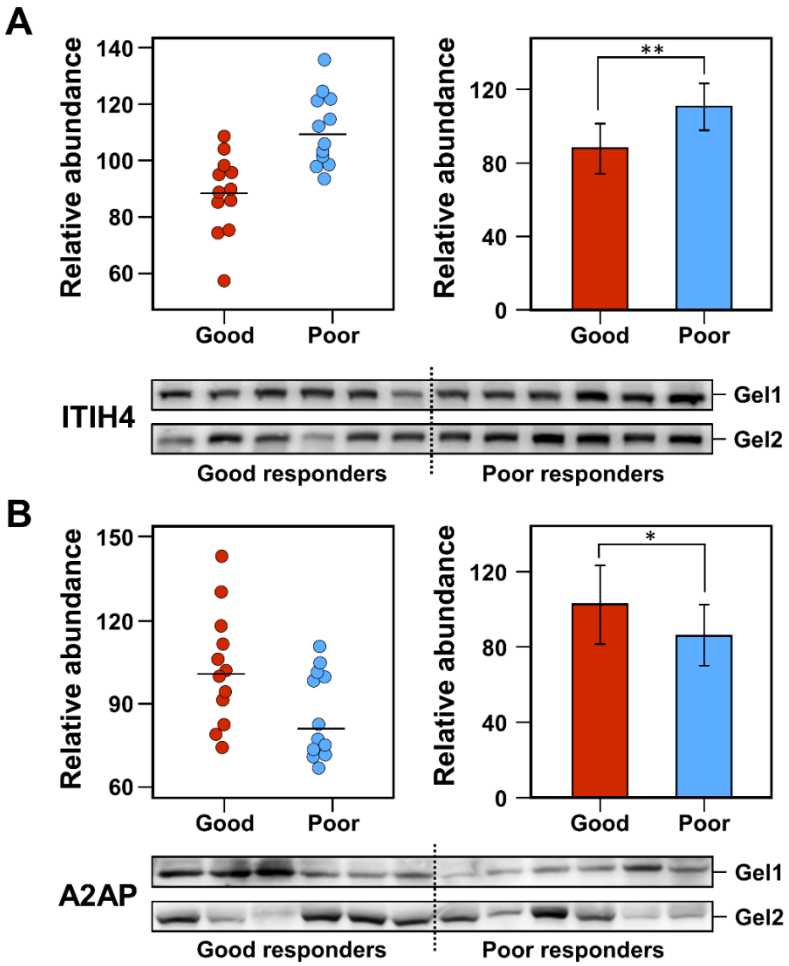


Figure 2-5. Validation by antibody based western blot

Random selected proteins were validated by western blot assay. (A) ITIH4 and (B) A2AP proteins were showed by dot plots and bar graphs. Red and blue dots mean protein abundance by western blots of each patients. Red and blue bar mean average protein abundance. P value was calculated by t-test. (* < *p*-value 0.05, ** < *p*-value 0.01)

The clinicopathologic characteristics to predict outcome in training set

Prior to MRM assay, we evaluated the correlations between outcome after TACE and the clinicopathologic characteristics of the training set of good responders (N=50) and poor responders (N=50) (Table 2-4). In univariate analysis, two clinicopathologic characteristics, number of lesions (OR=6.83, 95% CI=2.73 to 17.09) and concentration of PIVKA-II (OR=2.47, 95% CI=1.10 to 5.55), were significantly associated with outcome within 6 months after TACE. On the contrary, there were no significant association in these clinicopathologic characteristics with regard to albumin, prothrombin time, creatinine, platelet, ALT, bilirubin, and tumor size.

TABLE 2-4. Univariable analysis of clinical variables

Clinical variable	OR[a]	95% CI[b]	*p*-value
Albumin	0.73	0.3635 to 1.4727	0.38
Prothrombin time	0.18	0.0102 to 3.0339	0.22
Creatinine	1.71	0.4220 to 6.9491	0.44
Platelet (10^3/uL)	1.00	0.9907 to 1.0052	0.58
ALT, IU/L	1.01	0.9950 to 1.0303	0.14
Bilirubin, mg/dL	1.11	0.7381 to 1.6766	0.60
No. of lesions	6.83	2.7317 to 17.0935	P < 0.0001
Tumor size, cm	1.67	0.6818 to 4.0828	0.26
Pre-TACE AFP, ng/mL	1.91	0.8618 to 4.2198	0.11
Pre-TACE PIVKA-II mAU/mL	2.47	1.1003 to 5.5472	0.03

a)Odds ratio, estimated form logistic regression model.
b)Confidence interval of estimated OR

Also, we considered discriminant power of AFP and PIVKA-II, which are reported as early detection and prognosis markers, in training set for significant MCPs selection. In classification using ROC curve, AFP was shown an AUC of 0.60 and PIVKA-II was shown an AUC of 0.59 (Figure 2-6). Taken together, we selected MCPs that have more an AUC of 0.60 in the MRM assay.

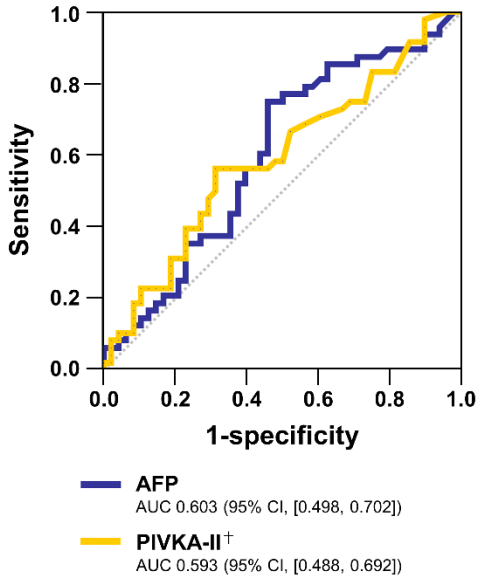


Figure 2-6. ROC curves of the level of AFP and PIVKA-II

Discrimination between good responders and poor responders in training sets. AUC values and 95% confidence interval (CI) were calculated by ROC curves. †Mark means that have 2 missing values in each group, respectively.

Combination for outcome prediction of 47 MCPs in training sets

To assess the prognostic potential of the 47 MCPs, we quantified in the training sets who are good (CR) or poor (PR, SD or PD) responders after TACE using MRM assay with labeled reference peptides. The relative protein abundance from MRM assay were calculated by MSstats linear mixed model with their multiple peptides, multiple transitions and two technical replicates (85).

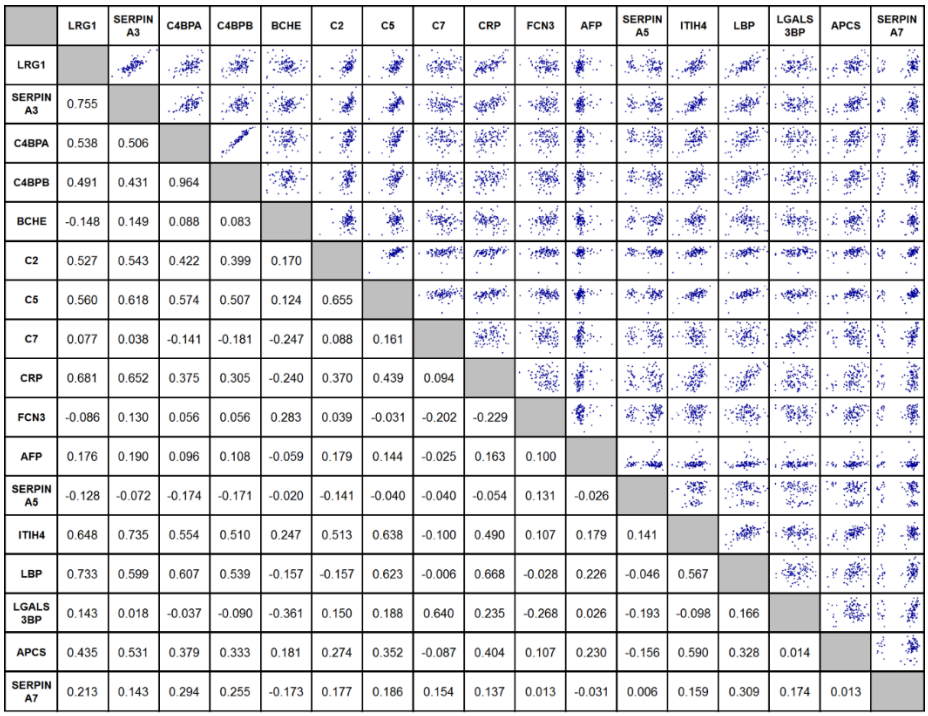
To suggest best-performing single marker, we performed a ROC analysis using the relative protein abundance of 47 MCPs. As a result, best-performing single marker proteins were LRG1 (AUC of 0.708) and C2 (AUC of 0.688). Also, we found that 17 proteins with AUC more than 0.60 were able to effectively discriminate poor responders from total patients with TACE (Table 2-5).

TABLE 2-5. Performance characteristics of the MCPs to predict prognosis after TACE

N	Uniprot ID	Gene Symbol	AUC[a]	95% CI[b]	P-value
1	A2GL	LRG1	0.708	0.702 to 0.713	3.4E-4
2	CO2	C2	0.688	0.682 to 0.693	1.2E-3
3	LBP	LBP	0.685	0.680 to 0.690	1.4E-3
4	C4BPA	C4BPA	0.685	0.680 to 0.690	1.4E-3
5	IPSP	SERPINA5	0.679	0.673 to 0.686	2E-3
6	AACT	SERPINA3	0.677	0.672 to 0.683	2.3E-3
7	CO5	C5	0.677	0.672 to 0.682	2.3E-3
8	C4BPB	C4BPB	0.665	0.660 to 0.670	4.4E-3
9	FCN3	FCN3	0.662	0.657 to 0.666	5.4E-3
10	SAMP	APCS	0.66	0.655 to 0.665	5.8E-3
11	CRP	CRP	0.656	0.652 to 0.660	7.2E-3
12	LG3BP	LGALS3BP	0.648	0.643 to 0.653	0.011
13	THBG	SERPINA7	0.645	0.641 to 0.650	0.012
14	CHLE	BCHE	0.636	0.631 to 0.642	0.019
15	CO7	C7	0.635	0.63 to 0.639	0.02
16	FETA	AFP	0.631	0.625 to 0.636	0.024
17	ITIH4	ITIH4	0.619	0.615 to 0.624	0.04

[a)]area under curve, estimated from ROC curve with 10-fold cross validation,
[b)]95% confidence interval, estimated from ROC curve with 10-fold cross validation

From logistic regression based multivariable analysis, the combination of 5 proteins, LRG1, APCS, BCHE, C7, and FCN3, showed that can discriminate more effective (AUC of 0.825) than single markers. Also, to keep redundancy of marker proteins that have similar abundance trend, we checked correlation coefficient (Figure 2-7). LRG1 were highly correlated ($r > 0.5$) with 7 proteins, SERPINA3, C4BPA, C2, C5, CRP, ITIH4, and LBP. However, our 5 proteins that used for combination panel showed low correlation coefficient, respectively.



	LRG1	SERPIN A3	C4BPA	C4BPB	BCHE	C2	C5	C7	CRP	FCN3	AFP	SERPIN A5	ITIH4	LBP	LGALS 3BP	APCS	SERPIN A7
LRG1																	
SERPIN A3	0.755																
C4BPA	0.538	0.506															
C4BPB	0.491	0.431	0.964														
BCHE	-0.148	0.149	0.088	0.083													
C2	0.527	0.543	0.422	0.399	0.170												
C5	0.560	0.618	0.574	0.507	0.124	0.655											
C7	0.077	0.038	-0.141	-0.181	-0.247	0.088	0.161										
CRP	0.681	0.652	0.375	0.305	-0.240	0.370	0.439	0.094									
FCN3	-0.086	0.130	0.056	0.056	0.283	0.039	-0.031	-0.202	-0.229								
AFP	0.176	0.190	0.096	0.108	-0.059	0.179	0.144	-0.025	0.163	0.100							
SERPIN A5	-0.128	-0.072	-0.174	-0.171	-0.020	-0.141	-0.040	-0.040	-0.054	0.131	-0.026						
ITIH4	0.648	0.735	0.554	0.510	0.247	0.513	0.638	-0.100	0.490	0.107	0.179	0.141					
LBP	0.733	0.599	0.607	0.539	-0.157	-0.157	0.623	-0.006	0.668	-0.028	0.226	-0.046	0.567				
LGALS 3BP	0.143	0.018	-0.037	-0.090	-0.361	0.150	0.188	0.640	0.235	-0.268	0.026	-0.193	-0.098	0.166			
APCS	0.435	0.531	0.379	0.333	0.181	0.274	0.352	-0.087	0.404	0.107	0.230	-0.156	0.590	0.328	0.014		
SERPIN A7	0.213	0.143	0.294	0.255	-0.173	0.177	0.186	0.154	0.137	0.013	-0.031	0.006	0.159	0.309	0.174	0.013	

Figure 2-7. Pearson's correlation coefficients between individual candidate and significant marker candidate proteins in training set

Correlation of coefficients of the proteins that have effective discriminant power (AUC > 0.6) were showed with correlation coefficient r and scatter plots.

Ensemble model analysis with protein markers and clinicopathologic characteristics

The MRM marker panel proteins, LRG1, APCS, BCHE, C7, and FCN3, were combined with the best-performing clinical variable panel, number of lesions, level of AFP, and level of PIVKA-II, using logistic regression modeling. Although level of AFP shown low significance in univariable analysis, we added in the panel because of having appropriate discriminant power regardless of multicollinearity with significant p value. Prior to combine, clinical variable panel were encoded as following; number of lesions = 0 if number <= 2, or 1 if number > 2; level of AFP 0 if level <= 20 ng/mL or 1 if level > 20 ng/mL; level of PIVKA-II = 0 if l <= 40 mAU/mL or 1 if level > 40 mAU/mL. Finally, the ensemble model with the MRM marker panel and clinical variable panel had an AUC of 0.881, whereas the MRM marker panel and clinical variable panel had the AUCs of 0.825 and 0.737. The ROC curves of the ensemble model and other panels are shown in figure 2-8.

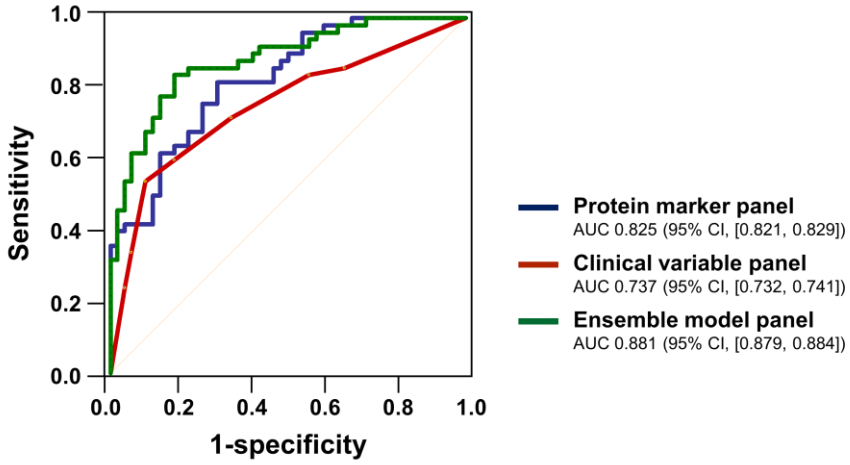
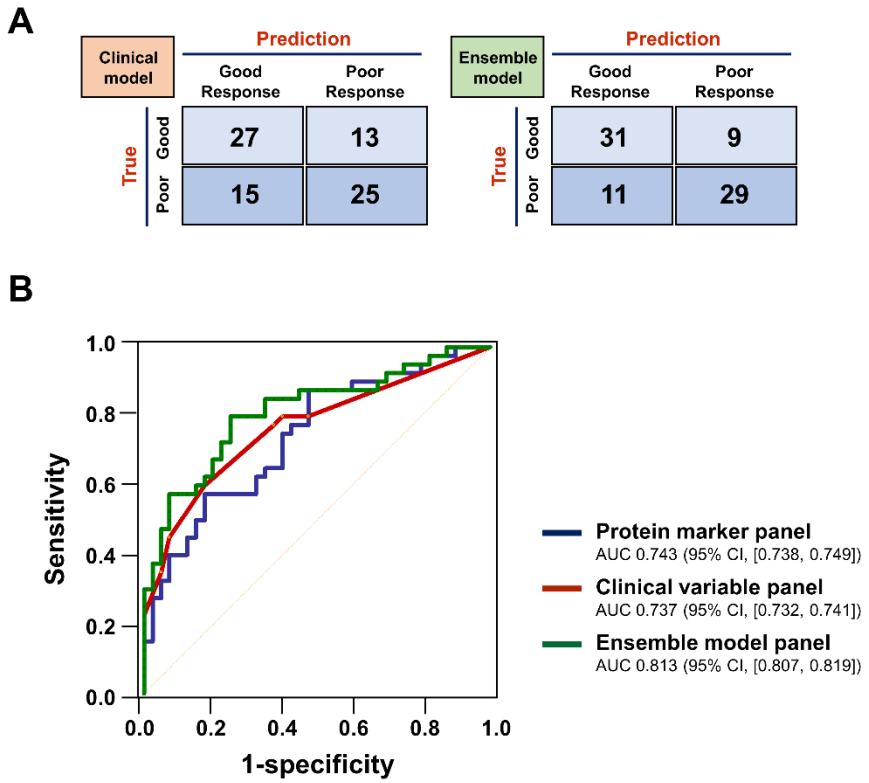


Figure 2-8. Performance characteristic of the best protein marker panel, clinical panel, and ensemble model panel to predict prognosis after TACE

Discrimination between good responders and poor responders in training sets. AUC values and 95% confidence interval (CI) were calculated by logistic regression model.

Model confirmation in validation set

To further evaluate the potential of the ensemble model identified in the training sets, we performed MRM assay in the validation set consisting of 40 good responders and 40 poor responders. From logistic regression modeling, ensemble model panel (3 proteins level and 3 clinical variable) showed that 31 of 40 good responders and 29 of 40 poor responders were correctly classified, whereas clinical model panel in training set showed that 27 of 40 good responders and 25 of 40 poor responders were correctly classified (Figure 2-9A). Also, the ensemble model panel was demonstrated an AUC of 0.813, similar to the training set. The ROC curves of the ensemble model in validation set are shown in figure 2-9B.



2-9. Comparison of the discriminatory power of the best single marker protein with ensemble model panel in validation cohort

(A) For comparisons between the clinical model panel and ensemble model panel, results are presented as confusion matrices. (B) AUC values and 95% confidence interval (CI) that were calculated by logistic regression model are represented with ROC curve.

Prognosis prediction power by ensemble model in TNM stages

To evaluate the prognosis prediction of our ensemble model panel by different Tumor Node Metastasis (TNM) stages (I, II, III, and IV), a total of 180 patients, including good responders (n=90) and poor responders (n=90) were segregated based upon TNM stage. The prediction scores of each patients were calculated from ensemble model equation. In each TNM stage, the prediction scores from ensemble model can significantly enhance the prognostic capability (Figure 2-10). Furthermore, the prediction scores showed no significant difference in good responders group without relevance to TNM stages. However, in poor responders, our prediction scores tended to increase a statistical significance in advanced stage.

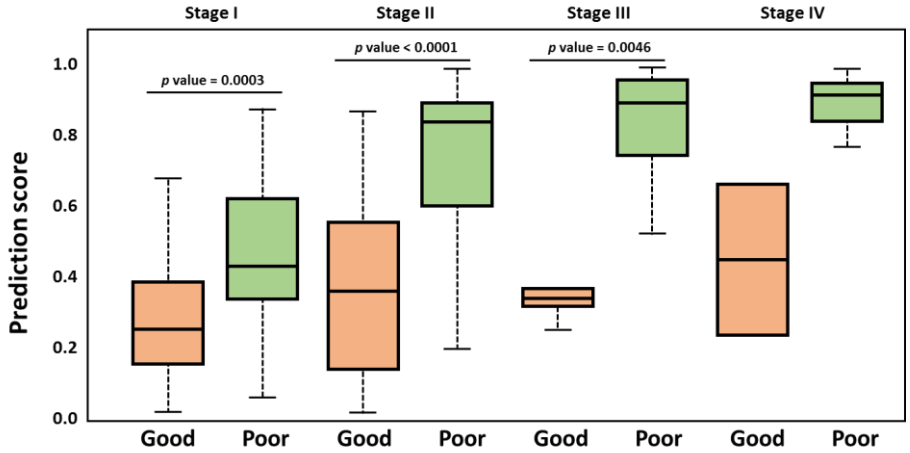


Figure 2-10. Prediction scores by TNM stages in 180 patient samples

Box plots represent prediction scores by logistic regression in 180 HCC patients. Boxes represent the interquartile range, and the horizontal line across each box indicates median values. Statistically significant differences were determined using the Mann-Whitney U test in each TNM stage. Also, statistically significant differences in each groups were determined using Kuskal-Wallis with post hoc Dunnett T3 test.

Longitudinal change in prognostic prediction marker

For observation of progression after TACE, the longitudinal cohort was composed of 100 patients at 2 time point (pre-TACE as baseline and between 6 to 12 months after TACE). We performed MRM assay with 47 MCPs in the longitudinal cohort. From linear mixed model analysis, longitudinal fold changes of the proteins were estimated. Among the proteins, we identified that 7 proteins showed significant longitudinal changes with the other side in each group (Figure 2-11).

As expected, the mean baseline of AFP in good responders was lower than that observed in poor responders group (adjusted p value < 0.001). However, within-person longitudinal change of AFP were no significant despite of a few increase/decrease in each group. We observed that CRP protein showed not only significant difference at baseline between two groups but also decreased longitudinally in good responders group. In addition, the mean baseline of CRAC1 protein were higher in good responders, and the protein was significantly increased in good responders after TACE. Interestingly, the mean baselines of APOF, APOC3, and BCHE in poor responders were lower than the mean baselines of good responders, and these proteins showed longitudinally decreased in poor responders group.

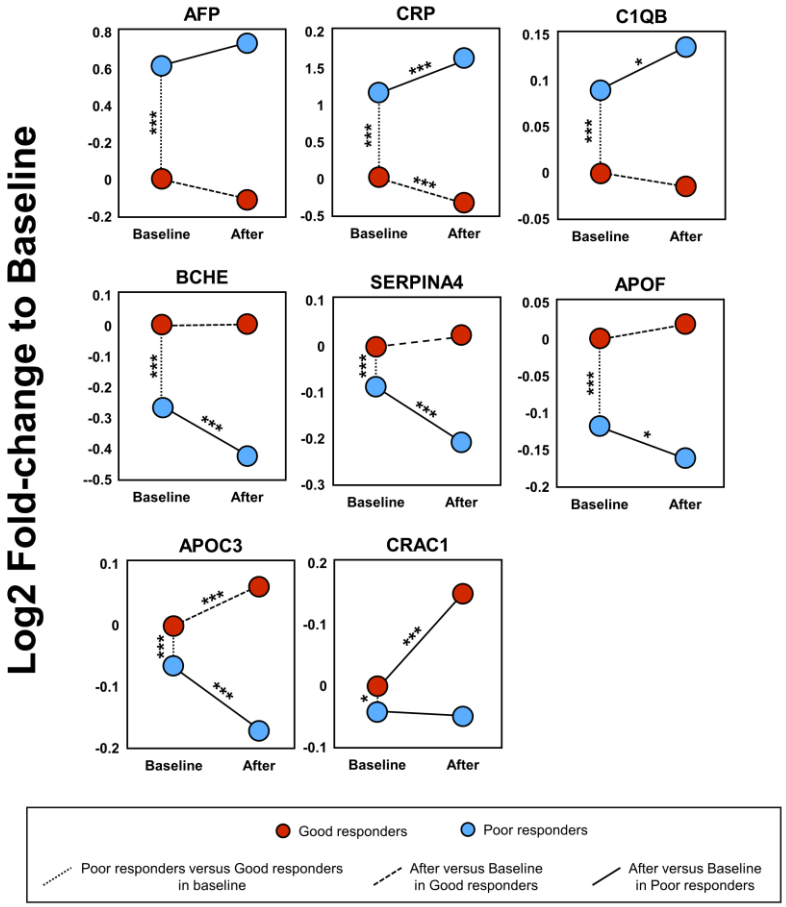


Figure 2-11. Evaluation of longitudinal changes of MCPs in good responders and poor responders

Relative fold change of selected proteins at baseline and after 6 to 12 month of each responder groups. Red and blue dots mean relative average abundance of good responders and poor responders. Linear mixed models by MSstats were used for calculation of significant fold changes. (* < adjusted p-value 0.01, ** < adjusted p-value 0.005, *** < adjusted p-value 0.001)

DISCUSSION

To the patients who cannot applicate curative treatment, such as surgical resection, local ablation, and liver transplantation, TACE may be an effective treatment option for improving survival. However, as TACE is palliative treatment, it needs repeated treatments every 3 to 6 month. Also, there are diverse outcomes after TACE in terms of treatment response and survival. Hence, prediction of outcomes before deciding on a TACE treatment is very important challenge.

In our previous study, we reported that HCC diagnosis markers, filamin-B (FNLB), and anillin (ANLN), were went back towards benign level range after HCC treatment (75). Also, typically over expressed protein in HCC state, CRP, was reported that showed different survival rate after TACE as baseline level of CRP (86). This suggests that HCC related proteins can be used in HCC prognosis prediction marker after treatment. Until a recent date, the most HCC prognosis marker studies were only performed by validation of discrimination power of AFP or PIVKA-II, which are reported diagnosis markers (69, 87). Although there was many marker candidates, it has limitation, because need to highly cost and effort for one by one validation without conviction. For overcoming this limitation, we performed the first study to identify new marker-candidate proteins (MCPs) from about 572 liver related proteins for prognosis/outcome prediction. Of the 572 MCPs, we could detect 89 quantitative proteins in serum using multi step MRM assay

without/with reference labeled peptides. First, 104 proteins were filtered by theoretical or experimental library from common dataset, and we checked detectability by mProphet analysis in pooled serum. Next, quantitative level of detected 104 MCPs were validated by their reference labeled peptides, and 89 MCPs can be measured in quantitative level. In the 89 quantitative proteins, 47 proteins showed significant difference expressions in small cohort set by linear mixed model analysis. Finally, we discovered five proteins marker panel (LRG1, APCS, BCHE, C7, and FCN3) from the training and the validation cohorts, and the panel can discriminate individuals who are versus are not good response after TACE.

In the proteins marker panel, cholinesterase (BCHE) was reported that appears to originate in the liver and is closely associated with the synthesis of serum albumin and coagulation factors (88). BCHE was also reported to reflect liver function in various clinical situations (89). In some liver disease conditions, such as severe chronic hepatitis, cirrhosis, and HCC states, BCHE was found to be very low and associated with increase mortality (90, 91). From out results, HCC patients who have low BCHE level tended to show poor response after TACE. Moreover, in our longitudinal study, we found that BCHE can confirm prognosis by change of up or down regulation. In poor responders group, BCHE level was significantly decreased compared with baseline level, whereas BCHE level of good responders group was not changed.

Importantly, clinicopathologic characteristic variables, level of AFP, level of PIVKA-II, and number of tumor lesions, were also identified as

significantly associated with prognosis in the multivariate analysis. In addition, these factors were reported about favorable performance in previous studies. However, the level of AFP and level of PIVKA-II were shown by an AUC of 0.603 and 0.593, respectively. Thus, we generated ensemble model with the proteins marker panel and clinicopathologic characteristic variables. So, our ensemble model panel be able to discriminate with high performance (an AUC of 0.881 in training cohorts and an AUC of 0.813 in validation cohorts).

In addition, our longitudinal study can support that some markers protein show progression state after treatment. The CRP protein showed longitudinally decrease in good responders after treatment. As mentioned, CRP protein was reported that over-expressed in HCC patients compared with healthy control. Our result showed that level of CRP was decrease in HCC patients who are recovered after treatment. Also, APOC3 and CRAC1 showed that can be used as progression marker after treatment in our longitudinal study. This is meaningful for simply trace of the progression without radiographic images.

Although our results is promising, there are several key limitations that should be acknowledged. As mentioned, we did not perform absolute quantitation assay. So our results can be depend on the instrument platform, sample preparation methods, and purity of reference peptides. In this study, we can suggest only marker panels, but cannot suggest final cut-off range for discriminant. Therefore, it required further absolute study with ELISA or stable isotope dilution MRM (SID-MRM) assay. Also, it required further external large validation with multicenter.

In conclusion, we discovered three new marker proteins that are associated with prognosis prediction after TACE in the first time. Also, we suggested that ensemble model (level of AFP, level of PIVKA-II, number of lesions, LRG1, APCS, BCHE, C7, and FCN3) can predict prognosis before TACE. Indeed, our results require more validation in large cohort and follow up study during long term. However, if validated, it ultimately can help as decision making guideline before TACE in future prospective studies.

REFERENCES

1. Zhang M, Zheng Y. [Analysis on the planning and developing population-based cancer registration in low- and middle-income settings]. Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi. 2014;35(9):1074.

2. Gupta S, Venkatesh A, Ray S, Srivastava S. Challenges and prospects for biomarker research: a current perspective from the developing world. Biochimica et biophysica acta. 2014;1844(5):899-908.

3. Cho WC. Proteomics in translational cancer research: biomarker discovery for clinical applications. Expert review of proteomics. 2014;11(2):131-3.

4. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis. 1998;19(11):1853-61.

5. Mishra A, Verma M. Cancer biomarkers: are we ready for the prime time? Cancers. 2010;2(1):190-208.

6. Kang UB, Ahn Y, Lee JW, Kim YH, Kim J, Yu MH, et al. Differential profiling of breast cancer plasma proteome by isotope-coded affinity tagging method reveals biotinidase as a breast cancer biomarker. BMC cancer. 2010;10:114.

7. Kim Y, Han D, Min H, Jin J, Yi EC, Kim Y. Comparative proteomic profiling of pancreatic ductal adenocarcinoma cell lines. Molecules and cells.

2014;37(12):888-98.

8. Han D, Moon S, Kim Y, Min H, Kim Y. Characterization of the membrane proteome and N-glycoproteome in BV-2 mouse microglia by liquid chromatography-tandem mass spectrometry. BMC genomics. 2014;15:95.

9. Moon S, Han D, Kim Y, Jin J, Ho WK, Kim Y. Interactome analysis of AMP-activated protein kinase (AMPK)-alpha1 and -beta1 in INS-1 pancreatic beta-cells by affinity purification-mass spectrometry. Scientific reports. 2014;4:4376.

10. Parkin DM, Fernandez LM. Use of statistics to assess the global burden of breast cancer. The breast journal. 2006;12 Suppl 1:S70-80.

11. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA: a cancer journal for clinicians. 2010;60(5):277-300.

12. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. Lancet. 2000;355(9202):479-85.

13. Tan F, Jiang Y, Sun N, Chen Z, Lv Y, Shao K, et al. Identification of isocitrate dehydrogenase 1 as a potential diagnostic and prognostic biomarker for non-small cell lung cancer by proteomic analysis. Molecular & cellular proteomics : MCP. 2012;11(2):M111 008821.

14. Tian T, Hao J, Xu A, Luo C, Liu C, Huang L, et al. Determination of metastasis-associated proteins in non-small cell lung cancer by comparative proteomic analysis. Cancer science. 2007;98(8):1265-74.

15. Hwang SJ, Seol HJ, Park YM, Kim KH, Gorospe M, Nam DH, et al. MicroRNA-146a suppresses metastatic activity in brain metastasis. Molecules and cells. 2012;34(3):329-34.

16. Lopez-Otin C, Bond JS. Proteases: multifunctional enzymes in life and disease. The Journal of biological chemistry. 2008;283(45):30433-7.

17. Dawson TM, Dawson VL. Molecular pathways of neurodegeneration in Parkinson's disease. Science. 2003;302(5646):819-22.

18. Opferman JT, Korsmeyer SJ. Apoptosis in the development and maintenance of the immune system. Nature immunology. 2003;4(5):410-5.

19. Rao JS. Molecular mechanisms of glioma invasiveness: the role of proteases. Nature reviews Cancer. 2003;3(7):489-501.

20. Nisman B, Biran H, Heching N, Barak V, Ramu N, Nemirovsky I, et al. 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. British journal of cancer. 2008;98(1):77-9.

21. Kawakami T, Hoshida Y, Kanai F, Tanaka Y, Tateishi K, Ikenoue T, et al. Proteomic analysis of sera from hepatocellular carcinoma patients after radiofrequency ablation treatment. Proteomics. 2005;5(16):4287-95.

22. Streckfus C, Bigler L, Dellinger T, Pfeifer M, Rose A, Thigpen JT. CA 15-3 and c-erbB-2 presence in the saliva of women. Clinical oral investigations. 1999;3(3):138-43.

23. Fanayan S, Smith JT, Lee LY, Yan F, Snyder M, Hancock WS, et al. Proteogenomic Analysis of Human Colon Carcinoma Cell Lines LIM1215, LIM1899 and LIM2405. Journal of proteome research. 2013.

24. Xue H, Lu B, Zhang J, Wu M, Huang Q, Wu Q, et al. Identification of serum biomarkers for colorectal cancer metastasis using a differential

secretome approach. Journal of proteome research. 2010;9(1):545-55.

25. Xie X, Feng S, Vuong H, Liu Y, Goodison S, Lubman DM. A comparative phosphoproteomic analysis of a human tumor metastasis model using a label-free quantitative approach. Electrophoresis. 2010;31(11):1842-52.

26. Wang C, Guo K, Gao D, Kang X, Jiang K, Li Y, et al. Identification of transaldolase as a novel serum biomarker for hepatocellular carcinoma metastasis using xenografted mouse model and clinic samples. Cancer letters. 2011;313(2):154-66.

27. Brown JR, Hartley BS. Location of disulphide bridges by diagonal paper electrophoresis. The disulphide bridges of bovine chymotrypsinogen A. The Biochemical journal. 1966;101(1):214-28.

28. Enoksson M, Li J, Ivancic MM, Timmer JC, Wildfang E, Eroshkin A, et al. Identification of proteolytic cleavage sites by quantitative proteomics. Journal of proteome research. 2007;6(7):2850-8.

29. Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas GR, et al. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nature biotechnology. 2003;21(5):566-9.

30. McDonald L, Beynon RJ. Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. Nature protocols. 2006;1(4):1790-8.

31. Anisowicz A, Huang H, Braunschweiger KI, Liu Z, Giese H, Wang H, et al. A high-throughput and sensitive method to measure global DNA

methylation: application in lung cancer. BMC cancer. 2008;8:222.

32. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nature methods. 2009;6(5):359-62.

33. Han D, Moon S, Kim Y, Ho WK, Kim K, Kang Y, et al. Comprehensive phosphoproteome analysis of INS-1 pancreatic beta-cells using various digestion strategies coupled with liquid chromatography-tandem mass spectrometry. Journal of proteome research. 2012;11(4):2206-23.

34. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. Journal of proteome research. 2006;5(9):2339-47.

35. Kim SJ, Jin J, Kim YJ, Kim Y, Yu HG. Retinal proteome analysis in a mouse model of oxygen-induced retinopathy. Journal of proteome research. 2012;11(11):5186-203.

36. Weijers RN. Amino acid sequence in bovine serum albumin. Clinical chemistry. 1977;23(7):1361-2.

37. Liu Y, Sun W, Zhang K, Zheng H, Ma Y, Lin D, et al. Identification of genes differentially expressed in human primary lung squamous cell carcinoma. Lung Cancer. 2007;56(3):307-17.

38. Domon B, Aebersold R. Mass spectrometry and protein analysis. Science. 2006;312(5771):212-7.

39. Liu H, Sadygov RG, Yates JR, 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Analytical chemistry. 2004;76(14):4193-201.

40. Freund DM, Prenni JE. Improved detection of quantitative differences using a combination of spectral counting and MS/MS total ion current. Journal of proteome research. 2013;12(4):1996-2004.

41. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Molecular & cellular proteomics : MCP. 2005;4(10):1487-502.

42. Ghosh D, Yu H, Tan XF, Lim TK, Zubaidah RM, Tan HT, et al. Identification of key players for colorectal cancer metastasis by iTRAQ quantitative proteomics profiling of isogenic SW480 and SW620 cell lines. Journal of proteome research. 2011;10(10):4373-87.

43. Willis ND, Cox TR, Rahman-Casans SF, Smits K, Przyborski SA, van den Brandt P, et al. Lamin A/C is a risk biomarker in colorectal cancer. PloS one. 2008;3(8):e2988.

44. Constantinescu D, Gray HL, Sammak PJ, Schatten GP, Csoka AB. Lamin A/C expression is a marker of mouse and human embryonic stem cell differentiation. Stem Cells. 2006;24(1):177-85.

45. Srivastava M, Bubendorf L, Nolan L, Glasman M, Leighton X, Miller G, et al. ANX7 as a bio-marker in prostate and breast cancer progression. Disease markers. 2001;17(2):115-20.

46. Pujol JL, Grenier J, Daures JP, Daver A, Pujol H, Michel FB. Serum fragment of cytokeratin subunit 19 measured by CYFRA 21-1 immunoradiometric assay as a marker of lung cancer. Cancer research. 1993;53(1):61-6.

47. Streckfus C, Bigler L, Tucci M, Thigpen JT. 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 investigation. 2000;18(2):101-9.

48. Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. Analytical chemistry. 2003;75(24):6843-52.

49. Prudova A, auf dem Keller U, Butler GS, Overall CM. Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. Molecular & cellular proteomics : MCP. 2010;9(5):894-911.

50. Dugaiczyk A, Law SW, Dennison OE. Nucleotide sequence and the encoded amino acids of human serum albumin mRNA. Proceedings of the National Academy of Sciences of the United States of America. 1982;79(1):71-5.

51. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. Nucleic acids research. 2004;32(Database issue):D115-9.

52. Brannan JM, Sen B, Saigal B, Prudkin L, Behrens C, Solis L, et al. EphA2 in the early pathogenesis and progression of non-small cell lung cancer. Cancer Prev Res (Phila). 2009;2(12):1039-49.

53. Bauer KM, Lambert PA, Hummon AB. Comparative label-free LC-MS/MS analysis of colorectal adenocarcinoma and metastatic cells treated with 5-fluorouracil. Proteomics. 2012;12(12):1928-37.

54. Li F, Glinskii OV, Zhou J, Wilson LS, Barnes S, Anthony DC, et al. Identification and analysis of signaling networks potentially involved in breast carcinoma metastasis to the brain. PloS one. 2011;6(7):e21977.

55. Yoon SY, Kim JM, Oh JH, Jeon YJ, Lee DS, Kim JH, et al. Gene expression profiling of human HBV- and/or HCV-associated hepatocellular carcinoma cells using expressed sequence tags. International journal of oncology. 2006;29(2):315-27.

56. Yoshimura K, Meckel KF, Laird LS, Chia CY, Park JJ, Olino KL, et al. Integrin alpha2 mediates selective metastasis to the liver. Cancer research. 2009;69(18):7320-8.

57. Shibue T, Weinberg RA. Integrin beta1-focal adhesion kinase signaling directs the proliferation of metastatic cancer cells disseminated in the lungs. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(25):10290-5.

58. Wu WS, Wu JR, Hu CT. Signal cross talks for sustained MAPK activation and cell migration: the potential role of reactive oxygen species. Cancer metastasis reviews. 2008;27(2):303-14.

59. Obchoei S, Weakley SM, Wongkham S, Wongkham C, Sawanyawisuth K, Yao Q, et al. Cyclophilin A enhances cell proliferation and tumor growth of liver fluke-associated cholangiocarcinoma. Molecular cancer. 2011;10:102.

60. Pratilas CA, Hanrahan AJ, Halilovic E, Persaud Y, Soh J, Chitale D, et al. Genetic predictors of MEK dependence in non-small cell lung cancer. Cancer research. 2008;68(22):9375-83.

61. Storr SJ, Carragher NO, Frame MC, Parr T, Martin SG. The calpain system and cancer. Nature reviews Cancer. 2011;11(5):364-74.

62. Kim YH, Kwei KA, Girard L, Salari K, Kao J, Pacyna-Gengelbach M, et al. Genomic and functional analysis identifies CRKL as an oncogene amplified in lung cancer. Oncogene. 2010;29(10):1421-30.

63. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA: a cancer journal for clinicians. 2005;55(2):74-108.

64. Bruix J, Sherman M, Practice Guidelines Committee AAftSoLD. Management of hepatocellular carcinoma. Hepatology. 2005;42(5):1208-36.

65. Llovet JM, Real MI, Montana X, Planas R, Coll S, Aponte J, et al. Arterial embolisation or chemoembolisation versus symptomatic treatment in patients with unresectable hepatocellular carcinoma: a randomised controlled trial. Lancet. 2002;359(9319):1734-9.

66. Park JW, Sherman M, Colombo M, Roberts LR, Schwartz ME, Degos F, et al. Observations of hepatocellular carcinoma (HCC) management patterns from the global HCC bridge study: First characterization of the full study population. J Clin Oncol. 2012;30(15).

67. Kim DY, Ryu HJ, Choi JY, Park JY, Lee DY, Kim BK, et al. Radiological response predicts survival following transarterial chemoembolisation in patients with unresectable hepatocellular carcinoma. Alimentary pharmacology & therapeutics. 2012;35(11):1343-50.

68. Georgiades C, Geschwind JF, Harrison N, Hines-Peralta A, Liapi E, Hong K, et al. Lack of response after initial chemoembolization for hepatocellular carcinoma: does it predict failure of subsequent treatment?

Radiology. 2012;265(1):115-23.

69. Wang Y, Chen Y, Ge N, Zhang L, Xie X, Zhang J, et al. Prognostic significance of alpha-fetoprotein status in the outcome of hepatocellular carcinoma after treatment of transarterial chemoembolization. Annals of surgical oncology. 2012;19(11):3540-6.

70. Park H, Park JY. Clinical significance of AFP and PIVKA-II responses for monitoring treatment outcomes and predicting prognosis in patients with hepatocellular carcinoma. BioMed research international. 2013;2013:310427.

71. Kitteringham NR, Jenkins RE, Lane CS, Elliott VL, Park BK. Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2009;877(13):1229-39.

72. Domanski D, Percy AJ, Yang J, Chambers AG, Hill JS, Freue GV, et al. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. Proteomics. 2012;12(8):1222-43.

73. Kennedy JJ, Abbatiello SE, Kim K, Yan P, Whiteaker JR, Lin C, et al. Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. Nature methods. 2014;11(2):149-55.

74. Surinova S, Huttenhain R, Chang CY, Espona L, Vitek O, Aebersold R. Automated selected reaction monitoring data analysis workflow for large-scale targeted proteomic studies. Nature protocols. 2013;8(8):1602-19.

75. Kim H, Kim K, Yu SJ, Jang ES, Yu J, Cho G, et al. Development of biomarkers for screening hepatocellular carcinoma using global data mining

and multiple reaction monitoring. PloS one. 2013;8(5):e63468.

76. Zhang Y, Yang C, Wang S, Chen T, Li M, Wang X, et al. LiverAtlas: a unique integrated knowledge database for systems-level research of liver and hepatic disease. Liver international : official journal of the International Association for the Study of the Liver. 2013;33(8):1239-48.

77. Chung JW, Kim HC, Yoon JH, Lee HS, Jae HJ, Lee W, et al. Transcatheter arterial chemoembolization of hepatocellular carcinoma: prevalence and causative factors of extrahepatic collateral arteries in 479 patients. Korean journal of radiology. 2006;7(4):257-66.

78. Yu SJ, Lee JH, Jang ES, Cho EJ, Kwak MS, Yoon JH, et al. Hepatocellular carcinoma: high hepatitis B viral load and mortality in patients treated with transarterial chemoembolization. Radiology. 2013;267(2):638-47.

79. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Seminars in liver disease. 2010;30(1):52-60.

80. Kim K, Yu J, Min H, Kim H, Kim B, Yu HG, et al. Online monitoring of immunoaffinity-based depletion of high-abundance blood proteins by UV spectrophotometry using enhanced green fluorescence protein and FITC-labeled human serum albumin. Proteome science. 2010;8:62.

81. Choi M, Chang CY, Clough T, Broudy D, Killeen T, MacLean B, et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics. 2014;30(17):2524-6.

82. Jeong SK, Na K, Kim KY, Kim H, Paik YK. PanelComposer: a web-based panel construction tool for multivariate analysis of disease biomarker

candidates. Journal of proteome research. 2012;11(12):6277-81.

83. Abbatiello SE, Mani DR, Keshishian H, Carr SA. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. Clinical chemistry. 2010;56(2):291-305.

84. Reiter L, Rinner O, Picotti P, Huttenhain R, Beck M, Brusniak MY, et al. mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nature methods. 2011;8(5):430-5.

85. Cerciello F, Choi M, Nicastri A, Bausch-Fluck D, Ziegler A, Vitek O, et al. Identification of a seven glycopeptide signature for malignant pleural mesothelioma in human serum by selected reaction monitoring. Clinical proteomics. 2013;10(1):16.

86. Sieghart W, Pinter M, Hucke F, Graziadei I, Schoniger-Hekele M, Muller C, et al. Single determination of C-reactive protein at the time of diagnosis predicts long-term outcome of patients with hepatocellular carcinoma. Hepatology. 2013;57(6):2224-34.

87. Pote N, Cauchy F, Albuquerque M, Voitot H, Belghiti J, Castera L, et al. Performance of PIVKA-II for early hepatocellular carcinoma diagnosis and prediction of microvascular invasion. Journal of hepatology. 2015;62(4):848-54.

88. McQueen MJ. Clinical and analytical considerations in the utilization of cholinesterase measurements. Clinica chimica acta; international journal of clinical chemistry. 1995;237(1-2):91-105.

89. Vorhaus LJ, Scudamore HH, Kark RM. Measurement of serum

cholinesterase activity in the study of diseases of the liver and biliary system. Gastroenterology. 1950;15(2):304-15.

90. Donadon M, Cimino M, Procopio F, Morenghi E, Montorsi M, Torzilli G. Potential role of cholinesterases to predict short-term outcome after hepatic resection for hepatocellular carcinoma. Updates in surgery. 2013;65(1):11-8.

91. Ohashi N, Tsuji N, Naito Y, Iwakura T, Isobe S, Ono M, et al. Relationship between urinary fractional excretion of sodium and life prognosis in liver cirrhosis patients. Hepatology research : the official journal of the Japan Society of Hepatology. 2013;43(11):1156-62.

ABSTRACT IN KOREAN

국 문 초 록

서론: 암은 전세계적으로 가장 많은 사망원인 중 하나이다. 이러한 암에 의한 사망 중 주요 요인은 초기 단계에서의 발견이 어렵기 때문이다. 이러한 암으로부터의 위협에 대처하기 위해 암 발생과정에 대한 이해 및 조기 발견과 치료 효과를 모니터링 하기 위한 방법이 필요 되고 있다. 프로테오믹스 기술이 발전함에 따라 이러한 표지자 단백질 발굴에 많은 도움을 주고 있으며, 최근에는 표지자 발굴뿐만 아니라 암 전이 메커니즘 연구에도 활발히 사용되고 있다.

방법: 1 장에서 전이에 관련된 단백질 변화를 관측하기 위하여 암 전이가 발생한 폐암 세포 (NCI-H1755)를 사용하였다. 이에 대조군으로 폐암 세포이며 전이가 발생하지 않은 세포 (NCI-H1703)를 사용하였다. 두 세포주의 단백질 발현 량 비교를 위하여 label-free 정량 분석을 시행하였다. 또한 세포 내에 비정상적으로 잘려진 단백질 파편을 찾기 위하여 N 말단 분석기법을 개발하였다. 2 장에서는 치료예후마커 발굴을 위한 데이터베이스 기반 마커 후보군을 선정하였다. 이를 기반으로 다중검지법을 적용하여 180 명의 간암환자에 대하여 마커후보군에 대한 정량분석을 시행하였다.

결과: 1 장에서는 질량분석기를 사용하여 총 2130 개의 단백질을 발견하였으며, 그 중에서 1355 개 단백질이 두 종류 세포에서 공통적으로 발견되었다. Label-free 정량 분석 기법에 의해 242 개의 단백질이 두 세포에서 유의적인 차이를 보이며 발현되는 것을 확인하였다. 또한 N 말단 분석기법을 통하여 325 개의 단백질 파편을 발견했으며, 45 개의 알려지지 않은 단백질 파편을 발견할 수 있었다. 위의 두 가지 실험 기법을 바탕으로 11 개의 정량 분석된 단백질과 8 개의 단백질 파편이 focal adhesion pathway 에 직접적으로 관련이 있음을 발견하였다. 2 장에서는 화학색전술을 받은 20 명의 간암환자에 대하여 47 개 단백질이 치료 예후 (6 개월동안 병소가 없는 상태가 유지된 그룹 또는 그렇지 못한 그룹)에 따라 유의적으로 차이를 보인 것을 확인하였다. 이를 기반으로 190 명의 환자에 적용하여 정량분석을 시행하였으며, 최종적으로 17 개의 단백질이 치료예후를 구분하는데 사용 가능함을 확인하였다. 이 중에서 5 개의 단백질 (LRG1, APCS, BCHE, C7, FCN3)과 3 개의 임상 정보 (AFP 수치, PIVKA-II 수치, 간암 병소 개수)를 조합한 다중 마커패널이 AUC 0.8 이상으로 구분력이 있음을 확인하였다.

결론: 1 장에서는 Label-free 정량 기법 및 N-말단 분석기법의 개발을 통하여 폐암 전이에 focal adhesion pathway 관련 단백질의 발현차이가 전이에 직접 또는 간접적으로 영향을 줄 수 있다는 것을 확인하였다. 이러한 기존의 프로테옴 분석뿐만 아니라 새로운 개념의 분석 방법을

사용한 단백질의 발현 정량 분석 및 단백질의 파편조각의 발견은 암 메커니즘 이해에 많은 도움을 줄 것으로 생각된다. 2 장에서는 환자 맞춤 치료 방법의 적용을 위한 다중마커패널을 개발하였다. 우리의 다중마커패널은 간암환자의 치료방법 선택에 있어서 좀 더 효과적으로 접근할 수 있는 가이드라인이 될 것이다. 따라서 이러한 프로테오믹스 연구 기법들은 암의 이해 및 치료 등에 사용 될 수 있을 것이다.

주요어 : 폐암, N 말단 분석, 암 전이, 프로테오믹스, 정량 분석, 다중검지법, 간암, 화학색전술, 치료예후마커

학 번 : 2008-21997

*본 내용은 Molecules and Cells 학술지에 출판 완료된 내용임

이학박사 학위논문

Cancer Biomarker Discovery Using N-terminal Peptides and Multiple Reaction Monitoring-MS Techniques

N-말단단백체 및 다중반응검지 질량분석기술을 이용한 암 표지자 개발 연구

2015 년 08 월

서울대학교 대학원

의과학과 의과학전공

민 호 필

A thesis of the Degree of Doctor of Philosophy

N-말단단백체 및 다중반응검지 질량분석기술을 이용한 암 표지자 개발 연구

Cancer Biomarker Discovery Using N-terminal Peptides and Multiple Reaction Monitoring-MS Techniques

August 2015

Major in Biomedical Sciences

Department of Biomedical Sciences

Seoul National University

Graduate School

Hophil Min

N-말단단백체 및 다중반응검지 질량분석기술을 이용한 암 표지자 개발 연구

지도교수 김 영 수

이 논문을 이학박사 학위논문으로 제출함

2015 년 04 월

서울대학교 대학원

의과학과 의과학전공

민 호 필

민호필의 이학박사 학위논문을 인준함

2015 년 07 월

위 원 장 정 구 홍 (인)

부위원장 김 영 수 (인)

위 원 박 태 성 (인)

위 원 김 윤 준 (인)

위 원 장 수 환 (인)

Cancer Biomarker Discovery Using N-terminal Peptides and Multiple Reaction Monitoring-MS Techniques

by

Hophil Min

A thesis submitted to the Department of Biomedical Sciences in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Biomedical Sciences at Seoul National University Graduate School

July 2015

Approved by Thesis Committee:

Professor _____Chairman

Professor _____Vice chairman

Professor _____

Professor _____

Professor _____

ABSTRACT

Cancer Biomarker Discovery Using N-terminal Peptides and Multiple Reaction Monitoring-MS Techniques

Hophil Min

Major in Biomedical Sciences

Department of Biomedical Sciences

Seoul National University

Graduate School

Introduction: Cancer is the leading cause of death in the worldwide, and the major cause of cancer death is the difficulty for early diagnosis. To overcome this problem, the discovery of cancer biomarkers is useful for early diagnosis, outcome monitoring, or predicting recurrence. For biomarker discovery, proteomics technique is powerful tools with high-throughput and high sensitivity. Thus, proteomics study can help variable cancer biomarker discovery and understand of cancer mechanisms in body.

Methods: In chapter I, to examine metastatic events 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). In chapter II, we performed to identify new marker-candidate proteins from LiverAtlas database. And abundance of marker-candidate proteins were quantified in individual patients by multiple reaction monitoring assay.

Results: In chapter I, 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. In chapter II, we selected quantitative 104 marker candidate proteins with reference labeled peptides. Among them, we found that 17 proteins with AUC more than 0.60 were able to effectively discriminate poor responders from total patients underwent TACE. Also, we discovered powerful ensemble model panel with protein markers and clinical variables.

Conclusions: In chapter I, our datasets of proteins and fragment peptides in lung cells might be valuable in discovering and validating lung cancer biomarkers and metastasis markers. This study increases our understanding of the NSCLC metastasis proteome. In chapter II, we discovered three new marker proteins that are associated with prognosis prediction after TACE in the first time. Our study can help to identify useful biomarkers for prediction of prognosis with multi-panel modeling.

Keywords: Non-small-cell lung cancer; label-free quantitative analysis; N-terminal analysis; Metastasis; Multiple Reaction Monitoring; transcatheter arterial chemoembolization; Hepatocellular carcinoma; Prognostic factor; Proteomics; Biomarker

Student Number: 2008-21997

*This work is published in Molecules and Cells journal. (H Min, D Han, Y Kim, J Cho, J Jin, and Y Kim. Label-Free Quantitative Proteomics and N-terminal Analysis of Human Metastatic Lung Cancer Cells. Mol. Cells 2014; 37(6): 457~466)

CONTENTS

LIST OF TABLES

Chapter I

Chapter II

LIST OF FIGURES

Chapter I

Chapter II

LIST OF ABBREVIATIONS

MS, Mass spectrometry

PTM, Post-translational modification

NSCLC, Non-small-cell lung cancer

CYFRA 21-1, serum cytokeratin 19 fragments

FASP, Filter-aided sample preparation

PMSF, Phenylmethanesulfonyl fluoride

SDS, Sodium dodecyl sulfate

TFA, Trifluoroacetic acid

CHCA, α-cyano-4-hydroxycinnamic acid

STD, StageTip desalting

BSA, Bovine serum albumin

TPP, Trans-Proteomics Pipeline

NSAF, Normalized spectral abundance factor

UniProtKB, Universal Protein Resource Knowledgebase

COFRADIC, Combined fractional diagonal chromatography

CRKL, v-crk sarcoma virus CT10 oncogene homolog (avian)-like

MRM, Multiple Reaction Monitoring

HCC, Hepatocellular carcinoma

AASLD, American Association for the Study of Liver

EASL, European Association for the Study of the Liver

AFP, Alpha-fetoprotein

PIVKA-II, Prothrombin induced by the absence of vitamin K or antagonist-II

MCPs, Marker candidate proteins

mRECIST, modified Response Evaluation Criteria in Solid Tumors

TACE, Transarterial chemoembolization

BCHE, Cholinesterase

ITIH4, Inter-alpha-trypsin inhibitor heavy chain H4

SERPINF2, Alpha-2-antiplasmin

C7, Complement component C7

ROC, Receiver operating characteristic

GENERAL INTRODUCTION

Cancer is the leading cause of death in the worldwide, and developed countries take the brunt of the disease with approximately 70% of deaths (1). The major cause of cancer death is the difficulty for early diagnosis and suitable treatment without major medical devices. To overcome this problem, the discovery of cancer biomarkers is useful for early diagnosis, monitoring how well a treatment is performed, or predicting recurrence (2, 3). Thus, the utility and importance of biomarkers are growing in both academic and industrial fields (4).

Most biomarkers are molecules that are secreted by tumor organ or specific responses to the presence of cancer (5). To detect the biomarkers, proteomics technique is one of the applicable tools with high-throughput and high sensitivity (6). As the development of mass spectrometry (MS), proteomics technique could analyze the relative protein abundance, occurrence site of post translational modifications (PTMs), protein-protein interactions, and cellular functions (7-9). So, knowledge of the proteome can be useful for cancer biomarker discovery.

In chapter I, to know cellular proteome changes, we performed label-free analysis and N-terminal peptides enrichments in non-small-cell lung cancer (NSCLC) cell lines. NSCLS is usually treated with surgery, but surgery is effective only in patients who are diagnosed at an early stage. Unfortunately, more than 70% of NSCLC patients are diagnosed at the late

stage with metastasis, resulting in a loss of opportunity for effective surgery. Thus, application of our new technique for biomarker discovery can help to develop novel and more effective molecular markers and therapeutic targets.

In chapter II, for application of quantitative proteomics tools in biomarker discovery, we performed multiple reaction monitoring (MRM) assay in serum of hepatocellular carcinoma patients underwent transarterial chemoembolization (TACE). TACE is an effective treatment option for reducing systemic toxicity, increasing local antitumor effects, and improving survival for late stage patients. But, unpredictable outcomes often occur after TACE in terms of treatment response and survival. So, in this study, we identified useful biomarkers for prediction of prognosis with multi-panel modeling.

CHAPTER I

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

INTRODUCTION

Lung cancer is the leading cause of cancer-related deaths worldwide (30%) but constitutes only 15% of new cancer diagnoses (10). 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 (11). Lung cancer is divided into 2 major histological types: small-cell lung cancer (SCLC) and non-small-cell lung cancer (NSCLC) (12). 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 (13).

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 (14, 15). During metastasis, irreversible protein fragmentation occurs (16). Dysregulation of protein fragment reactions in organs can cause pathological developmental disorders, such as cancer, inflammation, infection, and Alzheimer disease (17-19).

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 (20). Pro1708/Pro2044 (the C-terminal fragment of albumin) (21) and HER2 rb2 (the ectodomain of human epithelial growth factor receptor-2) (22) 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 (23). Many global quantitative proteomics studies have examined metastasis in various cancers, such as colorectal cancer (24), breast cancer (25), and hepatocellular carcinoma (26). 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 (14).

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 (27). To this end, several groups have established a method of identifying protease-generated (neo) peptides in cellular pathways, known as N-terminomics (28). 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 (29). 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 et al. developed a more rapid and simpler N-terminal peptide analysis method (positional proteomics) that is based on negative selection by chemical labeling of the α-amine in proteins (30).

In this study, to differentiate primary cancer cells from metastatic cells, we performed 2 parallel experiments: label-free quantification and novel 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 (31). 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 (Waltham, MA). Hydrochloric acid (HCl) and sodium chloride (NaCl) were purchased from DUKSAN (Gyungkido, Korea). Urea and dithiothreitol (DTT) were purchased from AMRESCO (Solon, OH). Phenylmethanesulfonyl fluoride (PMSF), sodium dodecyl sulfate (SDS), and Tris were obtained from USB (Cleveland, OH). Complete protease inhibitor cocktail tablets were acquired from ROCHE (Indianapolis, IN), and sequencing-grade modified trypsin was purchased from PROMEGA (Madison, WI). Sulfo-NHS acetate and NHS-Activated agarose slurry were obtained from Pierce (Rockford, IL). All other reagents—iodoacetamide, α-cyano-4-hydroxycinnamic acid (CHCA), and trifluoroacetic acid (TFA)—were purchased from Sigma-Aldrich (St. Louis, MO).

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, Daegu, Korea) with 10% fetal bovine serum (Gibco, Grand Island, NY), 100 U/mL penicillin and 100 μg/mL streptomycin (Gibco, Grand Island, NY) and 25 mM HEPES (Gibco, Grand Island, NY). The cultures were maintained in 95% humidified air and 5% CO_2 at 37℃.

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, 1x protease inhibitor cocktail, 0.1 M DTT, and 0.1 M HEPES. Lysates were incubated at 95℃ for 5 min and sonicated for 1 min. Supernatants were collected from the lysates by centrifugation at 15,000 × g for 20 min at 4℃. Protein concentrations were measured using the BCA Protein Assay Kit – reducing reagent-compatible (Pierce, Rockford, IL). Finally, each cell lysate was stored in 0.2-mg aliquot at -80℃ until use.

Filter-Aided Sample Preparation (FASP)

Cell lysates were processed by filter-aided sample preparation (FASP) (32) using a 10K molecular weight cutoff (MWCO) filter (Millipore, Pittsburgh, PA). Briefly, 200 μg of cell lysates in lysis buffer (4% SDS, 0.1 mM PMSF, 1x 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 × g at 20℃ 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 darkn 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 × g at 20℃ 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℃. 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 × g at 20℃ 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 × g at 20℃ 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℃. Digested peptides were collected by centrifugation and acidified with 2% TFA.

Desalting of peptides

Digested samples were desalted using in-house C_{18} StageTip desalting (STD) columns, as described (33). Briefly, in-house C_{18} STD columns were prepared by reversed-phase packing of POROS 20 R2 material into 0.2-mL yellow pipet tips that sat atop C_8 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, Rockford, IL). One milliliter of an NHS-agarose bead slurry (50% slurry in acetone) was prepared per the manufacturer's protocol (Pierce Rockford, IL). 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 hours at RT. Finally, the beads were centrifuged at 1000 × *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, Solon, OH) were N-terminally labeled as described above as control. The peptides were dissolved in 10 μl 0.1% TFA, and 0.5 μl of each sample was mixed with 0.5 l of a matrix solution that contained 5 mg/ml CHCA (Sigma, St. Louis, MO), 70% ACN, and 0.1% TFA. The peptides were spotted directly onto a MALDI plate (Opti-TOF™ 384-well Insert, Applied Biosystems, Foster City, CA) 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 x e^{-07} 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), adrenocorticotropic 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, Waltham, MA) that was coupled to an EasyLC II (Proxeon Biosystems, Denmark), equipped with a nanoelectrospray device and fitted with a 10-μm fused silica emitter tip (New Objective, Woburn, MA). Ten microliters of each samples was loaded onto a nano-LC trap column (ZORBAX 300SB-C_{18}, 5 μm, 0.3 × 5 mm, Agilent, Wilmington, DE), and peptides were separated on a C_{18} analytical column (75 μm × 15 cm) that was packed in-house with C_{18} resin (Magic C18-AQ 200 Å, 5-μm 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℃. 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 ±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, Milpitas, CA). Raw files from the LTQ-Orbitrap Velos were converted into mzXML files using Trans-Proteomics Pipeline (TPP, ISB, Seattle, WA). 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., Portland, OR).

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 (34). NSAF values were calculated as:

$$NSAF = (SpC/Mw)/\Sigma(SpC/Mw)n$$

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 Figure 1-1. 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_{18} 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_{18} in-house stage tips and filtered by NHS-activated beads that depleted the newly generated N-termini by trypsin. The supernatants of the N-terminal peptide samples were desalted using C_{18} 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.

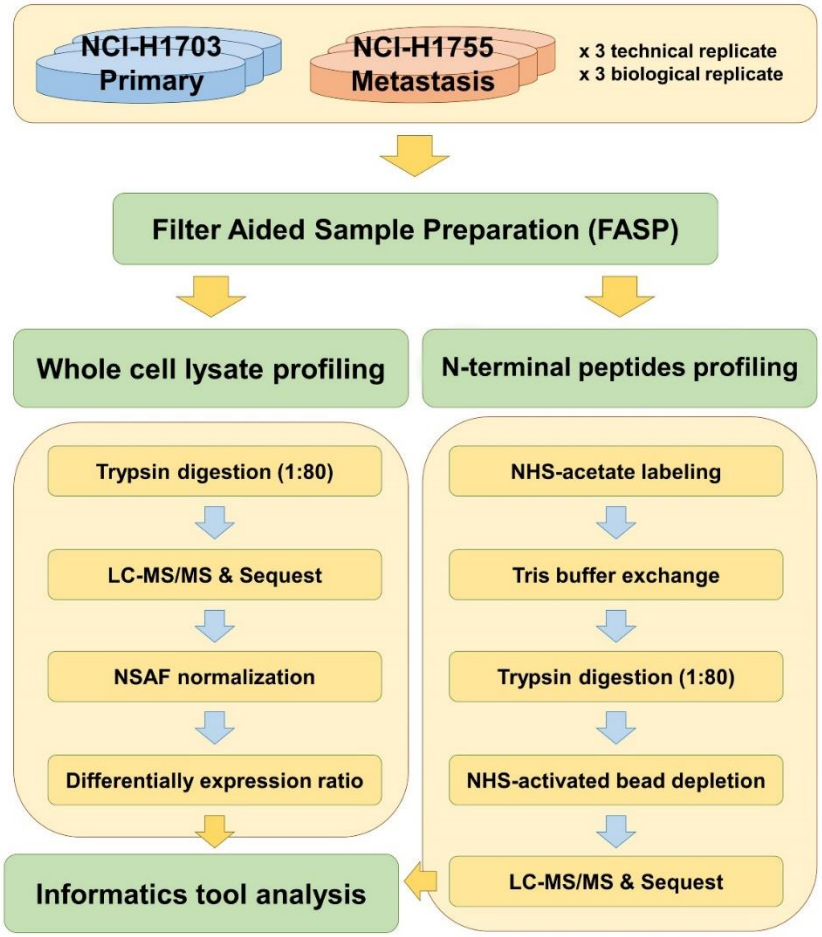


Figure 1-1. Overall scheme

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.

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 (Figure 1-2A), 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 (Figure 1-2B).

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) (Figure 1-2C). Notably, molecular functions were assigned many proteins: 493 proteins were annotate 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" (Figure 1-2D).

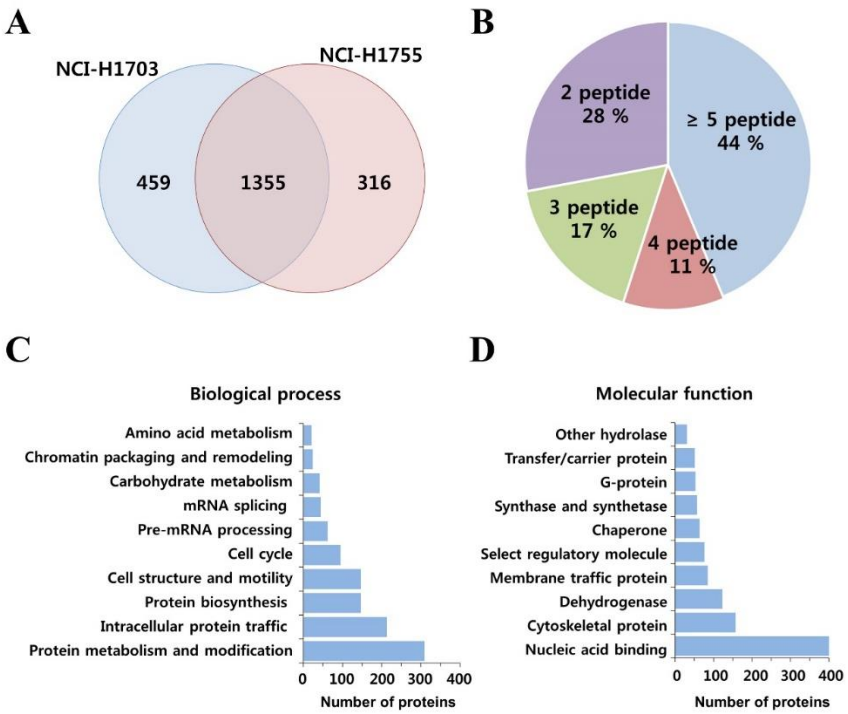


Figure 1-2. Identification and proteome analysis of two different cell lines

(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.

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 corresponding protein over a wide linear dynamic range (34). High-confidence proteins for label-free quantitation were selected with an average spectral count ≥ 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.

The distribution of the ratio correlation between NCI-H1703 and NCI-H1755 in the 3 biological replicates was selectively plotted, as shown in Figure 1-3A, 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 (Figure 1-3B) (35). The differential expression ratios for the 671 protein groups are shown in Figure 1-3C, in which ratios ≥ 1.5-fold are shadowed. The expression of 242 proteins changed ≥ 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 1-1).

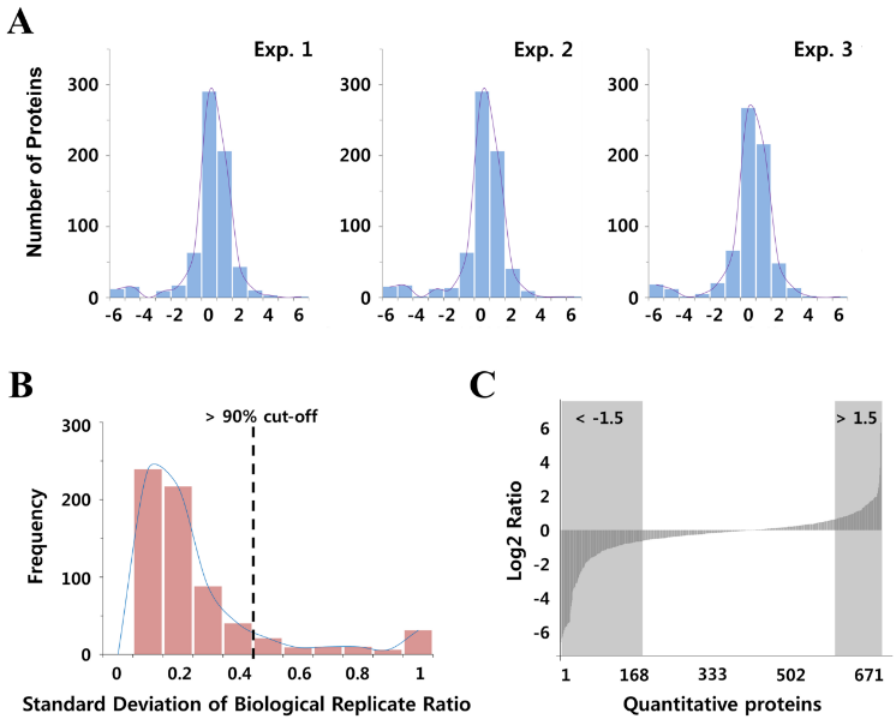


Figure 1-3. Distribution of log2 NSAF ratios and differentially expressed proteome

(A) The distributions of log2 NSAF ratios for primary cancer cells versus metastatic cancer cells were obtained by comparing 3 biological replicates from the label-free quantification experiments. (B) Fold-change cutoff of protein expression was considered the standard deviation of the 3 replicates. Ninety percent of all identified proteins were within less than 0.5 standard deviations. (C) Protein ratios are arranged in ascending order, resulting in a sigmoidal curve. The light shaded area represents unregulated protein groups with a less than 1.5-fold change in expression, and the dark shaded area represents protein groups that undergo more than a 1.5-fold change.

Table 1-1. *Up*- and *down*- regulated proteins

#	IPI	MW (kDa)	Expression ratio[a] Log_2(NCI-H1755/NCI-H1703)	*p* value[b]	Gene Symbol	Protein name
1	IPI00001453	55.3923	-7.47	0.00549	INA	Alpha-internexin
2	IPI00397526	230.7853	-6.87	0.00389	MYH10	Isoform 1 of Myosin-10
3	IPI00607787	58.6824	-6.64	0.00410	UAP1	Isoform 3 of UDP-N-acetylhexosamine pyrophosphorylase
4	IPI00856045	616.6283	-6.59	0.00284	AHNAK2	Isoform 1 of Protein AHNAK2
5	IPI00333619	54.8498	-6.52	0.00265	ALDH3A2	Isoform 1 of Fatty aldehyde dehydrogenase
6	IPI00178150	139.8838	-6.36	0.00830	KIF4A	Isoform 1 of Chromosome-associated kinesin KIF4A
7	IPI00237884	180.9821	-6.26	0.04158	AKAP12	Isoform 1 of A-kinase anchor protein 12
8	IPI00218775	51.2136	-6.12	0.00507	FKBP5	Peptidyl-prolyl cis-trans isomerase FKBP5
9	IPI00023972	50.6484	-6.11	0.00408	DDX47	Probable ATP-dependent RNA helicase DDX47
10	IPI00003505	48.5521	-5.96	0.00392	TRIP13	Isoform 1 of Pachytene checkpoint protein 2 homolog
11	IPI00396627	92.0913	-5.95	0.01184	ELAC2	Isoform 1 of Zinc phosphodiesterase ELAC protein 2
12	IPI00022977	42.6451	-5.89	0.00851	CKB	Creatine kinase B-type
13	IPI00294187	75.5654	-5.89	0.00076	PADI2	Protein-arginine deiminase type-2
14	IPI00017303	104.7458	-5.89	0.02007	MSH2	DNA mismatch repair protein Msh2
15	IPI00218922	87.9994	-5.77	0.01096	SEC63	Translocation protein SEC63 homolog
16	IPI00292894	91.8114	-5.72	0.00950	TSR1	Pre-rRNA-processing protein TSR1 homolog
17	IPI00553109	117.5145	-5.70	0.02743	PITRM1	Uncharacterized protein
18	IPI00165949	107.8444	-5.67	0.01092	ERAP1	Isoform 2 of Endoplasmic reticulum aminopeptidase 1
19	IPI00165092	53.201	-5.66	0.00146	YARS2	Tyrosyl-tRNA synthetase, mitochondrial
20	IPI00290439	72.3845	-5.63	0.00094	SRPK1	cDNA FLJ58405, highly similar to Serine/threonine-protein kinase SRPK1
21	IPI00554777	62.1702	-5.07	0.00007	ASNS	Asparagine synthetase [glutamine-hydrolyzing]
22	IPI00215893	32.8191	-4.87	0.00263	HMOX1	Heme oxygenase 1
23	IPI00294891	88.9752	-4.66	0.01774	NOP2	Isoform 1 of Putative ribosomal RNA methyltransferase NOP2
24	IPI00005024	148.8583	-4.45	0.00194	MYBBP1A	Isoform 1 of Myb-binding protein 1A
25	IPI00032158	101.2069	-4.27	0.00106	NAA15	Isoform 2 of N-alpha-acetyltransferase 15, NatA auxiliary subunit
26	IPI00550882	35.9808	-3.80	0.00734	PYCR1	Pyrroline-5-carboxylate reductase
27	IPI00033036	52.8923	-3.58	0.00087	METAP2	Methionine aminopeptidase 2
28	IPI00396203	132.6026	-3.56	0.00737	TBCD	Isoform 1 of Tubulin-specific chaperone D
29	IPI00218728	46.6374	-3.52	0.00492	PAFAH1B1	Isoform 1 of Platelet-activating factor acetylhydrolase IB subunit alpha
30	IPI00004534	144.7338	-3.47	0.00061	PFAS	Phosphoribosylformylglycinamidine synthase
31	IPI00024403	60.1316	-3.30	0.00526	CPNE3	Copine-3
32	IPI00829992	119.5254	-3.25	0.00027	MYO1C	Isoform 3 of Myosin-Ic
33	IPI00018350	82.2883	-3.25	0.00147	MCM5	DNA replication licensing factor MCM5

Continue.

34	IPI00217686	96.5605	-3.23	0.00267	FTSJ3	Putative rRNA methyltransferase 3
35	IPI00784414	88.0696	-3.15	0.00004	STAT3	Isoform 1 of Signal transducer and activator of transcription 3
36	IPI00014197	27.3347	-3.08	0.00551	CDV3	Isoform 1 of Protein CDV3 homolog
37	IPI00334907	31.5403	-2.88	0.00841	PITPNB	Isoform 1 of Phosphatidylinositol transfer protein beta isoform
38	IPI00178431	73.4589	-2.77	0.01296	RECQL	ATP-dependent DNA helicase Q1
39	IPI00384456	152.7899	-2.70	0.00265	MSH6	Isoform GTBP-N of DNA mismatch repair protein Msh6
40	IPI00001734	45.3561	-2.64	0.00001	PSAT1	Phosphoserine aminotransferase
41	IPI00015973	112.5878	-2.55	0.00056	EPB41L2	Band 4.1-like protein 2
42	IPI00016249	69.7209	-2.54	0.00811	FXR1	Isoform 1 of Fragile X mental retardation syndrome-related protein 1
43	IPI00004233	358.6286	-2.51	0.00200	MKI67	Isoform Long of Antigen KI-67
44	IPI00301263	236.0221	-2.40	0.00059	CAD	CAD protein
45	IPI01014863	41.3504	-2.38	0.01202	ACAT2	Acetyl-CoA acetyltransferase, cytosolic
46	IPI00011200	56.6506	-2.30	0.00111	PHGDH	D-3-phosphoglycerate dehydrogenase
47	IPI00306369	86.4728	-2.28	0.00412	NSUN2	tRNA (cytosine(34)-C(5))-methyltransferase
48	IPI00550385	838.3142	-2.28	0.00627	MACF1	Isoform 1 of Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5
49	IPI00744648	146.2052	-2.22	0.00719	SPAG9	Isoform 1 of C-Jun-amino-terminal kinase-interacting protein 4
50	IPI00220637	58.7787	-2.08	0.03169	SARS	Seryl-tRNA synthetase, cytoplasmic
51	IPI00333067	109.5883	-2.00	0.01436	HERC4	Isoform 1 of Probable E3 ubiquitin-protein ligase HERC4
52	IPI00376005	20.1709	-1.97	0.00290	EIF5A	Isoform 2 of Eukaryotic translation initiation factor 5A-1
53	IPI00216319	28.2196	-1.97	0.00063	YWHAH	14-3-3 protein eta
54	IPI00748303	117.6924	-1.95	0.00609	ZFR	Uncharacterized protein
55	IPI00299524	157.1863	-1.92	0.00019	NCAPD2	Condensin complex subunit 1
56	IPI00000030	66.1829	-1.88	0.00846	PPP2R5D	Isoform Delta-1 of Serine/threonine-protein phosphatase 2A 56 kDa regulatory subunit delta isoform
57	IPI00025273	107.7684	-1.84	0.00853	GART	Isoform Long of Trifunctional purine biosynthetic protein adenosine-3
58	IPI00013214	95.9103	-1.80	0.00011	MCM3	cDNA FLJ55599, highly similar to DNA replication licensing factor MCM3
59	IPI00441473	72.6851	-1.75	0.00033	PRMT5	Protein arginine N-methyltransferase 5
60	IPI00218606	15.8077	-1.74	0.04921	RPS23	40S ribosomal protein S23
61	IPI00002459	75.2808	-1.70	0.02534	ANXA6	Uncharacterized protein
62	IPI00783313	93.1375	-1.69	0.01078	PYGL	Glycogen phosphorylase, liver form
63	IPI00017334	29.8046	-1.67	0.00163	PHB	Prohibitin
64	IPI00003519	105.3856	-1.66	0.00540	EFTUD2	116 kDa U5 small nuclear ribonucleoprotein component
65	IPI00644431	53.6972	-1.65	0.01897	DDX39	cDNA FLJ55484, highly similar to ATP-dependent RNA helicase DDX39
66	IPI00015897	37.4897	-1.63	0.04521	CHORDC1	Isoform 1 of Cysteine and histidine-rich domain-containing protein 1
67	IPI00783097	83.1676	-1.62	0.02221	GARS	Glycyl-tRNA synthetase
68	IPI00219616	34.8347	-1.61	0.03823	PRPS1	Ribose-phosphate pyrophosphokinase 1
69	IPI00218914	54.8628	-1.60	0.00008	ALDH1A1	Retinal dehydrogenase 1
70	IPI00007928	273.6086	-1.59	0.00073	PRPF8	Pre-mRNA-processing-splicing factor 8
71	IPI00007334	150.5571	-1.57	0.02266	ACIN1	Isoform 1 of Apoptotic chromatin condensation inducer in the nucleus

Continue.

72	IPI00026569	40.8458	-1.56	0.00017	HLA-A	HLA class I histocompatibility antigen, A-1 alpha chain
73	IPI00101186	143.7051	-1.49	0.01579	RRP12	Isoform 1 of RRP12-like protein
74	IPI00385042	73.9673	-1.46	0.00283	GTPBP4	Nucleolar GTP-binding protein 1
75	IPI00290142	66.6907	-1.46	0.00027	CTPS	CTP synthase 1
76	IPI00219217	36.6386	-1.43	0.00026	LDHB	L-lactate dehydrogenase B chain
77	IPI00001159	292.7644	-1.41	0.00070	GCN1L1	Translational activator GCN1
78	IPI00298696	152.2035	-1.40	0.00625	TCOF1	Isoform 2 of Treacle protein
79	IPI00411559	147.1879	-1.36	0.00077	SMC4	Isoform 1 of Structural maintenance of chromosomes protein 4
80	IPI00219029	46.2481	-1.29	0.02572	GOT1	Aspartate aminotransferase, cytoplasmic
81	IPI00419979	58.0772	-1.28	0.01209	LOC646214	Serine/threonine-protein kinase PAK 2
82	IPI00329633	83.4378	-1.28	0.02555	TARS	Threonyl-tRNA synthetase, cytoplasmic
83	IPI00026781	273.4271	-1.27	0.00014	FASN	Fatty acid synthase
84	IPI00218830	48.1415	-1.27	0.00822	NMT1	Isoform Short of Glycylpeptide N-tetradecanoyltransferase 1
85	IPI00008433	22.877	-1.25	0.00300	RPS5	40S ribosomal protein S5
86	IPI00029629	70.9732	-1.23	0.02829	TRIM25	E3 ubiquitin/ISG15 ligase TRIM25
87	IPI00216694	67.6019	-1.22	0.00403	PLS3	Plastin-3
88	IPI00012462	67.8525	-1.22	0.01830	EIF2A	Eukaryotic translation initiation factor 2A
89	IPI00184330	101.8981	-1.19	0.00903	MCM2	DNA replication licensing factor MCM2
90	IPI00553185	60.5354	-1.15	0.00111	CCT3	T-complex protein 1 subunit gamma
91	IPI00234252	122.8674	-1.14	0.03174	SMARCC1	SWI/SNF complex subunit SMARCC1
92	IPI00299904	81.309	-1.14	0.02542	MCM7	Isoform 1 of DNA replication licensing factor MCM7
93	IPI00029019	114.5341	-1.13	0.00104	UBAP2L	Isoform 2 of Ubiquitin-associated protein 2-like
94	IPI00013683	50.4327	-1.13	0.02167	TUBB3	Tubulin beta-3 chain
95	IPI00024664	93.3095	-1.12	0.00654	USP5	Isoform Long of Ubiquitin carboxyl-terminal hydrolase 5
96	IPI00000816	29.175	-1.10	0.00323	YWHAE	Isoform 1 of 14-3-3 protein epsilon
97	IPI00022462	84.8736	-1.09	0.00103	TFRC	Transferrin receptor protein 1
98	IPI00031801	40.0894	-1.06	0.01178	CSDA	Isoform 1 of DNA-binding protein A
99	IPI00395865	47.82	-1.06	0.00044	RBBP7	Histone-binding protein RBBP7
100	IPI00964079	57.145	-1.05	0.03078	CCT5	Uncharacterized protein
101	IPI00909703	45.7302	-1.05	0.02712	ANXA11	Uncharacterized protein
102	IPI00000873	140.4675	-1.02	0.00746	VARS	Valyl-tRNA synthetase
103	IPI00176903	43.4765	-1.01	0.01776	PTRF	Isoform 1 of Polymerase I and transcript release factor
104	IPI00002214	57.8625	-0.98	0.01622	KPNA2	Importin subunit alpha-2
105	IPI00031517	92.8906	-0.98	0.02801	MCM6	DNA replication licensing factor MCM6
106	IPI00027626	58.0253	-0.97	0.03482	CCT6A	T-complex protein 1 subunit zeta
107	IPI00414676	83.2673	-0.96	0.00102	HSP90AB1	Heat shock protein HSP 90-beta
108	IPI00294536	38.4394	-0.94	0.01691	STRAP	cDNA FLJ51909, highly similar to Serine-threonine kinase receptor-associatedprotein
109	IPI00027252	33.2976	-0.93	0.02274	PHB2	Prohibitin-2

Continue.

110	IPI00028031	70.3916	-0.93	0.01170	ACADVL	cDNA FLJ56425, highly similar to Very-long-chain specific acyl-CoAdehydrogenase, mitochondrial
111	IPI00031461	50.6655	-0.92	0.02047	GDI2	cDNA FLJ60299, highly similar to Rab GDP dissociation inhibitor beta
112	IPI00012268	100.2022	-0.92	0.00741	PSMD2	26S proteasome non-ATPase regulatory subunit 2
113	IPI00301058	39.8291	-0.92	0.02523	VASP	Vasodilator-stimulated phosphoprotein
114	IPI00172656	52.6244	-0.91	0.02005	FAF2	FAS-associated factor 2
115	IPI00003768	66.0797	-0.88	0.03541	PES1	Isoform 1 of Pescadillo homolog
116	IPI00549248	32.5755	-0.86	0.03177	NPM1	Isoform 1 of Nucleophosmin
117	IPI00013452	170.5932	-0.85	0.00338	EPRS	Bifunctional aminoacyl-tRNA synthetase
118	IPI00018931	91.71	-0.85	0.01438	VPS35	Vacuolar protein sorting-associated protein 35
119	IPI00026202	18.1104	-0.84	0.00309	RPL18A	60S ribosomal protein L18a
120	IPI00291175	123.8013	-0.84	0.00704	VCL	Isoform 1 of Vinculin
121	IPI00029079	76.7167	-0.83	0.03810	GMPS	GMP synthase [glutamine-hydrolyzing]
122	IPI00290460	35.612	-0.83	0.04217	EIF3G	Eukaryotic translation initiation factor 3 subunit G
123	IPI00179473	47.6874	-0.82	0.00591	SQSTM1	Isoform 1 of Sequestosome-1
124	IPI00290204	51.5584	-0.81	0.02268	SNRNP70	Isoform 1 of U1 small nuclear ribonucleoprotein 70 kDa
125	IPI00550689	55.2102	-0.81	0.01408	C22orf28	tRNA-splicing ligase RtcB homolog
126	IPI00021435	48.6356	-0.80	0.01208	PSMC2	26S protease regulatory subunit 7
127	IPI00007074	59.1451	-0.80	0.00705	YARS	Tyrosyl-tRNA synthetase, cytoplasmic
128	IPI00926977	45.7991	-0.78	0.02952	PSMC6	26S protease regulatory subunit 10B
129	IPI00018274	128.6815	-0.77	0.01884	EGFR	Isoform 1 of Epidermal growth factor receptor
130	IPI00304596	54.2316	-0.77	0.03193	NONO	Non-POU domain-containing octamer-binding protein
131	IPI00026625	155.203	-0.76	0.01483	NUP155	Isoform 1 of Nuclear pore complex protein Nup155
132	IPI00478231	21.7684	-0.72	0.03730	RHOA	Transforming protein RhoA
133	IPI00021728	38.3897	-0.71	0.04327	EIF2S2	Eukaryotic translation initiation factor 2 subunit 2
134	IPI00008240	101.1175	-0.70	0.01495	MARS	Methionyl-tRNA synthetase, cytoplasmic
135	IPI00301936	38.997	-0.70	0.01963	ELAVL1	cDNA FLJ60076, highly similar to ELAV-like protein 1
136	IPI01019005	53.4908	-0.70	0.02830	ATXN10	Ataxin-10
137	IPI00297779	57.4899	-0.70	0.00365	CCT2	T-complex protein 1 subunit beta
138	IPI00017617	69.1497	-0.69	0.02957	DDX5	Probable ATP-dependent RNA helicase DDX5
139	IPI00002966	94.3319	-0.69	0.00451	HSPA4	Heat shock 70 kDa protein 4
140	IPI00306960	62.9443	-0.69	0.00241	NARS	Asparaginyl-tRNA synthetase, cytoplasmic
141	IPI00783271	157.9122	-0.69	0.01934	LRPPRC	Leucine-rich PPR motif-containing protein, mitochondrial
142	IPI00021187	50.2294	-0.68	0.00712	RUVBL1	Isoform 1 of RuvB-like 1
143	IPI00218342	107.4363	-0.68	0.04579	MTHFD1	C-1-tetrahydrofolate synthase, cytoplasmic
144	IPI00000690	66.2956	-0.66	0.01517	AIFM1	Isoform 1 of Apoptosis-inducing factor 1, mitochondrial
145	IPI00479262	158.521	-0.65	0.01142	EIF4G1	eukaryotic translation initiation factor 4 gamma 1 isoform 1
146	IPI00003881	45.6719	-0.64	0.02758	HNRNPF	Heterogeneous nuclear ribonucleoprotein F
147	IPI00290566	60.3452	-0.64	0.00387	TCP1	T-complex protein 1 subunit alpha

Continue.

148	IPI00012442	52.1628	-0.63	0.01410	G3BP1	Ras GTPase-activating protein-binding protein 1
149	IPI00011126	49.1864	-0.63	0.02509	PSMC1	26S protease regulatory subunit 4
150	IPI00440493	59.7521	-0.61	0.00859	ATP5A1	ATP synthase subunit alpha, mitochondrial
151	IPI00021700	28.7693	0.59	0.01195	PCNA	Proliferating cell nuclear antigen
152	IPI00021405	74.1407	0.61	0.00303	LMNA	Isoform A of Prelamin-A/C
153	IPI00783872	76.8599	0.61	0.02125	CAPRIN1	Isoform 1 of Caprin-1
154	IPI00016801	61.3997	0.61	0.03002	GLUD1	Glutamate dehydrogenase 1, mitochondrial
155	IPI00012074	70.9439	0.62	0.00173	HNRNPR	Isoform 1 of Heterogeneous nuclear ribonucleoprotein R
156	IPI00908881	59.9918	0.63	0.04730	GPI	Glucose-6-phosphate isomerase
157	IPI00783862	22.1187	0.63	0.00048	BLVRB	Flavin reductase
158	IPI00102069	42.5039	0.64	0.00908	EIF3M	Eukaryotic translation initiation factor 3 subunit M
159	IPI00219097	24.0346	0.65	0.00358	HMGB2	High mobility group protein B2
160	IPI00418471	53.6527	0.66	0.00015	VIM	Vimentin
161	IPI00031583	109.197	0.69	0.00668	USO1	Isoform 2 of General vesicular transport factor p115
162	IPI00783641	60.373	0.69	0.02486	TXNRD1	Isoform 3 of Thioredoxin reductase 1, cytoplasmic
163	IPI00217966	36.6892	0.73	0.00859	LDHA	Isoform 1 of L-lactate dehydrogenase A chain
164	IPI00219757	23.3567	0.73	0.00432	GSTP1	Glutathione S-transferase P
165	IPI00295851	107.146	0.74	0.00523	COPB1	Coatomer subunit beta
166	IPI00220642	28.3031	0.75	0.02477	YWHAG	14-3-3 protein gamma
167	IPI00011603	60.9796	0.75	0.02893	PSMD3	26S proteasome non-ATPase regulatory subunit 3
168	IPI00219420	141.5471	0.76	0.00661	SMC3	Structural maintenance of chromosomes protein 3
169	IPI00789155	37.1077	0.76	0.00830	CALU	calumenin isoform c precurosr
170	IPI00010105	26.5982	0.77	0.01270	EIF6	Eukaryotic translation initiation factor 6
171	IPI00026182	32.9492	0.77	0.02080	CAPZA2	F-actin-capping protein subunit alpha-2
172	IPI00479186	57.9375	0.82	0.00276	PKM2	Isoform M2 of Pyruvate kinase isozymes M1/M2
173	IPI00027230	92.4717	0.82	0.00096	HSP90B1	Endoplasmin
174	IPI00002460	50.3169	0.82	0.03805	ANXA7	Isoform 1 of Annexin A7
175	IPI00010796	57.1181	0.83	0.00017	P4HB	Protein disulfide-isomerase
176	IPI00256684	105.3642	0.83	0.02573	AP2A1	Isoform B of AP-2 complex subunit alpha-1
177	IPI00784154	61.0557	0.84	0.01822	HSPD1	60 kDa heat shock protein, mitochondrial
178	IPI00016613	45.9096	0.85	0.03114	CSNK2A1	Uncharacterized protein
179	IPI00031397	80.4231	0.89	0.04289	ACSL3	Long-chain-fatty-acid--CoA ligase 3
180	IPI00014424	50.4706	0.90	0.01501	EEF1A2	Elongation factor 1-alpha 2
181	IPI00028091	47.3719	0.92	0.00307	ACTR3	Actin-related protein 3
182	IPI00465439	39.4206	0.94	0.00688	ALDOA	Fructose-bisphosphate aldolase A
183	IPI00022334	48.5362	0.95	0.02032	OAT	Ornithine aminotransferase, mitochondrial
184	IPI00418262	48.4083	0.97	0.04258	ALDOC	Fructose-bisphosphate aldolase
185	IPI00550363	22.3919	0.98	0.01963	TAGLN2	Transgelin-2

Continue.

186	IPI00011284	30.0376	1.00	0.00434	COMT	Isoform Membrane-bound of Catechol O-methyltransferase
187	IPI00017375	82.9709	1.02	0.00545	SEC23A	Protein transport protein Sec23A
188	IPI00019912	79.6885	1.05	0.00673	HSD17B4	Peroxisomal multifunctional enzyme type 2
189	IPI00302592	280.7294	1.05	0.00001	FLNA	Isoform 2 of Filamin-A
190	IPI00014898	531.7839	1.09	0.00055	PLEC	Isoform 1 of Plectin
191	IPI00140420	101.2127	1.11	0.00275	SND1	Staphylococcal nuclease domain-containing protein 1
192	IPI00008982	87.3029	1.12	0.00604	ALDH18A1	Isoform Long of Delta-1-pyrroline-5-carboxylate synthase
193	IPI00024466	174.9825	1.12	0.02067	UGGT1	Isoform 1 of UDP-glucose:glycoprotein glucosyltransferase 1
194	IPI00141318	66.0222	1.15	0.02675	CKAP4	Cytoskeleton-associated protein 4
195	IPI00009904	72.934	1.16	0.00203	PDIA4	Protein disulfide-isomerase A4
196	IPI00012069	27.2967	1.16	0.00060	NQO1	NAD(P)H dehydrogenase [quinone] 1
197	IPI00006865	24.594	1.16	0.00606	SEC22B	Vesicle-trafficking protein SEC22b
198	IPI00219301	31.5542	1.18	0.01228	MARCKS	Myristoylated alanine-rich C-kinase substrate
199	IPI00289334	276.9334	1.19	0.00010	FLNB	Isoform 1 of Filamin-B
200	IPI00030781	87.3369	1.19	0.00498	STAT1	Isoform Alpha of Signal transducer and activator of transcription 1-alpha/beta
201	IPI00410067	101.4326	1.21	0.02741	ZC3HAV1	Isoform 1 of Zinc finger CCCH-type antiviral protein 1
202	IPI00004358	96.6983	1.24	0.02456	PYGB	Glycogen phosphorylase, brain form
203	IPI00414127	23.3108	1.34	0.00868	RANBP1	Ran-specific GTPase-activating protein
204	IPI00031131	46.4815	1.35	0.02408	C20orf3	Isoform 1 of Adipocyte plasma membrane-associated protein
205	IPI00000105	99.326	1.35	0.00260	MVP	Major vault protein
206	IPI00013070	90.2927	1.39	0.00222	HNRNPUL1	Isoform 1 of Heterogeneous nuclear ribonucleoprotein U-like protein 1
207	IPI00479722	28.7239	1.44	0.00328	PSME1	Proteasome activator complex subunit 1
208	IPI00018873	55.5225	1.49	0.00106	NAMPT	Nicotinamide phosphoribosyltransferase
209	IPI00008868	270.6344	1.53	0.00110	MAP1B	Microtubule-associated protein 1B
210	IPI00013808	104.8572	1.54	0.00101	ACTN4	Alpha-actinin-4
211	IPI00025084	28.3167	1.54	0.02978	CAPNS1	Calpain small subunit 1
212	IPI00442073	20.5671	1.57	0.00283	CSRP1	Cysteine and glycine-rich protein 1
213	IPI00017726	26.9231	1.59	0.00825	HSD17B10	Isoform 1 of 3-hydroxyacyl-CoA dehydrogenase type-2
214	IPI00329801	35.9386	1.60	0.00001	ANXA5	Annexin A5
215	IPI00030009	69.9763	1.61	0.01561	PAPSS2	Isoform A of Bifunctional 3'-phosphoadenosine 5'-phosphosulfate synthase 2
216	IPI00294578	77.3291	1.67	0.00018	TGM2	Isoform 1 of Protein-glutamine gamma-glutamyltransferase 2
217	IPI00742682	267.2893	1.68	0.02558	TPR	Nucleoprotein TPR
218	IPI00215687	65.4611	1.74	0.00215	GLS	Isoform 3 of Glutaminase kidney isoform, mitochondrial
219	IPI00005614	274.6134	1.76	0.00146	SPTBN1	Isoform Long of Spectrin beta chain, brain 1
220	IPI00550069	49.9744	1.76	0.01042	RNH1	Ribonuclease inhibitor
221	IPI00883655	73.503	1.77	0.00294	DPYSL2	dihydropyrimidinase-related protein 2 isoform 1
222	IPI00017283	113.7941	1.80	0.02110	IARS2	Isoleucyl-tRNA synthetase, mitochondrial
223	IPI00246975	26.5611	1.80	0.01699	GSTM3	Glutathione S-transferase Mu 3

Continue.

224	IPI00016862	51.7004	1.82	0.00510	GSR	Isoform Mitochondrial of Glutathione reductase, mitochondrial
225	IPI00289758	79.9991	1.86	0.00003	CAPN2	Calpain-2 catalytic subunit
226	IPI00182757	102.9031	1.94	0.04763	KIAA1967	Isoform 1 of Protein KIAA1967
227	IPI00844215	284.5427	1.96	0.00473	SPTAN1	Isoform 1 of Spectrin alpha chain, brain
228	IPI00003479	41.3919	2.00	0.00818	MAPK1	Mitogen-activated protein kinase 1
229	IPI00027223	46.6605	2.01	0.00096	IDH1	Isocitrate dehydrogenase [NADP] cytoplasmic
230	IPI00219525	51.8742	2.01	0.00009	PGD	6-phosphogluconate dehydrogenase, decarboxylating
231	IPI00414717	134.5539	2.03	0.02230	GLG1	Isoform 2 of Golgi apparatus protein 1
232	IPI00643920	68.8155	2.04	0.00156	TKT	cDNA FLJ54957, highly similar to Transketolase
233	IPI00744692	37.5417	2.22	0.00003	TALDO1	Transaldolase
234	IPI00292771	238.2597	2.23	0.00176	NUMA1	Isoform 1 of Nuclear mitotic apparatus protein 1
235	IPI00001539	41.9242	2.38	0.00085	ACAA2	3-ketoacyl-CoA thiolase, mitochondrial
236	IPI00215743	152.4706	2.60	0.00010	RRBP1	Isoform 3 of Ribosome-binding protein 1
237	IPI00017376	86.4811	2.86	0.00153	SEC23B	Protein transport protein Sec23B
238	IPI00216008	62.4697	3.35	0.00232	G6PD	Isoform Long of Glucose-6-phosphate 1-dehydrogenase
239	IPI00413641	35.8539	3.81	0.00223	AKR1B1	Aldose reductase
240	IPI00553131	38.282	5.90	0.00032	GALE	UDP-glucose 4-epimerase
241	IPI00006663	56.3814	6.57	0.00044	ALDH2	Aldehyde dehydrogenase, mitochondrial
242	IPI00013744	129.2979	6.60	0.00007	ITGA2	Integrin alpha-2

[a)] Significant difference expression $\log_2$ ratio of NCI-H1755/NCI-H1703 with NSAF value. [b)] Significant difference in t-test (p-value < 0.05).

Identification of N-terminal peptides using BSA as control

The scheme with which N-terminal peptides were identified is shown in Figure 3. The N-termini of proteins are characterized by an α-amine, as opposed to the ε-amines that are on lysine side chains. Thus, ε-amines on lysine side chains had to be blocked. We blocked the α-amine and ε-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 (30).

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 (36). Thus, our BSA had an aspartic acid at residue 25 as its natural N-terminus.

Acetylated BSA was digested with trypsin and analyzed by MALDI-MS (Figure 1-4A). 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 α-amine and ε-amine and confirmed with the peptide fingerprint by MS/MS analysis (Figure 1-4B).

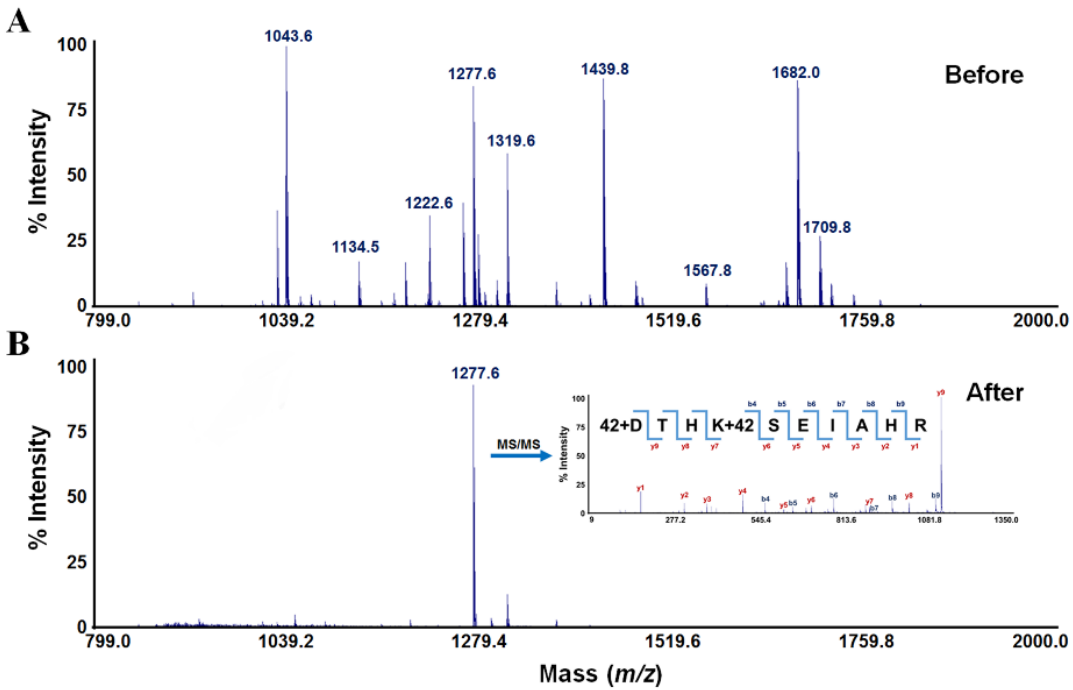


Figure 1-4. N-terminal peptide analysis of BSA control

(A) MS peaks are trypsin-digested peptides of acetylated BSA. (B) With our protocol, the labeled major ions correspond to the N-terminal peptides from BSA.

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 (30). 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 (Figure 1-5 A and B).

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) (Figure 1-5 C and D).

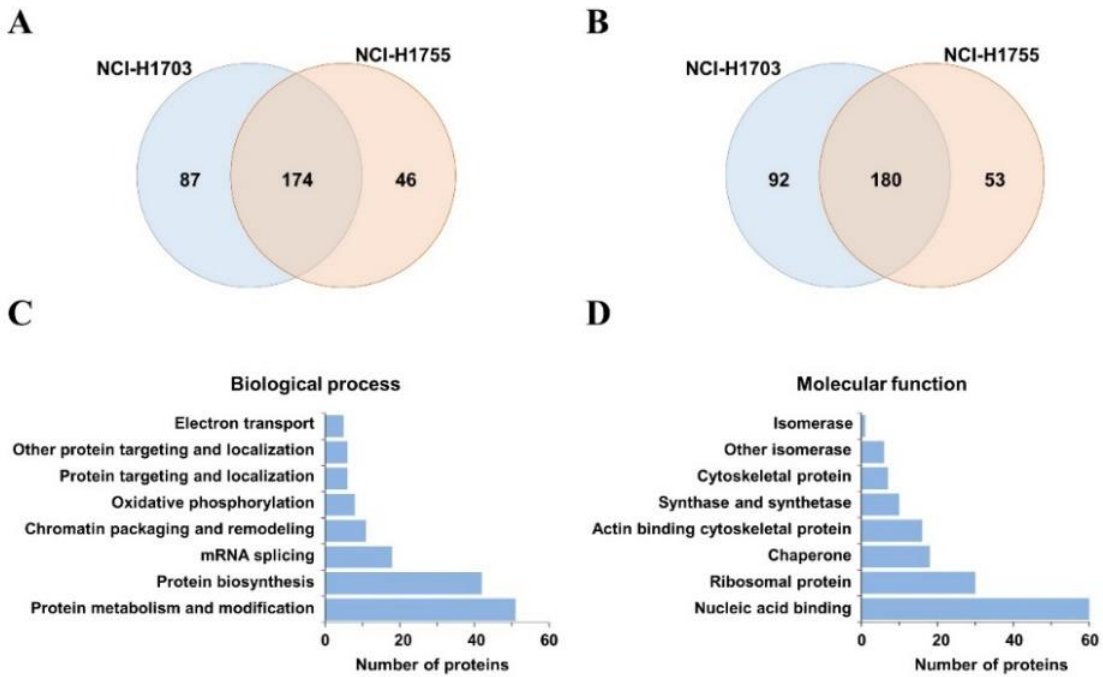


Figure 1-5. Summary of the identification of N-terminal peptides

(A) Numbers of all identified proteins and (B) peptides were shown in Venn diagrams. (C) Gene ontology (GO) for biological process and (D) molecular function of all identified proteins was performed using the DAVID bioinformatics tool.

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 (Figure 1-6 A and B): (1) initial methionine depletion, NCI-H1703 (169 peptides, 62.1%) and NCI-H1755 (148 peptides, 63.5%); (2) initial methionine nondepletion, NCI-H1703 (37 peptides, 13.6%) and NCI-H1755 (28 peptides, 12.1%); (3) signal peptide depletion, NCI-H1703 (15 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%).

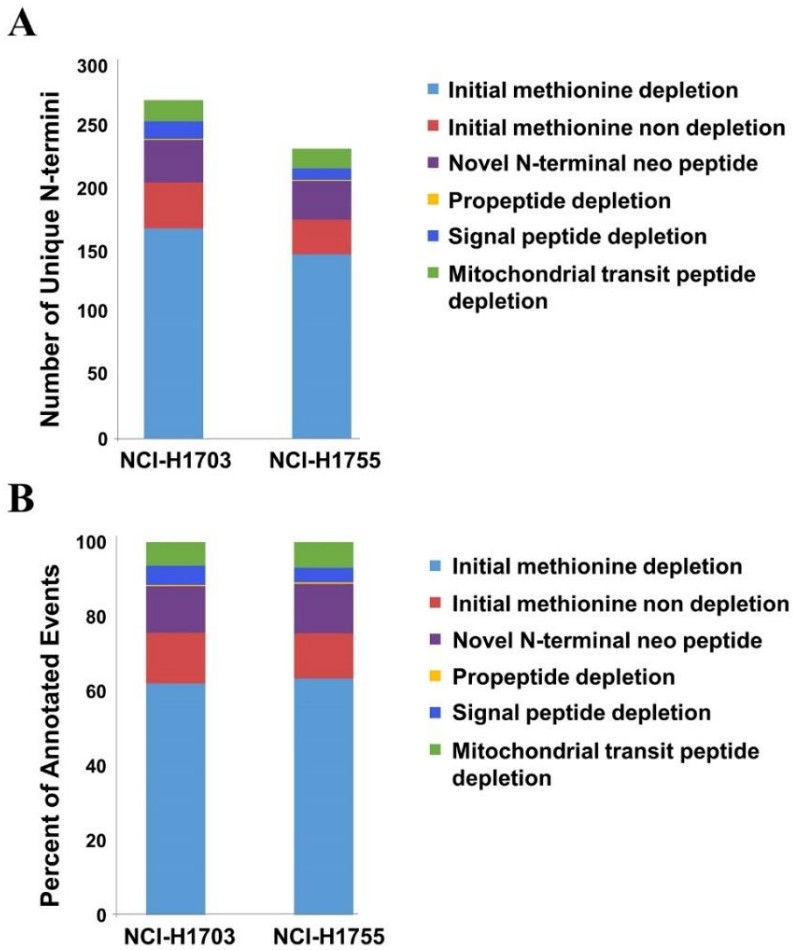


Figure 1-6. Site annotation of N-terminal peptides

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).

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 (Figure 1-7). Fourteen proteins were involved in the focal adhesion pathway in relation of cell invasion, growth, proliferation, and migration (Table 1-2), 5 of which (FLNA, FLNB, CAV1, MYL12B, and CAPN2) were common in the two parallel experiments.

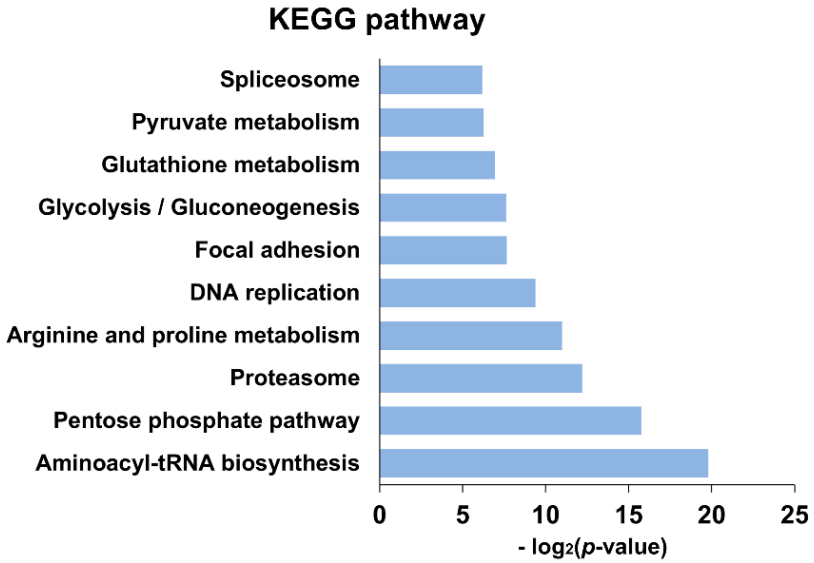


Figure 1-7. Pathways identified using differentially expressed proteins from both experiments

The numbers of significantly differentiated proteins associated with each pathway are shown in the bar graph.

Table 1-2. Focal adhesion pathway related protein list

#	IPI	peptide sequence	Peptide sequence site	Expression ratio	N-terminal peptides		Gene Symbol	Protein name
				log_2(NCI-H1755/NCI-H1703)	NCI-H1703	NCI-H1755		
1	IPI00302592	M.PATEKDLAEDAPWKKIQQNTFTR.W	Novel N-termini peptide	1.05	O	O	FLNA	Isoform 2 of Filamin-A
2	IPI00289334	M.PVTEKDLAEDAPWKKIQQNTFTR.W	Initial methionine depletion*	1.19	X	O	FLNB	Isoform 1 of Filamin-B
3	IPI00009236	M.SGGKYVDSEGHLYTVPIR.E	Initial methionine depletion	-0.04	X	O	CAV1	Isoform Alpha of Caveolin-1
4	IPI00033494	M.SSKKAKTKTTKKRPQR.A	Initial methionine depletion	0.04	X	O	MYL12B	Myosin regulatory light chain 12B
5	IPI00289758	M.AGIAAKLAKDR.E	Initial methionine depletion	1.86	X	O	CAPN2	calpain 2, (m/II) large subunit
6	IPI00004839	D.SSTCPGDYVLSVSENSR.V	Novel N-termini peptide	-	O	X	CRKL	v-crk sarcoma virus CT10 oncogene homolog (avian)-like
7	IPI00018195	M.AAAAAQGGGGGEPR.R	Initial methionine depletion	-	X	O	MAPK3	mitogen-activated protein kinase 3
8	IPI00218236	M.ADGELNVDSLITR.L	Initial methionine depletion	-	X	O	PPP1CB	protein phosphatase 1, catalytic subunit, beta isozyme
9	IPI00301058	-	-	-0.92	-	-	VASP	Vasodilator-stimulated phosphoprotein
10	IPI00291175	-	-	-0.84	-	-	VCL	Isoform 1 of Vinculin
11	IPI00478231	-	-	-0.72	-	-	RHOA	Transforming protein RhoA
12	IPI00013808	-	-	1.54	-	-	ACTN4	Alpha-actinin-4
13	IPI00003479	-	-	2.00	-	-	MAPK1	Mitogen-activated protein kinase 1
14	IPI00013744	-	-	6.60	-	-	ITGA2	Integrin alpha-2

* mark means that the peptide sequence site were not assigned in UniProtKB.

Focal adhesion pathway involved proteins were listed. List included the IPI accession number, gene symbol, peptide sequence and site from identified N-terminal peptide analysis. Expression ratio and p value were calculated by average NSAF value from label-free quantitative analysis. Information of peptide site by UniprotKB database was provided in this list.

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 (PATEKDLAEDAPWKKIQQNTFTR) in the NCI-H1703 and NCI-H1755 lines—showed differential expression in both cell lines in at least 1 experiments (Figure 1-8).

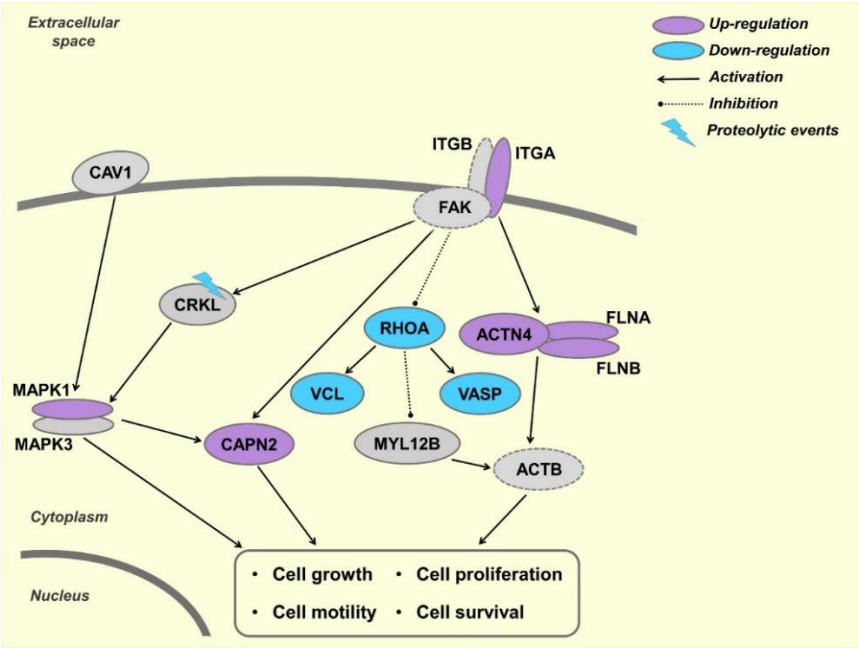


Figure 1-8. Deregulated focal adhesion pathway in NSCLC cell lines

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.

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 (37) 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.

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 (13). 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 (38) and MS/MS spectral counts (39), 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 (≤ 4 spectrum detected) and post translational modification proteins (40). However, the analysis is suitable for detection of subtle abundance changes in most proteins with high sensitivity and reproducibility (41).

In this study, we identified 2130 non-redundant 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 ≥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 (26). 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) (42).

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. In particular, LMNA is a metastatic biomarker of colorectal cancer cells (43) and a marker of embryonic stem cell differentiation (44), 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 (45); 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 (46, 47). 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 (30), dimethylation (48), iTRAQ (49), and PITC adman (50). 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 (51). 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 proteins were detected in NCI-H1703 cells, whereas 53 peptides from 46 proteins were identified in NCI-H1755 cells (Figure 1-5)—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 (52), 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 1-3). 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 (53-55). 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).

Table 1-3. Proteolytic events identified with less than 1.5 fold change

IPI	peptide sequence[a]	Ratio[b]	N-terminal analysis[c]	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.*HSNAAQTQTGEANR*.G	0.3	NCI-H1755	SHMT2	Serine hydroxymethyltransferase, mitochondrial

[a)] Observed peptide sequence from N-terminal peptide analysis is written by italics. [b)] Expression $\log_2$ ratio of NCI-H1755/NCI-H1703 with NSAF value by label-free analysis. [c)] Cell line with detected peptide sequences from N-terminal analysis.

We found 138 proteins that were common to both experiments. 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 (Figure 1-7). 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 (Figure 1-8). 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 (56). 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 (57). 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 (58).

MAP kinases mediate many processes in cancer cells, such as proliferation, migration, invasion, and metastasis (59, 60). Increased expression of MAPK1 promotes the expression of CAPN2, which functions in cell movement, migration, and invasion during metastasis (61). In the N-terminal peptide analysis, v-crk sarcoma virus CT10 oncogene 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 (62), we hypothesize that CRKL function is regulated by fragment events during metastasis.

In summary, we applied two proteome methods for biomarker discovery in lung cancer metastasis. Specially, N-terminal enrichment method was used for biomarker discovery for the first time. We can find that many of these quantitative proteins and N-terminal peptides are involved in pathways in cell migration, proliferation, and metastasis. Also, our datasets of proteins and fragment peptides in lung cells might be valuable in discovering and validating lung cancer biomarkers and metastasis markers.

CHAPTER II

Targeted proteomics predicts complete response after transarterial chemoembolization in hepatocellular carcinoma

INTRODUCTION

Hepatocellular carcinoma (HCC) is the fifth most common cancer and the third most common cause of cancer-related deaths worldwide (63). Recently, the treatment of HCC has well advanced after applications of curative therapeutic practices, such as surgical resection, liver transplantation, and local ablation (64). However, most HCC patients are diagnosed at advanced stage when curative treatment is no longer applicable. For these patients, transarterial chemoembolization (TACE) may be an effective treatment option for reducing systemic toxicity, increasing local antitumor effects, and improving survival (65). International Bridge study showed that TACE is the most widely used treatment for HCC worldwide, ahead of both surgical removal and systemic treatments (66). However, there are often shown unforeseeable outcomes after TACE in terms of treatment response and survival.

In real clinical practice, a high rate of recurrence and unsatisfactory treatment outcome after TACE remains troublesome and repeated TACE procedures are often needed, since the best response cannot always be achieved after one session of TACE, especially in large tumors (67). Georgiades et al. recommended that at least two TACE sessions should be performed before abandoning the procedure, on the basis of their observations that about half of patients who did not respond to initial TACE ultimately achieved response and that improved clinical outcomes were observed after

second course (68). Recently, Kim et al. reported that the complete response at initial TACE most strongly predicts survivals of patients with intermediate-stage HCC. However, it still remains unresolved which marker is the better for more accurate progsnostification in patients with HCC undergoing TACE.

Over the past decades, a large number of HCC diagnostic marker proteins including alpha-fetoprotein (AFP), *Lens culinaris* agglutinin A-reactive fraction of AFP (AFP-L3), and prothrombin induced by the absence of vitamin K or antagonist-II (PIVKA-II) have been discovered and , reported that dynamic change of these diagnostic marker proteins can predict outcome after TACE (69, 70). Therefore, identifying marker proteins that can help us to predict or prognosis of treatment outcomes before choosing this treatment option is an important endeavor in designing a treatment strategy.

Traditionally, the enzyme-linked immunosorbent assay (ELISA) utilizing antibodies is common quantitative assay for development of diagnostic marker proteins with high specificity and sensitivity (71). However, the immunoassay has major constraints that are the expensive and time-consuming development of specific antibodies, and the technical limitations for multiplex quantitation. In contrast, targeted proteomics approach through multiple reaction monitoring (MRM) assay suitable for multiplex quantitation of more than one hundreds of proteins with high accuracy and lower limit of quantitation (LLOQ) in efficiency cost (72). In addition, MRM assay has been shown consistent and reproducible data set across different laboratories in highly complex samples (73). More recently, Silvia *et al.* developed automated MRM data analysis workflow for validation of marker proteins in

large-scale clinical cohorts (74). In our previously study, we identified the HCC diagnostic markers using MRM and immunoassay from global data-mining. Additionally, these marker proteins showed difference level in HCC state and recovery state by treatments (75). Here, we applied marker-candidate proteins (MCPs) that have been previously reported as liver disease related proteins for TACE prognosis prediction.

Therefore, our aim was to identify pre-TACE marker proteins from the MCPs predicted to complete response after TACE, to ultimately suggest guideline for clinical decision making in future prospective studies.

MATERIALS AND METHODS

Study Population

This study was based on 180 HCC patients who were enrolled in a prospective cohort at Seoul National University Hospital (Seoul, Republic of Korea) as part of an ongoing study identifying the biomarkers associated with treatment response and prognosis in HCC. (Table 2-1). Patients with HCC who received TACE as the first-line therapy between 2008 and 2014 were considered eligible in this study. HCC was diagnosed by histological or radiological evaluation with reference to American Association for the Study of Liver (AASLD) or European Association for the Study of the Liver (EASL) guidelines.

TABLE 2-1. The clinicopathologic characteristics of the training and validation cohorts

		All cohort (*N*=180)				Training cohort (*N*=100)				Validation cohort (*N*=80)			
Variable		**Good Responders**	**(%)**	**Poor Responders**	**(%)**	**Good Responders**	**(%)**	**Poor Responders**	**(%)**	**Good Responders**	**(%)**	**Poor Responders**	**(%)**
Gender	Male	72	80.0%	82	91.1%	38	76.0%	47	94.0%	34	85.0%	35	87.5%
	Female	18	20.0%	8	8.9%	12	24.0%	3	6.0%	6	15.0%	5	12.5%
Age (years)	<60	32	35.6%	31	34.4%	19	38.0%	16	32.0%	13	32.5%	15	37.5%
	≥ 60	58	64.4%	59	65.6%	31	62.0%	34	68.0%	27	67.5%	25	62.5%
Etiology	Alcohol	4	4.4%	5	5.6%	2	4.0%	3	6.0%	2	5.0%	2	5.0%
	HBV	74	82.2%	67	74.4%	40	80.0%	39	78.0%	34	85.0%	28	70.0%
	HCV	9	10.0%	13	14.4%	6	12.0%	5	10.0%	3	7.5%	8	20.0%
	Others	3	3.3%	5	5.6%	**2**	4.0%	**3**	6.0%	1	2.5%	2	5.0%
Child-Pugh class	A	74	82.2%	63	70.0%	43	86.0%	36	72.0%	31	77.5%	27	67.5%
	B	16	17.8%	27	30.0%	7	14.0%	14	28.0%	9	22.5%	13	32.5%
MELD score	Mean ± SD	9.1 ± 2.6		9.4 ± 3.2		9.0 ± 2.4		8.8 ± 2.3		9.3 ± 2.9		10.2 ± 4.0	
Platelet (10^3/uL)	Mean ± SD	113.4 ± 53.1		116.9 ± 85.0		121.1 ± 56.3		115.0 ± 52.9		103.9 ± 47.9		120.3 ± 113.8	
ALT, IU/L	Mean ± SD	32.5 ± 19.9		39.9 ± 25.9		34.8 ± 18.3		41.8 ± 29.5		29.7 ± 21.7		36.9 ± 21.1	
Bilirubin, mg/dL	Mean ± SD	1.1 ± 0.7		1.1 ± 1.1		1.1 ± 0.5		1.2 ± 1.4		1.1 ± 0.8		1.1 ± 0.6	
Albumin	Mean ± SD	3.8 ± 0.55		3.7 ± 0.55		3.8 ± 0.58		3.7 ± 0.57		3.8 ± 0.52		3.7 ± 0.53	
Prothrombin time	Mean ± SD	1.2 ± 0.17		1.1 ± 0.13		1.2 ± 0.17		1.1 ± 0.11		1.2 ± 0.18		1.2 ± 0.15	
Creatinine	Mean ± SD	0.9 ± 0.31		1.0 ± 0.91		0.9 ± 0.29		0.9 ± 0.31		0.9 ± 0.33		1.2 ± 1.31	
No. of lesions	< 3	74	82.2%	39	43.3%	41	82.0%	20	40.0%	33	82.5%	19	47.5%
	≥ 3	16	17.8%	51	56.7%	9	18.0%	30	60.0%	7	17.5%	21	52.5%
Tumor size, cm	< 3	76	84.4%	64	71.1%	39	78.0%	34	68.0%	37	92.5%	30	75.0%
	≥ 3	14	15.6%	26	28.9%	11	22.0%	16	32.0%	3	7.5%	10	25.0%
BCLC stage	0	30	33.3%	17	18.9%	14	28.0%	9	18.0%	16	40.0%	8	20.0%
	A	43	47.8%	18	20.0%	25	50.0%	9	18.0%	18	45.0%	9	22.5%
	B	21	23.3%	40	44.4%	18	36.0%	24	48.0%	3	7.5%	16	40.0%
	C	5	5.6%	15	16.7%	2	4.0%	8	16.0%	3	7.5%	7	17.5%

Continue.

TNM stage	1	48	53.3%	23	25.6%	26	52.0%	12	24.0%	22	55.0%	11	27.5%
	2	35	38.9%	47	52.2%	20	40.0%	26	52.0%	15	37.5%	21	52.5%
	3	5	5.6%	17	18.9%	4	8.0%	11	22.0%	1	2.5%	6	15.0%
	4	2	2.2%	3	3.3%	0	0.0%	1	2.0%	2	5.0%	2	5.0%
Pre-TACE AFP, ng/mL	<20	57	63.3%	36	40.0%	28	56.0%	21	42.0%	29	72.5%	15	37.5%
	20-200	24	26.7%	38	42.2%	15	30.0%	21	42.0%	9	22.5%	17	42.5%
	>200	9	10.0%	16	17.8%	7	14.0%	8	16.0%	2	5.0%	8	20.0%
Pre-TACE PIVKA-II mAU/mL[†]	<40	65	72.2%	42	46.7%	33	68.8%	21	43.8%	32	80.0%	21	52.5%
	40-200	13	14.4%	22	24.4%	6	12.5%	14	29.2%	7	17.5%	8	20.0%
	>200	10	11.1%	24	26.7%	9	18.8%	13	27.1%	1	2.5%	11	27.5%
mRECIST[‡]	CR	90	100.0%	0	0.0%	50	100.0%	0	0.0%	40	100.0%	0	0.0%
	PR	0	0.0%	17	18.9%	0	0.0%	10	20.0%	0	0.0%	7	17.5%
	SD	0	0.0%	4	4.4%	0	0.0%	2	4.0%	0	0.0%	2	5.0%
	PD	0	0.0%	69	76.7%	0	0.0%	38	76.0%	0	0.0%	31	77.5%

AFP, alpha-fetoprotein; DCP, des gamma carboxy prothrombin; HBV, Hepatitis B virus; HCV, Hepatitis C virus; ECOG, Eastern Cooperative Oncology Group; TNM, tumor-node-metastasis, BCLC, Barcelona Clinic Liver Cancer; ALT, Alanine transaminase; mRECIST, modified response evaluation criteria in solid tumors.

[†] Missing values in training cohort, n=4 (Good responders=2, Poor responders=2)

[‡] Tumor response evaluation after 6 month with TACE

For candidate marker discovery, we adopted most recently established LiverAtlas (76) which included 19,801 genes and 50,265 proteins list related to the liver and various hepatic diseases by incorporating 53 database such as Hepatocellular carcinoma network database (HCC.net), Oncomine, Human Protein Atlas (HPA), and BiomarkerDigger. Of these databases, we selected MCPs for prognostic prediction marker discovery after TACE from pre-screening study. The training set consisted of 100 HCC patients and we collected paired samples before and 6 months after TACE. The validation set comprised 80 patients and we collected pre-TACE samples. Overall scheme of the study is summarized in Figure 1. This study protocol was in accordance with the ethical guidelines of the 1975 Declaration of Helsinki and written informed consent was obtained from each participant or responsible family member after possible complications of invasive procedures had been fully explained. This study procedure was approved by the Institutional Review Board of Seoul National University Hospital.

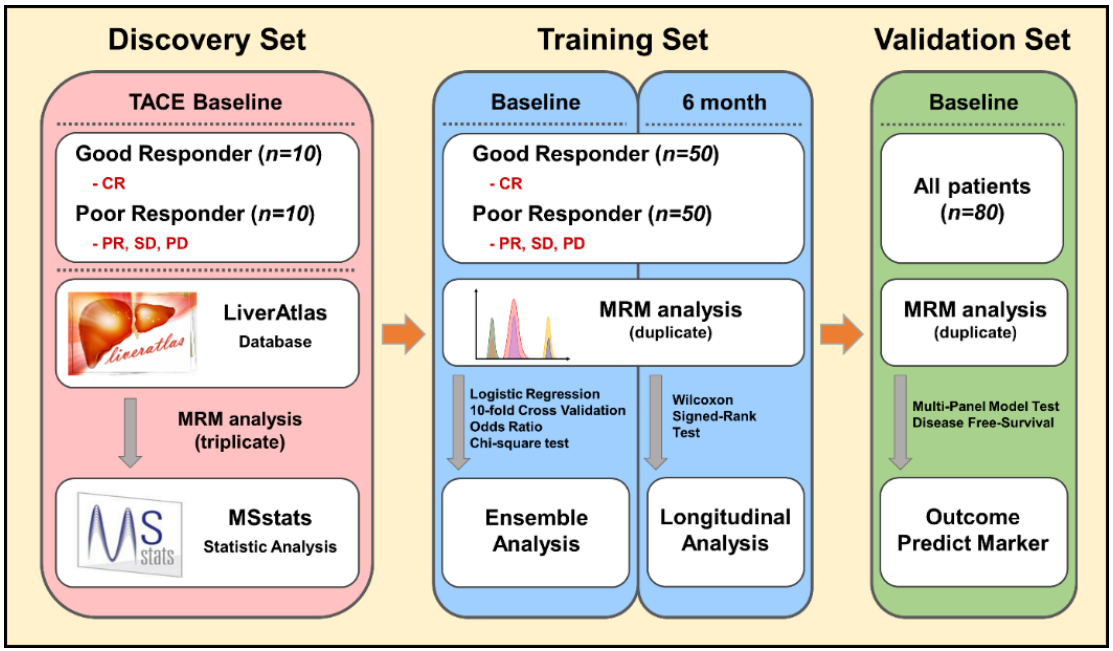


Figure 2-1. Workflow of prognostic prediction marker study

To develop prognostic prediction markers, MRM assay were performed by three strategy. First, marker candidate proteins were selected by LiverAtlas Database with MSstats statistical analysis. Next, the proteins were confirmed and combined as multi-panel model in training set by MRM assays. Finally, the model was validated in validation set.

Treatment modality

TACE was performed according to the Seoul National University Hospital protocol, as described previously. Chemoembolization was performed as selectively as possible via the lobar, segmental, or subsegmental arteries–depending on the tumor distribution and patient's hepatic functional reserve–by using a microcatheter (Microferret [Cook, Bloomington, Ind] or Progreat [Terumo, Tokyo, Japan]). The procedure was initially performed by infusing from 2 to 12 mL of iodized oil (Lipiodol; Andre Gurbet, Aulnay-sous-Bois, France) and from 10 to 60 mg of doxorubicin hydrochloride emulsion (Adriamycin RDF; Ildong Pharmaceutical, Seoul, Korea) until arterial flow stasis was achieved and/or iodized oil appeared in the portal branches. If the initial hepatic arterial blockade was insufficient because of arterioportal shunting or a large sized mass, then embolization was performed with absorbable gelatin sponge particles (1–2 mm in diameter; Gelfoam; Upjohn, Kalamazoo, MI) soaked in a mixture of from 4 to 6 mg of crystalline mitomycin (Mitomycin-C; Kyowa Hakko Kogyo, Tokyo, Japan) and 10 mL of nonionic contrast medium. The extent of chemoembolization was individually adjusted by using a superselective catheterization technique depending on the patient's hepatic functional reserve, similar to that used with surgical hepatectomy (77, 78).

Tumor Response Assessment after TACE

The tumor response evaluation for this study was assessed at CT or MRI by two expert abdominal radiologists by the modified Response Evaluation Criteria in Solid Tumors (mRECIST) for HCC (79). According to mRECIST criteria, complete response (CR) was defined as the complete disappearance of any intratumoral arterial enhancement in all recognizable tumors lesions. Partial response (PR) was defined as a decrease of at least 30 % in the sum of the longest diameter of viable (enhancement in the arterial phase) target lesions, taking as reference the baseline. Progressive disease (PD) was considered as the appearance of new lesions or as an increase of at least 20 % in the sum of the longest diameter of viable (enhancing) lesions, taking as reference the smallest sum of the longest diameters of viable (enhancing) lesions recorded since treatment started. Stable disease (SD) was defined as neither sufficient shrinkage to qualify for PR nor sufficient increase to qualify for PD criteria. Good responders were defined patients who maintained CR state for 6 months after TACE, but poor responders were defined as patients who did not.

Serum Protein Preparation for Selected Reaction Monitoring Assay

Serum depletion was performed using a Multiple Affinity Removal System Human-6 (MARS Hu-6, 4.6 mm x 100 mm, Agilent, CA, USA) affinity column on an HPLC system (Shimadzu, Kyoto, Japan) as described previously (80). Briefly, serum samples were centrifuged at 14,000 x g for 30 min at 4 ℃, and supernatants were transferred to fresh tubes. 40 µL of the supernatants were diluted with 160 µL MARS Buffer A (Agilent, CA, USA). The diluted sample was injected onto a MARS Hu-6 column and unbound fractions were collected into 1.5 mL tubes. Depleted serum was concentrated using 3000-MWCO centrifugal filter units (Amicon Ultra-4 3K, Millipore, MA, USA) and quantified by bicinchoninic acid (BCA) assay.

The depleted serum (0.1 mg) was denatured and reduced with 6 M urea, 20 mM dithiothreitol, 0.1 M Tris, pH 8 at 37 ℃ for 30 min, and alkylated with 50 mM iodoacetamide in the dark at room temperature for 30 min. To avoid trypsin compatibility concentration by urea, the alkylated sample was diluted 10-fold with 0.1 M Tris, pH 8 prior to incubation for 16 h at 37 ℃ with trypsin (Sequencing-grade modified, Promega, WI, USA) in a 1:50 enzyme to substrate ratio. After 16 h incubation, neat formic acid was added to 2% to quench the enzymatic reaction and desalted using Oasis® HLB 1cc (30 mg) extraction cartridges (Waters, MA, USA).

Desalting procedure was followed. Oasis cartridge was washed with 1 mL of 100% MeOH, washed with 3 mL of 100% ACN in 0.1% formic acid, and equilibrated with 3 mL of 0.1% formic acid, sequentially. After total

volume of the digested serum were loaded into the cartridge, the cartridge was washed with 3 mL of 0.1% formic acid, and eluted with 1 mL of 80% ACN in 0.1% formic acid. The eluted sample was lyophilized to vacuum centrifuged and stored at -80℃ until analysis. The sample was resolublized in 0.1% formic acid to 2 μg/μL prior to MRM analysis.

Quantification by multiple reaction monitoring assay

All samples were analyzed on an Agilent 6490 triple quadrupole (QQQ) mass spectrometry (Agilent Technologies, Santa Clara, CA) equipped with 1260 Infinity HPLC system (Agilent Technologies) and using a micro-flow (10 μL/min) gradient of 3 to 35% acetonitrile (ACN)/0.1% formic acid (FA) in 45 min. The analytical column was 150 mm × 0.5 mm id, packed with Agilent Zorbax SB-C18 (3.5-μm particle size), and maintained at 40℃. The MRM assay was conducted in the positive mode with 2500 V of the ion spray capillary and 2000 V of nozzle voltage. The drying gas and sheathe gas temperature was set to 250℃ at 15 L/min and 350℃ at 12 L/min, respectively. Delta EMV was set to 200 V, and the cell accelerator voltage and fragment voltage was 5 V and 380 V, respectively.

Statistical analysis

To analysis the MRM results, all raw files (.d format) were inputted in Skyline software. All transition signals were manually integrated with the Savitzky-Golay smoothing algorithm and exported MSstats format. And

protein significance and relative abundance were analyzed with MSstats package in R. We performed MSstats procedure as described in (81). Briefly, for data preprocessing, all transition intensities were transformed into log2 values. Then, we performed the equalizing of the median peak intensities of reference transitions between the runs. Finally, significant difference and relative abundance of the proteins were calculated by the linear mixed-effects model implemented in MSstats. Receiver operator characteristic (ROC) curves and logistic regression modeling were analyzed by panel composer web statistic tool (82) and MedCalc (Mariakerke, Belgium, ver12.2.1) with relative abundance of each proteins. Also, ROC curve was performed with 10-fold cross validation. Also, Mann-Whitney and Kuskal-Wallis with post-hoc Dunnett-T3 test were used for nonparametric group comparisons. Non-parametric statistics were done using SPSS 22.0 for Windows.

RESULTS

Selection of MCPs from LiverAtlas

From the LiverAtlas database, we focused on proteins that have reliability score more than 4 (27,410 proteins), liver specific proteins (162 proteins), or significant proteins in HCC (1,210 proteins). Out of 27,568 proteins, 948 proteins were reported as secretion proteins in UniProt Knowledgebase (UniProtKB, http://www.uniprot.org/). Then, 572 proteins were filtered with MS/MS spectrum from home-made and National Institute of Standards and Technology (NIST) MS/MS library for empirical evidence of MS detectability. From these proteins, we selected ten proteolytic peptides per protein, and 3,928 peptides were selected to represent the 572 MCPs for sorting detectable proteins in serum samples (Figure 2-2).

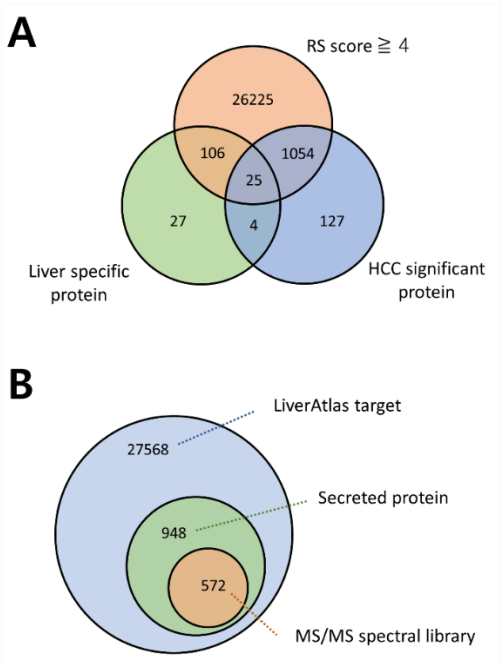


Figure 2-2. List of detectable marker candidate proteins from the LiverAtlas Database

(A) From the LiverAtlas Database, marker candidate proteins were selected by three criteria, RS score ≧ 4, liver specific proteins, and HCC significant proteins. (B) Experimental detectable proteins were selected by secretion DB and MS/MS spectral library.

Detection of MCPs in pooled serum using label-free MRM

Selection of true transition signals in complex samples is challenging due to numerous interfering (false) transition signals (83). To establish detectable proteins, we analyzed 572 MCPs with decoy peptides to pooled serum using label-free MRM. To minimize the number of MS run, we generated 393 decoy peptides (10% of total number of peptides), which is the minimal percentage acceptable for the mProphet tool in skyline software (84), by adding or subtracting a random integer to Q1 and Q3 m/z values. Total 186 MS runs were analyzed, and the results were evaluated by the mProphet tool. The mProphet tool suggests combined score of each peptide by intensity, co-elution count, library intensity dot-product, and peak shape of each peptide. As a result, 1,108 peptides corresponding to 109 proteins were detected with a false discovery rate (FDR) of 0.1%. Also, we manually selected 41 proteins that were missed from the mProphet tool but had high peak intensity and co-eluted transitions. Finally, total 175 peptides from 104 MCPs were synthesized for label MRM assays (Table 2-2).

TABLE 2-2. 104 marker candidate proteins list

N	Uniprot ID	Uniprot Accession	Gene Symbol	Protein name	HCC significant protein	RS score
1	1433S	P31947	SFN	14-3-3 protein sigma	N	4
2	A2AP	P08697	SERPINF2	Alpha-2-antiplasmin	N	5
3	A2GL	P02750	LRG1	Leucine-rich alpha-2-glycoprotein	Yes	5
4	A2MG	P01023	A2M	Alpha-2-macroglobulin	Yes	5
5	AACT	P01011	SERPINA3	Alpha-1-antichymotrypsin	Yes	4
6	ALS	P35858	IGFALS	Insulin-like growth factor-binding protein complex acid labile subunit	N	4
7	AMBP	P02760	AMBP	Protein AMBP	Yes	4
8	ANGT	P01019	AGT	Angiotensinogen	Yes	4
9	ANT3	P01008	SERPINC1	Antithrombin-III	Yes	5
10	APOA1	P02647	APOA1	Apolipoprotein A-I	Yes	5
11	APOA4	P06727	APOA4	Apolipoprotein A-IV	N	4
12	APOC1	P02654	APOC1	Apolipoprotein C-I	Yes	4
13	APOC2	P02655	APOC2	Apolipoprotein C-II	Yes	3
14	APOC3	P02656	APOC3	Apolipoprotein C-III	Yes	4
15	APOC4	P55056	APOC4	Apolipoprotein C-IV	N	4
16	APOE	P02649	APOE	Apolipoprotein E	Yes	5
17	APOF	Q13790	APOF	Apolipoprotein F	N	4
18	APOH	P02749	APOH	Beta-2-glycoprotein 1	Yes	4
19	APOL1	O14791	APOL1	Apolipoprotein L1	N	4
20	BGH3	Q15582	TGFBI	Transforming growth factor-beta-induced protein ig-h3	Yes	4
21	BTD	P43251	BTD	Biotinidase	Yes	4
22	C1QB	P02746	C1QB	Complement C1q subcomponent subunit B	N	4
23	C1QC	P02747	C1QC	Complement C1q subcomponent subunit C	N	4
24	C1RL	Q9NZP8	C1RL	Complement C1r subcomponent-like protein	Yes	3
25	C4BPA	P04003	C4BPA	C4b-binding protein alpha chain	Yes	4
26	C4BPB	P20851	C4BPB	C4b-binding protein beta chain	N	3
27	CATB	P07858	CTSB	Cathepsin B	Yes	4
28	CBPB2	Q96IY4	CPB2	Carboxypeptidase B2	N	5

Continue.

29	CD5L	O43866	CD5L	CD5 antigen-like	N	4
30	CETP	P11597	CETP	Cholesteryl ester transfer protein	Yes	5
31	CFAH	P08603	CFH	Complement factor H	Yes	4
32	CFAI	P05156	CFI	Complement factor I	Yes	4
33	CHLE	P06276	BCHE	Cholinesterase	Yes	4
34	CO2	P06681	C2	Complement C2	Yes	4
35	CO4A	P0C0L4	C4A	Complement C4-A	Yes	3
36	CO5	P01031	C5	Complement C5	N	4
37	CO6	P13671	C6	Complement component C6	Yes	4
38	CO7	P10643	C7	Complement component C7	Yes	4
39	CO8B	P07358	C8B	Complement component C8 beta chain	N	4
40	COL11	Q9BWP8	COLEC11	Collectin-11	N	4
41	CPN2	P22792	CPN2	Carboxypeptidase N subunit 2	N	5
42	CRAC1	Q9NQ79	CRTAC1	Cartilage acidic protein 1	N	4
43	CRP	P02741	CRP	C-reactive protein	Yes	5
44	CXCL7	P02775	PPBP	Platelet basic protein	N	4
45	FA10	P00742	F10	Coagulation factor X	N	5
46	FA11	P03951	F11	Coagulation factor XI	N	4
47	FA12	P00748	F12	Coagulation factor XII	N	4
48	FA9	P00740	F9	Coagulation factor IX	N	5
49	FBLN1	P23142	FBLN1	Fibulin-1	Yes	4
50	FCN3	O75636	FCN3	Ficolin-3	Yes	3
51	FETA	P02771	AFP	Alpha-fetoprotein	Yes	4
52	FETUA	P02765	AHSG	Alpha-2-HS-glycoprotein	Yes	4
53	FETUB	Q9UGM5	FETUB	Fetuin-B	N	4
54	FHR2	P36980	CFHR2	Complement factor H-related protein 2	N	4
55	FHR5	Q9BXR6	CFHR5	Complement factor H-related protein 5	N	4
56	FIBA	P02671	FGA	Fibrinogen alpha chain	Yes	4
57	FIBB	P02675	FGB	Fibrinogen beta chain	Yes	5
58	FIBG	P02679	FGG	Fibrinogen gamma chain	Yes	4
59	FINC	P02751	FN1	Fibronectin	Yes	4
60	HABP2	Q14520	HABP2	Hyaluronan-binding protein 2	N	5

Continue.

61	HEMO	P02790	HPX	Hemopexin	Yes	4
62	HGFA	Q04756	HGFAC	Hepatocyte growth factor activator	N	4
63	HPTR	P00739	HPR	Haptoglobin-related protein	Yes	4
64	IBP2	P18065	IGFBP2	Insulin-like growth factor-binding protein 2	Yes	4
65	IBP3	P17936	IGFBP3	Insulin-like growth factor-binding protein 3	Yes	3
66	IC1	P05155	SERPING1	Plasma protease C1 inhibitor	Yes	5
67	IGF2	P01344	IGF2	Insulin-like growth factor II	N	4
68	IGHG1	P01857	IGHG1	Ig gamma-1 chain C region	Yes	4
69	IGHG3	P01860	IGHG3	Ig gamma-3 chain C region	N	4
70	IGJ	P01591	IGJ	Immunoglobulin J chain	Yes	5
71	IPSP	P05154	SERPINA5	Plasma serine protease inhibitor	N	4
72	ISLR	O14498	ISLR	Immunoglobulin superfamily containing leucine-rich repeat protein	N	4
73	ITIH1	P19827	ITIH1	Inter-alpha-trypsin inhibitor heavy chain H1	Yes	4
74	ITIH2	P19823	ITIH2	Inter-alpha-trypsin inhibitor heavy chain H2	Yes	4
75	ITIH3	Q06033	ITIH3	Inter-alpha-trypsin inhibitor heavy chain H3	N	3
76	ITIH4	Q14624	ITIH4	Inter-alpha-trypsin inhibitor heavy chain H4	Yes	5
77	KAIN	P29622	SERPINA4	Kallistatin	Yes	4
78	KLKB1	P03952	KLKB1	Plasma kallikrein	N	5
79	LBP	P18428	LBP	Lipopolysaccharide-binding protein	N	4
80	LCAT	P04180	LCAT	Phosphatidylcholine-sterol acyltransferase	N	5
81	LG3BP	Q08380	LGALS3BP	Galectin-3-binding protein	Yes	4
82	LUM	P51884	LUM	Lumican	Yes	4
83	MBL2	P11226	MBL2	Mannose-binding protein C	N	5
84	NGAL	P80188	LCN2	Neutrophil gelatinase-associated lipocalin	Yes	4
85	PAPP1	Q13219	PAPPA	Pappalysin-1	N	5
86	PGRP2	Q96PD5	PGLYRP2	N-acetylmuramoyl-L-alanine amidase	N	4
87	PHLD	P80108	GPLD1	Phosphatidylinositol-glycan-specific phospholipase D	N	4
88	PLMN	P00747	PLG	Plasminogen	Yes	4
89	PON1	P27169	PON1	Serum paraoxonase/arylesterase 1	Yes	5
90	POSTN	Q15063	POSTN	Periostin	N	5
91	PROS	P07225	PROS1	Vitamin K-dependent protein S	N	4
92	PROZ	P22891	PROZ	Vitamin K-dependent protein Z	N	4

Continue.

93	PVR	P15151	PVR	Poliovirus receptor	N	5
94	QSOX1	O00391	QSOX1	Sulfhydryl oxidase 1	Yes	4
95	RET4	P02753	RBP4	Retinol-binding protein 4	Yes	5
96	SAMP	P02743	APCS	Serum amyloid P-component	Yes	5
97	SEPP1	P49908	SEPP1	Selenoprotein P	Yes	4
98	SODE	P08294	SOD3	Extracellular superoxide dismutase [Cu-Zn]	N	4
99	THBG	P05543	SERPINA7	Thyroxine-binding globulin	N	3
100	THRB	P00734	F2	Prothrombin	Yes	4
101	VTDB	P02774	GC	Vitamin D-binding protein	Yes	4
102	VTNC	P04004	VTN	Vitronectin	Yes	4
103	ZA2G	P25311	AZGP1	Zinc-alpha-2-glycoprotein	Yes	4
104	ZPI	Q9UK55	SERPINA10	Protein Z-dependent protease inhibitor	N	4

Selection of quantitative MCPs using MRM assay with labeled reference peptides

In MRM assay, measurement level such as limit of detection (LOD) and limit of quantitation (LOQ) is critical point (73). So, we selected quantitative MCPs by 2 steps as following; interference free transition using Automated Detection of Inaccurate and imprecise Transitions (AuDIT) algorithm and assay linearity using calibration curve.

First, in order to minimize interfering transition signals, pooled sample with 175 peptides (endogenous and reference peptide pairs) were analyzed with 5 or 6 transitions per peptide in triplicates. Of these peptides, 161 peptides were passed with more than 3 transitions having no interference signal, respectively (Figure 2-3A). On the contrary, 14 peptides that had less than 2 interference free transitions were excluded in the following step.

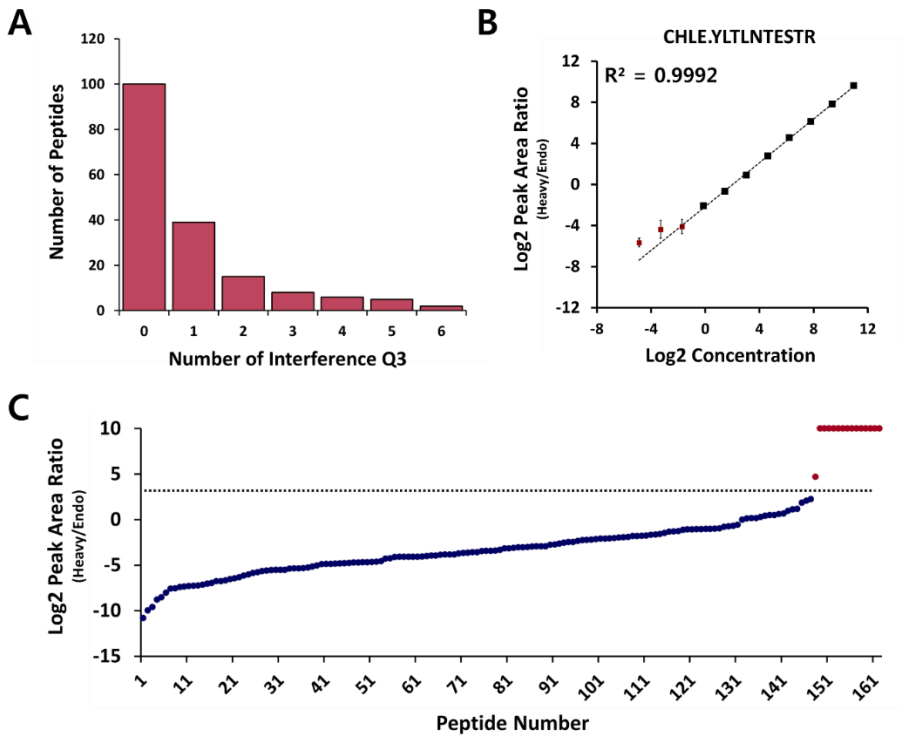


Figure 2-3. Selection of quantitative proteins/peptides by MRM assay

For selection of quantitative proteins/peptides, MRM assays were performed in pooled serum sample. (A) All peptides of the MCPs were considered with interference signal by AuDIT analysis. The peptides that have at least 3 transitions (Q3) were selected as first quantitative peptides. (B) Calibration curves were performed using each reference labeled peptide. Triplicate MRM assays were performed at 11 concentration points of each peptides. For example, calibration curve of "YLTLNTESTR" peptides of "BCHE" protein was showed. (C) Blue dots mean each protein that can be quantitate by MRM assays. Red dots mean each protein that cannot be quantitate by MRM assays.

Next, calibration curves were analyzed with series of diluted labeled reference peptide mixtures (in the range of approximate 0.01 – 2000 fmol/μL) in a pooled sample. For all peptides analysis per injection, we selected and analyzed 2 transitions that had best intensity per peptide from the AuDIT results. And then we performed MRM assay with technical triplicate for each concentration. Finally, the calibration curves were generated by linear regression analysis on the peak area ratio (reference/endogenous) versus spiked reference peptides concentration. Unfortunately, in this study, we used unpurified reference peptides. So, we determined only lower limit of quantitation (LLOQ) as minimum measurement (quantitation) level with linearity R^2 > 0.998 and 0.2 > coefficient of variation (CV). Therefore, we confirmed quantitative peptides with LLOQ less than 10 compared to endogenous peak area ratio in pooled sample. For example, "YLTLNTESTR" peptide of "BCHE" protein can be measured at about 1/5 lower level (0.23 in peak area ratio) compared to endogenous level in pooled sample (Figure 2-3B). According to the rules, we have 147 quantitative peptides from 89 MCPs (Figure 2-3C).

Feasible MCPs selection using pre-screening MRM and western blot

To confirm feasibility of 89 MCPs as prognosis prediction biomarkers after TACE, we performed pre-screening MRM using 10 patients selected blindly in each group with technical triplicate. From the analysis, we obtained the relative abundance of 89 MCPs in each samples. Statistical analysis of the relative abundance was performed using MSstats package in R. Significant differences (Fold change > 1.2 or < 0.83, and adjusted p value < 0.01) between good responders and poor responders were detected in 47 proteins; 24 proteins were highly expressed and 23 proteins were less expressed in poor responders (Table 2-3).

Table 2-3. Differentially expressed proteins from pre-screening MRM assay

N	Uniprot ID	Uniprot Accession	Gene Symbol	Log2 Fold Change	Adjust *p*-value	Standard Error
1	IPSP	P05154	SERPINA5	-1.65	< 0.005	0.019
2	CHLE	P06276	BCHE	-0.80	< 0.005	0.015
3	FCN3	O75636	FCN3	-0.53	< 0.005	0.029
4	FINC	P02751	FN1	-0.50	< 0.005	0.013
5	CPN2	P22792	CPN2	-0.49	< 0.005	0.125
6	APOA4	P06727	APOA4	-0.47	< 0.005	0.015
7	PON1	P27169	PON1	-0.46	< 0.005	0.014
8	IGHG1	P01857	IGHG1	-0.46	< 0.005	0.031
9	LCAT	P04180	LCAT	-0.45	< 0.005	0.022
10	PROZ	P22891	PROZ	-0.43	< 0.005	0.025
11	PGRP2	Q96PD5	PGLYRP2	-0.41	< 0.005	0.020
12	A2AP	P08697	SERPINF2	-0.40	< 0.005	0.016
13	CXCL7	P02775	PPBP	-0.40	< 0.005	0.017
14	KAIN	P29622	SERPINA4	-0.39	< 0.005	0.030
15	IBP3	P17936	IGFBP3	-0.36	< 0.005	0.025
16	APOC3	P02656	APOC3	-0.35	< 0.005	0.018
17	RET4	P02753	RBP4	-0.34	< 0.005	0.007
18	ALS	P35858	IGFALS	-0.33	< 0.005	0.013
19	FETUA	P02765	AHSG	-0.31	< 0.005	0.044
20	C1QB	P02746	C1QB	-0.31	< 0.005	0.029
21	KLKB1	P03952	KLKB1	-0.31	< 0.005	0.018

Continue.

22	APOA1	P02647	APOA1	-0.30	< 0.005	0.013
23	APOF	Q13790	APOF	-0.29	< 0.005	0.023
24	AACT	P01011	SERPINA3	0.28	< 0.005	0.020
25	ITIH4	Q14624	ITIH4	0.28	< 0.005	0.023
26	CO7	P10643	C7	0.32	< 0.005	0.019
27	CO5	P01031	C5	0.33	< 0.005	0.022
28	IC1	P05155	SERPING1	0.34	< 0.005	0.027
29	C4BPB	P20851	C4BPB	0.35	< 0.005	0.029
30	CO2	P06681	C2	0.36	< 0.005	0.015
31	IGJ	P01591	IGJ	0.39	< 0.005	0.023
32	APOE	P02649	APOE	0.40	< 0.005	0.015
33	LG3BP	Q08380	LGALS3BP	0.43	< 0.005	0.019
34	FETA	P02771	AFP	0.47	< 0.005	0.116
35	ITIH3	Q06033	ITIH3	0.49	< 0.005	0.019
36	C4BPA	P04003	C4BPA	0.51	< 0.005	0.019
37	CO4A	P0C0L4	C4A	0.54	< 0.005	0.033
38	SEPP1	P49908	SEPP1	0.63	< 0.005	0.023
39	FHR2	P36980	CFHR2	0.66	< 0.005	0.027
40	A2GL	P02750	LRG1	0.71	< 0.005	0.022
41	SAMP	P02743	APCS	0.77	< 0.005	0.023
42	LBP	P18428	LBP	0.83	< 0.005	0.017
43	FIBA	P02671	FGA	1.14	< 0.005	0.036
44	FIBG	P02679	FGG	1.19	< 0.005	0.041
45	FIBB	P02675	FGB	1.21	< 0.005	0.032
46	THBG	P05543	SERPINA7	1.38	< 0.005	0.138
47	CRP	P02741	CRP	1.96	< 0.005	0.034

For example, Inter-alpha-trypsin inhibitor heavy chain H4 (ITIH4), C-reactive protein (CRP), and AFP were highly expressed in poor responders. Conversely, Plasma serine protease inhibitor (SERPINA5), Cholinesterase (BCHE), and Alpha-2-antiplasmin (SERPINF2) significantly decreased in poor responders (Figure 2-4).

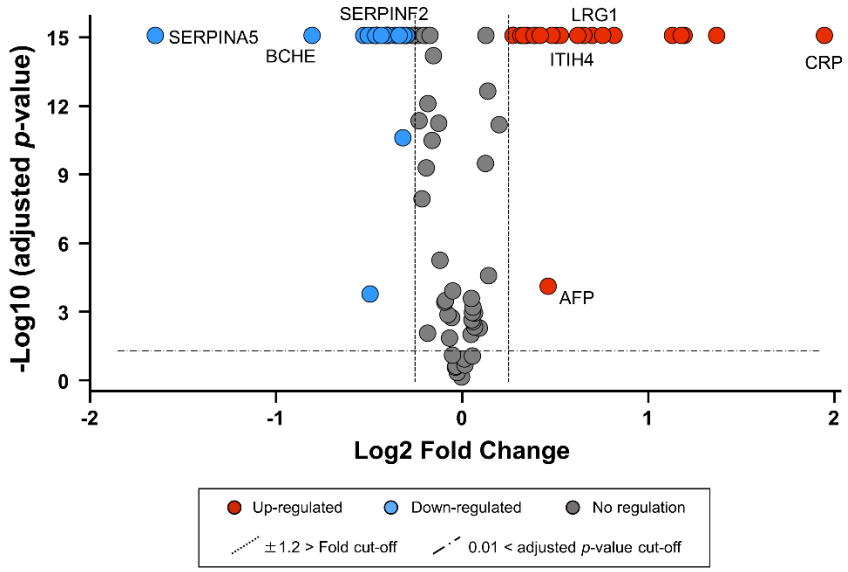


Figure 2-4. Quantification of MCPs by MSstats

Differential expression of MCPs in 20 HCC patients were calculated by MSstats. Log2 fold changes and the corresponding log10 adjusted *p*-values are summarized in a volcano plot. Significant proteins were considered by a fold change > ±1.2 and *p*-value < 0.01 and. Red dots mean up-regulation in poor responders and blue dots mean down-regulation in poor responders. Grey dots mean no regulation in both responders.

To verify our pre-screening MRM results, we performed antibody based western blot assay with 2 randomly selected proteins, ITIH4 and SERPINF2. Total 24 patients, 12 good responders and 12 poor responders, were randomly selected from training set cohorts. To normalize the variability between SDS-PAGE gels, we loaded 6 good responders and 6 poor responders per gel, and pooled sample was loaded on last lane of each gel as internal standard. As a result, ITIH4 protein showed significantly high expression in poor responders group (Figure 2-5A). In contrast, SERPINF2 protein showed low expression pattern in poor responders group (Figure 2-5B). These results were corresponded with pre-screening MRM results despite analysis using independent patients.

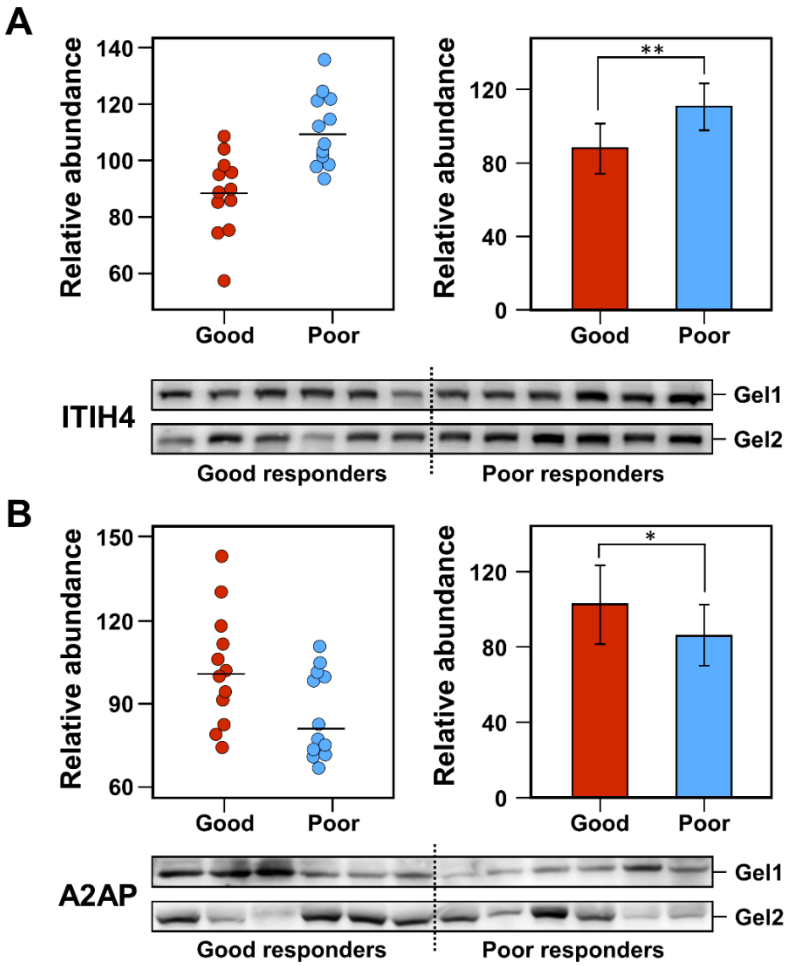


Figure 2-5. Validation by antibody based western blot

Random selected proteins were validated by western blot assay. (A) ITIH4 and (B) A2AP proteins were showed by dot plots and bar graphs. Red and blue dots mean protein abundance by western blots of each patients. Red and blue bar mean average protein abundance. P value was calculated by t-test. (* < *p*-value 0.05, ** < *p*-value 0.01)

The clinicopathologic characteristics to predict outcome in training set

Prior to MRM assay, we evaluated the correlations between outcome after TACE and the clinicopathologic characteristics of the training set of good responders (N=50) and poor responders (N=50) (Table 2-4). In univariate analysis, two clinicopathologic characteristics, number of lesions (OR=6.83, 95% CI=2.73 to 17.09) and concentration of PIVKA-II (OR=2.47, 95% CI=1.10 to 5.55), were significantly associated with outcome within 6 months after TACE. On the contrary, there were no significant association in these clinicopathologic characteristics with regard to albumin, prothrombin time, creatinine, platelet, ALT, bilirubin, and tumor size.

TABLE 2-4. Univariable analysis of clinical variables

Clinical variable	OR[a]	95% CI[b]	*p*-value
Albumin	0.73	0.3635 to 1.4727	0.38
Prothrombin time	0.18	0.0102 to 3.0339	0.22
Creatinine	1.71	0.4220 to 6.9491	0.44
Platelet (10^3/uL)	1.00	0.9907 to 1.0052	0.58
ALT, IU/L	1.01	0.9950 to 1.0303	0.14
Bilirubin, mg/dL	1.11	0.7381 to 1.6766	0.60
No. of lesions	6.83	2.7317 to 17.0935	P < 0.0001
Tumor size, cm	1.67	0.6818 to 4.0828	0.26
Pre-TACE AFP, ng/mL	1.91	0.8618 to 4.2198	0.11
Pre-TACE PIVKA-II mAU/mL	2.47	1.1003 to 5.5472	0.03

a)Odds ratio, estimated form logistic regression model.
b)Confidence interval of estimated OR

Also, we considered discriminant power of AFP and PIVKA-II, which are reported as early detection and prognosis markers, in training set for significant MCPs selection. In classification using ROC curve, AFP was shown an AUC of 0.60 and PIVKA-II was shown an AUC of 0.59 (Figure 2-6). Taken together, we selected MCPs that have more an AUC of 0.60 in the MRM assay.

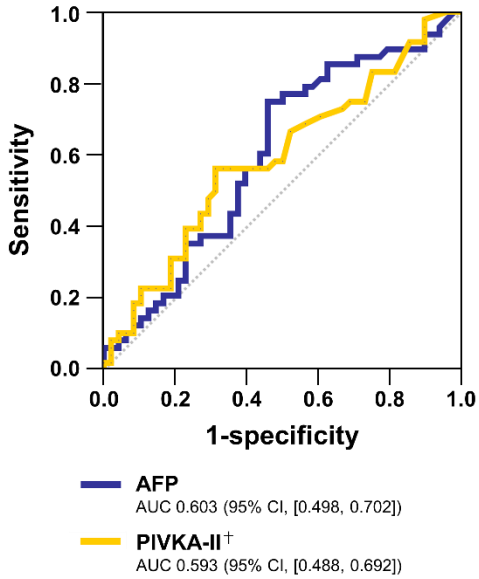


Figure 2-6. ROC curves of the level of AFP and PIVKA-II

Discrimination between good responders and poor responders in training sets. AUC values and 95% confidence interval (CI) were calculated by ROC curves. †Mark means that have 2 missing values in each group, respectively.

Combination for outcome prediction of 47 MCPs in training sets

To assess the prognostic potential of the 47 MCPs, we quantified in the training sets who are good (CR) or poor (PR, SD or PD) responders after TACE using MRM assay with labeled reference peptides. The relative protein abundance from MRM assay were calculated by MSstats linear mixed model with their multiple peptides, multiple transitions and two technical replicates (85).

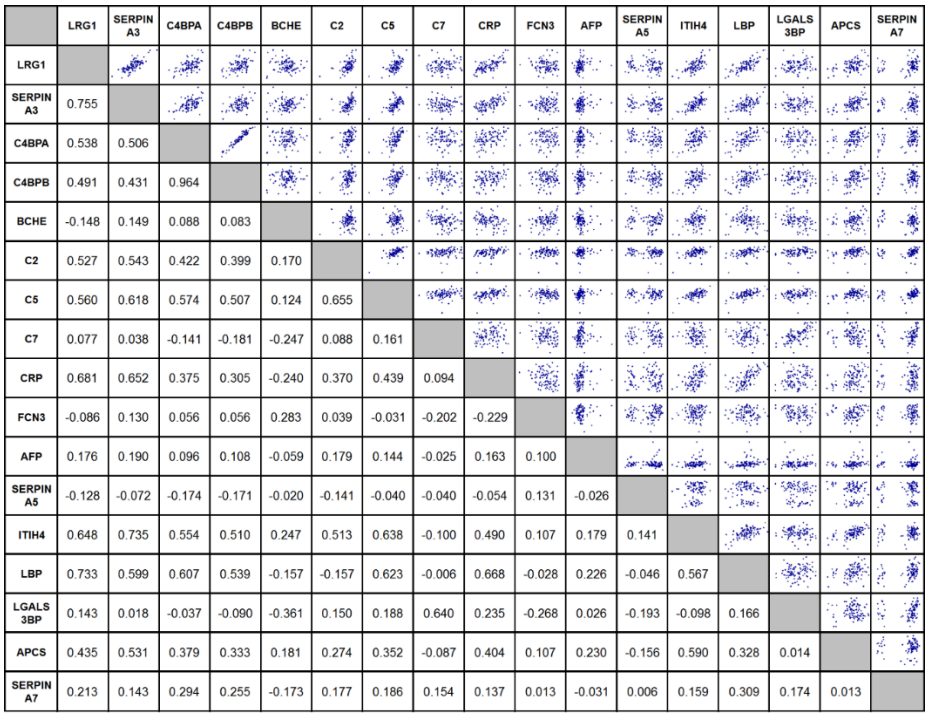
To suggest best-performing single marker, we performed a ROC analysis using the relative protein abundance of 47 MCPs. As a result, best-performing single marker proteins were LRG1 (AUC of 0.708) and C2 (AUC of 0.688). Also, we found that 17 proteins with AUC more than 0.60 were able to effectively discriminate poor responders from total patients with TACE (Table 2-5).

TABLE 2-5. Performance characteristics of the MCPs to predict prognosis after TACE

N	Uniprot ID	Gene Symbol	AUC[a]	95% CI[b]	P-value
1	A2GL	LRG1	0.708	0.702 to 0.713	3.4E-4
2	CO2	C2	0.688	0.682 to 0.693	1.2E-3
3	LBP	LBP	0.685	0.680 to 0.690	1.4E-3
4	C4BPA	C4BPA	0.685	0.680 to 0.690	1.4E-3
5	IPSP	SERPINA5	0.679	0.673 to 0.686	2E-3
6	AACT	SERPINA3	0.677	0.672 to 0.683	2.3E-3
7	CO5	C5	0.677	0.672 to 0.682	2.3E-3
8	C4BPB	C4BPB	0.665	0.660 to 0.670	4.4E-3
9	FCN3	FCN3	0.662	0.657 to 0.666	5.4E-3
10	SAMP	APCS	0.66	0.655 to 0.665	5.8E-3
11	CRP	CRP	0.656	0.652 to 0.660	7.2E-3
12	LG3BP	LGALS3BP	0.648	0.643 to 0.653	0.011
13	THBG	SERPINA7	0.645	0.641 to 0.650	0.012
14	CHLE	BCHE	0.636	0.631 to 0.642	0.019
15	CO7	C7	0.635	0.63 to 0.639	0.02
16	FETA	AFP	0.631	0.625 to 0.636	0.024
17	ITIH4	ITIH4	0.619	0.615 to 0.624	0.04

[a)]area under curve, estimated from ROC curve with 10-fold cross validation,
[b)]95% confidence interval, estimated from ROC curve with 10-fold cross validation

From logistic regression based multivariable analysis, the combination of 5 proteins, LRG1, APCS, BCHE, C7, and FCN3, showed that can discriminate more effective (AUC of 0.825) than single markers. Also, to keep redundancy of marker proteins that have similar abundance trend, we checked correlation coefficient (Figure 2-7). LRG1 were highly correlated ($r > 0.5$) with 7 proteins, SERPINA3, C4BPA, C2, C5, CRP, ITIH4, and LBP. However, our 5 proteins that used for combination panel showed low correlation coefficient, respectively.



	LRG1	SERPIN A3	C4BPA	C4BPB	BCHE	C2	C5	C7	CRP	FCN3	AFP	SERPIN A5	ITIH4	LBP	LGALS 3BP	APCS	SERPIN A7
LRG1																	
SERPIN A3	0.755																
C4BPA	0.538	0.506															
C4BPB	0.491	0.431	0.964														
BCHE	-0.148	0.149	0.088	0.083													
C2	0.527	0.543	0.422	0.399	0.170												
C5	0.560	0.618	0.574	0.507	0.124	0.655											
C7	0.077	0.038	-0.141	-0.181	-0.247	0.088	0.161										
CRP	0.681	0.652	0.375	0.305	-0.240	0.370	0.439	0.094									
FCN3	-0.086	0.130	0.056	0.056	0.283	0.039	-0.031	-0.202	-0.229								
AFP	0.176	0.190	0.096	0.108	-0.059	0.179	0.144	-0.025	0.163	0.100							
SERPIN A5	-0.128	-0.072	-0.174	-0.171	-0.020	-0.141	-0.040	-0.040	-0.054	0.131	-0.026						
ITIH4	0.648	0.735	0.554	0.510	0.247	0.513	0.638	-0.100	0.490	0.107	0.179	0.141					
LBP	0.733	0.599	0.607	0.539	-0.157	-0.157	0.623	-0.006	0.668	-0.028	0.226	-0.046	0.567				
LGALS 3BP	0.143	0.018	-0.037	-0.090	-0.361	0.150	0.188	0.640	0.235	-0.268	0.026	-0.193	-0.098	0.166			
APCS	0.435	0.531	0.379	0.333	0.181	0.274	0.352	-0.087	0.404	0.107	0.230	-0.156	0.590	0.328	0.014		
SERPIN A7	0.213	0.143	0.294	0.255	-0.173	0.177	0.186	0.154	0.137	0.013	-0.031	0.006	0.159	0.309	0.174	0.013	

Figure 2-7. Pearson's correlation coefficients between individual candidate and significant marker candidate proteins in training set

Correlation of coefficients of the proteins that have effective discriminant power (AUC > 0.6) were showed with correlation coefficient r and scatter plots.

Ensemble model analysis with protein markers and clinicopathologic characteristics

The MRM marker panel proteins, LRG1, APCS, BCHE, C7, and FCN3, were combined with the best-performing clinical variable panel, number of lesions, level of AFP, and level of PIVKA-II, using logistic regression modeling. Although level of AFP shown low significance in univariable analysis, we added in the panel because of having appropriate discriminant power regardless of multicollinearity with significant p value. Prior to combine, clinical variable panel were encoded as following; number of lesions = 0 if number <= 2, or 1 if number > 2; level of AFP 0 if level <= 20 ng/mL or 1 if level > 20 ng/mL; level of PIVKA-II = 0 if l <= 40 mAU/mL or 1 if level > 40 mAU/mL. Finally, the ensemble model with the MRM marker panel and clinical variable panel had an AUC of 0.881, whereas the MRM marker panel and clinical variable panel had the AUCs of 0.825 and 0.737. The ROC curves of the ensemble model and other panels are shown in figure 2-8.

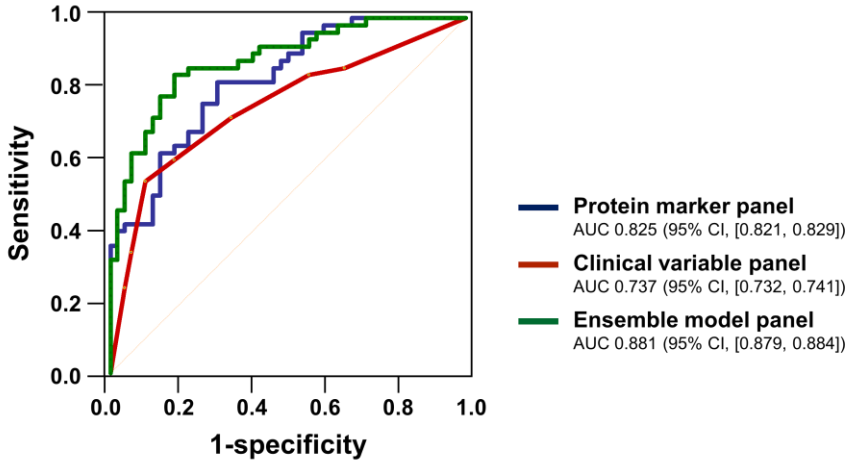
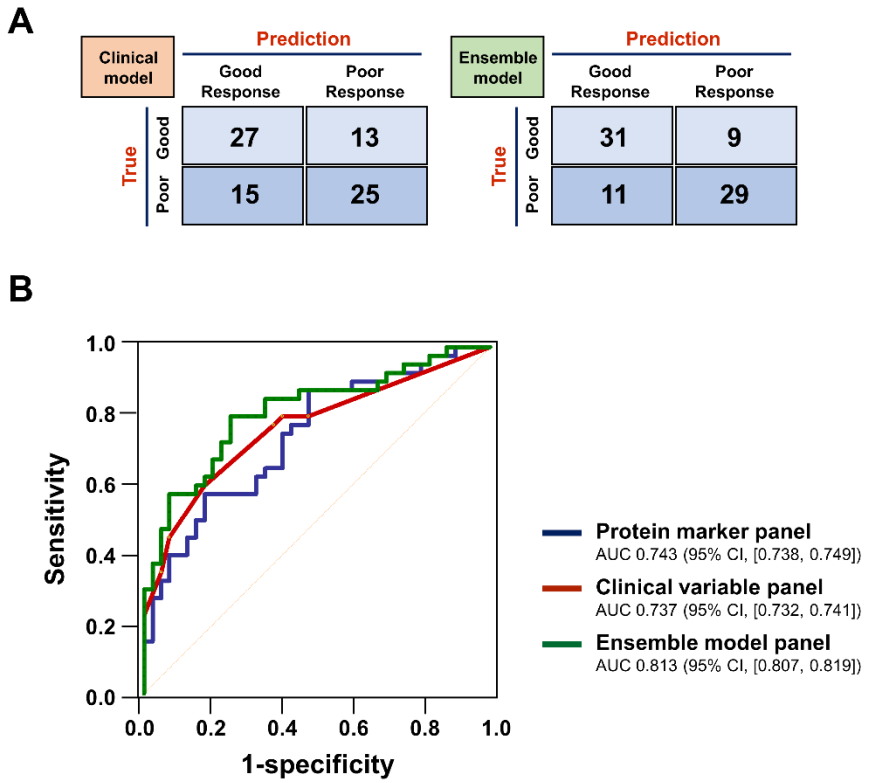


Figure 2-8. Performance characteristic of the best protein marker panel, clinical panel, and ensemble model panel to predict prognosis after TACE

Discrimination between good responders and poor responders in training sets. AUC values and 95% confidence interval (CI) were calculated by logistic regression model.

Model confirmation in validation set

To further evaluate the potential of the ensemble model identified in the training sets, we performed MRM assay in the validation set consisting of 40 good responders and 40 poor responders. From logistic regression modeling, ensemble model panel (3 proteins level and 3 clinical variable) showed that 31 of 40 good responders and 29 of 40 poor responders were correctly classified, whereas clinical model panel in training set showed that 27 of 40 good responders and 25 of 40 poor responders were correctly classified (Figure 2-9A). Also, the ensemble model panel was demonstrated an AUC of 0.813, similar to the training set. The ROC curves of the ensemble model in validation set are shown in figure 2-9B.



2-9. Comparison of the discriminatory power of the best single marker protein with ensemble model panel in validation cohort

(A) For comparisons between the clinical model panel and ensemble model panel, results are presented as confusion matrices. (B) AUC values and 95% confidence interval (CI) that were calculated by logistic regression model are represented with ROC curve.

Prognosis prediction power by ensemble model in TNM stages

To evaluate the prognosis prediction of our ensemble model panel by different Tumor Node Metastasis (TNM) stages (I, II, III, and IV), a total of 180 patients, including good responders (n=90) and poor responders (n=90) were segregated based upon TNM stage. The prediction scores of each patients were calculated from ensemble model equation. In each TNM stage, the prediction scores from ensemble model can significantly enhance the prognostic capability (Figure 2-10). Furthermore, the prediction scores showed no significant difference in good responders group without relevance to TNM stages. However, in poor responders, our prediction scores tended to increase a statistical significance in advanced stage.

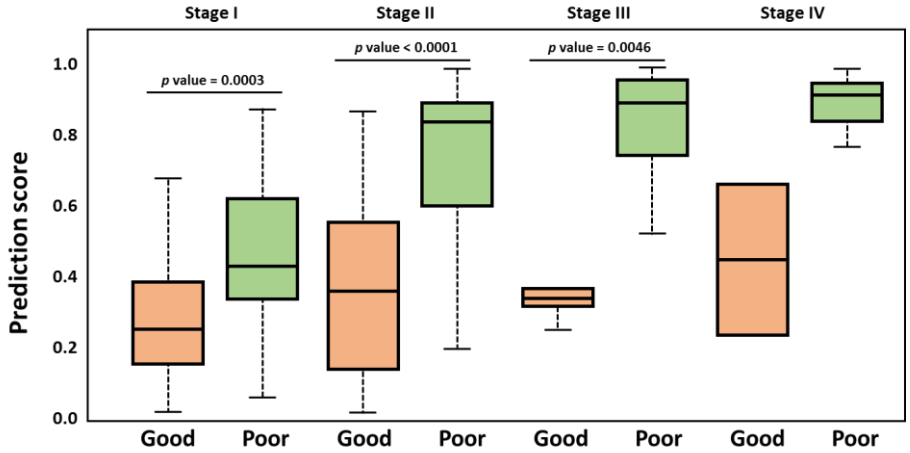


Figure 2-10. Prediction scores by TNM stages in 180 patient samples

Box plots represent prediction scores by logistic regression in 180 HCC patients. Boxes represent the interquartile range, and the horizontal line across each box indicates median values. Statistically significant differences were determined using the Mann-Whitney U test in each TNM stage. Also, statistically significant differences in each groups were determined using Kuskal-Wallis with post hoc Dunnett T3 test.

Longitudinal change in prognostic prediction marker

For observation of progression after TACE, the longitudinal cohort was composed of 100 patients at 2 time point (pre-TACE as baseline and between 6 to 12 months after TACE). We performed MRM assay with 47 MCPs in the longitudinal cohort. From linear mixed model analysis, longitudinal fold changes of the proteins were estimated. Among the proteins, we identified that 7 proteins showed significant longitudinal changes with the other side in each group (Figure 2-11).

As expected, the mean baseline of AFP in good responders was lower than that observed in poor responders group (adjusted p value < 0.001). However, within-person longitudinal change of AFP were no significant despite of a few increase/decrease in each group. We observed that CRP protein showed not only significant difference at baseline between two groups but also decreased longitudinally in good responders group. In addition, the mean baseline of CRAC1 protein were higher in good responders, and the protein was significantly increased in good responders after TACE. Interestingly, the mean baselines of APOF, APOC3, and BCHE in poor responders were lower than the mean baselines of good responders, and these proteins showed longitudinally decreased in poor responders group.

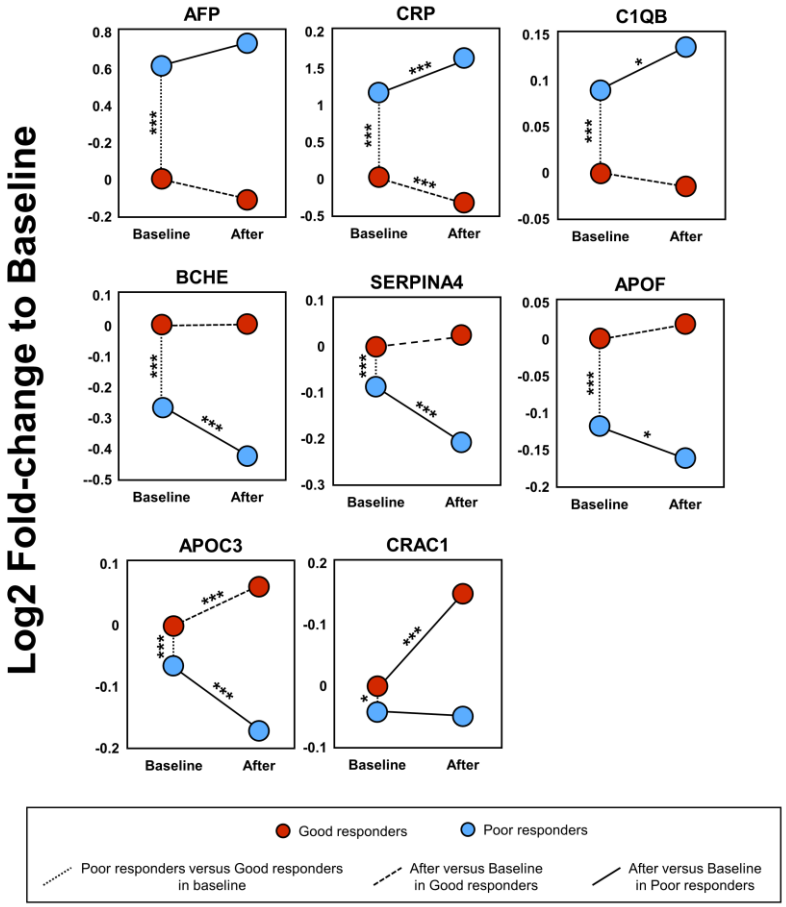


Figure 2-11. Evaluation of longitudinal changes of MCPs in good responders and poor responders

Relative fold change of selected proteins at baseline and after 6 to 12 month of each responder groups. Red and blue dots mean relative average abundance of good responders and poor responders. Linear mixed models by MSstats were used for calculation of significant fold changes. (* < adjusted p-value 0.01, ** < adjusted p-value 0.005, *** < adjusted p-value 0.001)

DISCUSSION

To the patients who cannot applicate curative treatment, such as surgical resection, local ablation, and liver transplantation, TACE may be an effective treatment option for improving survival. However, as TACE is palliative treatment, it needs repeated treatments every 3 to 6 month. Also, there are diverse outcomes after TACE in terms of treatment response and survival. Hence, prediction of outcomes before deciding on a TACE treatment is very important challenge.

In our previous study, we reported that HCC diagnosis markers, filamin-B (FNLB), and anillin (ANLN), were went back towards benign level range after HCC treatment (75). Also, typically over expressed protein in HCC state, CRP, was reported that showed different survival rate after TACE as baseline level of CRP (86). This suggests that HCC related proteins can be used in HCC prognosis prediction marker after treatment. Until a recent date, the most HCC prognosis marker studies were only performed by validation of discrimination power of AFP or PIVKA-II, which are reported diagnosis markers (69, 87). Although there was many marker candidates, it has limitation, because need to highly cost and effort for one by one validation without conviction. For overcoming this limitation, we performed the first study to identify new marker-candidate proteins (MCPs) from about 572 liver related proteins for prognosis/outcome prediction. Of the 572 MCPs, we could detect 89 quantitative proteins in serum using multi step MRM assay

without/with reference labeled peptides. First, 104 proteins were filtered by theoretical or experimental library from common dataset, and we checked detectability by mProphet analysis in pooled serum. Next, quantitative level of detected 104 MCPs were validated by their reference labeled peptides, and 89 MCPs can be measured in quantitative level. In the 89 quantitative proteins, 47 proteins showed significant difference expressions in small cohort set by linear mixed model analysis. Finally, we discovered five proteins marker panel (LRG1, APCS, BCHE, C7, and FCN3) from the training and the validation cohorts, and the panel can discriminate individuals who are versus are not good response after TACE.

In the proteins marker panel, cholinesterase (BCHE) was reported that appears to originate in the liver and is closely associated with the synthesis of serum albumin and coagulation factors (88). BCHE was also reported to reflect liver function in various clinical situations (89). In some liver disease conditions, such as severe chronic hepatitis, cirrhosis, and HCC states, BCHE was found to be very low and associated with increase mortality (90, 91). From out results, HCC patients who have low BCHE level tended to show poor response after TACE. Moreover, in our longitudinal study, we found that BCHE can confirm prognosis by change of up or down regulation. In poor responders group, BCHE level was significantly decreased compared with baseline level, whereas BCHE level of good responders group was not changed.

Importantly, clinicopathologic characteristic variables, level of AFP, level of PIVKA-II, and number of tumor lesions, were also identified as

significantly associated with prognosis in the multivariate analysis. In addition, these factors were reported about favorable performance in previous studies. However, the level of AFP and level of PIVKA-II were shown by an AUC of 0.603 and 0.593, respectively. Thus, we generated ensemble model with the proteins marker panel and clinicopathologic characteristic variables. So, our ensemble model panel be able to discriminate with high performance (an AUC of 0.881 in training cohorts and an AUC of 0.813 in validation cohorts).

In addition, our longitudinal study can support that some markers protein show progression state after treatment. The CRP protein showed longitudinally decrease in good responders after treatment. As mentioned, CRP protein was reported that over-expressed in HCC patients compared with healthy control. Our result showed that level of CRP was decrease in HCC patients who are recovered after treatment. Also, APOC3 and CRAC1 showed that can be used as progression marker after treatment in our longitudinal study. This is meaningful for simply trace of the progression without radiographic images.

Although our results is promising, there are several key limitations that should be acknowledged. As mentioned, we did not perform absolute quantitation assay. So our results can be depend on the instrument platform, sample preparation methods, and purity of reference peptides. In this study, we can suggest only marker panels, but cannot suggest final cut-off range for discriminant. Therefore, it required further absolute study with ELISA or stable isotope dilution MRM (SID-MRM) assay. Also, it required further external large validation with multicenter.

In conclusion, we discovered three new marker proteins that are associated with prognosis prediction after TACE in the first time. Also, we suggested that ensemble model (level of AFP, level of PIVKA-II, number of lesions, LRG1, APCS, BCHE, C7, and FCN3) can predict prognosis before TACE. Indeed, our results require more validation in large cohort and follow up study during long term. However, if validated, it ultimately can help as decision making guideline before TACE in future prospective studies.

REFERENCES

1. Zhang M, Zheng Y. [Analysis on the planning and developing population-based cancer registration in low- and middle-income settings]. Zhonghua liu xing bing xue za zhi = Zhonghua liuxingbingxue zazhi. 2014;35(9):1074.

2. Gupta S, Venkatesh A, Ray S, Srivastava S. Challenges and prospects for biomarker research: a current perspective from the developing world. Biochimica et biophysica acta. 2014;1844(5):899-908.

3. Cho WC. Proteomics in translational cancer research: biomarker discovery for clinical applications. Expert review of proteomics. 2014;11(2):131-3.

4. Anderson NL, Anderson NG. Proteome and proteomics: new technologies, new concepts, and new words. Electrophoresis. 1998;19(11):1853-61.

5. Mishra A, Verma M. Cancer biomarkers: are we ready for the prime time? Cancers. 2010;2(1):190-208.

6. Kang UB, Ahn Y, Lee JW, Kim YH, Kim J, Yu MH, et al. Differential profiling of breast cancer plasma proteome by isotope-coded affinity tagging method reveals biotinidase as a breast cancer biomarker. BMC cancer. 2010;10:114.

7. Kim Y, Han D, Min H, Jin J, Yi EC, Kim Y. Comparative proteomic profiling of pancreatic ductal adenocarcinoma cell lines. Molecules and cells.

2014;37(12):888-98.

8. Han D, Moon S, Kim Y, Min H, Kim Y. Characterization of the membrane proteome and N-glycoproteome in BV-2 mouse microglia by liquid chromatography-tandem mass spectrometry. BMC genomics. 2014;15:95.

9. Moon S, Han D, Kim Y, Jin J, Ho WK, Kim Y. Interactome analysis of AMP-activated protein kinase (AMPK)-alpha1 and -beta1 in INS-1 pancreatic beta-cells by affinity purification-mass spectrometry. Scientific reports. 2014;4:4376.

10. Parkin DM, Fernandez LM. Use of statistics to assess the global burden of breast cancer. The breast journal. 2006;12 Suppl 1:S70-80.

11. Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA: a cancer journal for clinicians. 2010;60(5):277-300.

12. Hoffman PC, Mauer AM, Vokes EE. Lung cancer. Lancet. 2000;355(9202):479-85.

13. Tan F, Jiang Y, Sun N, Chen Z, Lv Y, Shao K, et al. Identification of isocitrate dehydrogenase 1 as a potential diagnostic and prognostic biomarker for non-small cell lung cancer by proteomic analysis. Molecular & cellular proteomics : MCP. 2012;11(2):M111 008821.

14. Tian T, Hao J, Xu A, Luo C, Liu C, Huang L, et al. Determination of metastasis-associated proteins in non-small cell lung cancer by comparative proteomic analysis. Cancer science. 2007;98(8):1265-74.

15. Hwang SJ, Seol HJ, Park YM, Kim KH, Gorospe M, Nam DH, et al. MicroRNA-146a suppresses metastatic activity in brain metastasis. Molecules and cells. 2012;34(3):329-34.

16. Lopez-Otin C, Bond JS. Proteases: multifunctional enzymes in life and disease. The Journal of biological chemistry. 2008;283(45):30433-7.

17. Dawson TM, Dawson VL. Molecular pathways of neurodegeneration in Parkinson's disease. Science. 2003;302(5646):819-22.

18. Opferman JT, Korsmeyer SJ. Apoptosis in the development and maintenance of the immune system. Nature immunology. 2003;4(5):410-5.

19. Rao JS. Molecular mechanisms of glioma invasiveness: the role of proteases. Nature reviews Cancer. 2003;3(7):489-501.

20. Nisman B, Biran H, Heching N, Barak V, Ramu N, Nemirovsky I, et al. 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. British journal of cancer. 2008;98(1):77-9.

21. Kawakami T, Hoshida Y, Kanai F, Tanaka Y, Tateishi K, Ikenoue T, et al. Proteomic analysis of sera from hepatocellular carcinoma patients after radiofrequency ablation treatment. Proteomics. 2005;5(16):4287-95.

22. Streckfus C, Bigler L, Dellinger T, Pfeifer M, Rose A, Thigpen JT. CA 15-3 and c-erbB-2 presence in the saliva of women. Clinical oral investigations. 1999;3(3):138-43.

23. Fanayan S, Smith JT, Lee LY, Yan F, Snyder M, Hancock WS, et al. Proteogenomic Analysis of Human Colon Carcinoma Cell Lines LIM1215, LIM1899 and LIM2405. Journal of proteome research. 2013.

24. Xue H, Lu B, Zhang J, Wu M, Huang Q, Wu Q, et al. Identification of serum biomarkers for colorectal cancer metastasis using a differential

secretome approach. Journal of proteome research. 2010;9(1):545-55.

25. Xie X, Feng S, Vuong H, Liu Y, Goodison S, Lubman DM. A comparative phosphoproteomic analysis of a human tumor metastasis model using a label-free quantitative approach. Electrophoresis. 2010;31(11):1842-52.

26. Wang C, Guo K, Gao D, Kang X, Jiang K, Li Y, et al. Identification of transaldolase as a novel serum biomarker for hepatocellular carcinoma metastasis using xenografted mouse model and clinic samples. Cancer letters. 2011;313(2):154-66.

27. Brown JR, Hartley BS. Location of disulphide bridges by diagonal paper electrophoresis. The disulphide bridges of bovine chymotrypsinogen A. The Biochemical journal. 1966;101(1):214-28.

28. Enoksson M, Li J, Ivancic MM, Timmer JC, Wildfang E, Eroshkin A, et al. Identification of proteolytic cleavage sites by quantitative proteomics. Journal of proteome research. 2007;6(7):2850-8.

29. Gevaert K, Goethals M, Martens L, Van Damme J, Staes A, Thomas GR, et al. Exploring proteomes and analyzing protein processing by mass spectrometric identification of sorted N-terminal peptides. Nature biotechnology. 2003;21(5):566-9.

30. McDonald L, Beynon RJ. Positional proteomics: preparation of amino-terminal peptides as a strategy for proteome simplification and characterization. Nature protocols. 2006;1(4):1790-8.

31. Anisowicz A, Huang H, Braunschweiger KI, Liu Z, Giese H, Wang H, et al. A high-throughput and sensitive method to measure global DNA

methylation: application in lung cancer. BMC cancer. 2008;8:222.

32. Wisniewski JR, Zougman A, Nagaraj N, Mann M. Universal sample preparation method for proteome analysis. Nature methods. 2009;6(5):359-62.

33. Han D, Moon S, Kim Y, Ho WK, Kim K, Kang Y, et al. Comprehensive phosphoproteome analysis of INS-1 pancreatic beta-cells using various digestion strategies coupled with liquid chromatography-tandem mass spectrometry. Journal of proteome research. 2012;11(4):2206-23.

34. Zybailov B, Mosley AL, Sardiu ME, Coleman MK, Florens L, Washburn MP. Statistical analysis of membrane proteome expression changes in Saccharomyces cerevisiae. Journal of proteome research. 2006;5(9):2339-47.

35. Kim SJ, Jin J, Kim YJ, Kim Y, Yu HG. Retinal proteome analysis in a mouse model of oxygen-induced retinopathy. Journal of proteome research. 2012;11(11):5186-203.

36. Weijers RN. Amino acid sequence in bovine serum albumin. Clinical chemistry. 1977;23(7):1361-2.

37. Liu Y, Sun W, Zhang K, Zheng H, Ma Y, Lin D, et al. Identification of genes differentially expressed in human primary lung squamous cell carcinoma. Lung Cancer. 2007;56(3):307-17.

38. Domon B, Aebersold R. Mass spectrometry and protein analysis. Science. 2006;312(5771):212-7.

39. Liu H, Sadygov RG, Yates JR, 3rd. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. Analytical chemistry. 2004;76(14):4193-201.

40. Freund DM, Prenni JE. Improved detection of quantitative differences using a combination of spectral counting and MS/MS total ion current. Journal of proteome research. 2013;12(4):1996-2004.

41. Old WM, Meyer-Arendt K, Aveline-Wolf L, Pierce KG, Mendoza A, Sevinsky JR, et al. Comparison of label-free methods for quantifying human proteins by shotgun proteomics. Molecular & cellular proteomics : MCP. 2005;4(10):1487-502.

42. Ghosh D, Yu H, Tan XF, Lim TK, Zubaidah RM, Tan HT, et al. Identification of key players for colorectal cancer metastasis by iTRAQ quantitative proteomics profiling of isogenic SW480 and SW620 cell lines. Journal of proteome research. 2011;10(10):4373-87.

43. Willis ND, Cox TR, Rahman-Casans SF, Smits K, Przyborski SA, van den Brandt P, et al. Lamin A/C is a risk biomarker in colorectal cancer. PloS one. 2008;3(8):e2988.

44. Constantinescu D, Gray HL, Sammak PJ, Schatten GP, Csoka AB. Lamin A/C expression is a marker of mouse and human embryonic stem cell differentiation. Stem Cells. 2006;24(1):177-85.

45. Srivastava M, Bubendorf L, Nolan L, Glasman M, Leighton X, Miller G, et al. ANX7 as a bio-marker in prostate and breast cancer progression. Disease markers. 2001;17(2):115-20.

46. Pujol JL, Grenier J, Daures JP, Daver A, Pujol H, Michel FB. Serum fragment of cytokeratin subunit 19 measured by CYFRA 21-1 immunoradiometric assay as a marker of lung cancer. Cancer research. 1993;53(1):61-6.

47. Streckfus C, Bigler L, Tucci M, Thigpen JT. 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 investigation. 2000;18(2):101-9.

48. Hsu JL, Huang SY, Chow NH, Chen SH. Stable-isotope dimethyl labeling for quantitative proteomics. Analytical chemistry. 2003;75(24):6843-52.

49. Prudova A, auf dem Keller U, Butler GS, Overall CM. Multiplex N-terminome analysis of MMP-2 and MMP-9 substrate degradomes by iTRAQ-TAILS quantitative proteomics. Molecular & cellular proteomics : MCP. 2010;9(5):894-911.

50. Dugaiczyk A, Law SW, Dennison OE. Nucleotide sequence and the encoded amino acids of human serum albumin mRNA. Proceedings of the National Academy of Sciences of the United States of America. 1982;79(1):71-5.

51. Apweiler R, Bairoch A, Wu CH, Barker WC, Boeckmann B, Ferro S, et al. UniProt: the Universal Protein knowledgebase. Nucleic acids research. 2004;32(Database issue):D115-9.

52. Brannan JM, Sen B, Saigal B, Prudkin L, Behrens C, Solis L, et al. EphA2 in the early pathogenesis and progression of non-small cell lung cancer. Cancer Prev Res (Phila). 2009;2(12):1039-49.

53. Bauer KM, Lambert PA, Hummon AB. Comparative label-free LC-MS/MS analysis of colorectal adenocarcinoma and metastatic cells treated with 5-fluorouracil. Proteomics. 2012;12(12):1928-37.

54. Li F, Glinskii OV, Zhou J, Wilson LS, Barnes S, Anthony DC, et al. Identification and analysis of signaling networks potentially involved in breast carcinoma metastasis to the brain. PloS one. 2011;6(7):e21977.

55. Yoon SY, Kim JM, Oh JH, Jeon YJ, Lee DS, Kim JH, et al. Gene expression profiling of human HBV- and/or HCV-associated hepatocellular carcinoma cells using expressed sequence tags. International journal of oncology. 2006;29(2):315-27.

56. Yoshimura K, Meckel KF, Laird LS, Chia CY, Park JJ, Olino KL, et al. Integrin alpha2 mediates selective metastasis to the liver. Cancer research. 2009;69(18):7320-8.

57. Shibue T, Weinberg RA. Integrin beta1-focal adhesion kinase signaling directs the proliferation of metastatic cancer cells disseminated in the lungs. Proceedings of the National Academy of Sciences of the United States of America. 2009;106(25):10290-5.

58. Wu WS, Wu JR, Hu CT. Signal cross talks for sustained MAPK activation and cell migration: the potential role of reactive oxygen species. Cancer metastasis reviews. 2008;27(2):303-14.

59. Obchoei S, Weakley SM, Wongkham S, Wongkham C, Sawanyawisuth K, Yao Q, et al. Cyclophilin A enhances cell proliferation and tumor growth of liver fluke-associated cholangiocarcinoma. Molecular cancer. 2011;10:102.

60. Pratilas CA, Hanrahan AJ, Halilovic E, Persaud Y, Soh J, Chitale D, et al. Genetic predictors of MEK dependence in non-small cell lung cancer. Cancer research. 2008;68(22):9375-83.

61. Storr SJ, Carragher NO, Frame MC, Parr T, Martin SG. The calpain system and cancer. Nature reviews Cancer. 2011;11(5):364-74.

62. Kim YH, Kwei KA, Girard L, Salari K, Kao J, Pacyna-Gengelbach M, et al. Genomic and functional analysis identifies CRKL as an oncogene amplified in lung cancer. Oncogene. 2010;29(10):1421-30.

63. Parkin DM, Bray F, Ferlay J, Pisani P. Global cancer statistics, 2002. CA: a cancer journal for clinicians. 2005;55(2):74-108.

64. Bruix J, Sherman M, Practice Guidelines Committee AAftSoLD. Management of hepatocellular carcinoma. Hepatology. 2005;42(5):1208-36.

65. Llovet JM, Real MI, Montana X, Planas R, Coll S, Aponte J, et al. Arterial embolisation or chemoembolisation versus symptomatic treatment in patients with unresectable hepatocellular carcinoma: a randomised controlled trial. Lancet. 2002;359(9319):1734-9.

66. Park JW, Sherman M, Colombo M, Roberts LR, Schwartz ME, Degos F, et al. Observations of hepatocellular carcinoma (HCC) management patterns from the global HCC bridge study: First characterization of the full study population. J Clin Oncol. 2012;30(15).

67. Kim DY, Ryu HJ, Choi JY, Park JY, Lee DY, Kim BK, et al. Radiological response predicts survival following transarterial chemoembolisation in patients with unresectable hepatocellular carcinoma. Alimentary pharmacology & therapeutics. 2012;35(11):1343-50.

68. Georgiades C, Geschwind JF, Harrison N, Hines-Peralta A, Liapi E, Hong K, et al. Lack of response after initial chemoembolization for hepatocellular carcinoma: does it predict failure of subsequent treatment?

Radiology. 2012;265(1):115-23.

69. Wang Y, Chen Y, Ge N, Zhang L, Xie X, Zhang J, et al. Prognostic significance of alpha-fetoprotein status in the outcome of hepatocellular carcinoma after treatment of transarterial chemoembolization. Annals of surgical oncology. 2012;19(11):3540-6.

70. Park H, Park JY. Clinical significance of AFP and PIVKA-II responses for monitoring treatment outcomes and predicting prognosis in patients with hepatocellular carcinoma. BioMed research international. 2013;2013:310427.

71. Kitteringham NR, Jenkins RE, Lane CS, Elliott VL, Park BK. Multiple reaction monitoring for quantitative biomarker analysis in proteomics and metabolomics. Journal of chromatography B, Analytical technologies in the biomedical and life sciences. 2009;877(13):1229-39.

72. Domanski D, Percy AJ, Yang J, Chambers AG, Hill JS, Freue GV, et al. MRM-based multiplexed quantitation of 67 putative cardiovascular disease biomarkers in human plasma. Proteomics. 2012;12(8):1222-43.

73. Kennedy JJ, Abbatiello SE, Kim K, Yan P, Whiteaker JR, Lin C, et al. Demonstrating the feasibility of large-scale development of standardized assays to quantify human proteins. Nature methods. 2014;11(2):149-55.

74. Surinova S, Huttenhain R, Chang CY, Espona L, Vitek O, Aebersold R. Automated selected reaction monitoring data analysis workflow for large-scale targeted proteomic studies. Nature protocols. 2013;8(8):1602-19.

75. Kim H, Kim K, Yu SJ, Jang ES, Yu J, Cho G, et al. Development of biomarkers for screening hepatocellular carcinoma using global data mining

and multiple reaction monitoring. PloS one. 2013;8(5):e63468.

76. Zhang Y, Yang C, Wang S, Chen T, Li M, Wang X, et al. LiverAtlas: a unique integrated knowledge database for systems-level research of liver and hepatic disease. Liver international : official journal of the International Association for the Study of the Liver. 2013;33(8):1239-48.

77. Chung JW, Kim HC, Yoon JH, Lee HS, Jae HJ, Lee W, et al. Transcatheter arterial chemoembolization of hepatocellular carcinoma: prevalence and causative factors of extrahepatic collateral arteries in 479 patients. Korean journal of radiology. 2006;7(4):257-66.

78. Yu SJ, Lee JH, Jang ES, Cho EJ, Kwak MS, Yoon JH, et al. Hepatocellular carcinoma: high hepatitis B viral load and mortality in patients treated with transarterial chemoembolization. Radiology. 2013;267(2):638-47.

79. Lencioni R, Llovet JM. Modified RECIST (mRECIST) assessment for hepatocellular carcinoma. Seminars in liver disease. 2010;30(1):52-60.

80. Kim K, Yu J, Min H, Kim H, Kim B, Yu HG, et al. Online monitoring of immunoaffinity-based depletion of high-abundance blood proteins by UV spectrophotometry using enhanced green fluorescence protein and FITC-labeled human serum albumin. Proteome science. 2010;8:62.

81. Choi M, Chang CY, Clough T, Broudy D, Killeen T, MacLean B, et al. MSstats: an R package for statistical analysis of quantitative mass spectrometry-based proteomic experiments. Bioinformatics. 2014;30(17):2524-6.

82. Jeong SK, Na K, Kim KY, Kim H, Paik YK. PanelComposer: a web-based panel construction tool for multivariate analysis of disease biomarker

candidates. Journal of proteome research. 2012;11(12):6277-81.

83. Abbatiello SE, Mani DR, Keshishian H, Carr SA. Automated detection of inaccurate and imprecise transitions in peptide quantification by multiple reaction monitoring mass spectrometry. Clinical chemistry. 2010;56(2):291-305.

84. Reiter L, Rinner O, Picotti P, Huttenhain R, Beck M, Brusniak MY, et al. mProphet: automated data processing and statistical validation for large-scale SRM experiments. Nature methods. 2011;8(5):430-5.

85. Cerciello F, Choi M, Nicastri A, Bausch-Fluck D, Ziegler A, Vitek O, et al. Identification of a seven glycopeptide signature for malignant pleural mesothelioma in human serum by selected reaction monitoring. Clinical proteomics. 2013;10(1):16.

86. Sieghart W, Pinter M, Hucke F, Graziadei I, Schoniger-Hekele M, Muller C, et al. Single determination of C-reactive protein at the time of diagnosis predicts long-term outcome of patients with hepatocellular carcinoma. Hepatology. 2013;57(6):2224-34.

87. Pote N, Cauchy F, Albuquerque M, Voitot H, Belghiti J, Castera L, et al. Performance of PIVKA-II for early hepatocellular carcinoma diagnosis and prediction of microvascular invasion. Journal of hepatology. 2015;62(4):848-54.

88. McQueen MJ. Clinical and analytical considerations in the utilization of cholinesterase measurements. Clinica chimica acta; international journal of clinical chemistry. 1995;237(1-2):91-105.

89. Vorhaus LJ, Scudamore HH, Kark RM. Measurement of serum

cholinesterase activity in the study of diseases of the liver and biliary system. Gastroenterology. 1950;15(2):304-15.

90. Donadon M, Cimino M, Procopio F, Morenghi E, Montorsi M, Torzilli G. Potential role of cholinesterases to predict short-term outcome after hepatic resection for hepatocellular carcinoma. Updates in surgery. 2013;65(1):11-8.

91. Ohashi N, Tsuji N, Naito Y, Iwakura T, Isobe S, Ono M, et al. Relationship between urinary fractional excretion of sodium and life prognosis in liver cirrhosis patients. Hepatology research : the official journal of the Japan Society of Hepatology. 2013;43(11):1156-62.

ABSTRACT IN KOREAN

국 문 초 록

서론: 암은 전세계적으로 가장 많은 사망원인 중 하나이다. 이러한 암에 의한 사망 중 주요 요인은 초기 단계에서의 발견이 어렵기 때문이다. 이러한 암으로부터의 위협에 대처하기 위해 암 발생과정에 대한 이해 및 조기 발견과 치료 효과를 모니터링 하기 위한 방법이 필요 되고 있다. 프로테오믹스 기술이 발전함에 따라 이러한 표지자 단백질 발굴에 많은 도움을 주고 있으며, 최근에는 표지자 발굴뿐만 아니라 암 전이 메커니즘 연구에도 활발히 사용되고 있다.

방법: 1 장에서 전이에 관련된 단백질 변화를 관측하기 위하여 암 전이가 발생한 폐암 세포 (NCI-H1755)를 사용하였다. 이에 대조군으로 폐암 세포이며 전이가 발생하지 않은 세포 (NCI-H1703)를 사용하였다. 두 세포주의 단백질 발현 량 비교를 위하여 label-free 정량 분석을 시행하였다. 또한 세포 내에 비정상적으로 잘려진 단백질 파편을 찾기 위하여 N 말단 분석기법을 개발하였다. 2 장에서는 치료예후마커 발굴을 위한 데이터베이스 기반 마커 후보군을 선정하였다. 이를 기반으로 다중검지법을 적용하여 180 명의 간암환자에 대하여 마커후보군에 대한 정량분석을 시행하였다.

결과: 1 장에서는 질량분석기를 사용하여 총 2130 개의 단백질을 발견하였으며, 그 중에서 1355 개 단백질이 두 종류 세포에서 공통적으로 발견되었다. Label-free 정량 분석 기법에 의해 242 개의 단백질이 두 세포에서 유의적인 차이를 보이며 발현되는 것을 확인하였다. 또한 N 말단 분석기법을 통하여 325 개의 단백질 파편을 발견했으며, 45 개의 알려지지 않은 단백질 파편을 발견할 수 있었다. 위의 두 가지 실험 기법을 바탕으로 11 개의 정량 분석된 단백질과 8 개의 단백질 파편이 focal adhesion pathway 에 직접적으로 관련이 있음을 발견하였다. 2 장에서는 화학색전술을 받은 20 명의 간암환자에 대하여 47 개 단백질이 치료 예후 (6 개월동안 병소가 없는 상태가 유지된 그룹 또는 그렇지 못한 그룹)에 따라 유의적으로 차이를 보인 것을 확인하였다. 이를 기반으로 190 명의 환자에 적용하여 정량분석을 시행하였으며, 최종적으로 17 개의 단백질이 치료예후를 구분하는데 사용 가능함을 확인하였다. 이 중에서 5 개의 단백질 (LRG1, APCS, BCHE, C7, FCN3)과 3 개의 임상 정보 (AFP 수치, PIVKA-II 수치, 간암 병소 개수)를 조합한 다중 마커패널이 AUC 0.8 이상으로 구분력이 있음을 확인하였다.

결론: 1 장에서는 Label-free 정량 기법 및 N-말단 분석기법의 개발을 통하여 폐암 전이에 focal adhesion pathway 관련 단백질의 발현차이가 전이에 직접 또는 간접적으로 영향을 줄 수 있다는 것을 확인하였다. 이러한 기존의 프로테옴 분석뿐만 아니라 새로운 개념의 분석 방법을

사용한 단백질의 발현 정량 분석 및 단백질의 파편조각의 발견은 암 메커니즘 이해에 많은 도움을 줄 것으로 생각된다. 2 장에서는 환자 맞춤 치료 방법의 적용을 위한 다중마커패널을 개발하였다. 우리의 다중마커패널은 간암환자의 치료방법 선택에 있어서 좀 더 효과적으로 접근할 수 있는 가이드라인이 될 것이다. 따라서 이러한 프로테오믹스 연구 기법들은 암의 이해 및 치료 등에 사용 될 수 있을 것이다.

주요어 : 폐암, N 말단 분석, 암 전이, 프로테오믹스, 정량 분석, 다중검지법, 간암, 화학색전술, 치료예후마커

학 번 : 2008-21997

*본 내용은 Molecules and Cells 학술지에 출판 완료된 내용임