Identification of Gene Signatures and Molecular Markers for Human Lung Cancer Prognosis using an In vitro Lung Carcinogenesis System

Humam Kadara,1 Ludovic Lacroix,1 Carmen Behrens,1,2 Luisa Solis,1,2 Xuemin Gu,3 J. Jack Lee,3 Eiji Tahara,1 Dafna Lotan,1 Waun Ki Hong,1 Ignacio I. Wistuba1,2 and Reuben Lotan1

Abstract
Lung cancer continues to be a major deadly malignancy. The mortality of this disease could be reduced by improving the ability to predict cancer patients’ survival. We hypothesized that genes differentially expressed among cells constituting an in vitro human lung carcinogenesis model consisting of normal, immortalized, transformed, and tumorigenic bronchial epithelial cells are relevant to the clinical outcome of non–small cell lung cancer (NSCLC). Multidimensional scaling, microarray, and functional pathways analyses of the transcriptomes of the above cells were done and combined with integrative genomics to incorporate the microarray data with published NSCLC data sets. Up-regulated (n = 301) and down-regulated genes (n = 358) displayed expression level variation across the in vitro model with progressive changes in cancer-related molecular functions. A subset of these genes (n = 584) separated lung adenocarcinoma clinical samples (n = 361) into two clusters with significant survival differences. Six genes, UBE2C, TPX2, MCM2, MCM6, FEN1, and SFN, selected by functional array analysis, were also effective in prognosis. The mRNA and protein levels of one these genes—UBE2C—were significantly up-regulated in NSCLC tissue relative to normal lung and increased progressively in lung lesions. Moreover, stage I NSCLC patients with positive UBE2C expression exhibited significantly poorer overall and progression-free survival than patients with negative expression. Our studies with this in vitro model have lead to the identification of a robust six-gene signature, which may be valuable for predicting the survival of lung adenocarcinoma patients. Moreover, one of those genes, UBE2C, seems to be a powerful biomarker for NSCLC survival prediction.

In 2008, 215,020 new cases, and 161,840 deaths due to lung cancer were expected in the United States, accounting for 31% of all cancer deaths (1). Lung cancer mortality is high because most cancers are diagnosed after regional or distant spread of the disease had already occurred (2, 3). It is noteworthy that even the 5-year survival rate of stage I lung cancer is among the worst for early-stage disease of all other malignancies (1, 4). It is plausible to assume that the mortality of patients will decrease if progress is made in identification of effective prognostic molecular biomarkers.

Lung carcinogenesis involves the accumulation of genetic and epigenetic alterations that occur over a long course due to chronic exposure to carcinogens such as tobacco smoke or to genetic susceptibility factors (2). Several early changes that occur during lung carcinogenesis have been identified including mutations of TP53 and KRAS (5, 6), silencing of retinoic acid receptor β (7), inactivation of the cyclin-dependent kinase inhibitor p16/CDKN2A (8), epidermal growth factor receptor amplification and mutations in adenocarcinomas, and amplification and mutations of the tyrosine kinase receptors, HER2 and MET (2, 3). However, all of the abovementioned changes account for <60% of human lung cancers. Moreover, compared with substantial knowledge about the malignant stage, our understanding of the molecular changes occurring early in lung carcinogenesis is still lacking.

One well-characterized system for studying changes that occur at different stages of lung carcinogenesis consists of normal human bronchial epithelial (NHBE) cells, NHBE cells...
immortalized with SV40 T/Adeno12 virus (BEAS-2B), and three cell lines derived from BEAS-2B after s.c. growth as xenotransplants in nude mice; immortalized (1799), transformed (1198), and tumorigenic (1170-I). The latter two were isolated after exposure of BEAS-2B transplants to cigarette smoke condensate in vitro (9). The study of this in vitro human lung carcinogenesis model offers opportunities to identify different progressive molecular changes that are relevant to human lung cancer development.

In this study, we identified genes that are expressed differentially and progressively among the cells that constitute the in vitro model that helped us identify a six-gene signature that is capable of predicting the survival of lung adenocarcinoma. Furthermore, we showed the sequential increase of ubiquitin-conjugating enzyme E2C (UBE2C) protein levels in lung lesions of various stages as well as its up-regulation in non–small cell lung cancer (NSCLC) tissue specimens and potential as a powerful molecular marker of early stage NSCLC prognosis.

Materials and Methods

Cell culture and tissue specimens

NHBE cells and normal small airway epithelial cells (SAEC) were purchased from Cambrex/Clonetics and used at the second passage in our laboratory. The in vitro lung carcinogenesis model that includes SV40 large T-immortalized NHBE cells (BEAS-2B and 1799), transformed (1198), and tumorigenic (1170-I) cells derived from BEAS-2B by exposure to cigarette smoke condensate during in vitro growth as xenotransplants (9) were obtained from Dr. Klein-Szanto (Fox Chase Cancer Center, Philadelphia, PA). All of the above cells were grown in Keratinocyte Serum-Free Medium (Life Technologies, Inc.) containing epidermal growth factor and bovine pituitary extract at 37°C in a humidified atmosphere of 95% air and 5% CO2.

RNA extraction

Total RNA was purified from cultured cells or frozen tissues using the RNeasy Mini kit according to the manufacturer's instruction (Qiagen, Inc.). RNA was treated with DNase provided by the manufacturer for elimination of genomic DNA. Extracted total RNA was quantified using the Nanodrop ND-1000 spectrophotometer (Thermo Fisher Scientific). RNA quality, based on the 28S/18S rRNAs ratio, was assessed using the Experion automated electrophoresis system (Bio-Rad Laboratories) according to the manufacturer's instructions. A total of 1 μg of RNA was reverse-transcribed using the Quantitect Reverse Transcription kit (Qiagen, Inc.) for first-strand cDNA synthesis with random primers according to the manufacturer's instructions and diluted in nuclease-free water.

Microarray sample preparation, hybridization, scanning, and analysis

All steps leading to generation of raw microarray data were processed at the University of Texas MD Anderson Murine Microarray and Affymetrix Facility. Array synthesis and cleanup of cRNA from double-stranded cDNA, fragmented cRNAs (15 μg) were hybridized to 12 (six cell lines in duplicate) GeneChip Human Genome U133A arrays (Affymetrix), according to the manufacturer's instructions and as previously described (10). The arrays were scanned with a GeneChip Scanner 3000 from Affymetrix and raw image files were converted to probe set data (*.CEL files), using the Affymetrix GeneChip Operating Software. Microarray analysis is detailed in the supplementary material accompanying this manuscript. Expression data were deposited into the Gene Expression Omnibus (GSE accession #17073).

Immunohistochemistry analysis

Immunohistochemistry was done on histology sections of formalin-fixed paraffin-embedded tissue samples, using the purified primary rabbit polyclonal anti-human UBE2C/UBCH10 antibody (A-650; Boston Biochem) at a dilution of 1:500. The sections were deparaffinized, hydrated, subjected to antigen retrieval by heating in a steamer for 20 min with 10 mmol/L sodium citrate (pH 6.0), and then incubated in peroxidase blocking reagent (DAKO). Sections were then washed with Tris-containing buffer and incubated overnight at 4°C with the primary anti-UBE2C antibody. Subsequently, the sections were washed and incubated with the secondary antibody (goat anti-rabbit) using the Evison plus labeled polymer kit (DAKO) for 30 min followed by incubation with avidin-biotin-peroxidase complex (DAKO) and development with diaminobenzidine chromogen for 5 min. Finally, the sections were rinsed in distilled water, counterstained with hematoxylin (DAKO), and mounted on glass slides before evaluation under the microscope. Formalin-fixed and paraffin-embedded pellets from lung cancer cell lines displaying UBE2C expression by Western blot analysis were used as a positive control, whereas samples processed similarly except for the omission of the primary antibody were used as negative controls. A lung cancer pathologist examined the UBE2C immunostaining using light microscopy. Only nuclear UBE2C expression was evaluated because most of the UBE2C immunoreactivity was detected in the nucleus. Nuclear UBE2C immunostaining was quantified using a range of 0 to 100 according to the percentage of positive nuclei present among all tumor or epithelium cells present in the tissue microarray (TMA) cores.

Statistical analyses

The data were summarized using standard descriptive statistics. The rank-based nonparametric Wilcoxon rank-sum test was used to assess the significance of the differences in nuclear UBE2C score among all normal, preneoplastic, and malignant lung tissue histologic sections. The association between nuclear UBE2C score and NSCLC patients' smoking status was also assessed using the Wilcoxon rank-sum test. The association between nuclear UBE2C expression and patient survival was analyzed as previously described (11). Continuous UBE2C nuclear scores were dichotomized into two categories (UBE2C negative and UBE2C positive) using the Classification and Regression Tree algorithm for both the overall and progression-free survival of patients, which were compared by the Kaplan-Meier method for estimation of survival probability using the R 2.6.0 statistical package.

Results

Identification of differential gene expression in normal, immortalized, transformed, and tumorigenic lung epithelial cells

Multidimensional scaling analysis of the transcriptome of the in vitro lung carcinogenesis cell constituents revealed that normal (NHBE and SAEC), immortalized (BEAS-2B and derived 1799), transformed (1198), and tumorigenic (1170-I) cells were positioned in different coordinates within the planes, with the normal and tumorigenic cells being the farthest apart (Fig. 1A). The two normal cell strains (NHBE and SAEC) were almost super-imposable, with the 1799 cells resembling more closely the transformed and tumorigenic cells, which were in close proximity to each other. The initial analysis of the Affymetrix U133A chips gene expression data identified 1221 gene features that were differentially expressed by at least 1.65-fold in any of the indicated cells compared with the NHBE cells (Fig. 1B). SOM analysis and unsupervised clustering by Pearson correlation and average

4 www.r-project.org
linkage identified 659 genes (301 upregulated and 358 downregulated; red and green, respectively) displaying progressive variation from the NHBE to the 1170-I cells (Fig. 1C). Genes with progressive variation by at least 2-fold across cells of the in vitro human lung carcinogenesis model were further functionally analyzed using Ingenuity Pathways Analysis (IPA®). Cell growth and proliferation, cell death, cell cycle, DNA recombination and repair, and cell to cell interaction gene sets were estimated by IPA® to be altered in function significantly (measured as -log of the P value) and progressively across cells of the in vitro model (Fig. 1D). In addition, the number of genes differentially expressed relative to NHBE cells and in four molecular gene sets (indicated above the bars) was highest in the 1170-I tumorigenic cells and lowest in the normal SAEC (Fig. 1D).

Significant modulation of high-interaction gene networks in the tumorigenic 1170-I cells relative to the NHBE cells

We next tried to identify genes predicted to be important in the progression of the in vitro human lung carcinogenesis model through both their modulation in expression and molecular interactions. Gene interaction network analysis of the 1221 gene features differentially expressed between the NHBE cells and tumorigenic 1170-I lung epithelial cells by IPA® revealed the significant (score; -log of the P value) modulation of functional networks related to DNA recombination and repair, cell cycle, cell death, cellular assembly and organization, and cell-to-cell signaling and interaction (Fig. 2A). The expression of Minichromosome maintenance (MCM)2 and 6, Stratifin (SFN), Flap structure-specific endonuclease 1 (FEN1), Targeting protein for Xklp2 (TPX2), and UBE2C is depicted in a representative gene network in the 1799, 1198, and 1170-I cells compared with NHBE cells (Fig. 2B). The expression levels of UBE2C, TPX2, MCM2, MCM6, as well as FEN1 were increased (indicated by the red color) in the premalignant and tumorigenic cells relative to the NHBE cells, whereas the expression level of SFN was decreased (indicated by the green color).

Integrative genomics analysis of the association of differentially and progressively expressed genes within the human in vitro lung carcinogenesis model with NSCLC gene expression patterns and clinical outcome

To explore the relevance of this in vitro model to NSCLC, we developed a gene signature composed of 584 genes based
on a differential expression of at least 2-fold between the 1170-I tumorigenic and the NHBE cells and with a statistical significance of $P$ value of $<0.001$ of an applied univariate $t$ test with permutation, as well as displaying progressive variation across all cells of the in vitro model. The expression of these genes was then analyzed in 361 adenocarcinomas from the study by Shedden et al. (12) as described in the Supplementary Materials and Methods and integrated with their expression in the NHBE and 1170-I cells to create a composite gene expression data set depicted in Fig. 3A. Hierarchical cluster analysis of the integrated data revealed that the 361 patients could be divided into two groups or clusters comprising either the NHBE cells (NHBE cluster, highlighted in blue) or the 1170-I tumorigenic lung cells (1170-I cluster, highlighted in red; Fig. 3B). In addition, Kaplan-Meier plots and log-rank survival statistics showed that the lung adenocarcinoma patients from the 1170-I expression cluster displayed significantly poorer overall survival ($P = 0.0009$) and progression-free survival ($P = 0.03$) than patients from the NHBE expression subgroup (Fig. 3C). These findings show that the gene expression patterns of cells of the human in vitro lung carcinogenesis model used in this study are highly evident in NSCLC clinical samples and associated with lung adenocarcinoma prognosis.

Analysis of the prognostic potential of the six selected genes in lung adenocarcinoma patients

The functional pathways analysis of the transcriptome of the human in vitro lung carcinogenesis model enabled us to select six differentially expressed genes prominent in high gene interactions (UBE2C, MCM2, MCM6, TPRX2, FEN1, and SFN). We next assessed the expression of these genes in published microarray data sets of lung adenocarcinoma cohorts. Hierarchical cluster analysis revealed that lung adenocarcinoma patients in each of the three cohorts (Shedden et al., Bhattacharjee et al., and Bild et al.; refs. 12–14) were each divided into two subgroups or clusters based on the expression of the selected genes alone (Fig. 4A, B, and C). In addition, Kaplan-Meier plots and log-rank statistics showed that the two divided lung adenocarcinoma patient clusters of the analyzed cohort published by Shedden et al. ($n = 361$; ref. 12) exhibited significant differences in overall survival ($P = 0.004$) and progression-free survival ($P = 0.04$; Fig. 4A). Consistently, the identified clusters in each of both the Bhattacharjee et al. (13) and Bild et al. (14) lung adenocarcinoma data sets exhibited significant differences in overall survival ($P = 0.004$ and $P = 0.02$, respectively; Fig. 4B and C). These findings show that the human in vitro lung carcinogenesis model is powerful for generation of gene classifiers associated with prognosis of lung adenocarcinoma.

Differential expression of UBE2C mRNA in NSCLC compared with adjacent normal lung tissues as revealed by in silico and quantitative real-time PCR analyses

Using Oncomine, a publicly available microarray analysis tool and platform (15), we found that the mRNA one of the six selected genes, UBE2C, was significantly elevated in human lung adenocarcinomas relative to adjacent normal lung tissue in four microarray databases and cohorts (all $P < 0.05$; Supplementary Fig. S1A). Moreover, in silico analysis of gene expression data from the Garber et al. cohort (16) also revealed the significant up-regulation of UBE2C in squamous cell carcinomas (SCC) relative to adjacent normal lung tissues ($P < 0.001$; Supplementary Fig. S1A). We also analyzed the mRNA levels of UBE2C in a set of frozen human lung adenocarcinomas tissues ($n = 26$) and adjacent normal lung tissues ($n = 24$) and found that they were significantly higher in the lung adenocarcinomas relative to adjacent normal lung ($P < 0.001$; Supplementary Fig. S1B).

Analysis of UBE2C protein expression by immunohistochemistry in normal, preneoplastic, and malignant lung tissue specimens and assessment of its value in lung cancer prognosis

We analyzed UBE2C expression at the protein level by immunohistochemistry using histologic tissue sections that included normal bronchial epithelia, preneoplastic lesions, SCCs ($n = 98$), and adenocarcinomas ($n = 141$). UBE2C protein was localized to the cell nucleus in most cases and its nuclear expression was very low in normal bronchial epithelia but was higher in preneoplastic and malignant lung lesions (Fig. 5A and B). In addition, UBE2C expression was statistically significantly different in SCCs ($n = 98$), or adenocarcinomas ($n = 141$), when compared with normal bronchial epithelia ($n = 62$; *, $P < 0.001$) or hyperplasias ($n = 61$; **, $P < 0.001$; Fig. 6A). In addition, squamous metaplasias ($n = 15$), dysplasias ($n = 9$), and carcinoma in situ ($n = 26$) exhibited significantly higher levels of nuclear UBE2C when compared with normal bronchial epithelia (*, $P < 0.001$) or hyperplasias (**, $P < 0.001$). Moreover, nuclear UBE2C levels were statistically significantly higher in hyperplasias relative to the levels in normal bronchial epithelia ($P < 0.001$; Fig. 6A).

Our analyses also showed that nuclear UBE2C protein expression was statistically different between SCCs and lung adenocarcinomas ($P < 0.001$; Fig. 6A). We also analyzed the correlation of nuclear UBE2C expression with clinicopathologic variables examined for all NSCLC histologic tissue specimens. There were no statistically significant correlations between nuclear UBE2C expression and age, gender, or disease stage in all NSCLC tissues or when adenocarcinomas or SCCs were analyzed alone (data not shown). However, nuclear UBE2C levels were significantly higher in NSCLC tissue resected from smokers ($n = 205$) relative to those resected from nonsmokers ($n = 34$; $P < 0.001$; Fig. 6B).

We next assessed the clinical relevance of nuclear UBE2C expression in early-stage lung cancer. Log-rank statistics and Kaplan-Meier plots revealed that stage I NSCLC patients with positive UBE2C expression ($n = 118$) exhibited significantly poorer overall survival and progression-free survival than patients with negative nuclear UBE2C expression ($n = 53$; Fig. 6C and D). Moreover, multivariate Cox proportional hazard regression analyses revealed that nuclear UBE2C expression was an independent predictor of progression-free survival ($P = 0.05$) but not overall survival (data not shown). These results show the potential role of aberrant nuclear UBE2C expression in early-stage NSCLC progression and clinical outcome.

Discussion

Lung cancer is the leading cause of cancer deaths in the United States with the overall 5-year survival rate of <15%
### Table

<table>
<thead>
<tr>
<th>Top categories</th>
<th>No. focus genes</th>
<th>Score</th>
<th>Genes in the network (modulated in 1170-I compared to NHBE)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA Replication, recombination, Repair, Cell Cycle, Cellular assembly and organization</td>
<td>35</td>
<td>47</td>
<td>A9S4A4, AURKA, CCNA2, CCNB1, CCNB2, CDC2, CDC5, CDC7, CDC20, CDC25B, CDC25C, CDC45L, CDT1, CES1B, CSTA, FOS, FOXM1, P231, NMRB, LAMP2, LTBP2, MCM2, MCM3, MCM4, MCM5, MCM6, MCM7, MPP10, PTTG1, RB1, RUNBL2, SERPINB7, TGR11, TPX2, UBE2C</td>
</tr>
<tr>
<td>Cancer, Cell-To-Cell Signaling and Interaction, Cellular Compromise</td>
<td>35</td>
<td>47</td>
<td>BUB1, BUB1B, CD44, CD7, CENPE, CENPF, CSEI1, C3P2G, C3S8, DKK1, EIF5A, F3, FBLN4, FLNB, FLOT1, FN1, GB33, ITGB6, KPN2, KPNB1, LTBP1, NOX1, PTX4, PTX1, PTX3, PTX5, RANIP1, REC1 (includes EG1104), S100A10, S100A11, SFPI, TGF1, TGM1, THBS1, THY1, TNC, T2X1, TXN</td>
</tr>
<tr>
<td>DNA Replication, Recombination, and Repair, Cell Death</td>
<td>35</td>
<td>47</td>
<td>APLP2, APP, ATAD2, BIK, CT2, CT5, CD9, CDN2A, DENRF, DONSON, FABPS, FEN1, G1P2, GLUL, HLA-C, IER3, IFI16, ITTM1, ITM2B, IIAA0101, MAP1B, MCL1, MKL1, MSH5, PCNA, RAB3B, RAE4C, S100A7, SNCA, TSC1, TMEER3, TNSF10, TNC1, U0K1L</td>
</tr>
<tr>
<td>Cell cycle, and cellular assembly organization</td>
<td>26</td>
<td>27</td>
<td>AURKB, BRCA5, BRN1, CDC6, CDH2, CENPA, CENP1, DES, DGA1A, FSCN1, HCAP30, HCAP-G, HCAP-H2, HSP90B1, HSP90B2, JUND, KIF23, KIF4A, LMA1, LMB1, LUCP5, NASP, PRC1, RNR1, PGE3, RACAP1, SMC1L1, SMC4L1, SDI1, STIP1, TNF, TOP2A, TWIST1, UPP1, VIM, YWHAZ</td>
</tr>
</tbody>
</table>

### Diagram

![Diagram of gene network models](attachment:image.png)

**NHBE**

**1799**

**1198**

**1170-I**
However, the survival of stage I or early stage NSCLC patients is significantly higher than that of patients with advanced lung cancers (2, 3). Therefore, a considerable effort has been mounted over the last few years to apply the progress in various high throughput methods for the identification of potential molecular markers for diagnosis and prognosis (2, 17). Most of these studies were based on comparisons of gene expression in tumors relative to normal appearing tissue and almost no studies addressed changes in premalignant tissue, most likely due to the paucity of material available for high throughput analysis. Previously, several in vitro approaches have been applied to address the shortage of premalignant tissue. For example, ectocervical keratinocytes and mammary epithelial cells have been immortalized using human papillomavirus (18, 19), and normal bronchial epithelial cells were immortalized using telomerase plus CDK4 (20).

NSCLC develops by a multistep process that involves the accumulation of many genetic aberrations over a long period of time. Interestingly, most of these genetic changes (such as p53 loss or mutation or epidermal growth factor receptor...

---

**Fig. 3.** The association of genes differentially expressed in the in vitro model of lung carcinogenesis with lung adenocarcinoma gene expression patterns and prognosis. A and B, the 584 genes were selected as described in the Supplementary Materials and Methods. The genes were median centered independently in the lung epithelial cells and adenocarcinomas (n = 361) before integration, further filtered to include those with modulation in expression of at least 2-fold in at least eight observations and then analyzed by hierarchical cluster analysis with average linkage. Data are represented in a matrix format in which individual columns represent single gene features and rows represent experiments. High (red) or low (green) gene expression levels are indicated, respectively, as indicated by the log2 transformed scale bar. C, Kaplan-Meier plots for the overall survival (OS) and progression-free survival (PFS) of lung adenocarcinoma patients separated according to clustering patterns (blue, NHBE cluster; red, 1170-I cluster). The number of analyzed samples of each cluster in both plots is indicated as N next to the plotted arms. P values were obtained by the log-rank test.

**Fig. 2.** Gene interaction network analysis of genes differentially expressed in the human in vitro lung carcinogenesis model. A, top four gene networks generated from IPA and significantly modulated between the 1170-I tumorigenic and NHBE cells. Gene symbols (blue) are presented; arrows, variation of their gene expression in the 1170-I tumorigenic cells relative to NHBE cells. The network score was calculated by the inverse log of the P value and indicates the likelihood of focus genes in a network being found together than due to chance. The number of focus genes (in bold) refers to the genes differentially regulated and within an IPA network composed of a maximum number of 35 genes. B, functional pathway analysis by IPA of selected genes and their interaction nodes in 1799, 1198, and 1170-I cells relative to NHBE cells. The selected genes (MCM2, MCM6, SFN, TPX2, UBE2C, and FEN1) are highlighted by a blue border. Gene expression variation by at least 2-fold is depicted by color (red, up-regulated; green, down-regulated; gray, no significant change).
mutation and amplification) occur very early in the premalignant stages and persist during the carcinogenesis process (2). Our study was aimed to gain information on genetic changes that occur at different early stages of lung carcinogenesis and can be relevant to the clinical outcome of the disease. Towards this, we used an in vitro lung carcinogenesis cell model composed of normal, immortalized, transformed, and tumorigenic lung epithelial cells, the latter two developed in vivo in mice that had the immortalized BEAS-2B transplanted and exposed to cigarette smoke condensate (9). We did global gene expression analysis on the transcriptome of these cells and showed the gradual modulation of genes and their associated cancer-related molecular functions and gene interaction networks. Using integrative gene expression analysis methodologies, we showed the relevance of genes differentially expressed within the in vitro model to lung adenocarcinoma gene expression patterns and clinical outcome. In addition and specifically, we showed the progressive increase in UBE2C protein expression in lung lesions and lastly unraveled an association between its expression and poor NSCLC prognosis.

Functional pathways and topological gene interaction network analyses helped us select genes with both differential expression as well as predicted deregulated signaling function. Our six gene signature includes two genes (MCM2 and TPX2) that had been previously implicated independently in NSCLC progression and prognosis (21–23). We also found that the FEN1 endonuclease was progressively differentially expressed in the premalignant lung cells. Interestingly, transgenic mice harboring mutant FEN1 exhibit substantial pulmonary hypoplasia (24). The 14-3-3 protein family member, SFN (14-3-3 σ), was also found in our analyses to be differentially expressed in the premalignant lung cells. It is worthy to note that CpG hypermethylation of SFN promoter has been implicated in many cancers including bladder, breast, lymphoid, and lung tumors (25–28). Our findings suggest that epigenetic silencing of SFN may occur very early in lung cancer development.

Previous gene expression profiling of cultured cells have led to the discovery of individual genes or a group of genes for the study of distinct breast cancer subclasses (29), and predicting...
survival of cancer patients, namely ovarian and breast cancers (17, 30). For example, a series of human mammary epithelial cells expressing different transfected and activated oncogenes were used to develop gene pathway signatures that distinguished between human cancer subtypes based on distinct oncogenic pathways and predicted clinical outcome in several cancer patient subsets (14). Interestingly, our list of 584 differentially expressed genes included 112 genes that were present in two gene classifier lists capable of predicting clinical outcome without the incorporation of clinical covariables and compiled in the multiblinded microarray validation study by Shedden et al. (12), which provides the largest available set of microarray data with extensive pathologic and clinical annotation for lung adenocarcinomas (supporting material). Our findings show that by analyzing the transcriptomes of normal, premalignant, and tumorigenic lung epithelial cells in culture, we were able to highlight genes prominent in NSCLC clinical samples and capable of predicting clinical outcome. Interestingly, six of these genes, selected after IPA® analysis, were also associated with lung adenocarcinomas of poor prognosis, which was reproducible in more independent cohorts than the total 584 genes. It is worthwhile to note that our study also incorporates one of the largest in silico validations of the prognostic potential of an identified gene expression signature.

It is worthy to note that both the 584-gene and the 6-gene signatures were specific for adenocarcinoma because, when analyzed in data sets of lung SCCs, they were not significantly associated with poor prognosis. The specificity of the signatures for adenocarcinomas may reflect the distinct molecular and biological profiles of lung adenocarcinomas and SCCs (2). Although restricted to adenocarcinomas, these findings are of value as there has been a major global trend with a sharp increase in adenocarcinoma and a decrease in SCC (31, 32). In addition and in the past five decades, adenocarcinoma has become the predominant type of lung cancer cell in smokers (33). The selective association of the signatures with adenocarcinoma survival seemed to be unexpected given that the in vitro cell system was derived from bronchial epithelial cells that are considered to harbor precursors of SCC. However, if one considers the finding of Klein-Szanto et al. (9) that the 1170-I lung tumorigenic cells derived from the bronchial BEAS-2B cells formed invasive lung adenocarcinomas rather than SCCs in xenotransplanted mice, then the specificity for adenocarcinoma survival is not counterintuitive.

To our knowledge, this is the first demonstration of the potential in predicting NSCLC prognosis of an in vitro system emulating the various lung carcinogenesis phases. The immortalization of the human bronchial epithelial cells that served as
the basis for this model was accomplished by infection with an adenovirus 12-SV40 hybrid virus (34). The molecular basis for the immortalization is most likely due to the inactivation of p53 and retinoblastoma tumor suppressor genes by the SV40 large tumor antigen (35). These molecular abnormalities led to changes in gene expression that seem to be shared by many NSCLC tumors because concurrent abnormalities in both p53 and retinoblastoma pathways have been identified in between 28% (36) and 37% (37) of NSCLC and the majority of the rest had an abnormality at least one of the two pathways. It is also noteworthy that a previously identified SV40 T/t-Antigen cancer gene signature, developed and integrated from mouse models, was highly evident in the gene expression patterns of clinical samples of human breast, prostate, and lung cancers and was significantly associated with poor prognosis in these tumors (38). Moreover, we have previously validated aberrant gene expression within the vitro system we used at the mRNA level in 11 established NSCLC cell lines and at the protein level in clinical NSCLC samples (39, 40).

In this much more comprehensive study using more advanced technology, we further established that this particular cell system provides a powerful tool to predict outcomes of NSCLC cancers and aids in the exploration of the molecular mechanisms of lung carcinogenesis that could be translated into both chemoprevention and treatment of lung cancer.

Our network analysis showed that the E2 ubiquitin-conjugating enzyme, UBE2C, was part of the most significantly activated gene interaction network. UBE2C up-regulation and marked nuclear expression is associated with esophageal adenocarcinoma progression (41). More recently, UBE2C was shown to be crucial for activation of the anaphase promoting complex for subsequent ubiquitination and deactivation of spindle checkpoint substrates (42, 43). Despite intriguing observations that UBE2C mRNA levels were higher in a variety of primary tumors compared with corresponding normal tissues (44), its expression in premalignant stages nor its potential as a prognostic marker for lung cancer has been investigated. We have shown that

![Fig. 6. Association of UBE2C expression with NSCLC progression and clinical outcome.](https://example.com/fig6.png)

A, box-plot depicting statistical analysis by the Wilcoxon-rank test of nuclear UBE2C score in normal bronchial epithelia (white box), preneoplastic lesions (light gray boxes), as well as lung SCC and adenocarcinoma (dark gray boxes). P values representing significance of pair-wise comparison between different lung lesions when compared with normal bronchial epithelia (*) or hyperplasias (**) are marked, respectively. Significance of other pair-wise comparisons is also indicated on the figure (NS, not significant). B, box-plot representing pair-wise comparison between NSCLC nonsmokers (no) and smokers (yes) for nuclear UBE2C score. Kaplan-Meier plots for the overall survival (OS; C) and progression-free survival (PFS; D) of stage I NSCLC patients stratified according to positive (n = 118) or negative (n = 53) nuclear UBE2C immunoreactivity. P values were obtained by the log-rank test.
UBE2C protein expression is up-regulated in preneoplastic lung lesions and NSCLC samples, its expression was significantly associated with smoking status and that stage I NSCLC patients who expressed nuclear UBE2C exhibited significantly poorer survival than patients who did not express it. It is possible that NSCLC patients with higher UBE2C expression are more resistant to Taxol-based chemotherapies because cells overexpressing UBE2C exhibit compromised mitotic arrest induced by the microtubule disrupting and spindle checkpoint activating Taxol (42).

In conclusion, our study using the human in vitro lung carcinogenesis model identified novel gene signatures that are effective in the prognostic evaluation of lung cancers. Moreover, we unraveled a potential novel role for UBE2C expression in NSCLC pathogenesis and prognosis.

References


Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

www.aacrjournals.org

711


Downloaded from cancerpreventionresearch.aacrjournals.org on July 3, 2017. © 2009 American Association for Cancer Research.
Identification of Gene Signatures and Molecular Markers for Human Lung Cancer Prognosis using an *In vitro* Lung Carcinogenesis System

Humam Kadara, Ludovic Lacroix, Carmen Behrens, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1940-6207.CAPR-09-0084

Supplementary Material
Access the most recent supplemental material at:
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2009/07/31/1940-6207.CAPR-09-0084.DC1
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2009/09/15/1940-6207.CAPR-09-0084.DC2

Cited articles
This article cites 44 articles, 21 of which you can access for free at:
http://cancerpreventionresearch.aacrjournals.org/content/2/8/702.full#ref-list-1

Citing articles
This article has been cited by 7 HighWire-hosted articles. Access the articles at:
http://cancerpreventionresearch.aacrjournals.org/content/2/8/702.full#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.