

Research Article

Identification of a Novel Biomarker, *SEMA5A*, for Non-Small Cell Lung Carcinoma in Nonsmoking WomenTzu-Pin Lu¹, Mong-Hsun Tsai^{3,5}, Jang-Ming Lee⁶, Chung-Ping Hsu¹⁰, Pei-Chun Chen⁵, Chung-Wu Lin⁸, Jin-Yuan Shih⁷, Pan-Chyr Yang⁷, Chuhsing Kate Hsiao^{4,5}, Liang-Chuan Lai^{2,5,9}, and Eric Y. Chuang^{1,5,9}

Abstract

Background: Although cigarette smoking is the major risk factor for lung cancer, only 7% of female lung cancer patients in Taiwan have a history of smoking. The genetic mechanisms of carcinogenesis in nonsmokers are unclear, but semaphorins have been suggested to play a role as lung tumor suppressors. This report is a comprehensive analysis of the molecular signature of nonsmoking female lung cancer patients in Taiwan, with a particular focus on the semaphorin gene family.

Methods: Sixty pairs of tumor and adjacent normal lung tissue specimens were analyzed by using Affymetrix U133plus2.0 expression arrays. Differentially expressed genes in tumor tissues were identified by a paired *t* test and validated by reverse transcriptase-PCR and immunohistochemistry. Functional analysis was conducted by using Ingenuity Pathway Analysis as well as gene set enrichment analysis and sigPathway algorithms. Kaplan-Meier survival analyses were used to evaluate the association of *SEMA5A* expression and clinical outcome.

Results: We identified 687 differentially expressed genes in non-small cell lung carcinoma (NSCLC). Many of these genes, most notably the semaphorin family, were participants in the axon guidance signaling pathway. The downregulation of *SEMA5A* in tumor tissue, both at the transcriptional and translational levels, was associated with poor survival among nonsmoking women with NSCLC.

Conclusions: In summary, several semaphorin gene family members were identified as potential therapeutic targets, and *SEMA5A* may be useful as a prognostic biomarker for NSCLC, which may also be gender specific in Taiwanese patients.

Impact: A novel biomarker for NSCLC is identified. *Cancer Epidemiol Biomarkers Prev*; 19(10); 2590–7. ©2010 AACR.

Introduction

Lung cancer is the leading cause of cancer-related death worldwide and brings significant socioeconomic impact to patients, their families, and society as a whole.

Authors' Affiliations: ¹Graduate Institute of Biomedical Electronics and Bioinformatics, ²Graduate Institute of Physiology, ³Institute of Biotechnology, ⁴Department of Public Health, and ⁵Bioinformatics and Biostatistics Core, Research Center for Medical Excellence, National Taiwan University; Departments of ⁶Surgery, ⁷Internal Medicine, and ⁸Pathology; ⁹National Clinical Trial and Research Center, National Taiwan University Hospital, Taipei, Taiwan; and ¹⁰Division of Thoracic Surgery, Taichung Veterans General Hospital, Taichung, Taiwan

Note: Supplementary data for this article are available at Cancer Epidemiology, Biomarkers & Prevention Online (<http://cebp.aacrjournals.org>).

T-P. Lu, M-H. Tsai, and J-M. Lee contributed equally to this work.

The microarray data of this study have been submitted to the Gene Expression Omnibus database (accession number GSE19804).

Corresponding Authors: Liang-Chuan Lai, Graduate Institute of Physiology, National Taiwan University, Taipei 110, Taiwan. Phone: 886-2-2312-3456, ext. 88241; Fax: 886-2-2396-4350. E-mail: llai@ntu.edu.tw and Eric Y. Chuang, Department of Electrical Engineering, Graduate Institute of Biomedical Electronics and Bioinformatics, National Taiwan University, Taipei 106, Taiwan. Phone: 886-2-3366-3660; Fax: 886-2-3366-3682. E-mail: chuangey@ntu.edu.tw

doi: 10.1158/1055-9965.EPI-10-0332

©2010 American Association for Cancer Research.

Non-small cell lung carcinoma (NSCLC) accounts for the majority of lung tumors (1). Among NSCLCs, adenocarcinoma is the major histologic type of lung carcinoma in Taiwan (52.5%).

Smoking is the major risk factor for lung cancer (2), although other factors, such as environmental exposure (e.g., chemicals, physical agents, and radiation), clinical history of lung diseases (e.g., chronic bronchitis, emphysema, pneumonia, and tuberculosis; ref. 3), familial tumor history (4), or diet (5, 6), may also be associated with the development of lung cancer (7). In Western countries, 70% to 90% of lung cancers are attributable to cigarette smoking, whereas in Taiwan, only 7% of female lung cancer cases are associated with smoking (8, 9). Many genes [e.g., *TP53* (10, 11), *EGFR* (12, 13), *KRAS* (14), *PIK3CA* (15), and *EML4-ALK* (16)] have been reported in association with lung cancer in never smokers, although the molecular mechanisms of NSCLC in nonsmoking women still remain unclear.

In the United States and other Western countries, the 5-year overall survival rate of lung cancer is only 15% and has not improved over several decades. In Taiwan, lung cancer mortality rates have become the highest in the world (17, 18). The high mortality of lung cancer worldwide is largely attributable to the difficulty of

obtaining an early diagnosis and the lack of effective therapeutic methods. To improve survival rates in nonsmoking lung cancer patients, a comprehensive analysis of the molecular signature of the carcinogenic processes in NSCLC is needed to identify better biomarkers for diagnosis and new molecular targets for drug development. One potential set of biomarkers for NSCLC is the semaphorin gene family.

Semaphorins are a large family of secreted, transmembrane, and glycosyl-phosphatidylinositol-linked proteins that were initially discovered to attract/repel growing axons or migrating neurons (19, 20). Recent reports showed that the semaphorin family is expressed not only in the nervous system but also in many other tissues. In addition, this family has been correlated to various carcinogenic processes, particularly angiogenesis and metastasis (19, 21-23), and may play a role as lung tumor suppressors (24-26). However, most studies to date have only focused on the semaphorin class 3 genes, and the details of the broader involvement of semaphorins in tumorigenesis are still unknown.

We conducted a genome-wide analysis of gene expression in nonsmoking female lung cancer patients. Several semaphorin (SEMA) genes in the axonal guidance signaling pathway (e.g., *SEMA5A*, *SEMA6A*) were identified as potential therapeutic targets, and *SEMA5A* can be further used as a prognostic biomarker for NSCLC, which may also be gender specific in Taiwanese patients.

Materials and Methods

Sample collection

One hundred and twenty pairs of tumor and adjacent normal lung tissue specimens were collected from nonsmoking females admitted to National Taiwan University Hospital or Taichung Veterans General Hospital. Written informed consent was obtained from all subjects and/or guardians for the use of their tissue samples. Lung tissue specimens were immediately immersed in RNAlater

buffer (Applied Biosystems), snap-frozen in liquid N₂, and stored at -80°C for RNA extraction. Only those samples in which both specimens in the pair passed quality controls ($n = 60$) were processed for gene expression profiling. The institutional review boards of National Taiwan University Hospital and Taichung Veterans General Hospital both approved the sample acquisition and its subsequent use.

Isolation and amplification of total RNA for gene expression profiling

Total RNAs from the tissue specimens were isolated using TRIzol reagent (Invitrogen) and purified with RNeasy mini kit (Qiagen) according to the manufacturers' instructions. Labeled double-strand cDNA and cRNA were synthesized by using purified total RNA (1-15 µg) as template following Affymetrix standard synthesis protocols. GeneChip Human Genome U133 Plus 2.0 expression arrays (Affymetrix, Inc.) were hybridized to the biotinylated cRNA targets. After 16 hours of hybridization at 45°C, the arrays were washed by a Fluidics Station 450 and scanned by a GeneChip Scanner 3000.

Data mining and statistical analysis

After scanning, the intensity data of GeneChip Human Genome U133 Plus 2.0 expression arrays (Affymetrix) were analyzed by Partek (Partek, Inc.) for mRNA expression levels. Probe-level data were preprocessed, including background correction, quantile normalization, and summarization, using robust multiarray average analysis. After preprocessing of data, principal component analysis (PCA), which reduces higher-dimensional data into a two-dimensional graph, was used to evaluate the similarity of gene expression profiles. To identify differentially expressed genes, paired t tests ($P < 10^{-16}$) and Bonferroni P value adjustments were used. Hierarchical clustering analysis and the Genesis program (27) were used to generate a visual representation of the expression profiles. The similarity of gene expression profiles

Table 1. Clinical characteristics of patients in this study

Characteristics	Microarray		Immunohistochemistry	
	Sample size, n (%)	Age (mean \pm SD, y)	Sample size, n (%)	Age (mean \pm SD, y)
Female	60 (100)	61 \pm 10	49 (100)	63 \pm 10
Tumor type				
Adenocarcinoma	56 (93)	61 \pm 10	41 (84)	63 \pm 11
Bronchioloalveolar carcinoma	3 (5)	66 \pm 9	1 (2)	59
Squamous carcinoma	1 (2)	59	4 (8)	66 \pm 3
Others	—	—	3 (6)	57 \pm 8
Tumor stage				
I + II	47 (78)	61 \pm 11	13 (27)	61 \pm 14
III + IV	13 (22)	61 \pm 7	32 (65)	63 \pm 9
Unknown	—	—	4 (8)	62 \pm 9

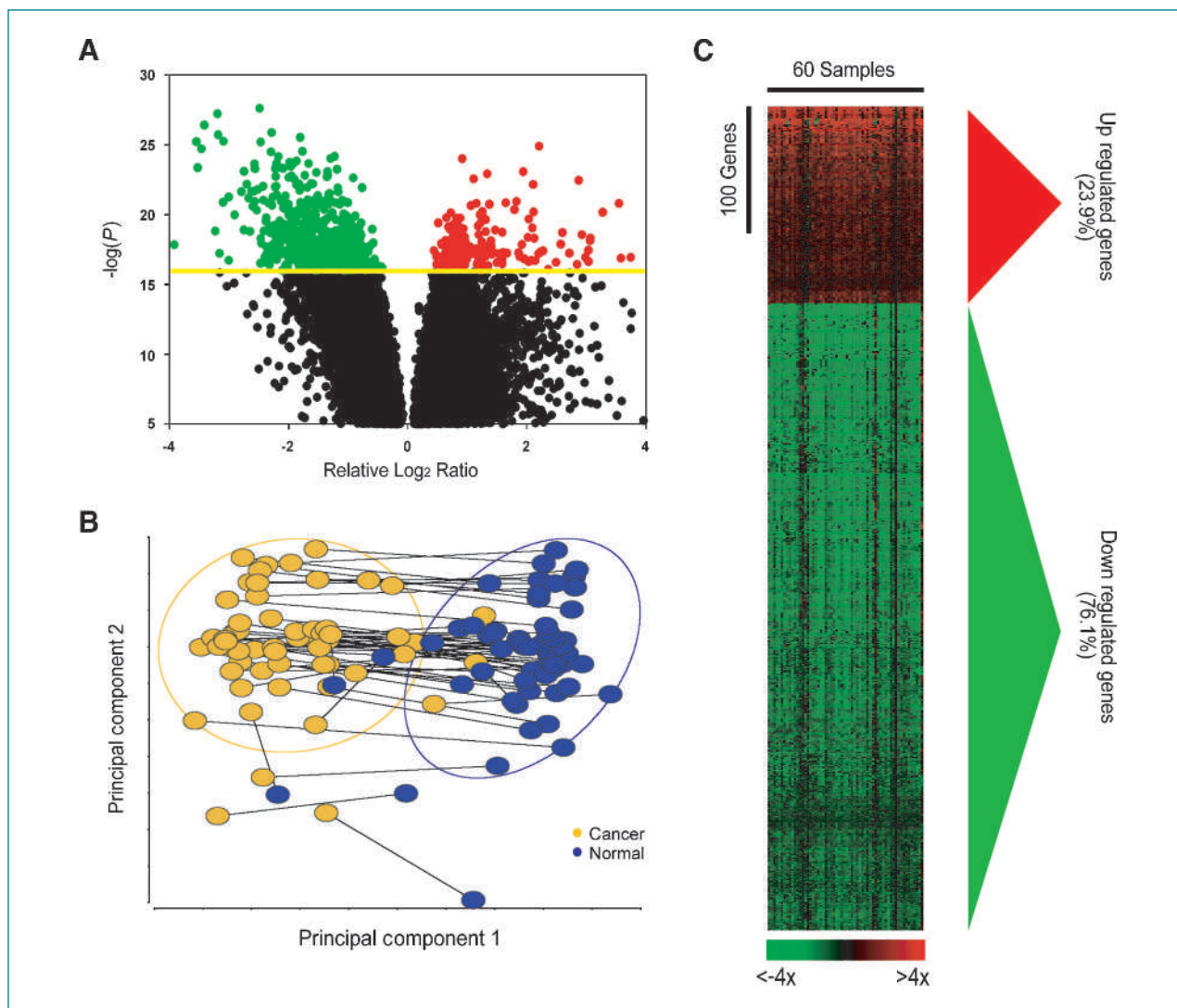


Figure 1. Expression profiles of differentially expressed genes in lung cancer tissues. A, volcano plot of all genes. The $-\log(P)$ of each gene is plotted against the log₂ ratio of cancer intensity to normal intensity. Yellow line indicates the significance level at $P = 10^{-16}$. Red dots indicate genes that were upregulated in cancer tissues; green dots indicate downregulated genes. B, PCA of the 60 paired cancer and normal samples. Gray lines indicate the paired samples from the same individual. Yellow dots denote cancer tissues; blue dots denote normal tissues. Each dot represents the expression values of the significant genes that were summarized at the first two principal component coordinates. C, hierarchical clustering of the 687 significant genes. Input data are the log₂ ratios of cancer intensity over normal intensity. Red indicates genes that were upregulated in cancer tissues; green indicates downregulated genes.

between samples was assessed by Euclidean distance of the log-transformed expression ratios for differentially expressed genes. Finally, Ingenuity Pathway Analysis (Ingenuity Systems, Inc.) was applied to describe gene-gene interaction networks, biological functions, and canonical pathways of differentially expressed genes.

Comparison of axon guidance pathway with independent studies

Two independent studies, GSE7670 (28) and GSE10072 (29), were retrieved from Gene Expression Omnibus (30). To reduce individual variances, only samples with paired

tumor and normal microarray data were selected to conduct further comparison. Gene set enrichment analysis (GSEA; refs. 31, 32) and sigPathway (33) were used to analyze differences in genome-wide expression pathway patterns between tumor and normal tissues within the same individual.

Quantitative reverse transcriptase-PCR

After reverse transcription of total RNA, real-time PCR was done using ABI 7300 (Applied Biosystems) with SYBR Green (Sigma) according to the standard protocols. The PCR primers were as follows: forward primer *SEMA5A*,

5'-TGGAAGACACCTGGACCACATTCA-3', reverse primer *SEMA5A*, 5'-ATCCAGCTCAGGCAGGAA-GAAAGT-3'. Each measurement was made in triplicate and normalized to a *GAPDH* control to ensure comparable amounts of cDNA in all wells.

Immunohistochemical staining

Forty-nine lung tumor tissues, which were not used for gene expression profiling, were used for immunohistochemical analysis. Tissue sections (4- μ m thick) were constructed from paraffin-embedded tissues and used for immunostaining. The tissue sections were deparaffinized in xylene and rehydrated in graded alcohol solutions. The tissue sections were then boiled in citrate buffer (pH 9) for 10 minutes and treated with 3% hydrogen peroxidase to block endogenous peroxidase activity. After washing in PBS solution, the tissue sections were incubated with normal goat serum (dilution 1:500; Dako) for 30 minutes and incubated with SEMA5A antibody (dilution 1:50; Abgent) overnight at 4°C. The next morning, biotin- and streptavidin-labeled antibodies were used for 3,3'-diaminobenzidine staining. More than 50% of the tumor cells that were immunoreactive to SEMA5A antibody were scored as having high SEMA5A expression.

Survival analyses of two independent cohorts

Two independent cohorts (34, 35) with microarray data were selected to further elucidate whether SEMA5A was a prognosis biomarker in Caucasian lung cancer patients. To reduce the variations resulting from different NSCLC subtypes, only patients with adenocarcinoma were examined and were divided into two groups: (a) the "high SEMA5A" group, in which RNA expression levels of SEMA5A were higher than the median SEMA5A expression in all samples, and (b) the "low SEMA5A" group, in which RNA expression levels of SEMA5A were lower than the median SEMA5A expression in all samples. Kaplan-Meier survival analyses were conducted on patients in the high

SEMA5A and low SEMA5A groups to evaluate the association between SEMA5A expression and clinical outcomes.

Results

Clinical characteristics of patients

In this study, we collected 120 pairs of cancer and adjacent normal lung tissue specimens from nonsmoking female lung cancer patients who were admitted to National Taiwan University Hospital or Taichung Veterans General Hospital. Among them, 60 paired samples with good RNA quality were subjected to microarray experiments (Table 1). The mean \pm SD age of patients used for microarray experiments was 61 \pm 10 years. Most of the tumors were adenocarcinomas (93%), and 78% of the samples were in stage I or II.

Gene expression profiling in cancer and normal tissues

Affymetrix GeneChip Human Genome U133 plus 2.0 expression arrays were used to identify differentially expressed genes from cancerous and normal tissues. Because the cancer and normal tissues were from the same individual, paired *t* tests and Bonferroni post hoc *P* value adjustment were used. There were 687 differentially expressed genes with *P* values of $<10^{-16}$. Among them, 523 genes (76.1%) were downregulated and 164 genes (23.9%) were upregulated in cancer tissues (Fig. 1). As shown in Fig. 1A, the $-\log(P)$ of each gene was plotted against the \log_2 ratio of cancer intensity to normal intensity. The number of upregulated genes (ratio >0) was similar to that of downregulated ones (ratio <0), but more downregulated genes (green spots) had *P* values of $<10^{-16}$ than upregulated ones (red spots). Next, PCA was applied to examine whether the differentially expressed genes could be used to distinguish cancer from normal tissues (Fig. 1B). The results of PCA showed that cancer tissues aggregated to the left side, whereas normal tissues clustered to the right, indicating that the

Table 2. Dysregulation of axonal guidance signaling pathway in lung cancer using GSEA and sigPathway algorithms

Algorithm	Pathway	Study	Significance level*	Gene number
GSEA (31, 32)	Axon guidance (HSA04360 in KEGG)	Lu et al.	0.0046 [†]	128
		Su et al. (28)	0.0061 [†]	113
		Landi et al. (29)	0.0988	113
sigPathway (33)	Axon guidance (GO:0007411 in GO)	Lu et al.	0.0001 [†]	106
		Su et al. (28)	0.0080 [†]	56
		Landi et al. (29)	0.0016 [†]	56

Abbreviations: KEGG, Kyoto Encyclopedia of Genes and Genomes; GO, Gene Ontology.

*The significance level was determined by the *P* value for the GSEA algorithm and the *q* value for the sigPathway algorithm.

[†]*P* < 0.01.

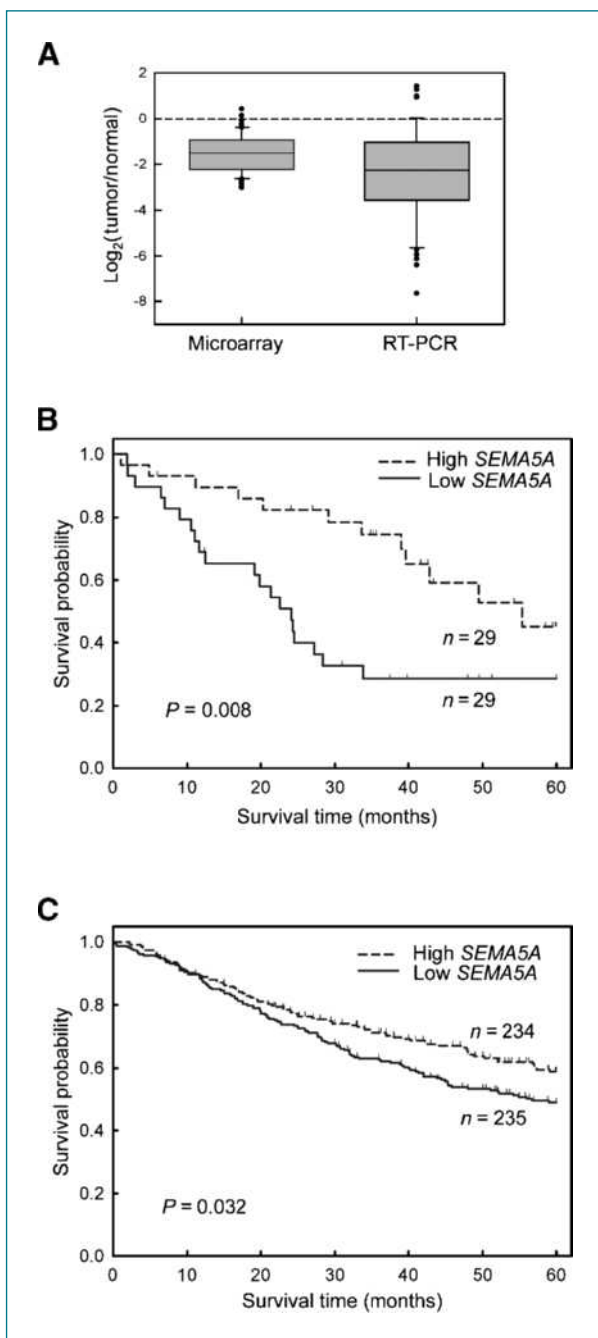


Figure 2. Downregulation of *SEMA5A* in lung cancer tissues is associated with poor overall survival. **A**, box plots of *SEMA5A* expression ratios of tumor to normal tissue by microarray ($n = 60$) and quantitative RT-PCR ($n = 58$). The median for each method is indicated by the center line of the gray box; the first and third quartiles (the interquartile range, IQR) are indicated by the edges of the gray area. The values within 1.5 times the IQR from the upper or lower quartile are indicated by the ends of the lines extending from the IQR, and those >1.5 times the IQR are indicated by dots. *GAPDH* was used for an internal control in quantitative RT-PCR. **B** and **C**, Kaplan-Meier estimates of *SEMA5A* expression with clinical outcome in two independent cohorts of lung adenocarcinoma patients. **B**, the Bild et al. (35) cohort of adenocarcinoma ($n = 58$). **C**, the Shedden et al. (34) cohort of adenocarcinoma ($n = 469$).

differentially expressed genes could be used to separate the tissue samples into two distinct groups.

To validate the expression pattern of these significant genes, two public data sets with tumor and normal tissues from the same individual and with the same microarray platform were compared with our data (28, 29). To reduce variations caused by different versions, probes present in both Affymetrix U133A and U133 plus2.0 arrays ($n = 391$) were selected for further analyses. As shown in Supplementary Fig. S1, these 391 genes showed highly similar expression patterns among our data and that of Su et al. (28) and Landi et al. (29), despite differences in population, gender, and smoking history. This result suggests that the expression profile of differentially expressed genes may not be specific to nonsmoking females with Asian ethnicity, but a general gene expression pattern observed in tumor and normal tissues from the same individual.

Dysregulation of axon guidance signaling pathway in lung cancer

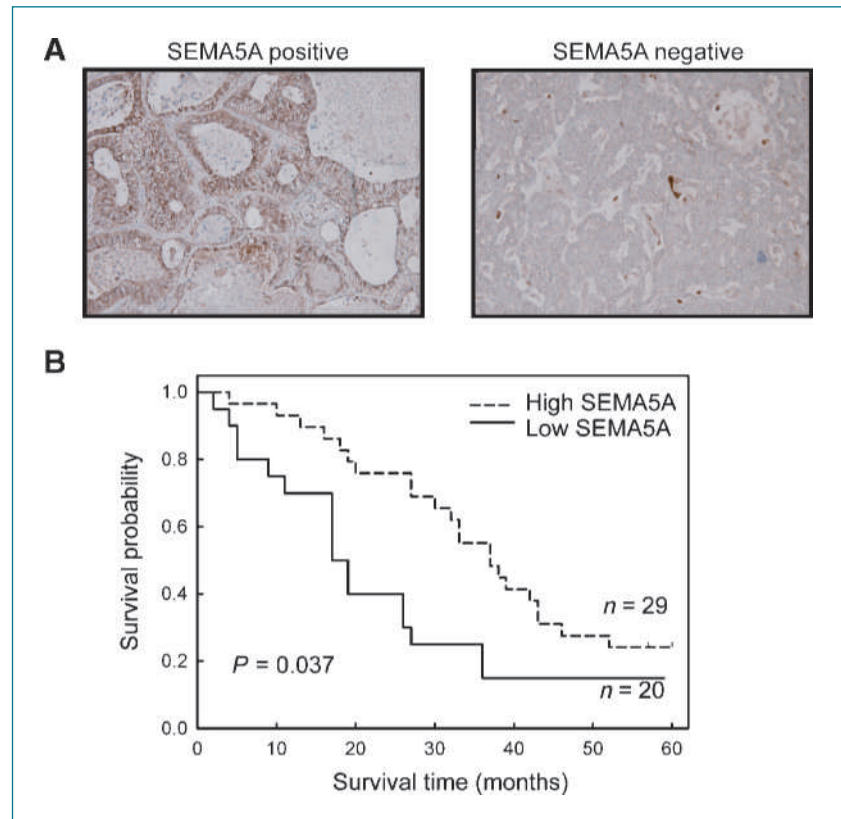
To investigate which functional categories and canonical pathways were significantly dysregulated in tumors compared with normal tissues, Ingenuity Pathway Analysis was carried out. Fisher's exact test identified 16 canonical pathways that were significantly [$-\log(P) > 1.3$] enriched in tumor tissues (Supplementary Table S1). The three pathways with the most significant P value include axonal guidance signaling, ephrin receptor signaling, and angiopoietin signaling (Supplementary Table S1). Angiopoietin signaling molecules have been reported to be associated with lung cancer (36), but it was not clear why genes in the axon guidance and ephrin receptor signaling pathways would be differentially expressed in lung cancer. Because ephrin receptors are a subset of axon guidance signaling molecules (20), further investigation was focused on genes that were involved in the axon guidance signaling pathway.

Once again, the expression levels of genes involved in the axon guidance signaling pathway were compared with Su et al.'s (28) and Landi et al.'s (29) data. As shown in Supplementary Fig. S2, these axon guidance genes ($n = 208$) had very similar expression patterns among the three studies. In addition, considering that the significant genes were identified by an arbitrary P value threshold, we used two other algorithms, GSEA (31, 32) and sig-Pathway (33), which use genome-wide pathway expression patterns, to analyze our data alongside that of Su et al. (28) and Landi et al. (29). Both algorithms identified the axon guidance pathway as significantly dysregulated in both our data and those of the two other studies (Table 2).

Downregulation of *SEMA5A* in tumor tissues is associated with poor clinical outcome

To choose candidate genes for further investigation, the expression ratio and P value of each axon guidance pathway gene were examined. Notably, for the four

Figure 3. Low expression of SEMA5A protein in lung tumor tissues is associated with poor overall survival. A, representative images of immunohistochemistry staining for SEMA5A protein on primary tumor specimens from NSCLC cancer patients ($n = 49$). B, patients were divided into the high SEMA5A group or the low SEMA5A group, with tumor sections showing immunoreactivity to SEMA5A antibody in $\geq 50\%$ or $< 50\%$ of the cancer cells, respectively. Kaplan-Meier estimates of overall survival in an independent cohort of 49 NSCLC cancer patients according to their SEMA5A protein expression levels ($P = 0.037$).



major axon guidance signaling families (20), many genes were significantly downregulated, especially *SEMA5A* ($P < 10^{-19}$; lowest $P = 1.03 \times 10^{-21}$) in the SEMA gene family (Supplementary Table S2). Therefore, we focused on *SEMA5A* for further analysis.

First, to validate the microarray results, quantitative reverse transcriptase-PCR (RT-PCR) was conducted in 58 paired tumor and normal tissues. As shown in Fig. 2A, quantitative RT-PCR results were in good agreement with the microarray data, and *SEMA5A* expression was significantly ($P = 4.1 \times 10^{-13}$) downregulated in tumor tissues.

Second, because most of the patients who provided samples for our microarray experiments are still alive, we chose two other independent cohorts to further investigate the association between *SEMA5A* expression and survival rate (34, 35). To reduce the variations resulting from different NSCLC subtypes, only patients with adenocarcinoma were examined. They were divided into two groups as described in Materials and Methods: the high *SEMA5A* group and the low *SEMA5A* group. Kaplan-Meier survival curves showed that the low *SEMA5A* group had significantly ($P < 0.05$) poorer survival rate than the high *SEMA5A* group (Fig. 2B and C).

Third, in addition to examine the association at the transcriptional level, immunohistochemistry was used

on tissues that had not been used for microarray experiments ($n = 49$) to examine the amount of SEMA5A protein. High SEMA5A expression in tissue samples was defined as being immunoreactive to SEMA5A antibody in $> 50\%$ of the tumor cells (Fig. 3A). The results of Kaplan-Meier survival analysis showed that patients with low SEMA5A protein expression had poor overall survival ($P < 0.05$) compared with the high expression group (Fig. 3B).

Discussion

The high mortality of lung cancer is largely attributable to difficulties in early diagnosis and the lack of effective therapeutic methods. Recent studies suggest that lung cancers arising in nonsmokers are more often adenocarcinoma, are more likely to have epidermal growth factor receptor mutations, and have a better natural history and prognosis with therapy (37). However, the majority of molecular analyses of lung cancer have focused on genetic profiling of pathways responsible for metabolism of tobacco carcinogens. Limited research has been conducted among nonsmokers. In this study, we report a comprehensive analysis of the genetic expression profile from paired tumor and normal tissues among nonsmoking female lung cancer patients in Taiwan. Results revealed that the axon guidance signaling pathway was

significantly dysregulated in lung cancers not attributable to smoking and that the expression of *SEMA5A* was associated with clinical outcome.

It is known that tumor tissues are highly heterogeneous between different individuals. For this reason, this study investigated paired tumor and normal tissues from the same individual, which can dramatically decrease confounding factors from interindividual differences, as shown by the conspicuously low *P* values ($<10^{-16}$) in our results. Further functional and pathway analyses unexpectedly identified the axon guidance signaling pathway as being strongly dysregulated in lung cancer compared with normal cells. Although there is little literature on the topic, a few recent studies have indicated that axon guidance molecules, especially semaphorins, are closely related to angiogenesis, metastasis, and apoptosis in cancer (20, 38, 39). Therefore, exploring lung carcinogenesis by paired tumor and normal samples may help better identify distinct pathways and genes compared with investigating tumor samples only.

Semaphorins are a large family of secreted, transmembrane, glycosyl-phosphatidylinositol-linked proteins initially characterized in the development of the nervous system and axonal guidance. In addition to commanding the outgrowth and regeneration of neurons, many studies indicated that semaphorin genes were strongly associated with cell mobility by regulating the dynamics of actin formation and cytoskeletal changes in cancer tissues (20, 21, 39). However, the functions of semaphorins are promiscuous in that they have been reported as putative tumor suppressors and antiangiogenic factors, and at the same time as mediators of tumor angiogenesis, invasion, and metastasis. Semaphorins may even display divergent activities in different cell types (40). These multifaceted functions may be explained by the involvement of different kinds of semaphorin receptor complexes and by the consequent activation of multiple signaling pathways in different cells or at different stages of the cell cycle.

In lung cancer, two genes in the semaphorin III family have been suggested as potential tumor suppressors: *SEMA3B*, which induces apoptosis of lung cancer cells (24), and *SEMA3F*, the loss of which correlates with advanced grade and stage of lung cancer (41). However, as shown in Supplementary Table S1, many semaphorin genes, especially the semaphorin V family, were significantly downregulated in lung tumor tissues. The functions of the other semaphorins are mainly unknown. In addition to the amino-terminal SEMA domain, the semaphorin V family was characterized by seven thrombospondin repeats functionally important for tumorigenicity and metastasis (21). Although some studies reported that *SEMA5A* and its receptor plexin B3 improved angiogenesis (42, 43) and promoted tumor invasiveness in gastric, pancreatic, and prostate cancer cells (44-46), our results showed that *SEMA5A* was significantly ($P < 10^{-19}$) downregulated in lung tumor tissues. Two previous studies (34, 35) that

conducted survival analyses in Caucasians indicated that adenocarcinomas with higher expression of *SEMA5A* correlated with significantly lower mortality rates, whereas squamous lung cancers did not show such a correlation. Furthermore, immunohistochemistry results revealed that female patients with lower amounts of *SEMA5A* protein had poorer survival rates. Therefore, the functional role of *SEMA5A* in the metastatic progression of different cells and functional stages deserves further investigation.

Importantly, we observed that the association between the *SEMA5A* expression levels and clinical outcome existed only in women, not men (data not shown). This might suggest that *SEMA5A* protein is a potential gender-specific prognostic biomarker of lung cancer patients in Taiwan. Yet, in the transcriptional data, the significance level of the Cox hazard regression model was not changed in Shedden's study (34) after adjusting by gender, suggesting that *SEMA5A* was also an effective prognostic biomarker of lung cancer in Caucasian males. This gender discrepancy between Taiwanese and Caucasians might arise not only from ethnic differences but also from the higher proportion of smokers among Caucasian women. However, as described in some review articles (7, 37, 47), the definition of nonsmokers is ambiguous and inconsistent among different studies. More studies will be required to further evaluate whether smoking and gender are important parameters related to *SEMA5A*.

In conclusion, a comprehensive analysis of the genetic expression profile from paired samples of normal and cancerous lung tissue was conducted. *SEMA5A* was identified as a novel biomarker of lung cancer in nonsmoking women. Furthermore, survival correlated with both transcriptional and translational levels of *SEMA5A* in NSCLC patients in different ethnicities. Thus, *SEMA5A* may be a potential new prognostic biomarker and treatment target for NSCLC.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank the Division of Genomic Medicine, National Taiwan University Research Center for Medical Excellence for financial support; the technicians at National Center of Excellence for Clinical Trials and Research for technical assistance; and Melissa Stauffer, Ph.D., for editing the manuscript.

Grant Support

Department of Health, Taiwan (grant no. DOH98-TD-G-111-014) and National Science Council, Taiwan (grant no. 98-2320-B-002-044-MY3).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Received 03/30/2010; revised 05/29/2010; accepted 06/14/2010; published OnlineFirst 08/27/2010.

References

- Nugent WC, Edney MT, Hammerness PG, et al. Non-small cell lung cancer at the extremes of age: impact on diagnosis and treatment. *Ann Thorac Surg* 1997;63:193-7.
- Wood ME, Kelly K, Mullineaux LG, Bunn PAJ. The inherited nature of lung cancer: a pilot study. *Lung Cancer* 2000;30:135-44.
- Wu AH, Fonham ET, Reynolds P, et al. Previous lung disease and risk of lung cancer among lifetime nonsmoking women in the United States. *Am J Epidemiol* 1995;141:1023-32.
- Wu AH, Yu MC, Thomas DC, Pike MC, Henderson BE. Personal and family history of lung disease as risk factors for adenocarcinoma of the lung. *Cancer Res* 1988;48:7279-84.
- Alavanja MC, Brown CC, Swanson C, Brownson RC. Saturated fat intake and lung cancer risk among nonsmoking women in Missouri. *J Natl Cancer Inst* 1993;85:1906-16.
- De Stefani E, Deneo-Pellegrini H, Mendilaharsu M, Carzoglio JC, Ronco A. Dietary fat and lung cancer: a case-control study in Uruguay. *Cancer Causes Control* 1997;8:913-21.
- Samet JM, Avila-Tang E, Boffetta P, et al. Lung cancer in never smokers: clinical epidemiology and environmental risk factors. *Clin Cancer Res* 2009;15:5626-45.
- Ger LP, Liou SH, Shen CY, Kao SJ, Chen KT. Risk factors of lung cancer. *J Formos Med Assoc* 1992;91 Suppl 3:S222-31.
- Hirayasu T, Iwamasa T, Kamada Y, et al. Human papillomavirus DNA in squamous cell carcinoma of the lung. *J Clin Pathol* 1996;49:810-7.
- Hainaut P, Pfeifer GP. Patterns of p53 G→T transversions in lung cancers reflect the primary mutagenic signature of DNA-damage by tobacco smoke. *Carcinogenesis* 2001;22:367-74.
- Toyooka S, Tsuda T, Gazdar AF. The TP53 gene, tobacco exposure, and lung cancer. *Hum Mutat* 2003;21:229-39.
- Shigematsu H, Lin L, Takahashi T, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. *J Natl Cancer Inst* 2005;97:339-46.
- Martin P, Kelly CM, Carney D. Epidermal growth factor receptor-targeted agents for lung cancer. *Cancer Control* 2006;13:129-40.
- Eberhard DA, Johnson BE, Amler LC, et al. Mutations in the epidermal growth factor receptor and in KRAS are predictive and prognostic indicators in patients with non-small-cell lung cancer treated with chemotherapy alone and in combination with erlotinib. *J Clin Oncol* 2005;23:5900-9.
- Yamamoto H, Shigematsu H, Nomura M, et al. PIK3CA mutations and copy number gains in human lung cancers. *Cancer Res* 2008;68:6913-21.
- Wong DW, Leung EL, So KK, et al. The EML4-ALK fusion gene is involved in various histologic types of lung cancers from nonsmokers with wild-type EGFR and KRAS. *Cancer* 2009;115:1723-33.
- Chen CJ, Wu HY, Chuang YC, et al. Epidemiologic characteristics and multiple risk factors of lung cancer in Taiwan. *Anticancer Res* 1990;10:971-6.
- Yang SP, Luh KT, Kuo SH, Lin CC. Chronological observation of epidemiological characteristics of lung cancer in Taiwan with etiological consideration—a 30-year consecutive study. *Jpn J Clin Oncol* 1984;14:7-19.
- Potiron VA, Roche J, Drabkin HA. Semaphorins and their receptors in lung cancer. *Cancer Lett* 2008;273:1-14.
- Chedotal A, Kerjan G, Moreau-Fauvarque C. The brain within the tumor: new roles for axon guidance molecules in cancers. *Cell Death Differ* 2005;12:1044-56.
- Kruger RP, Aurandt J, Guan KL. Semaphorins command cells to move. *Nat Rev Mol Cell Biol* 2005;6:789-800.
- Nagai H, Sugito N, Matsubara H, et al. CLCP1 interacts with semaphorin 4B and regulates motility of lung cancer cells. *Oncogene* 2007;26:4025-31.
- Nasarre P, Kusy S, Constantin B, et al. Semaphorin SEMA3F has a repulsing activity on breast cancer cells and inhibits E-cadherin-mediated cell adhesion. *Neoplasia* 2005;7:180-9.
- Castro-Rivera E, Ran S, Thorpe P, Minna JD. Semaphorin 3B (SEMA3B) induces apoptosis in lung and breast cancer, whereas VEGF165 antagonizes this effect. *Proc Natl Acad Sci U S A* 2004;101:11432-7.
- Tomizawa Y, Sekido Y, Kondo M, et al. Inhibition of lung cancer cell growth and induction of apoptosis after reexpression of 3p21.3 candidate tumor suppressor gene SEMA3B. *Proc Natl Acad Sci U S A* 2001;98:13954-9.
- Brambilla E, Constantin B, Drabkin H, Roche J. Semaphorin SEMA3F localization in malignant human lung and cell lines: a suggested role in cell adhesion and cell migration. *Am J Pathol* 2000;156:939-50.
- Sturm A, Quackenbush J, Trajanoski Z. Genesis: cluster analysis of microarray data. *Bioinformatics* 2002;18:207-8.
- Su LJ, Chang CW, Wu YC, et al. Selection of DDX5 as a novel internal control for Q-RT-PCR from microarray data using a block bootstrap re-sampling scheme. *BMC Genomics* 2007;8:140.
- Landi MT, Dracheva T, Rotunno M, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One* 2008;3:e1651.
- Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002;30:207-10.
- Mootha VK, Lindgren CM, Eriksson KF, et al. PGC-1 α -responsive genes involved in oxidative phosphorylation are coordinately down-regulated in human diabetes. *Nat Genet* 2003;34:267-73.
- Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545-50.
- Tian L, Greenberg SA, Kong SW, et al. Discovering statistically significant pathways in expression profiling studies. *Proc Natl Acad Sci U S A* 2005;102:13544-9.
- Shedden K, Taylor JM, Enkemann SA, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008;14:822-7.
- Bild AH, Yao G, Chang JT, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006;439:353-7.
- Shibuya M. Vascular endothelial growth factor-dependent and -independent regulation of angiogenesis. *BMB Rep* 2008;41:278-86.
- Rudin CM, Avila-Tang E, Harris CC, et al. Lung cancer in never smokers: molecular profiles and therapeutic implications. *Clin Cancer Res* 2009;15:5646-61.
- Neufeld G, Shraga-Heled N, Lange T, et al. Semaphorins in cancer. *Front Biosci* 2005;10:751-60.
- Zhou Y, Gunput RA, Pasterkamp RJ. Semaphorin signaling: progress made and promises ahead. *Trends Biochem Sci* 2008;33:161-70.
- Rizzolio S, Tamagnone L. Semaphorin signals on the road to cancer invasion and metastasis. *Cell Adh Migr* 2007;1:62-8.
- Kusy S, Nasarre P, Chan D, et al. Selective suppression of *in vivo* tumorigenicity by semaphorin SEMA3F in lung cancer cells. *Neoplasia* 2005;7:457-65.
- Sadanandam A, Rosenbaugh EG, Singh S, Varney M, Singh RK. Semaphorin 5A promotes angiogenesis by increasing endothelial cell proliferation, migration, and decreasing apoptosis. *Microvasc Res* 2009;79:1-9.
- Fiore R, Rahim B, Christoffels VM, Moorman AF, Puschel AW. Inactivation of the SEMA5A gene results in embryonic lethality and defective remodeling of the cranial vascular system. *Mol Cell Biol* 2005;25:2310-9.
- Sadanandam A, Varney ML, Kinarsky L, et al. Identification of functional cell adhesion molecules with a potential role in metastasis by a combination of *in vivo* phage display and *in silico* analysis. *OMICS* 2007;11:41-57.
- Pan GQ, Ren HZ, Zhang SF, Wang XM, Wen JF. Expression of semaphorin 5A and its receptor plexin B3 contributes to invasion and metastasis of gastric carcinoma. *World J Gastroenterol* 2009;15:2800-4.
- Artigiani S, Conrotto P, Fazzari P, et al. Plexin-B3 is a functional receptor for semaphorin 5A. *EMBO Rep* 2004;5:710-4.
- Rudin CM, Avila-Tang E, Samet JM. Lung cancer in never smokers: a call to action. *Clin Cancer Res* 2009;15:5622-5.