A Gene Expression Signature from Peripheral Whole Blood for Stage I Lung Adenocarcinoma

Melissa Rotunno1, Nan Hu1, Hua Su1, Chaoyu Wang1, Alisa M. Goldstein1, Andrew W. Bergen1,3, Dario Consonni4, Angela C. Pesatori4, Pier Alberto Bertazzi4, Sholom Wacholder1, Joanna Shih2, Neil E. Caporaso1, Phil R. Taylor1, and Maria Teresa Landi1

Abstract

Affordable early screening in subjects with high risk of lung cancer has great potential to improve survival from this deadly disease. We measured gene expression from lung tissue and peripheral whole blood (PWB) from adenocarcinoma cases and controls to identify dysregulated lung cancer genes that could be tested in blood to improve identification of at-risk patients in the future. Genome-wide mRNA expression analysis was conducted in 153 subjects (73 adenocarcinoma cases, 80 controls) from the Environment And Genetics in Lung cancer Etiology study using PWB and paired snap-frozen tumor and noninvolved lung tissue samples. Analyses were conducted using unpaired t tests, linear mixed effects, and ANOVA models. The area under the receiver operating characteristic curve (AUC) was computed to assess the predictive accuracy of the identified biomarkers. We identified 50 dysregulated genes in stage 1 adenocarcinoma versus control PWB samples (false discovery rate \( \leq 0.1 \), fold change \( \geq 1.5 \) or \( \leq 0.66 \)). Among them, eight \( (TGFBR3, RUNX3, TRGC2, TRGV9, TARP, ACP1, VCAN, \) and \( TSTA3) \) differentiated paired tumor versus noninvolved lung tissue samples in stage I cases, suggesting a similar pattern of lung cancer–related changes in PWB and lung tissue. These results were confirmed in two independent gene expression analyses in a blood-based case–control study \( (n = 212) \) and a tumor–nontumor paired tissue study \( (n = 54) \). The eight genes discriminated patients with lung cancer from healthy controls with high accuracy \( (AUC = 0.81, 95\% CI = 0.74–0.87) \). Our finding suggests the use of gene expression from PWB for the identification of early detection markers of lung cancer in the future. Cancer Prev Res; 4(10); 1599–608. ©2011 AACR.

Introduction

Lung cancer causes more deaths than any other cancer in both men and women, with more than 160,000 deaths annually in the United States and 1 million worldwide (1). Unfortunately, the average 5-year survival rate has remained relatively stable at 15% over many decades because of minimal improvements in early detection and treatment. Noninvasive assays for detection of lung cancer at a curable stage could offer the best therapeutic option for these patients. Although promising, imaging techniques such as low-dose helical computed tomography are expensive and potentially associated to increased risk due to ionizing radiation exposure. Blood-based biomarker assays are a potentially important alternative noninvasive method to screen for lung cancer. Technological advances in methods of blood collection and RNA stabilization have only recently increased our ability to detect transcript levels in gene expression studies of human blood samples. Recent studies of gene expression from blood cells have successfully identified gene signatures for diverse exposures [e.g., tobacco smoking (2) or benzene (3)], and health conditions, including autoimmune disorders (4–6), inflammatory diseases (7), and cancer (8–11).

In our study, we first compared gene expression changes in blood between adenocarcinoma cases and noncancer controls to select the genes whose expression mostly differentiated cases from controls. We then compared this signature in paired adenocarcinoma versus noninvolved lung tissue samples to identify the subset of genes differentiating both cases–controls (blood samples) and tumor–nontumor (tissue samples). These expression changes could be specifically due to early development of cancer. Finally, we validated the overlapping gene expression signature in additional blood-based and tissue-based...
independent studies. If confirmed in prospective studies, gene expression changes from blood tests can provide a useful tool for the early detection of cancer in at-risk individuals.

Materials and Methods

Overview of strategy
Our study design included 3 phases. (i) First, we aimed to identify molecular changes in blood due to cancer, by comparing stage I adenocarcinoma cases ($n = 26$) to controls ($n = 80$). We restricted the analysis to stage I cases to focus on early molecular changes not affected by systematic metabolic disruption, such as weight loss or other sequelae of advanced disease. We then verified whether these gene changes in peripheral whole blood (PWB) were present also in later stages ($n = 47$). Because tobacco smoking is the most important risk factor for lung cancer (12) and has been associated with lung cancer progression (13), we also explored potentially distinct gene signatures by smoking groups. (ii) We then compared the blood-related gene expression signature distinguishing stage I cases from controls with the gene expression signature differentiating fresh frozen paired tumor versus noninvolved tissue samples in a subgroup ($n = 15$) of the same stage I cases. With this comparison, we aimed to identify expression changes in PWB due to lung cancer that paralleled changes in the target organ. (iii) Finally, we sought to validate the main results using (a) quantitative reverse transcriptase PCR (qRT-PCR) analysis from PWB for all identified genes in additional 82 stage I adenocarcinoma patients and 130 controls from the same population and (b) microarray gene expression data for all identified genes in 54 lung adenocarcinoma and noninvolved paired tissue samples from a previously published independent study (14).

Population
Individuals with lung adenocarcinoma ($n = 73$ for the microarray experiment; $n = 82$ for qRT-PCR validation) and healthy controls ($n = 80$ for the microarray experiment; $n = 130$ for qRT-PCR validation) were randomly sampled from a large, well-defined population-based case–control study, the Environment And Genetics in Lung cancer Etiology (EAGLE) study (15–21), including 1,000 consecutive incident lung cancer cases and 1,200 controls (all Caucasians) from Italy. Selected cases had histologically confirmed primary adenocarcinoma of the lung, including all stages, and controls were matched to cases by age, sex, and smoking status (never, former, and current smoking). For the validation set, we focused on current smoker stage I cases and controls. Detailed characteristics of subjects are described in Table 1.

The study was approved by the Institutional Review Board of each participating institutions in Italy and by the National Cancer Institute, Bethesda, MD. All participants signed an informed consent.

Blood and tissue collection for RNA extraction
PWB was collected for all EAGLE participants (after lung cancer diagnosis and before treatment for cases, and at enrollment for controls) using the PAXgene Blood RNA System (PreAnalytiX) containing a proprietary solution that reduces RNA degradation and gene induction (22, 23). Fresh lung tissue samples were snap frozen within 20 minutes of surgical resection.

Microarray gene expression data
Data on microarray gene expression from PWB were obtained using the Affymetrix GeneChip HG-U133A v2.0. After exclusion of 2 samples with poor quality profile (see quality assessment in Supplementary Material S1), the remaining 162 samples were processed and normalized with the Robust Multichip Average (RMA) method. Corresponding CEL files and information conform to the MIAME guidelines and are publicly available on the GEO database (accession number GSE20189). Nine subjects were excluded after data normalization because of reclassification to nonadenocarcinoma morphology during histologic review. The final analyses were based on 73 adenocarcinoma cases and 80 controls. All 22,277 probe sets based on RMA summary measures were used in the analyses.

The detailed description of the gene expression examination of lung tissues in lung cancer cases in EAGLE (also based on the Affymetrix HG-U133A GeneChip) and sample inclusion and exclusion criteria have been published previously (24). For this study, we used data from paired tumor and noninvolved lung tissue samples from 15 of the same stage I adenocarcinoma cases included in the PWB-based study.

The validation lung tissue set consisted of 27 tumor and 27 noninvolved paired lung tissue samples from a previously published independent study (14). Details of the specimens, mRNA processing and hybridization, and data access are described in the relative publication (14).

qRT-PCR gene expression data
We followed the procedure described in Hu N and colleagues (25). Briefly, RNA quality and quantity was determined using the RNA 6000 LabChip/Agilent 2100 Bioanalyzer. RNA purification was done according to the manufacturer’s instructions (Qiagen Inc.). After reverse transcription of RNA, all real-time PCRs were conducted using an ABI Prism 7000 Sequence Detection System with the designed primers and probes for target genes and an internal control gene, GAPDH. Each sample for each gene was run in triplicate. Quantitative methods require that PCR efficiencies be similar for all genes and 90% or more. Efficiency was measured using a standard curve generated by serial dilutions of the RNA as described in http://docs.appliedbiosystems.com/search.

Statistical analysis
(i) A 2-sample t test was conducted to test whether blood RNA expression differed between cases and
controls (overall and stratified by stage and by smoking status). Age, sex, and smoking variables were similarly distributed across the groups (Table 1) and were not associated with the expression of the 61 selected gene targeting probes (gene probes) among controls or cases. Analyses adjusted or unadjusted for these factors provided almost identical results. Unadjusted results are shown throughout the article. We used the Benjamini–Hochberg procedure (26) to calculate the false discovery rate (FDR) to adjust for the approximately 22,000 comparisons and only further considered results with a maximum FDR/C20 ≤ 0.1 (based on single gene probe P value threshold of 0.001). In addition, only gene probes with a fold change (FC) ≤ 0.66 for downregulated gene probes or ≥ 1.5 for

### Table 1. Characteristics of lung adenocarcinoma patients and controls used for the gene expression analysis from peripheral whole blood, Italy, 2002–2005

<table>
<thead>
<tr>
<th></th>
<th>Discovery set</th>
<th>Validation set</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Controls</td>
<td>Cases</td>
<td></td>
</tr>
<tr>
<td></td>
<td>P&lt;sup&gt;a&lt;/sup&gt;</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>40</td>
<td>37</td>
<td>111</td>
</tr>
<tr>
<td>Female</td>
<td>40</td>
<td>36</td>
<td>19</td>
</tr>
<tr>
<td>Age</td>
<td>0.73</td>
<td></td>
<td>0.46</td>
</tr>
<tr>
<td>40–44</td>
<td>2</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>45–49</td>
<td>3</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>50–54</td>
<td>7</td>
<td>7</td>
<td>5</td>
</tr>
<tr>
<td>55–59</td>
<td>11</td>
<td>9</td>
<td>16</td>
</tr>
<tr>
<td>60–64</td>
<td>13</td>
<td>13</td>
<td>27</td>
</tr>
<tr>
<td>65–69</td>
<td>13</td>
<td>16</td>
<td>31</td>
</tr>
<tr>
<td>70–74</td>
<td>21</td>
<td>19</td>
<td>33</td>
</tr>
<tr>
<td>75–79</td>
<td>10</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>Cigarette smoking</td>
<td>0.71</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>21</td>
<td>22</td>
<td>0</td>
</tr>
<tr>
<td>Former</td>
<td>32</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>Current</td>
<td>27</td>
<td>24</td>
<td>130</td>
</tr>
<tr>
<td>Cumulative pack-years</td>
<td>0.44</td>
<td></td>
<td>0.76</td>
</tr>
<tr>
<td>Below median</td>
<td>31</td>
<td>23</td>
<td>63</td>
</tr>
<tr>
<td>Above median</td>
<td>28</td>
<td>28</td>
<td>67</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>33.7 (21.9)</td>
<td>34.7 (20.5)</td>
<td>51.2 (21.9)</td>
</tr>
<tr>
<td>Cigarette per day</td>
<td>0.93</td>
<td></td>
<td>0.56</td>
</tr>
<tr>
<td>Below median</td>
<td>29</td>
<td>25</td>
<td>48</td>
</tr>
<tr>
<td>Above median</td>
<td>30</td>
<td>25</td>
<td>82</td>
</tr>
<tr>
<td>Missing</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Mean (SD)</td>
<td>17.9 (10.1)</td>
<td>17.5 (9.1)</td>
<td>21.1 (7.8)</td>
</tr>
<tr>
<td>Tumor stage&lt;sup&gt;b&lt;/sup&gt;</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IA</td>
<td>11</td>
<td>36</td>
<td></td>
</tr>
<tr>
<td>IB</td>
<td>15</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>IIA</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>IIB</td>
<td>15</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>IIIA</td>
<td>9</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>IIIIB</td>
<td>8</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>12</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Tumor grade</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>6</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>28</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>33</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>Missing</td>
<td>6</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup>P value based on a 2-sided Wald test.

<sup>b</sup>Stage definition was based on the International System for Staging Lung Cancer by the AJCC and UICC, Mountain CF 1997.
upregulated gene probes were considered for follow-up in subsequent analyses.

(ii) Because significantly fewer hypotheses (61 probes) were tested in the following analyses, less stringent significance criteria were applied ($P < 0.005$). For analyses of tumor versus noninvolved paired tissues from the same subjects, a linear mixed effects model was used to account for intraperson correlation. Gene probes with $P < 0.005$ and same FC direction and intensity (i.e., FC $\leq 0.66$ or $\geq 1.5$) as in the case–control blood RNA comparison were selected for validation analyses.

(iii) To validate the significant results, we analyzed: (a) the qRT-PCR gene expression PWB-based data using the $2^{-\Delta\Delta Ct}$ method (27) to compare cases with controls and (b) the microarray gene expression tissue-based data using linear mixed effects model to compare tumor with noninvolved paired lung tissue samples. In addition, receiver operating characteristic (ROC) analysis was done on the PWB-based validation data and the area under the curve (AUC) was estimated to assess the accuracy of the identified biomarkers, both individually and combined, in discriminating between lung cancer patients and controls.

All statistical analyses were conducted using R program language v2.10.

Results

(i) We compared mRNA expression from PWB in stage I adenocarcinoma cases versus controls, in the overall sample and stratified by smoking categories (Table 2). Two significant gene signatures in stage I cases were detected: one in the combined smokers and nonsmokers (FDR $= 0.10$), and the second among current smokers only (FDR $= 0.15$). No significant results were found within subsets of former or never-smokers (FDR $= 0.97$ and $1.00$, respectively). However, gene expression changes significant in the analysis among current smokers showed similar, although not significant, trends in the analyses among never and former smokers.
Gene Expression in Blood and Tissue in Stage I Lung Cancer

Figure 1. Heat map of genes differentially expressed in stage I lung adenocarcinoma patients versus controls. Gene symbols and Affymetrix gene probe names are reported on the vertical axis and sample color codes are shown on the horizontal axis. Samples and gene probes are ordered according to unconditional clustering on the basis of linkage method and Spearman correlation distance. Samples are colored coded as follows: stage I cases (magenta), controls (cyan), current smokers (blue), former smokers (orange), and never-smokers (yellow). The first 36 gene probes from the top are upregulated in stage I cases compared with controls. The last 25 gene probes are downregulated in stage I cases compared with controls. Subjects with different smoking status are homogeneously distributed along the horizontal axis, which indicates that smoking status does not contribute to the differential expression between cases and controls in these genes. Detailed information on these gene probes and related results are described in Supplementary Material S2.

(data not shown). At the same time, the analysis among current smokers revealed distinct alterations, which might be particularly important for individuals who smoke. Thus, for the comparison of stage I cases versus controls, we considered both results from all subjects and from current smokers only (221 and 144 gene probes, respectively, 81 overlapping between the two). To increase specificity, we restricted the successive analyses to gene probes with FC ≤ 0.66 or ≥ 1.5. The resulting 25 downregulated gene probes (20 genes), and 36 upregulated genes probes (30 genes), are shown in the heat map of Figure 1 and Supplementary Material S2. In general, FCs were stronger in the analysis restricted to current smokers than in the overall analysis. Because there was no significant difference between cigarette per day or cumulative pack-years between cases and controls (Table 1) and the analysis adjusted by these covariates provided almost identical results, our findings are unlikely due to differences in smoking quantity between cases and controls. We verified whether the identified 61 gene probes were also differentially expressed between cases and controls in late-stage disease. FCs were consistently stronger in the analysis limited to stage I cases but had concordant directions in all groups analyzed (Fig. 2).

(ii) We aimed to identify changes in gene expression related to early-stage lung cancer that are detectable in both blood cells and lung tissue cells. Thus, for the 61 gene probes (50 genes) in the analysis of stage I patients and controls (Fig. 1 and Supplementary Material S2), we compared gene expression in tumor versus paired noninvolved lung tissue samples in 15 stage I adenocarcinoma cases. We found that 10 probes from 8 genes (TGFB3, RUNX3, TRGC2, TRGV9, TARP, ACP1, VCAN, and TSTA3) were differentially expressed (P ≤ 0.003) in tumor compared with noninvolved lung tissue samples and in the same direction and intensity...
Figure 2. Comparison of gene expression changes in stage I cases versus in cases with advanced stages. Plot of FCs by case-control status for the gene probes that differed significantly (i.e., FDR \( < 0.1 \) based on \( P < 0.001 \)) and had large FC (i.e., FC \( \geq 0.66 \) or \( \leq 1.5 \)) in the analysis comparing stage I cases versus controls, among current smokers or among all smoking categories. Probes are listed by chromosomal location. FCs for downregulated probes are shown in the lower side of the graph (FC \( \leq 1 \), open symbols) and corresponding gene symbols in the bottom x-axis. FCs for upregulated probes are shown on the upper side of the graph (FC \( > 1 \), filled symbols) and corresponding gene symbols in the top x-axis. FCs were obtained comparing cases with stage I (black triangles), cases with stage II, III, and IV (dark grey squares), and cases with all stages (light grey circles) to controls. For consistency, FCs of gene probes that were significant in the analysis including later stages. Analogously, FCs of gene probes that were significant in the analysis comparing stage I cases to controls among current smokers were computed based on current smokers also in the analyses including later stages. Analogously, FCs of gene probes that were significant in the analysis comparing stage I cases to controls among all smoking categories also in the analyses including advanced stages. Gene probes showed generally higher upregulated FCs and lower downregulated FCs in the analysis based on stage I cases (triangles) with respect to the analysis based on advanced stage cases (squares), although with concordant direction. As expected, FCs in the analyses based on all stage cases (circles) were in between those based on stage I (triangles) and those based on advanced stage (squares) cases.

(iii) We validated the PWB-based gene expression differences in stage I cases compared with controls (Fig. 3).

as in stage I adenocarcinoma cases compared with controls (Fig. 3).

genes using microarray gene expression data from a previously published dataset (14), which included 27 adenocarcinoma and noninvolved paired lung tissue samples. The direction of changes was 100% consistent with our original finding: RUNX3, TGFBR3, TRGV9, TARP, and TRGC2 were significantly downregulated in tumor compared with noninvolved tissues (FCs = 0.7, 0.2, 0.3, 0.6, 0.5, \( P = 0.06, 3.0 \times 10^{-7}, 4.0 \times 10^{-7}, 3.4 \times 10^{-3}, 2.5 \times 10^{-6} \)), respectively) and VCAN, ACP1, and TSTA3 were significantly upregulated in tumor compared with noninvolved tissues (FCs = 2.6, 1.5, 2.5, \( P = 0.002, 5.7 \times 10^{-5}, 3.8 \times 10^{-9} \)), respectively). We evaluated the ability of PWB-based expression of each gene to discriminate lung cancer patients from controls in the validation set by means of ROC curves (Fig. 4). The AUC ranged from 0.55 (95% CI = 0.46–0.64) for ACP1 to 0.73 (95% CI = 0.66–0.81) for TRGV9 (Fig. 4), thus indicating a reasonable discrimination power between lung cancer cases and controls for most genes when considered individually. In addition, a combination of all
markers based on a logistic regression model showed the best diagnostic accuracy with an AUC of 0.81 (95% CI = 0.74–0.87, red ROC curve in Fig. 4).

Discussion

We identified a gene expression signature from blood samples consisting of 8 genes (RUNX3, TGFBR3, TRGC2, TRGV9, TARP, ACPI, VCAN, and TSTA3) that differentiates stage I lung adenocarcinoma cases from controls and mirrors cancer-related gene expression changes in the target tissue. Results were validated in additional independent sets of tissue-based and blood-based gene expression analyses of adenocarcinoma cases and controls. Although present in all stages, expression changes were weaker in advanced stages, possibly because of secondary changes due to the spread of the disease. Similarly, the changes were stronger in current smokers but present in all smoking categories. The accuracy in differentiating between stage I lung adenocarcinoma cases and controls was good for most genes when considered separately; in particular, those that were downregulated between cases and controls. A multiplex model on the basis of expression of all 8 genes combined showed a high diagnostic accuracy of 81% (Fig. 4). If further validated in prospective studies using PWB of cases drawn prior to lung cancer diagnosis (28), this gene expression signature may be used as a blood-based biomarker for early detection of lung adenocarcinoma in heavy smokers at high risk of lung cancer. We validated its use in current smokers. Further study in never- and former smokers is warranted. Moreover, it will be important to test the identified biomarkers in other lung cancer histologies.

The identified genes are promising with regard to potential mechanistic relevance. RUNX3 (runt-related transcription factor 3), downregulated in our analyses and with an AUC of 0.69, is involved in the negative regulation of epithelial cell proliferation, functions as a tumor suppressor, and is frequently deleted or transcriptionally silenced in cancer. Hypermethylation of RUNX3 has also been associated with the evolution of lung cancer (29) and, specifically, of lung adenocarcinoma (30). In addition, higher protein expression of RUNX3 has been associated with increased survival from lung adenocarcinoma (31). TGFBR3 (TGF-beta receptor III) encodes a glycoprotein that binds TGFβ, a cytokine that modulates several tissue development and repair processes. TGFBR3 is the TGF-beta component most commonly downregulated at both the message and protein levels in several cancers (32–36), including non–small cell lung cancer (37). Our study is the first to show downregulation of TGFBR3 mRNA expression in both blood and tumor tissue cells of lung adenocarcinoma patients. TGFBR3 showed the highest accuracy among the single-gene models in discriminating cases from controls (AUC = 0.73). TRGC2 (T-cell receptor gamma constant 2), TRGV9 (T-cell receptor gamma variable 9), and TARP (T-cell receptor gamma alternate reading frame protein) are colocated at chromosome locus 7p14.1, close to the 7p14.3 chromosomal region that frequently shows allelic loss in non–small cell lung cancer.
TARP is embedded within the TCR gamma locus and cDNA that detect TCR gamma mRNA also detect TARP mRNA. Accordingly, probes in TRGC2, TRGV9, and TARP showed very similar results in our study. TRGV9 cells have been shown to contribute to the natural immune surveillance against colon cancers (39). TARP has been previously studied as a prostate-specific gene and an androgen-regulated protein that may carry out its biological functions via action on mitochondria (40). Down-regulation in cases with respect to controls and in tumor compared with noninvolved tissues of TRGC2, TRGV9, and TARP points to an immune-related alteration as a possible contribution to lung adenocarcinoma development. Case-control discrimination on the basis of TRGC2, TRGV9, and TARP was also good (average AUC = 0.70). ACP1 (acid phosphatase 1) gene, upregulated in our analysis, is polymorphic and encodes at least 2 electrophoretically different isozymes. An increase of fast isozyme concentration increases invasiveness of cancer cells, whereas a decrease of slow isozyme concentration in cancer results in cancer cell proliferation (41). In the validation set, ACP1 showed the poorest accuracy in discriminating cases and controls (AUC = 0.55). VCAN (versican) encodes a protein involved in cell adhesion, proliferation, migration, angiogenesis, tissue morphogenesis, and maintenance. VCAN was initially identified in cultures of lung fibroblasts (42) and has been recognized to play a role in the invasion of several cancers (43) including lung cancer (44). VCAN mRNA expression was upregulated in both lung tumor tissue and PWB of adenocarcinoma cases in our study. TSTA3 (tissue-specific transplantation antigen P35B) gene, also upregulated in our analysis, is involved in the expression of many glycoconjugates. Intriguingly, TSTA3 is located in chromosomal region 8q24, which contains several polymorphic variants recently associated with several cancers (45–47). VCAN and TSTA3 also showed a reasonable performance in discriminating between cases and controls (AUC = 0.61 and 0.59, respectively). In addition to the described 8 genes, we also identified 42 additional genes whose expression in PWB distinguishes stage I lung adenocarcinoma from controls (Fig. 1 and Supplementary Material S2) and was stronger among subjects who currently smoked. If further confirmed in additional blood-based analyses, these genes could also contribute to the detection of early lung adenocarcinoma lesions.

In conclusion, gene expression changes from peripheral blood samples can differentiate early-stage lung adenocarcinoma cases from controls and resemble gene expression changes in early-stage lung adenocarcinoma tissue. This finding suggests that early processes of lung adenocarcinoma development may lead to systemic alterations that can be detected in peripheral blood tests. Gene expression from PWB can provide an important tool for the identification of early detection markers of cancer in the future.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Acknowledgments

We thank the EAGLE participants and study collaborators listed on the EAGLE website (http://eagle.cancer.gov/).

Grant Support

This research was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics, National Cancer Institute, NIH, DHHS, Bethesda, MD.

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 July 23, 2010; revised May 25, 2011; accepted June 21, 2011; published OnlineFirst July 8, 2011.

References

Gene Expression in Blood and Tissue in Stage I Lung Cancer

blood mononuclear cell gene expression in benzene-exposed works.


Zimmermann DI, Ruoslahti E. Multiple domains of the large fibroblast proteoglycan, versican, EMBO J 1989;8:2975–81.


A Gene Expression Signature from Peripheral Whole Blood for Stage I Lung Adenocarcinoma

Melissa Rotunno, Nan Hu, Hua Su, et al.


Updated version
Access the most recent version of this article at:
doi:10.1158/1940-6207.CAPR-10-0170

Supplementary Material
Access the most recent supplemental material at:
http://cancerpreventionresearch.aacrjournals.org/content/suppl/2011/09/27/1940-6207.CAPR-10-0170.DC1

Cited articles
This article cites 47 articles, 19 of which you can access for free at:
http://cancerpreventionresearch.aacrjournals.org/content/4/10/1599.full.html#ref-list-1

Citing articles
This article has been cited by 6 HighWire-hosted articles. Access the articles at:
/content/4/10/1599.full.html#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.