

Anticancer drug clustering based on proteomic profiles and a sensitivity database in a lung cancer cell line panel

MITSUNORI HINO¹, KUNIKO MATSUDA¹, AKIHIKO MIYANAGA¹, HIDEHIKO KURIBAYASHI¹,
HIDEAKI MIZUTANI¹, RINTARO NORO¹, YUJI MINEGISHI¹, TETSUYA OKANO¹,
MASAHIRO SEIKE¹, AKIKO KAWAKAMI², AKINOBU YOSHIMURA¹, NAOKI OGAWA²,
HARUKA UESAKA³, SHOJI KUDOH¹ and AKIHIKO GEMMA¹

¹Division of Pulmonary Medicine, Infectious Diseases and Oncology, Department of Internal Medicine,
Nippon Medical School, Bunkyo-ku, Tokyo 113-8602; ²Genetic Lab Co., Ltd., Sapporo,
Hokkaido 001-0027; ³MediBic, Chiyoda-ku, Tokyo 100-0013, Japan

Received July 29, 2009; Accepted October 6, 2009

DOI: 10.3892/etm_00000007

Abstract. Previously, we performed a molecular pharmacological study that applied a combination of DNA microarray-based gene expression profiling and drug sensitivity tests *in vitro* with a view to designing an improved chemotherapeutic strategy for advanced lung cancer. Utilizing recent key technological advances in proteomics, particularly antibody array-based methodologies, the current study aimed to examine the benefit of protein expression profiling in an analogous molecular pharmacological context. We performed protein expression analysis in a panel of lung cancer cell lines via an antibody array approach. Using a modified NCI program, we related cell line-specific proteomic profiles to the previously determined cytotoxic activity of a selection of commonly used anticancer agents, namely docetaxel, paclitaxel, gemcitabine, vinorelbine, 5-fluorouracil (5-FU), SN38, cisplatin (CDDP) and carboplatin (CBDCA). In addition, we compared these results with those obtained from our prior DNA microarray-based transcriptomic study. In our expression-drug correlation analysis using antibody array, gemcitabine consistently belonged to an isolated cluster. Docetaxel, paclitaxel, 5-FU, SN38, CBDCA and CDDP were gathered together into one large cluster. These results coincided with those generated by the prior transcriptomic study. Various genes were commonly listed that differentiated gemcitabine from the others. The identified factors associated with drug sensitivities were different between both analyses. Our proteomic profiling data provided confirmation of the

previous transcript expression-drug sensitivity correlation analysis. These results suggest that chemotherapy regimens that include gemcitabine should be evaluated in second-line chemotherapy in cases where the first-line chemotherapy did not include this drug. Protein expression-drug sensitivity correlations in lung cancer cells *in vitro* may provide useful information in determining the most appropriate therapeutic options for lung cancer patients.

Introduction

There are many anticancer agents being used in the treatment of lung cancer, such as cisplatin (CDDP), carboplatin (CBDCA), docetaxel, paclitaxel, vinorelbine, gemcitabine, 5-fluorouracil (5-FU) and CPT-11. A number of combination therapy regimens employing platinum compounds have proven to be effective (1) and are widely applied to initial treatment for unresected non-small cell lung cancer (NSCLC) (2). In addition, docetaxel and pemetrexed have been reported to be effective in the context of second-line chemotherapy for NSCLC (3,4). However, at present, the effect of these therapies on improving patient survival remains far from satisfactory (1-3). Therefore, it is desirable to find more appropriate therapeutic opportunities based on informed insight. Previously, we performed a gene expression profile analysis of a series of lung cancer cell lines using both cDNA filter and high-density oligonucleotide arrays (5). In parallel, we also examined the sensitivity of these cell lines to commonly used anticancer agents (docetaxel, paclitaxel, gemcitabine, vinorelbine, 5-FU, SN38, CDDP and CBDCA) *in vitro* via an MTT assay (5). We related the cytotoxic activity of each of these agents to the corresponding transcriptomic expression profile in each of the cell lines using a modified NCI program to determine an improved chemotherapeutic strategy for advanced lung cancer (5). In this prior mRNA expression-drug correlation analysis, gemcitabine consistently belonged to an isolated cluster in a reproducible manner (5).

Recent technological advances in protein analysis, such as antibody array approaches, allow for large-scale analysis

Correspondence to: Dr Akihiko Gemma, Department of Internal Medicine, Division of Pulmonary Medicine, Infectious Diseases and Oncology, Nippon Medical School, 1-1-5 Sendagi, Bunkyo-ku, Tokyo 113-8602, Japan
E-mail: agemma@nms.ac.jp

Key words: antibody array, cDNA array, anticancer drug, drug sensitivity test, clustered image map, lung cancer, chemotherapy

of protein expression (6-8). We related the known cytotoxic activity of the above-mentioned anticancer agents to corresponding proteomic profiling patterns, as determined using an antibody array approach, across our lung cancer cell line panel. In addition, we compared transcriptomic and proteomic data from the same set of lung cancer cell lines. The essential aim of this study was to aid in the selection of appropriate drug combinations for the treatment of lung cancer.

Materials and methods

Cell lines. The following panel of cell lines was used: PC9, PC7, PC14, A549, LK-2, RERF-LC-KJ, RERF-LC-MS, RERF-LC-AI, PC1, PC3, PC10, ABC-1, EBC-1 and QG56. The PC1, PC3, PC7, PC9, PC10, PC14 and QG56 cell lines were obtained from IBL (Gumma, Japan). The A549 cell line was obtained from the American Type Culture Collection (Rockville, MD, USA) (9). The LK-2 cell line was obtained from the Health Science Research Resources Bank (Osaka, Japan), from which RERF-LC-MS, EBC-1 and ABC-1 cell lines were also purchased. PC1, PC3 and PC10 cell lines were provided by S. Hirohashi (National Cancer Center Research Institute, Tokyo, Japan). RERF-LC-KJ and RERF-LC-AI cell lines were obtained from the RIKEN Cell Bank (Ibaraki, Japan). PC7, PC9, PC14, A549, RERF-LC-KJ, RERF-LC-MS, PC3 and ABC-1 are adenocarcinoma cell lines. LK-2, RERF-LC-AI, PC1, PC10, EBC-1 and QG56 are squamous cell cancer cell lines.

Antibody microarray analysis. Antibody microarray assays were performed using an Antibody Microarray system (cat #AA-001; Lab Vision Co., CA, USA) according to the manufacturer's instructions. In brief, whole cell extracts were prepared and protein samples biotinylated using appropriate buffers from the kit. Antibody microarray slides (containing 722 antibodies in triplicate) were initially blocked by soaking in blocking buffer. Hybridization buffer was added to biotin-labeled protein samples. Antibody microarray slides were set in a humidity chamber, coverslips were slowly placed onto the slides and the samples were applied to the microarray surface. After incubation for 2 h, the slides were washed three times with washing buffer from the kit. Following this, the slides were applied to a second detection chamber and incubated with Cy3-tagged detection antibody for 45 min at room temperature. The slides were washed three times with washing buffer. The slides were imaged with a GenePix 4000B microarray scanner (Axon Instruments/Molecular Devices, ON, Canada), with protein expression levels quantified by measuring the intensity of the signals using Array-Pro Analyzer (Media Cybernetics, MD, USA). Signal intensities were normalized by comparison with the average values of all probes. All the experiments were repeated at least twice.

Data analysis. Data analysis for the correlation coefficients that related the drug activity patterns to protein expression profiles was principally performed by a modified NCI program (10). The symbol (A) (GI_{50}) refers to the drug activity matrix in which the rows represent the anticancer drugs and the columns represent the human lung cancer cell lines. The symbol (T) (protein expression) refers to the protein expression matrix in

which the rows represent individual protein and the columns represent the cell lines. In order to analyze the relationship between protein expression and drug activity, we generated a protein-drug correlation matrix (AT) (correlation coefficient) in which the rows represent the proteins and the columns represent the drugs. Firstly, we subtracted its mean value from the matrix (A) in the direction of the row and columns for a pre-treatment. Secondly, we normalized each element in the matrix (A) by subtracting its row-wise mean and dividing by its row-wise standard deviation; normalized (T) was generated in a similar way. Finally, we took the inner product of the matrix (A) and the transpose of the matrix (T). The resulting matrix (AT) implied the Pearson correlation coefficients that reflected the relationship between drug activity and protein expression.

The Pearson correlation coefficient r is given by the following formula:

$$r = \frac{1}{n-1} \sum_{k=1}^n \left(\frac{A_k - \bar{A}}{\Phi_A} \right) \left(\frac{T_k - \bar{T}}{\Phi_T} \right) \\ = \frac{\frac{1}{n-1} \sum_{k=1}^n (A_k - \bar{A})(T_k - \bar{T})}{\sqrt{\frac{1}{n-1} \sum_{k=1}^n (A_k - \bar{A})^2} \sqrt{\frac{1}{n-1} \sum_{k=1}^n (T_k - \bar{T})^2}}$$

where Φ_X is the standard deviation of X; and \bar{X} is the mean of X.

Hierarchical clustering helps to comprehend huge volumes of data. With cluster analysis, the elements are divided into groups that show similar patterns by calculating the distances between their respective rows and columns. The AT-clustered image map (CIM), indicating the correlation coefficients between protein expression and drug sensitivity in the human lung cancer cell lines, was obtained by the linkage-average clustering method, also known as UPGMA (un-weighted pair-group method using arithmetic average). The statistical algorithms and the graphical outputs described here were implemented in MATLAB 6.5 Release 13 (MathWorks Inc., USA).

Results

Clustering on the basis of drug activity and proteomic expression patterns. We used an antibody array, representing 722 antibodies, to perform proteomic profile analysis of 14 human lung cancer cell lines. To avoid the influence of cell culture artifacts, we separately cultured each cell line in six bottles (11). Signal intensities were normalized by comparison with the average values of all probes. Drug sensitivity tests, namely via MTT analysis, had been previously performed on the same panel of cell lines (5). Eight anticancer drugs currently used for lung cancer chemotherapy, namely docetaxel, paclitaxel, gemcitabine, vinorelbine, 5-FU, SN38, CDDP and CBDCA, were selected for these analyses. We then analyzed the protein profiling data in relation to the

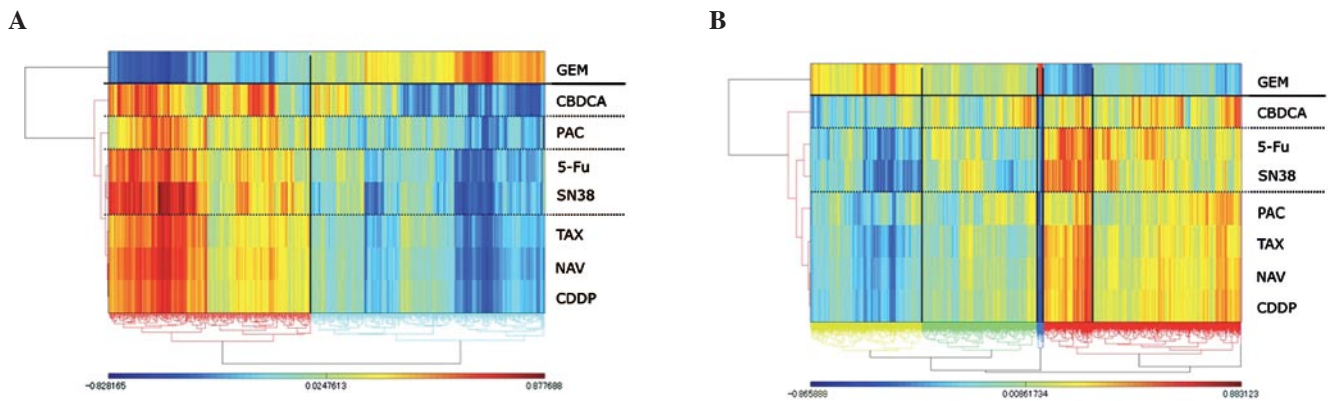


Figure 1. Clustered image map (CIM) relating drug activity to gene/protein expression in lung cancer cell lines using either an antibody array (A), or DNA microarray-based transcriptomic profiling approach (B). The cluster trees of drugs (y-axis) and gene expression (x-axis) are shown in the CIM. Each block of colors represents correlations between clusters of genes and drugs; red (high-positive correlation) and blue (negative correlation). In this analysis, gemcitabine belongs to an isolated cluster. GEM, gemcitabine; CBDCA, carboplatin; 5-Fu, 5-fluorouracil; PAC, paclitaxel; TAX, docetaxel; NAV, vinorelbine; CDDP, cisplatin.

Table I. Factors associated with gemcitabine sensitivity based on proteomic profiles and sensitivity database in the lung cancer cell line panel.

Correlation	Name	Correlation	Name	Correlation	Name
0.773651	MMP-7	-0.62520	IgM	-0.68526	CD165
0.707979	MMP-2	-0.62673	CD2	-0.68530	Ornithine decarboxylase
0.706151	HSP 27	-0.62777	PHAS-I	-0.68611	CD1b
0.695709	p18INK4c	-0.63198	Horseradish peroxidase	-0.68972	FADD
0.684783	MHC II	-0.63323	Troponin T	-0.69125	Pax-5
0.681702	Histone H1	-0.63554	Laminin B2	-0.69198	Claudin 5
0.675647	ERCC1	-0.63606	Streptavidin	-0.69200	ADP-ribosylation factor
0.650884	Mucin 5AC	-0.63876	DcR2/TRAIL-R4/TRUNDD	-0.69726	Fli-1
0.645762	Thrombospondin-1	-0.64042	SHP-1	-0.69890	HSP 90a
0.642413	c-myc	-0.64597	mGluR5	-0.70024	S100A4
0.627175	Cdk1/p34cdc2	-0.64859	Bim/BOD	-0.70111	eNOS
0.618493	MHC I (HLA-A, B, C)	-0.65018	Collagen IV	-0.70265	IGF-I
0.601166	CD8	-0.65352	PDGFR, β	-0.72938	MHC I (HLA-A)
-0.60455	Insulin Ab-4	-0.65827	Factor XIIIa	-0.73024	Topoisomerase II β
-0.60675	Itk/Emt/Tsk	-0.66410	IP10/CRG2	-0.73035	GRIP1
-0.60859	Catenin γ	-0.66641	Cytochrome c	-0.73614	GABAA receptor $\alpha 1$
-0.60876	HDAC1	-0.66869	CD105	-0.73734	TIMP-1
-0.61024	Synaptophysin	-0.66992	Raf-1	-0.73959	Neurofilament
-0.61305	TrxR2	-0.67275	Actin β	-0.74252	Desmin
-0.61367	Neutrophil elastase	-0.67412	E-cadherin	-0.77936	Acinus
-0.61938	Fra2	-0.67550	Survivin	-0.68526	CD165

activity profiles of the eight drugs examined. The drugs were clustered on the basis of Pearson correlation coefficients that related their activity patterns across the 14 cell lines to the proteomic profile information obtained from the same cell lines (10). The AT-matrix CIM summarized the relationship between drug sensitivity and protein profiling data, as it allows the visualization of patterns of similarity in large sets of high-

dimensional data (Fig. 1A) (10). In this analysis, gemcitabine was located in a separate cluster (Fig. 1A). Various proteins were commonly listed that were associated with gemcitabine sensitivity (Table I). Most of these factors were different from our previous transcriptome analysis. Catenin γ was commonly listed as a factor associated with gemcitabine sensitivity using proteomic and transcription analyses.

Clustering on the basis of drug activity and gene expression patterns. We previously performed an analogous gene expression profiling screen using Affymetrix GeneChip arrays (5). We re-analyzed the transcript expression profiling data in relation to the activity profiles of the eight drugs within the same panel of lung cancer cell lines, receiving similar results with respect to drug clusters (Fig. 1B). In this analysis, gemcitabine was again located in separate clusters (Fig. 1B).

Discussion

The NCI-60 set of cell lines, which are representative of tumors from nine different tissue origins, has been profiled by various laboratories at the DNA, RNA and protein levels, as well as correlated with pharmacologic sensitivity (8,10,12,13). In order to aid in the selection of appropriate drug combinations for the treatment of lung cancer, we previously established a human lung cancer cell panel (5). In our previous study, we used a DNA microarray-based gene expression profiling approach, together with assessment of the cytotoxic activity of several widely applied anticancer agents (5). In this report, we related gene expression and drug sensitivity patterns in our cell line panel (5). Our present analysis using proteomic profiling confirmed this prior transcript expression-drug sensitivity correlation analysis. Overall, our data might provide a rational basis, at the molecular pharmacological level, for the combination of chemotherapeutic regimens involving drugs associating with discrete gene-protein expression clusters. Indeed, gemcitabine was deemed, in particular, from our studies, to be a good candidate for the treatment of recurrent or refractory NSCLC.

Using transcriptomic profiling techniques and clinical response data, it is sometimes difficult to consistently reproduce gene-drug sensitivity correlation data. Such data are often influenced by sampling methods, sample preservation status, tumor size and tumor environment status including tumor vessels and inflammation. In our study, the influence of these confounding factors would be expected to be minimal as cancer cell lines were used. However, cell lines differ from tumor cells and should therefore be considered as surrogates.

A question might be raised as to which technologies should be used for determining gene-drug correlations using the NCI approach. Current methods for proteomic profiling include two-dimensional PAGE, antibody microarrays and other mass spectrometry-based techniques (7,8,14,15). Transcript expression profiling via use of DNA microarrays is technologically more advanced and easier than protein profiling. As most biological functions are executed by proteins, it is important to evaluate the utility of transcript profiling for prediction of protein expression. There have been several reports in which gene expression and protein profiling technologies have been used in order to compare the same samples (16,17). These prior reports have shown modest correlation between transcript and protein profiling data. In our study, drug clustering by transcript profiling coincided with that of protein profiling. Transcript profiling may continue to be useful for these analyses. The identified factors associated with drug sensitivities were different between both analyses. There are several possible explanations for this difference

including rates of translation, post-translational modification and protein degradation.

In the treatment of lung cancer, a number of combination therapy regimens employing platinum compounds have proven to be effective (1) and are widely applied as first-line treatment for unresected NSCLC; for example, CDDP + docetaxel, CBDCA + paclitaxel, CDDP + gemcitabine, CDDP + CPT-11, CDDP + paclitaxel and CDDP + vinorelbine (2). In addition, docetaxel and pemetrexed have been reported to be effective in the context of second-line chemotherapy for NSCLC (3,4). It is consequently desirable to find more appropriate therapeutic opportunities based on informed insight. The gene-drug correlations using the NCI program in these studies may be a valuable tool for the identification of determinants of anticancer drug activity in tumors and for the design of cancer chemotherapy and may be useful for improved therapeutic options in the treatment of specific tumor types.

Our present analysis using proteomic profiling confirmed the results of the transcript expression-drug sensitivity correlation analysis. The molecular pharmacological approach taken here suggests that chemotherapy regimens involving gemcitabine should be evaluated in second-line chemotherapy when the initial chemotherapy does not include this drug. A total design approach to cancer chemotherapy through gene-drug correlations using the NCI program may yield improved therapeutic options.

Acknowledgements

This study was supported in part by a Grant-in-Aid from the Ministry of Education, Culture, Sports, Science and Technology of Japan, and the Japan Society for the Promotion of Science. Authors contribution: A.G., M.H. and S.K. designed this study, analyzed and interpreted the data. Y.S. and K.M. carried out the sensitivity test, Y.S., A.M., H.M., J.S., R.N., A.Y., M.N. and A.Y. carried out the cell culture and antibody array work. A.S., A.K. and N.O. carried out acquisition of cDNA array data, while T.O. and H.U. carried out the statistical analysis. All authors read and approved the final manuscript.

References

- Schiller JH, Harrington D, Belani CP, Langer C, Sandler A, Krook J, Zhu J and Johnson DH; Eastern Cooperative Oncology Group: Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. *N Engl J Med* 346: 92-98, 2002.
- Bonomi P, Kim K, Fairclough D, Cella D, Kugler J, Rowinsky E, Jirouek M and Johnson D: Comparison of survival and quality of life in advanced non-small-cell lung cancer patients treated with two dose levels of paclitaxel combined with cisplatin versus etoposide with cisplatin: results of an Eastern Cooperative Oncology Group trial. *J Clin Oncol* 18: 623-631, 2000.
- Shepherd FA, Dancey J, Ramlau R, Mattson K, Gralla R, O'Rourke M, Levitan N, Gressot L, Vincent M, Burkes R, Coughlin S, Kim Y and Berille J: Prospective randomized trial of docetaxel versus best supportive care in patients with non-small-cell lung cancer previously treated with platinum-based chemotherapy. *J Clin Oncol* 18: 2095-2103, 2000.
- Smit EF, Mattson K, von Pawel J, Manegold C, Clarke S and Postmus PE: ALIMTA (pemetrexed disodium) as second-line treatment of non-small-cell lung cancer: a phase II study. *Ann Oncol* 14: 455-460, 2003.

5. Gemma A, Li C, Sugiyama Y, Matsuda K, Seike Y, Kosaihiira S, Minegishi Y, Noro R, Nara M, Seike M, Yoshimura A, Shionoya A, Kawakami A, Ogawa N, Uesaka H and Kudoh S: Anticancer drug clustering in lung cancer based on gene expression profiles and sensitivity database. *BMC Cancer* 6: 174, 2006.
6. Nishizuka S, Charboneau L, Young L, Major S, Reinhold WC, Waltham M, Kouros-Mehr H, Bussey KJ, Lee JK, Espina V, Munson PJ, Petricoin E III, Liotta LA and Weinstein JN: Proteomic profiling of the NCI-60 cancer cell lines using new high-density reverse-phase lysate microarrays. *Proc Natl Acad Sci USA* 100: 14229-14234, 2003.
7. Seike M, Kondo T, Fujii K, Yamada T, Gemma A, Kudoh S and Hirohashi S: Proteomic signature of human cancer cells. *Proteomics* 4: 2776-2788, 2004.
8. Shankavaram UT, Reinhold WC, Nishizuka S, Major S, Morita D, Chary KK, Reimers MA, Scherf U, Kahn A, Dolginow D, Cossman J, Kaldjian EP, Scudiero DA, Petricoin E, Liotta L, Lee JK and Weinstein JN: Transcript and protein expression profiles of the NCI-60 cancer cell panel: an integromic microarray study. *Mol Cancer Ther* 6: 820-832, 2007.
9. Gemma A, Seike M, Seike Y, Uematsu K, Hibino S, Kurimoto F, Yoshimura A, Shibuya M, Harris CC and Kudoh S: Somatic mutation of the hBUB1 mitotic checkpoint gene in primary lung cancer. *Genes Chromosomes Cancer* 29: 213-218, 2000.
10. Scherf U, Ross DT, Waltham M, Smith LH, Lee JK, Tanabe L, Kohn KW, Reinhold WC, Myers TG, Andrews DT, Scudiero DA, Eisen MB, Sausville EA, Pommier Y, Botstein D, Brown PO and Weinstein JN: A gene expression database for the molecular pharmacology of cancer. *Nat Genet* 24: 236-244, 2000.
11. Gemma A, Takenaka K, Hosoya Y, Matuda K, Seike M, Kurimoto F, Ono Y, Uematsu K, Takeda Y, Hibino S, Yoshimura A, Shibuya M and Kudoh S: Altered expression of several genes in highly metastatic subpopulations of a human pulmonary adenocarcinoma cell line. *Eur J Cancer* 37: 1554-1561, 2001.
12. Weinstein JN, Myers TG, O'Connor PM, Friend SH, Fornace AJ Jr, Kohn KW, Fojo T, Bates SE, Rubinstein LV, Anderson NL, Buolamwini JK, van Osdol WW, Monks AP, Scudiero DA, Sausville EA, Zaharevitz DW, Bunow B, Viswanadhan VN, Johnson GS, Wittes RE and Paull KD: An information-intensive approach to the molecular pharmacology of cancer. *Science* 275: 343-349, 1997.
13. Bussey KJ, Chin K, Lababidi S, Reimers M, Reinhold WC, Kuo WL, Gwadry F, Ajay, Kouros-Mehr H, Fridlyand J, Jain A, Collins C, Nishizuka S, Tonon G, Roschke A, Gehlhaus K, Kirsch I, Scudiero DA, Gray JW and Weinstein JN: Integrating data on DNA copy number with gene expression levels and drug sensitivities in the NCI-60 cell line panel. *Mol Cancer Ther* 5: 853-867, 2006.
14. Seike M, Kondo T, Mori Y, Gemma A, Kudoh S, Sakamoto M, Yamada T and Hirohashi S: Proteomic analysis of intestinal epithelial cells expressing stabilized beta-catenin. *Cancer Res* 63: 4641-4647, 2003.
15. Okano T, Kondo T, Kakisaka T, Fujii K, Yamada M, Kato H, Nishimura T, Gemma A, Kudoh S and Hirohashi S: Plasma proteomics of lung cancer by a linkage of multi-dimensional liquid chromatography and two-dimensional difference gel electrophoresis. *Proteomics* 6: 3938-3948, 2006.
16. Anderson L and Seilhamer J: A comparison of selected mRNA and protein abundances in human liver. *Electrophoresis* 18: 533-537, 1997.
17. Chen G, Gharib TG, Huang CC, Taylor JM, Misek DE, Kardias SL, Giordano TJ, Iannettoni MD, Orringer MB, Hanash SM and Beer DG: Discordant protein and mRNA expression in lung adenocarcinomas. *Mol Cell Proteomics* 1: 304-313, 2002.