

Comparison of Genomic Signatures of Non-small Cell Lung Cancer Recurrence between Two Microarray Platforms

JOHN WEN-CHENG CHANG¹, NIEN-CHIH WEI², HUNG-JU SU³, JIE-LEN HUANG³,
TSE-CHING CHEN⁴, YI-CHENG WU⁵, CHIH-TENG YU⁶, MING-MO HOU¹, CHIA-HSUN HSIEH¹,
JIA-JUAN HSIEH^{1,7}, CHIAO-EN WU¹, HSIN-YI CHENG¹, TODD HSU⁷ and TZU-HAO WANG⁸

Divisions of ¹Hematology-Oncology, ⁴Pathology, ⁵Thoracic and Cardiovascular Surgery, ⁶Thoracic Medicine, and
⁸Obstetrics and Gynecology, Genomic Medicine Research Core Lab, Chang Gung Memorial Hospital,

Chang Gung University College of Medicine, Taoyuan, Taiwan, R.O.C.;

²Auspex Diagnostics, Taiwan, R.O.C.;

³Microarray Laboratory, Biomedical Engineering Research Laboratory,
Industrial Technology Research Institute, Hsinchu, Taiwan, R.O.C.;

⁷Institute of Bioscience and Biotechnology and Center of Excellence for Marine Bioenvironment and Biotechnology,
National Taiwan Ocean University, Keelung, Taiwan, R.O.C.

Abstract. *Background:* Cancer genomic signatures may vary using different platforms. We compared the differential gene expression in non-small cell lung cancer (NSCLC) between two platforms in order to find the most relevant genomic signatures of tumor recurrence. *Materials and Methods:* We analyzed gene expression in frozen lung cancer tissue from 59 selected patients who had undergone surgical resection of NSCLC. These patients were divided into two groups: group R, patients who had a tumor recurrence within four years, n=37; group NR, patients who remained disease-free four years following initial surgery, n=22. Each RNA sample was assayed twice using both Affymetrix and Illumina GeneChip. Data were analyzed by principal component analysis and leave-one-out cross-validation. *Results:* Using the same filtering criteria, 13 genes that were differentially expressed between R and NR were identified by Affymetrix, while 21 genes were identified by Illumina GeneChip. In common, a total of six genes were detected by both systems. Using univariate analysis, four (lipocalin 2, LCN2; parathyroid hormone-like hormone, PTHLH; ras-related protein Rab-38, RAB38; and four jointed box 1, FJX1) of these six genes were associated with survival. A risk score of survival was calculated according to the four-

gene expression. There was a significant difference in overall survival between low- and high-risk groups. *Conclusion:* A four-gene signature is associated with survival among patients with early-stage NSCLC. Further validation of these findings is warranted.

Lung cancer is the leading cause of cancer death. Non-small cell lung cancer (NSCLC) accounts for 80% to 85% of all lung cancer cases. Most patients present with advanced disease (1). Despite recent advances in multi-modal therapy, the overall 5-year survival rate for NSCLC remains of the order of 8 to 12%. Surgery is still the first choice of treatment for localized NSCLC if the patient's physical condition allows it. However, the result of surgical treatment remains unsatisfactory, and 35-50% of such patients will experience disease relapse within 5 years.

The microarray techniques, first described in 1994 by Drmanac and Drmanac (2), use the method of hybridization of large-scale cDNA with a mass probe to identify the expression of individual genes. A method by Affymetrix GeneChip utilizes silicon chips where more than 400,000 oligonucleotides can be synthesized on a single 1.6-cm² microscopic glass slide (3). The major advantage of DNA microarray is the screening of large numbers of genes with greater sensitivity using a smaller amount of sample. Currently, up to 30,000 cDNA probes can be placed on a small microscope glass slide, with the future goal to screen the entire human genome (approximately 30,000 expressed genes, that almost cover the entire genome) in one experiment. It will be a powerful tool in studying the effects of changes made in cellular signaling pathways and identifying the changes that occur in the function of other

Correspondence to: Tzu-Hao Wang, MD, Department of Obstetrics and Gynecology, Chang Gung Memorial Hospital, 5 Fu-Shin Street, Kwei-Shan, Taoyuan 333, Taiwan, R.O.C. Tel: +886 33281200 ext.5402, Fax: +886 33288252, e-mail: knoxtn@cgmh.org.tw

Key Words: Microarray, genomic signature, tumor recurrence, lung cancer, NSCLC, LCN2, PTHLH, RAB38, FJX1.

genes downstream of various genetic alterations. Gene expression analysis using microarrays, combined with supervised clustering data analysis, could provide insight into the molecular signature of lung cancer and its treatment and prognosis. This would provide the basis for more effective therapeutic intervention.

There have been several advances in the molecular prediction of individual clinical outcome by microarray technologies. In breast cancer, the success was manifested in the commercial gene tests, such as Oncotype DX (4) and MammaPrint (5, 6). In current studies of biomarker identification, genes are ranked according to their association with the clinical outcome. Nevertheless, each individual gene selection algorithm has different strengths and limitations. A modified model combining multiple gene selection might provide a better method to identify novel biomarkers. In 2007, Chen *et al.* used computer-generated random numbers to assign 185 frozen specimens for microarray analysis, real-time reverse-transcriptase polymerase chain reaction (RT-PCR) analysis, or both (7). They studied gene expression in frozen specimens of lung cancer tissue from 125 randomly selected patients who had undergone surgical resection of NSCLC and evaluated the association between the level of expression and survival. A five-gene signature (including dual-specificity phosphatase 6, *DUSP6*; monocyte-to-macrophage differentiation associated protein, *MMD*; signal transducer and activator of transcription 1, *STAT1*; v-erb-b2 avian erythroblastic leukemia viral oncogene homolog 3, *ERBB3*; and lymphocyte-specific protein tyrosine kinase, *LCK*) was an independent predictor of relapse-free and overall survival. However, there is no fully-validated and clinically applied model for predicting lung cancer recurrence after surgery.

To our knowledge, ours is the first study to use two microarray platforms (Affymetrix and Illumina) to analyze the genomic signatures of tumor recurrence in resectable lung cancer.

Materials and Methods

Selection of adequate samples from the Chang Gung Memorial Hospital (CGMH) tissue bank. We used the lung tumor tissues collected by the CGMH tissue bank according to its standard procedures, including written informed consent obtained from patients. A pathologist reviewed a tissue section of the archived fresh frozen tissue. Only samples containing at least 50% cancer cells were used. We collected 549 lung tumor tissues from the CGMH tissue bank during the period between January 2004 and December 2007. According to the clinical database obtained from the Cancer Registry in CGMH-Linkou Medical Center, from these further selected 183 lung tumor specimens for RNA extraction. A total of 59 acceptable samples of RNA were isolated for microarray analysis with Affymetrix and Illumina GeneChip.

Identification of clinical information. The accuracy of the clinical database had been confirmed by two doctors from the Division of Thoracic Surgery and Chest Medicine in CGMH. All of the cases

had been selected by one doctor at first, then the selected patients were cross-confirmed by the other doctor, both followed the same standard operating procedure (Table I) to confirm the criteria needed.

Tumor classification based on tumor recurrence within four years and RNA extraction. The tumor and non-tumor parts of each sample were identified and sampled by a pathologist. The patients were divided into two groups: group R, patients who had a tumor recurrence within four years of surgery; group NR, patients who remained disease-free four years following initial surgery. Tumor samples from the 37 recurrent (R) and 22 non-recurrent (NR) NSCLC cases were utilized. A pathologist reviewed the tissue section of the archived fresh frozen tissue. Only samples containing at least 50% cancer cells were used. One section was stained by hematoxylin and eosin (H&E) to confirm the adequacy of the tumor tissue. The rest of the samples were used for total RNA extraction. Following mechanical tissue disruption, total tumor RNA was extracted using the RNeasy Mini kit (QIAGEN, Hilden, Germany). RNA integrity was assessed with the total RNA Nano Chip Assay on an Agilent 2100 Bioanalyzer (Agilent Technologies GmbH, Berlin, Germany). The RNA integrity number (RIN score) was generated for each sample on a scale of 1-10 as an indication of RNA quality. Total RNA with RIN score >5 was used for microarrays. Prior to array analysis, one round of T7 promotor-based RNA amplification was performed.

Microarrays, microarray data processing and normalization. Affymetrix® HG-U133 Plus 2.0 mRNA expression arrays (Affymetrix, Santa Clara, CA, USA) were used in order to determine the expression of 47,400 transcripts, corresponding to 38,500 human genes (8). These arrays have proven high reproducibility for mRNA expression analysis (9). Briefly, 1-15 µg total RNA was reverse-transcribed into cDNA, followed by RNase H-mediated second-strand cDNA synthesis. The double-stranded cDNA was purified and served as a template in the subsequent *in vitro* transcription (IVT) reaction. The IVT reaction was carried out in the presence of T7 RNA polymerase and a biotinylated nucleotide analog/ribonucleotide mix for complementary RNA (cRNA) amplification and biotin labeling. The biotinylated cRNA targets were then cleaned up, fragmented, and hybridized to GeneChip expression arrays. A hybridization cocktail was prepared, including the fragmented target, probe array controls, bovine serum albumin (BSA) and herring sperm DNA. It was then hybridized to the probe array during 16 hours' incubation. Specific experimental information was defined using Affymetrix® Microarray Suite on a PC-compatible workstation.

Illumina BeadChip Ref-6 (Ambion, Inc., Austin, TX, USA) containing 48,804 probes was used to perform a whole human gene expression profile. Biotin-labeled cRNA for hybridization was generated by *in vitro* transcription based on Eberwine protocol using Illumina's recommended commercial kits. Briefly, 500 ng total RNA was reverse-transcribed into cDNA, followed by linear amplification steps according to the Illumina TotalPrep RNA Amplification kit for Illumina. Hybridization was performed with 1.5 µg biotin-labeled cRNA in each array of BeadChip Ref-6. After 16 hours incubation at 58°C, BeadChips were washed, stained and scanned according to the manufacturer's manual. Scanning used the Illumina BeadArray Reader software together with the Illumina BeadStation 500 platform. Processing and analysis of the microarray data were performed with the Illumina BeadStudio v3.3.7 software. The data were normalized using the quintile option.

Table I. *Recurrence definition.*

1.	Cut-off of 48 months
2.	Should not have received adjuvant chemotherapy
3.	The surgical and pathologic reports should be reviewed to ensure accurate staging.
4.	After surgery, regularly followed up with systematic history-taking and physical examinations, serum tumor markers, computed tomography of the chest, abdominal sonography, and bone scintigraphy at 3-month intervals as needed.
5.	The exact cause of death and the date of death should be documented.
6.	If the patient had disease recurrence, then the pattern of failure should be documented with exact date and sites, as well as the diagnostic method of the recurrence.
7.	The local or distant recurrence should be documented separately. Local recurrence was defined as tumor recurrence in a contiguously anatomical site including ipsilateral hemithorax and mediastinum after surgical resection. Distant recurrence was defined as tumor recurrence in the contralateral lung or outside the hemithorax and mediastinum after surgical resection.
8.	The overall survival time is defined from the date of operation to the date of death, including surgical mortality. The disease-free survival time is defined from the date of operation to the date of first noted recurrence.

Table II. *Top 20 differentially expressed genes in the non-recurrence group by the Affymetrix platform in 16 cases.*

Gene ID	Symbol	Gene name	<i>p</i> -Value	NR/R (log ratio)
6947	<i>TCN1</i>	Transcobalamin I (vitamin B12 binding protein, R binder family)	0.012	3.01
79152	<i>FA2H</i>	Fatty acid 2-hydroxylase	0.004	2.64
6340	<i>SCNN1G</i>	Sodium channel, nonvoltage-gated 1, gamma	0.046	2.59
3934	<i>LCN2</i>	Lipocalin 2	0.035	2.56
3960	<i>LGALS4</i>	Lectin, galactoside-binding, soluble, 4	0.044	2.41
344905	<i>ATP13A5</i>	ATPase type 13A5	0.013	2.28
54997	<i>TESC</i>	Tescalcin	0.030	2.15
10170	<i>DHRS9</i>	Dehydrogenase/reductase (SDR family) member 9	0.023	2.08
9194	<i>SLC16A7</i>	Solute carrier family 16, member 7 (monocarboxylic acid transporter 2)	0.037	2.07
150696	<i>PROM2</i>	Prominin 2	0.040	1.93
4897	<i>NRCAM</i>	Neuronal cell adhesion molecule	0.041	1.77
57419	<i>SLC24A3</i>	Solute carrier family 24 (sodium/potassium/calcium exchanger), member 3	0.028	1.71
10257	<i>ABCC4</i>	ATP-binding cassette, sub-family C (CFTR/MRP), member 4	0.035	1.67
26298	<i>EHF</i>	Ets homologous factor	0.036	1.66
956	<i>ENTPD3</i>	Ectonucleoside triphosphate diphosphohydrolase 3	0.030	1.65
9568	<i>GABBR2</i>	Gamma-aminobutyric acid (GABA) B receptor, 2	0.033	1.63
10267	<i>RAMP1</i>	Receptor (G protein-coupled) activity modifying protein 1	0.011	1.59
4319	<i>MMP10</i>	Matrix metalloproteinase 10 (stromelysin 2)	0.029	1.53
4586	<i>MUC5AC</i>	Mucin 5AC, oligomeric mucus/gel-forming	0.044	1.49
84825	<i>MGC5370</i>	CDNA FLJ37519 fis, clone BRCAN2004699	0.009	1.49

NR/R: The log ratio of the expressed genes in non-recurrent tumors to recurrent tumors.

Identification of predictive transcripts by leave-one-out cross-validation. To identify predictive transcripts, a leave-one-out process was used. Predictive gene signatures were generated using the expression profiles and sensitivity data of all 59 test tumors as a training set. Leave-one-out cross-validation (LOOCV) involved removing a single tumor from the original training set of 59 tumors and using the remaining 59 tumors as the training set and the removed tumor for validation. This procedure was repeated in away that each tumor in the original training set was used once for validation.

Network visualization and analysis. Network analyses of differentially expressed genes were performed using the MetaCore Analytical Suite (GeneGo Inc., St Joseph, MI, USA; <http://www.genego.com>) (10, 11). MetaCore is a web-based computational platform and useful for analyzing a cluster of genes

in the context of regulatory networks and signaling pathways. For the network analysis of a group of genes, MetaCore can be used to calculate the statistical significance (*p*-value) based on the probability of assembly from a random set of nodes (genes) of the same size as the input list (11, 12).

Results

Functional analysis of gene expression signature by the Affymetrix platform. After the differentially expressed genes were selected under the criteria ($p < 0.05$ and fold change > 2), principal component analysis (PCA) on the resulted 97 genes could clearly separate R tumors from NR tumors. All of 97 genes were ranked by fold changes and the top 20 genes were

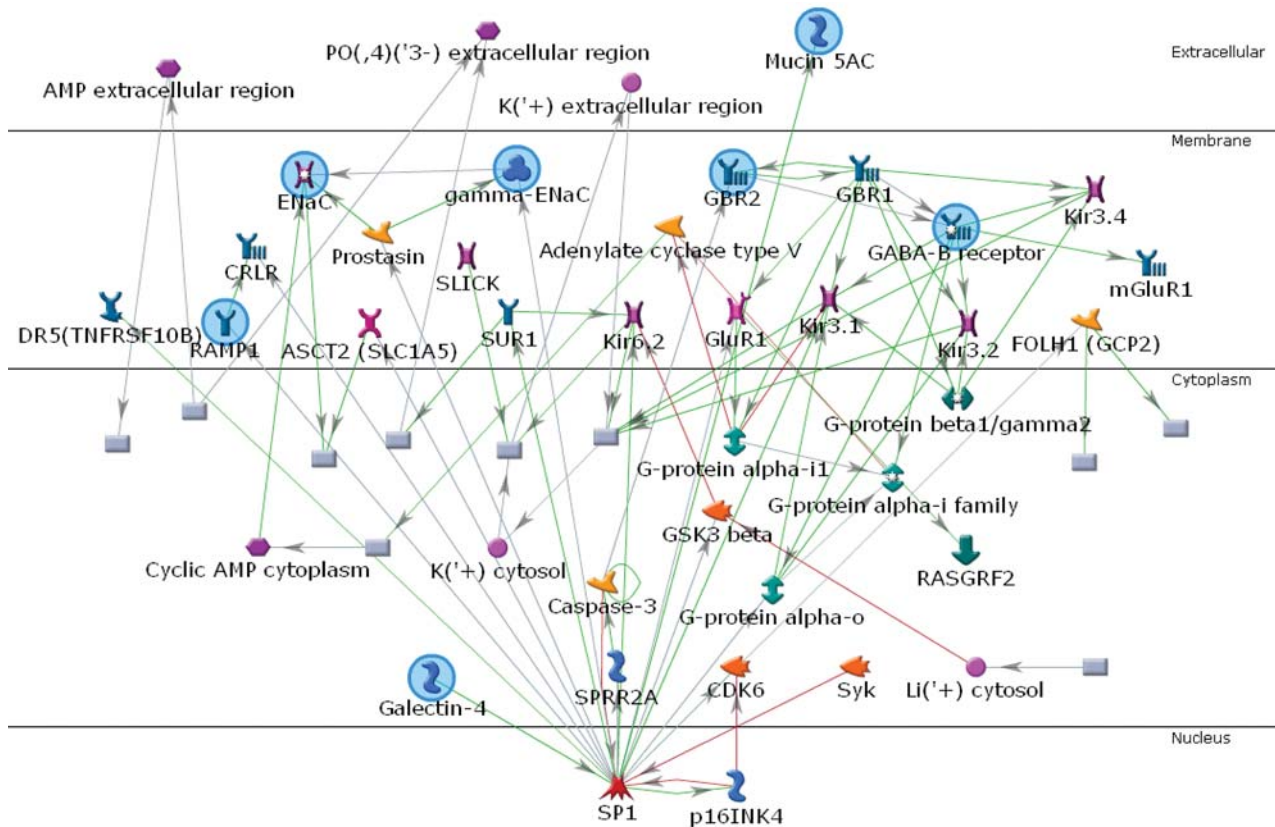


Figure 1. Functional networks of 16 individual lung cancer tumors. Some genes are labeled by aliases. The genes (blue circles) that were up-regulated in non-recurrent tumors (also listed in Table II) were generalized into the networks using the MetaCore program.

listed in Table II. The differentially expressed genes listed in Table II represent the genes that were expressed higher in NR tumors, *i.e.*, they were the down-regulated genes in R tumors. In order to explore the potential biological functions of these genes, we uploaded these 20 genes into the MetaCore database and algorithm and derived the following functional networks that contained 7 root genes (denoted as blue circles in Figure 1). The up-regulated genes in NR tumors included *lipocalin 2 (LCN2)*, which was involved in induction of the epithelial to mesenchymal transition.

Comparison and correlation between Affymetrix and Illumina platforms. Following validation by LOOCV, we analyzed those genes which were either up-regulated or down-regulated by two-fold using Student *t*-test (NR vs. R, $p < 0.05$, $|\text{NR-R}| > 1$), a total of 13 genes were selected and identified with the Affymetrix (Figure 2A) platform. Furthermore, using the same filtering criteria, we were able to identify 21 genes in the parallel Illumina database (Figure 2B). There were six common genes identified in both the Affymetrix and Illumina platforms. Using univariate

analysis, four (lipocalin 2, *LCN2*; parathyroid hormone-like hormone, *PTHLH*; ras-related protein Rab-38, *RAB38*; and four jointed box 1, *FJX1*) out of these six genes were associated with survival.

Survival prediction using the four-gene prognostic model. A four-gene signature was further analyzed for correlation of gene expression and the disease-free (DFS) and overall survival (OS) from initial surgery by Kaplan-Meier analysis. The risk score was generated for each sample on a scale of 0-4 as an indication of expression of these four genes. A score of 1 was assigned for each gene if it is up-regulated in the tumor. A tumor risk score ≥ 3 was identified as high-risk (patients who had low tumor expression of these four genes) and ≤ 2 as low-risk (those who had high tumor expression of these four genes). As a result, a four-gene signature was identified, which was well-associated with DFS and OS. The four-gene signature stratified patients into high- and low-risk groups with distinct postoperative survival in Kaplan-Meier analyses (log-rank $p < 0.05$). The median DFS (Figure 3A) was not reached in the low-risk group and was 20.4 months

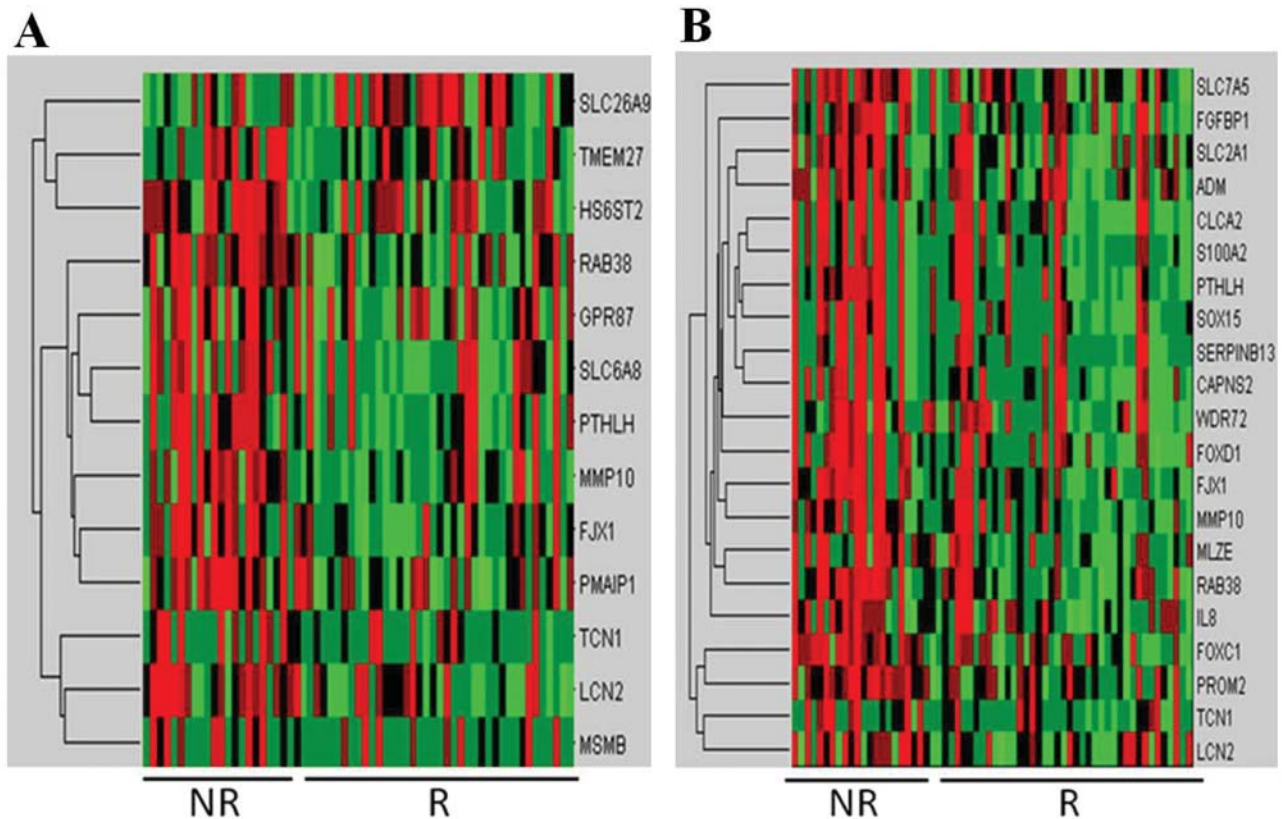


Figure 2. Different gene expression between recurrence (R) and non-recurrence (NR) groups in Affymetrix (A) and Illumina (B) platforms. There were 13 genes selected and identified in the Affymetrix platform, while 21 genes were found in the Illumina platform. Student *t* test ($p < 0.05$, $|NR-R| > 1$) was used.

(95% confidence interval (CI)=3.6 to 37.2 months) in the high-risk group ($p=0.001$). The median OS (Figure 3B) was not reached in the low-risk group and was 37 months (95% CI=2.6 to 71.4 months) in the high-risk group ($p=0.014$). The four-gene risk score was also correlated with tumor recurrence. Thirty-one (78%) of 40 high-risk patients developed a recurrence, as compared to 6 (32%) of 19 low-risk patients ($p=0.001$). There was thus a significant difference in overall survival between the low- and high-risk groups using the four-gene prediction model.

Discussion

Lung cancer continues to be a major deadly malignancy. The mortality caused by this disease could be reduced by improving the ability to predict cancer patients' survival. It is also important to identify clinically relevant prognostic biomarkers in order to develop personalized treatment. In this study, we identified a four-gene signature, differently expressed between the R and NR groups in NSCLC using both Affymetrix and Illumina microarray platforms. The identification of four genes that can predict the clinical

outcome in patients with NSCLC may reveal targets for the development of therapy for lung cancer. LCN2, a member of the lipocalin family that transports small lipophilic ligands, has gained recent attention as both a potential biomarker and a modulator of human cancer (13). LCN2 has been shown to induce the epithelial to mesenchymal transition in breast cancer cells and to promote breast tumor invasion (14). PTHLH is an important chondrogenic regulator; however, the gene has not been directly linked to human disease (15). RAB38 is a member of the RAB small GTPase family that regulates intracellular vesicle trafficking (16). RAB proteins and their effectors coordinate multiple stages of membrane transport, such as vesicle formation, motility, and tethering to their target compartments. These proteins can also be found in both membrane-bound and cytosolic forms and are prenylated at their carboxyl termini (17).

Several studies of NSCLC have reported the ability to generate expression signatures for identification of grouping subjects according to their survival outcomes. However, most studies are small and typically collect data from a single institution. Shedden *et al.* reported a large, training-testing, multisite blinded validation study aiming to characterize the

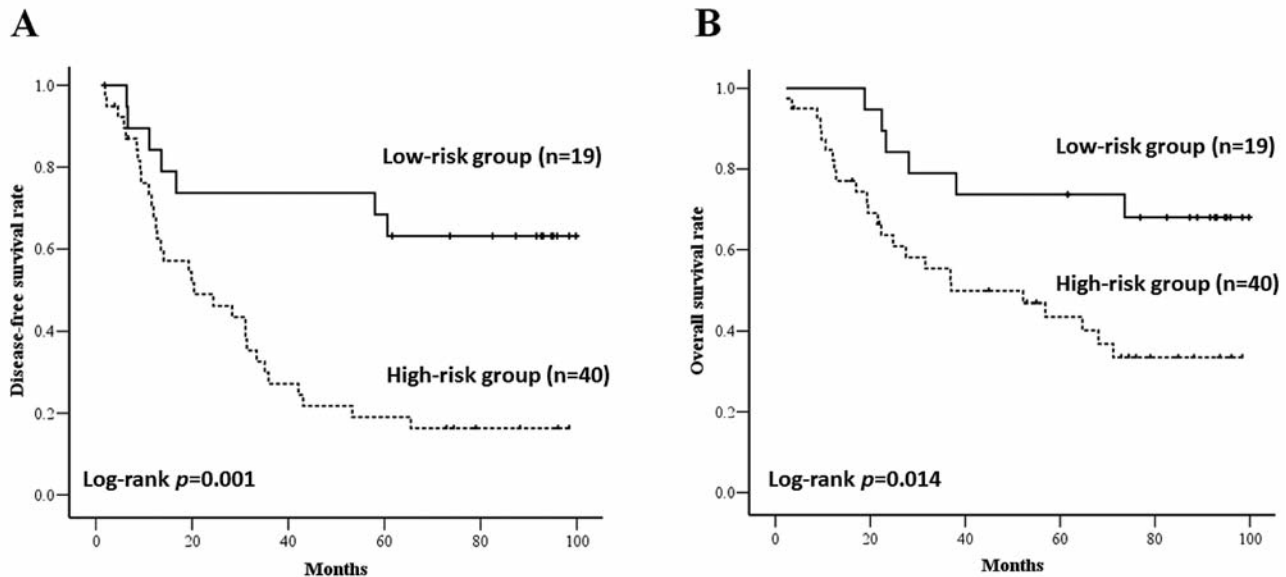


Figure 3. Kaplan-Meier plot of disease-free survival (DFS) (A) and overall survival (OS) (B) for patients with high-risk versus low-risk tumors. The median DFS was 20.4 (95% CI=3.6 to 37.2) months and not reached, respectively. The median OS was 37 (95% CI=2.6 to 71.4) months and not reached, respectively.

performance of several prognostic models based on gene expression for 442 lung adenocarcinomas (18). They examined whether microarray measurements of gene expression, either alone or combined with basic clinical covariates (stage, age, sex), can be used to predict OS of patients with lung cancer. Most methods performed better with clinical data, supporting the combined use of clinical and molecular information when building prognostic models for early-stage lung cancer. They also provided the largest available set of microarray data, including four institutions (University of Michigan Cancer Center (UM), Moffitt Cancer Center (HLM), Memorial Sloan-Kettering Cancer Center (MSK) and the Dana-Farber Cancer Institute (DFCI)), with extensive pathological and clinical annotation for lung adenocarcinomas. They used several methods to analyze these data sets, including gene clustering (method A), univariate testing (methods B to G) and analysis on a mechanistic basis (method H). Method A, which worked with all tumor samples or with stage I samples alone, both with and without clinical covariates, showed the best overall predictive ability. In the MSK test, if sensitivity was 0.9, specificity was 0.3. In DFCI test, if sensitivity was 0.9, specificity was 0.2. To date, a microarray model with high sensitivity and specificity to predict tumor recurrence and patient survival has yet to be developed.

In this study, we utilized two microarray platforms in order to analyze the correlation of genomic signature and tumor recurrence. Although there were six genes to both systems, a further selected four-gene signature was able to

predict the DFS and OS of patients who might have good prognosis from initial surgery. This four-gene signature also stratifies stage II lung adenocarcinoma patients into two distinct DFS (log-rank $p=0.014$) and OS (log-rank $p=0.036$) groups (data not shown). The next step is the validation of the prediction of tumor recurrence within four years using this scoring system (four-gene signature) in the clinic.

In conclusion, we identified a four-gene signature of differently expressed genes between R and NR groups in NSCLC by both Affymetrix and Illumina platforms. The selected four-gene signature was able to predict the DFS and OS from surgical therapy. Further investigation of functional analysis and clinical validation is warranted.

Acknowledgements

Authors are grateful to Dr. Lee Yun-Shien for data evaluation and Dr. Hong Ji-Hong for administrative support. This study was supported by a grant to John W.C. Chang from CGMH (CMRPG380761 and CMRPG3B0081), Taiwan.

References

- 1 Silvestri GA and Rivera MP: Targeted therapy for the treatment of advanced non-small cell lung cancer: a review of the epidermal growth factor receptor antagonists. *Chest* 128: 3975-3984, 2005.
- 2 Drmanac S and Drmanac R: Processing of cDNA and genomic kilobase-size clones for massive screening, mapping and sequencing by hybridization. *Biotechniques* 17: 328-329 and 332-336, 1994.

- 3 Lockhart DJ, Dong H, Byrne MC, Follettie MT, Gallo MV, Chee MS, Mittmann M, Wang C, Kobayashi M, Horton H and Brown EL: Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14: 1675-1680, 1996.
- 4 Paik S, Shak S, Tang G, Kim C, Baker J, Cronin M, Baehner FL, Walker MG, Watson D, Park T, Hiller W, Fisher ER, Wickerham DL, Bryant J and Wolmark N: A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med* 351: 2817-2826, 2004.
- 5 van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R and Friend SH: Gene expression profiling predicts clinical outcome of breast cancer. *Nature* 415: 530-536, 2002.
- 6 van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH and Bernards R: A gene expression signature as a predictor of survival in breast cancer. *N Engl J Med* 347: 1999-2009, 2002.
- 7 Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, Cheng CL, Wang CH, Terng HJ, Kao SF, Chan WK, Li HN, Liu CC, Singh S, Chen WJ, Chen JJW and Yang PC: A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med* 356: 11-20, 2007.
- 8 URL: http://www.affymetrix.com/support/technical/datasheets/human_datasheet.pdf
- 9 Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JC, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW, Goodsaid FM, Herman D, Jensen RV, Johnson CD, Lobenhofer EK, Puri RK, Schrf U, Thierry-Mieg J, Wang C, Wilson M, Wolber PK, Zhang L, Amur S, Bao W, Barbacioru CC, Lucas AB, Bertholet V, Boysen C, Bromley B, Brown D, Brunner A, Canales R, Cao XM, Cebula TA, Chen JJ, Cheng J, Chu TM, Chudin E, Corson J, Corton JC, Croner LJ, Davies C, Davison TS, Delenstarr G, Deng X, Dorris D, Eklund AC, Fan XH, Fang H, Fulmer-Smentek S, Fuscoe JC, Gallagher K, Ge W, Guo L, Guo X, Hager J, Haje PK, Han J, Han T, Harbottle HC, Harris SC, Hatchwell E, Hauser CA, Hester S, Hong H, Hurban P, Jackson SA, Ji H, Knight CR, Kuo WP, LeClerc JE, Levy S, Li QZ, Liu C, Liu Y, Lombardi MJ, Ma Y, Magnuson SR, Maqsoodi B, McDaniel T, Mei N, Myklebost O, Ning B, Novoradovskaya N, Orr MS, Osborn TW, Papallo A, Patterson TA, Perkins RG, Peters EH, Peterson R, Philips KL, Pine PS, Pusztai L, Qian F, Ren H, Rosen M, Rosenzweig BA, Samaha RR, Schena M, Schroth GP, Shchegrova S, Smith DD, Staedtler F, Su Z, Sun H, Szallasi Z, Tezak Z, Thierry-Mieg D, Thompson KL, Tikhonova I, Turpaz Y, Vallanat B, Van C, Walker SJ, Wang SJ, Wang Y, Wolfinger R, Wong A, Wu J, Xiao C, Xie Q, Xu J, Yang W, Zhang L, Zhong S, Zong Y and Slikker W Jr.: The MicroArray Quality Control (MAQC) project shows interplatform reproducibility of gene expression measurements. *Nat Biotechnol* 24: 1151-1161, 2006.
- 10 Nikolsky Y, Ekins S, Nikolskaya T and Bugrim A: A novel method for generation of signature networks as biomarkers from complex high throughput data. *Toxicol Lett* 158: 20-29, 2005.
- 11 Mason CW, Swaan PW, and Weiner CP: Identification of interactive gene networks: A novel approach in gene array profiling of myometrial events during guinea pig pregnancy. *Am J Obstet Gynecol* 194: 1513-1523, 2006.
- 12 Guo L, Lobenhofer EK, Wang C, Shippy R, Harris SC, Zhang L, Mei N, Chen T, Herman D, Goodsaid FM, Hurban P, Phillips KL, Xu J, Deng X, Sun YA, Tong W, Dragan YP and Shi L: Rat toxicogenomic study reveals analytical consistency across microarray platforms. *Nat Biotechnol* 24: 1162-1169, 2006.
- 13 Yang J and Moses MA: Lipocalin 2: A multifaceted modulator of human cancer. *Cell Cycle* 8: 2347-2352, 2009.
- 14 Yang J, Bielenberg DR, Rodig SJ, Doiron R, Clifton MC, Kung AL, Strong RK, Zurakowski D and Moses MA: Lipocalin 2 promotes breast cancer progression. *Proc Natl Acad Sci USA* 106: 3913-3918, 2009.
- 15 Maass PG, Wirth J, Aydin A, Rump A, Stricker S, Tinschert S, Otero M, Tsuchimochi K, Goldring MB, Luft FC, and Bähring S: A *cis*-regulatory site downregulates PTHLH in translocation t(8;12)(q13;p11.2) and leads to brachydactyly type E. *Hum Mol Genet* 19: 848-860, 2010.
- 16 Loftus SK, Larson DM, Baxter LL, Antonellis A, Chen Y, Wu X, Jiang Y, Bittner M, Hammer JA, and Pavan WJ: Mutation of melanosome protein RAB38 in chocolate mice. *Proc Natl Acad Sci USA* 99: 4471-4476, 2002.
- 17 Osanai K, Oikawa R, Higuchi J, Kobayashi M, Tsuchihara K, Iguchi M, Jongsu H, Toga H, and Voelker DR: A mutation in *Rab38* small GTPase causes abnormal lung surfactant homeostasis and aberrant alveolar structure in mice. *Am J Pathol* 173: 1265-1274, 2008.
- 18 Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, Eschrich S, Jurisica I, Giordano TJ, Misek DE, Chang AC, Zhu CQ, Strumpf D, Hanash S, Shepherd FA, Ding K, Seymour L, Naoki K, Pennell N, Weir B, Verhaak R, Ladd-Acosta C, Golub T, Gruidl M, Sharma A, Szoke J, Zakowski M, Rusch V, Kris M, Viale A, Motoi N, Travis W, Conley B, Seshan VE, Meyerson M, Kuick R, Dobbin KK, Lively T, Jacobson JW and Beer DG: Gene expression-based survival prediction in lung adenocarcinoma: a multisite, blinded validation study. *Nat Med* 14: 822-827, 2008.

Received January 16, 2012

Revised February 23, 2012

Accepted February 27, 2012