The expression of four genes as a prognostic classifier for stage I lung adenocarcinoma in 12 independent cohorts

Hirokazu Okayama*1, Aaron J. Schetter*1, Teruhide Ishigame1, Ana I. Robles1, Takashi Kohno2, Jun Yokota3, Seiichi Takenoshita4 and Curtis C. Harris1

1Laboratory of Human Carcinogenesis, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA; 2Division of Genome Biology, National Cancer Center Research Institute, Tokyo 104-0045, Japan; 3 Genomics and Epigenomics of Cancer Prediction Program, Institute of Predictive and Personalized Medicine of Cancer, Barcelona, Spain; 4Department of Organ Regulatory Surgery, Fukushima Medical University School of Medicine, Fukushima 960-1295, Japan. *authors contributed equally.

Corresponding author:

Curtis C. Harris
Laboratory of Human Carcinogenesis, CCR, NCI
37 Convent Dr. MSC4258
Building 37, Room 3068A
Bethesda, MD 20892-4258
(301) 496 2048

Conflict of Interest Statement:
The authors have no conflicts of interest to declare.

This research was supported by the Intramural Research Program of the National Cancer Institute, NIH and a Department of Defense Congressionally Directed Medical Research Program Grant PR093793

Running Title:
Prognostic classifier for stage I lung ADC in 12 cohorts
Abstract

Background: We previously developed a prognostic classifier using the expression levels of BRCA1, HIF1A, DLC1, and XPO1 that identified stage I lung adenocarcinoma patients with a high risk of relapse. That study evaluated patients in five independent cohorts from various regions of the world. In an attempt to further validate the classifier, we have used a meta-analysis based approach to study 12 cohorts consisting of 1069 TNM stage I lung adenocarcinoma patients from every suitable, publically available dataset.

Materials and Methods: Cohorts were obtained through a systematic search of public gene expression datasets. These data were used to calculate the risk score using the previously published 4-gene risk model. A fixed effects meta-analysis model was used to generate a pooled estimate for all cohorts.

Results: The classifier was associated with prognosis in ten of the twelve cohorts (p<0.05). This association was highly consistent regardless of the ethnic diversity or microarray platform. The pooled estimate demonstrated that patients classified as high risk had worse overall survival for all stage I (Hazard Ratio [HR], 2.66; 95% Confidence Interval [CI], 1.93-3.67; P<0.0001) patients and in stratified analyses of stage IA (HR, 2.69; 95%CI, 1.66-4.35; P<0.0001) and stage IB (HR, 2.69; 95%CI, 1.74-4.16; P<0.0001) patients.

Conclusions: The 4-gene classifier provides independent prognostic stratification of stage IA and stage IB patients beyond conventional clinical factors

Impact: Our results suggest that the 4-gene classifier may assist clinicians in decisions regarding postoperative management of early stage lung adenocarcinoma patients.
Introduction

Lung cancer is the leading cause of cancer-death in the world, accounting for more than one-fourth of all cancer-deaths (1). Approximately 85% of lung cancers are non-small cell lung cancer (NSCLC). The most common histology for NSCLC is adenocarcinoma (ADC), followed by squamous cell carcinoma (SQC) and large cell carcinoma. Despite therapeutic advances, prognosis remains considerably poor relative to other solid cancers, even in early stage patients (1). Thus, more refined treatment strategies are needed.

TNM staging is the best prognostic factor for NSCLC. TNM staging is used by clinicians to guide treatment options for NSCLC. Early stage patients, including TNM stage I and II, are typically approached with curative surgery as the optimal treatment. Among such patients with completely resected NSCLC, adjuvant chemotherapy is recommended only for stage II patients based on several randomized trials that demonstrated survival benefit of platinum-based chemotherapy (2-4). By contrast, clinical trials have revealed no survival advantage and potential deleterious side-effects of adjuvant chemotherapy for stage IA patients (2, 5). With regard to stage IB patients, the evidence supporting routine use of adjuvant chemotherapy is controversial (2, 6, 7). A more detailed histological subtyping of lung cancer may improve on TNM classification system. For example, the presents of the micropapillary histologic subtype has been found to be associated with cancer recurrence after limited resection of peripheral lung ADC and may help guide treatment strategies (8).

Approximately 30% of stage I lung cancer patients will relapse and ultimately die of this disease. The majority of these patients are being treated by surgery alone because of the lack of clear evidence of benefit from adjuvant chemotherapy. Consequently, 5-year overall survival rates for pathological stage IA and IB are 73% and 58%, respectively, based on the recently-revised, 7th edition of TNM staging (9). One simple and critical question is how clinicians can distinguish the approximately 30% of stage I patients who have higher risk of relapse from the other 70% of patients who have excellent prognosis. High-risk patients might have undetectable micrometastases at the time of surgery. Hence their outcome could potentially be improved by postoperative systemic therapy with the primary goal of eliminating residual occult metastases that lead to disease recurrence. There is a substantial need to identify stage IB patients who are unlikely to benefit from adjuvant chemotherapy and/or immunotherapy as well as stage IA patients who have the highest risk of relapse. In view of that, it is vital to develop prognostic
biomarkers that can help clinicians determine appropriate postoperative management for each individual patient. The demand for such clinical prognostic tests is now undoubtedly increasing, as the extensive use of computed tomography (CT) screening becomes widely accepted, in which the majority of patients are diagnosed at stage I (10).

Numerous studies have identified prognostic biomarkers for NSCLC based on multigene expression by using qRT-PCR and/or microarray technology (11-26). However, associations reported in single studies often failed to provide sufficient validation in additional populations (12, 26, 27). A recent review criticized prognostic gene signatures for their unspecified clinical utility as well as the lack of reproducibility, and suggested that no lung cancer signatures are ready for clinical application (27). Taking into account the guidelines suggested in that review, we started to develop a gene expression-based prognostic signature that was intended to be used for early stage lung ADCs, especially for stage I patients. Our goal was to make a classifier based on a few key genes that would be a simple and robust classifier for prognosis of stage I lung cancer patients. A similar strategy based on analyzing important 31 cell proliferation genes has shown to be a robust prognostic classifier for lung cancer (28). Our resulting signature, namely the 4-gene classifier, was found to be highly robust in all 5 cohorts that we analyzed. Its prognostic significance was independent of other clinical factors, including age, gender, TNM stage and smoking status (29). These results suggest that the 4-gene classifier may be useful in guiding therapeutic decisions for early stage lung cancer patients. We have now set out to test the 4-gene classifier in as many independent populations as we could identify from publically available gene expression data. As a result, we have evaluated more than 1000 stage I ADC patients from 12 independent cohorts using different gene expression platforms. We use a meta-analytic approach to measure the association of the 4-gene classifier with prognosis and evaluate its reproducibility across those cohorts. We focus on stage IA and IB patients separately to further evaluate the clinical usefulness of this classifier.

Materials and Methods

Selection of studies

We searched GEO (Gene Expression Omnibus; http://www.ncbi.nlm.nih.gov/geo/) in June 2013 with the search terms “lung cancer”, “non-small cell lung cancer”, “lung adenocarcinoma”, “lung adenocarcinomas” and
"NSCLC". The retrieved GEO series were filtered by Organism (Homo sapiens) and Series Type (Expression Profiling by Array) as well as sorted by the number of samples (series that have at least 30 samples). Ninety-two GEO series identified by the initial GEO search were screened on the basis of their Title, Summary and Overall Design as described in GEO Accession Display. Datasets were excluded if they analyzed only cell lines/xenograft samples, only non-tumor specimens (e.g., bronchial epithelial cells, blood, fluid), or contained no primary ADC tumors. Also, several superseries that consisted of one or more subseries were excluded (due to duplicate data) and the corresponding subseries with gene expression data were retrieved, leaving 46 GEO datasets of lung cancer-related clinical studies. In parallel with this search, we used ONCOMINE (Compendia Bioscience, Ann Arbor, MI; http://www.oncomine.com) to identify public microarray datasets that had ADC patients with survival status. ONCOMINE search identified 10 datasets, 5 of which were not deposited in GEO. The resulting 51 datasets containing primary ADC samples were further reviewed based on the Sample Characteristics in Series Matrix File, or the Dataset Detail in ONCOMINE. Selection criteria for all publicly available datasets required each dataset to include survival information for more than 30 TNM stage I patients of ADC and have expression data for BRCA1, HIF1A, DLC1 and XPO1. After removing 40 datasets that did not fit our criteria, we found 11 independent microarray datasets, including the Botling (GSE37745) (30), Tang (GSE42127) (24), Rousseaux (GSE30219) (31), Matsuyama (GSE11969) (32), Wilkerson (GSE26939) (33), Lee (GSE8894/ONCOMINE) (17), Bild (GSE3141/ONCOMINE) (34) cohorts as well as the Bhattacharjee (ONCOMINE) (35), Directors (ONCOMINE) (12), Japan (GSE31210) (36), Tomida (GSE13213) (16) cohorts. Among them, the former 7 cohorts were newly obtained from GEO or ONCOMINE (if available) for this present study, whereas the latter 4 datasets were the original cohorts that we analyzed in our initial study (29). The selection flowchart and the list of retrieved datasets are presented in Figure 1 and Supplementary Table 1. All of the data from the stage I lung adenocarcinoma cohorts are available in the supplemental data.

For the 4-coding gene analyses in SQC patients, we used multiple cohorts of stage I SQC. Six cohorts, including the Botling, Rousseaux, Tang, Matsuyama, Lee and Bild datasets among the ADC datasets mentioned above, were included, as these cohorts also contained expression data for SQC patients with survival information. We obtained one SQC dataset from GEO (GSE17710) deposited by Wilkerson et al. (37), separated from their ADC data (GSE26939, the Wilkerson ADC cohort) (33). Additionally, among 3 SQC datasets with survival information
which were found in ONCOMINE, including the Raponi (SQC only) (38), Larsen (SQC only) (39) and Zhu (ADC and SQC) (11) cohort, the Raponi and the Zhu cohorts were included in SQC analyses. For the Zhu cohort, only SQC patients were analyzed, while ADC patients (n=14, Stage I) were disregarded, since considerable number of ADC patients were already used within (CAN/DF) of the Directors cohort (11). The Larsen cohort was excluded because BRCA1 gene was not available in their platform.

**Gene expression data analysis**

In this study, we focused only on stage I patients. The 4-coding gene analysis of five original cohorts used AJCC TNM 6th edition as described previously (29). Concerning 7 new cohorts, although the TNM edition was not specified as either 6th or 7th in each of original papers, we assumed that they were based on AJCC TNM 6th edition, since most tumors were collected before the development of TNM 7th edition in 2009. For the Rousseaux cohort, T1N0 tumors were defined as stage IA, while T2N0 tumors were defined as stage IB, according to the provided TNM classification for each patient. Among all available stage I cases obtained from the public datasets, 2 ADCs and 1 SQC in the Tang cohort, 3 ADCs and 4 SQCs in the Lee cohort were excluded from the analysis, since survival information was not provided for those cases.

For all analyses, the normalized expression values were obtained from each dataset and were not processed further. We then generated criteria to select the most reliable, informative probes. In brief, if there were multiple probes to a single gene, pairwise correlation of each probe was analyzed in each cohort. Probes were removed if they showed no correlation (R<0.5) between any of the other probes for that gene. If there were only 2 probes to a single gene that didn’t correlate with each other, the probe with the highest expression was selected. Probes that were selected are shown in Supplementary Table 2. If more than one probe was selected, they were averaged and no further processing was performed. The 4-coding gene classifier \[(0.104 \times BRCA1) + (0.133 \times HIF1A) + (-0.246 \times DLC1) + (0.378 \times XPO1)\] was applied to all newly-obtained cohorts using microarray expression data, and the resulting classifier score was categorized as low, medium, or high based on tertiles within each cohort separately. The within cohort categorization was performed to standardize risk scores across all cohorts. This will compensate for the fact that each study used different methodologies to measure the expression of each of the four genes and these expression values are not directly comparable to one another across cohorts. The association between
the 4-gene classifier and survival was assessed by the Kaplan–Meier log-rank test for trend using Graphpad Prism v5.0 (Graphpad Software Inc). Cox regression analyses were carried out using SPSS 11.0 (SPSS Inc), and all univariate and multivariate models were adjusted for cohort membership where appropriate. Forest plot analyses and calculations for the nomograms were performed using Stata 11.2 (Stata-Corp LP). Heterogeneity test for the combined HR was carried out using the I-squared statistics (40). Nomograms were developed based on coefficients from multivariable Cox regression models on 5-year overall survival using all variables that were significantly associated with patient outcome.

Results

Identification of eligible studies

Since the purpose of this gene expression-based classifier is to identify high-risk, stage I ADC patients who may benefit from additional intervention after surgery, we limited all the analyses in this study to stage I primary ADC tumors. The systematic search identified 11 microarray datasets consisting each of more than 30 cases of stage I ADC patients with sufficient survival information with gene expression data for all 4 genes, including BRCA1, HIF1A, DLC1 and XPO1, as described in Figure 1. Four of the 11 datasets were previously analyzed for our initial paper describing the signature, in which 5 independent cohorts of stage I ADCs were each analyzed by qRT-PCR and/or microarrays (29). Hence, 7 independent cohorts were newly obtained through this systematic search and a total of 12 cohorts were included in this study. The characteristics of the 12 cohorts are summarized in Table 1. This analysis includes 1069 patients in total, consisting of 546 stage IA and 518 stage IB cases (5 cases were not specified as stage IA or IB). These cohorts were derived from 6 different countries, including Japan, Norway, Sweden, France, South Korea, as well as at least 8 different institutions in the United States. Nine of 12 cohorts reported overall survival information, 2 cohorts reported relapse-free survival and 1 cohort reported cancer-specific survival. In each cohort, RNA samples were isolated from frozen tumor specimens and were subjected to gene expression analysis based on various platforms, including qRT-PCR and Affymetrix, Illumina, or Agilent microarrays. (Table 1)

The 4-gene classifier is tested in 12 independent cohorts

The 4-gene classifier was applied to each of the 7 newly-obtained cohorts using microarray expression data for 4 genes, and then cases were categorized as high, medium or low, based on tertiles for each cohort separately. Similar to our
previous results, highly concordant associations were found between the 4-gene classifier and prognosis in all 7 newly-obtained cohorts, including the Tang (p=0.046), Rousseaux (p=0.044), Wilkerson (p=0.014), Matsuyama (p=0.028), Lee (p=0.010), Botling (p=0.058) and Bild (p=0.120) cohorts by the Kaplan-Meier analysis (Figure 2). Overall, the 4-gene classifier was significantly associated with prognosis in 10 cohorts, while the remaining 2 cohorts had marginal associations in the proper direction.

The 9 cohorts that had overall survival information were analyzed in a fixed effects meta-analysis model, which included 817 stage I cases. There was no evidence for heterogeneity or inconsistency across multiple cohorts (I-squared=0.0%, p=0.980), suggesting that these results are representative of most lung ADCs and not a result of selection bias (Figure 3). Patients that were classified as high risk had significantly worse overall survival (HR, 1.73; 95%CI, 1.47-2.02) in stage I analysis (Figure 3A). The corresponding Kaplan-Meier analysis for the combined stage I patients demonstrated a significant association with overall survival and the 4-gene classifier (p<0.0001; Figure 3A). Furthermore, stratified analyses were performed for stage IA and IB separately, to address the prognostic impact of this classifier in these subgroups. Significant associations between the 4-gene classifier and overall survival were found in both stage IA (HR, 1.61; 95%CI, 1.27-2.06) and stage IB (HR, 1.76; 95%CI, 1.41-2.19) analyses, respectively (Figure 3B and C).

**The 4-gene classifier is an independent prognostic biomarker for stage IA as well as stage IB patients**

Given that the classifier is significantly associated with survival in stage IA and IB subgroups, Cox regression analysis was conducted using the combined cohort with respect to each stage (Table 2). All univariate and multivariate Cox analyses were adjusted for cohort membership and multivariate models were adjusted for age, gender and TNM stage. Since most of the public datasets did not provide complete clinical information, we could not apply other parameters, such as smoking status or adjuvant chemotherapy to the Cox analysis. In univariate analysis, older age, male gender, TNM stage IB and the 4-gene classifier were each significantly associated with worse outcome. Multivariate models revealed that the 4-gene classifier was significantly associated with poor overall survival, independent of other parameters, in stage I patients (HR, 2.66; 95%CI, 1.93-3.67; P<0.0001) and in stratified analyses of stage IA (HR, 2.69; 95%CI, 1.66-4.35; P<0.0001) and stage IB (HR, 2.69; 95%CI, 1.74-4.16; P<0.0001) patients. We
have also performed an analysis of the risk score as a linear variable (rather than a ordered categorical variable) to demonstrate these associations are highly robust and do not rely on using tertiles as cutpoints for the data (Supplementary Table 3)

The potential use of the 4-gene classifier to predict prognosis for stage I lung adenocarcinoma

In order for a prognostic classifier to be clinically useful, it has to provide actionable information to the physician. To demonstrate this potential for the 4-gene classifier, we developed a nomogram to predict 5-year survival rates in patients diagnosed with stage I lung ADC (Figure 4). This nomogram is based on the 9 cohorts with overall survival data and includes all variables that were significantly associated with 5-year overall survival. The points assigned to each variable are weighted based on Cox regression coefficients. This nomogram demonstrates that the 4-gene classifier could be used with clinical staging and other clinical parameters to predict the probability of 5-year survival. Subgroup analysis within stage is important to demonstrate to show clinical utility, therefore we created nomograms stratified by TNM stage IA and IB separately as another example of how this classifier can be integrated with TNM staging to help determine patient prognosis (Supplementary Figure 1).

The 4-gene classifier is only applicable to adenocarcinoma patients

Up to now we have focused only on lung cancer patients with ADC histology. To determine if this association could be observed across different histologies, we examined the 4-gene classifier in SQC, which is another major histological type of NSCLC. Nine independent cohorts, consisting of 337 stage I SQC patients, were obtained and the 4-gene classifier was applied to each cohort (Supplementary Table 4). In a combined analysis, the 8 cohorts with overall survival information were combined (n=292). However, no significant association was found in any of the SQC analyses (Supplementary Figure 2), indicating that the 4-gene classifier is specific to ADC. This was not completely unexpected because we built this classifier using ADC gene expression data only. SQC and ADC are also considered to be molecularly distinct entities (41). Therefore, this classifier appears to be only useful for lung ADC and not SQC. Sufficient numbers were not available to examine other histologies.

Discussion
Translating prognostic gene signatures into clinical use is a major challenge in the field of lung cancer research. There is little doubt that prognostic tests are necessary for stage I lung cancer patients after complete resection. In breast cancer, it is striking to note that several multigene assays are already commercially available and are currently used by clinical oncologists and supported by the NCCN and other guidelines (42, 43). As for lung cancer, no prognostic biomarkers have been incorporated into the current guidelines despite a large number of published gene signatures. This may be at least in part due to insufficient reproducibility as well as the lack of large-scale validation (27). Also, it has been suggested that many signatures were developed without clear focus on specific clinical contexts (27). In order to address those issues, our study has set out to test whether the 4-gene classifier that we previously identified is a robust prognostic classifier for stage I lung ADC using every publically available dataset. The 4-gene classifier was a robust classifier for over 1000 TNM Stage I lung ADC cases from 12 cohorts regardless of ethnic difference in the genetic background of the patients. When each cohort was separately analyzed, the 4-gene classifier showed highly consistent results for its association with survival. The classifier was highly reproducible across multiple platforms for gene expression measurement, including qRT-PCR and commercial/custom microarrays from Affymetrix, Agilent and Illumina. There is no evidence of selection bias in any of the 12 cohorts, which suggests these results presented here are representative of stage I lung ADC. However, the classifier had no prognostic impact on SQC patients, indicating its limited utility only to lung ADCs.

The 4-gene classifier has potential to be used as a prognostic biomarker for the management of stage IA and stage IB ADC patients in the current clinical setting. The pooled estimate revealed significant association between the 4-gene classifier and prognosis in stage IA and IB lung ADC subgroups, independent of other clinical variables. This suggests that the classifier can add additional discriminative value to identify high-risk patients beyond conventional clinical characteristics. Hence, “low-risk” stage IB patients identified by the classifier and predicted to have excellent survival probabilities may be recommended to forgo adjuvant therapy. Likewise, the classifier may also identify “high-risk” stage IA patients for whom intensive postoperative intervention is considered. Future work should explore and identify the optimal cutpoint for this assay that should distinguish "high-risk" and "low-risk patients". For convenience, we have used tertile as a cutpoint for our studies to distinguish between high/medium/low risk. It
is likely that optimal cutpoints can be found that will improve the clinical utility of this classifier.

Many published multigene signatures for NSCLC had utilized dozens to hundreds, or even thousands of genes along with complex classification models that are difficult to understand. We believe that this is an obstacle to the rapid development and feasibility of the clinical tests. By contrast, the 4-gene classifier uses the expression values of only 4 genes, potentially providing an opportunity to develop a simple and practical laboratory test. The classifier is composed of biologically relevant genes that are mechanistically important and are each significantly associated with prognosis in early stage lung ADC (29). We consider that our strategy of focusing on only biologically relevant genes have improved the chances of developing a robust classifier that will be generalizable to lung ADC.

A potential limitation for this analysis is that there was incomplete data on smoking status and types of chemotherapy that were retrieved for several of the cohorts used for the meta-analysis. Therefore, we were not able to include these covariates in the models. Smoking history was considered as a covariate within the discovery cohorts and found to not contribute to the risk association model (29). Adjuvant chemotherapy is not recommended for stage I patients, thus the majority of the patients included in our analysis would not have received any. In fact, only 4% of patients in the Japan discovery cohort received adjuvant chemotherapy (29). Still, these factors are important for survival after lung cancer surgery and it is possible that their inclusion may modulate the association between the 4-gene classifier and prognosis.

We have shown that the classifier is robust and that using qRT-PCR data and microarray data from a variety of labs provides similar results. In order to turn the 4-gene classifier into a clinical test, future work should focus on developing a standardized assay which will include developing methods for measuring each of the four genes and recommending methods for tissue handling, processing and RNA isolation. Assays will have to be designed to minimize the potential batch effects and inter-laboratory differences. Another future possibility is the use of RNA samples extracted from formalin-fixed paraffin-embedded (FFPE) tissues. This could extend the practical utility of the classifier to readily-available archived specimens. Furthermore, our previous study demonstrated an improved prognostic association by combining multiple, validated classifiers, namely the combination of the 4-coding gene classifier with non-coding microRNA-21 (miR-21) classifier for stage I lung ADC (29). In addition to the present results, recent meta-analysis
studies demonstrating miR-21 as a promising prognostic biomarker for NSCLC may be supportive to the potential combined use of the 4-gene and miR-21 classifiers as validated biomarkers (44, 45). This demonstrates that the integration of multiple, independent classifiers can further improve prognostic predictions and suggests that these classifiers can be combined with additional biomarkers and histological subtyping data to improve decision making capabilities for early stage lung cancer.

In conclusion, the 4-gene classifier that we recently developed was rigorously validated in large-scale, multiple cohorts with a meta-analytic approach consisting of more than 1000 stage I patients. To our knowledge, this is the first report of an RNA-based classifier in lung ADC to be tested and validated this extensively. The reproducibility of its performance was clearly demonstrated in the intended clinical context based on unbiased approaches. Particularly, the classifier provides additional prognostic stratification beyond the current risk factors, namely, stage IA and stage IB subgroups, highlighting the potential of this classifier in personalized management for early stage patients. These results support the development of standardized tests for the 4-gene assays and the incorporation of these assays into prospective studies.

References

<table>
<thead>
<tr>
<th>Cohorts</th>
<th>Country</th>
<th>n</th>
<th>TNM Stage</th>
<th>Age</th>
<th>Gender</th>
<th>Smoker (%)</th>
<th>Postoperative Therapy</th>
<th>Outcome</th>
<th>Platform</th>
<th>GEO ID</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>IA</td>
<td>IB</td>
<td>IAorIB</td>
<td>Mean</td>
<td>M</td>
<td>F</td>
<td>CT/RT</td>
<td>None</td>
</tr>
<tr>
<td>Five cohorts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Japan</td>
<td>Japan</td>
<td>149</td>
<td>100</td>
<td>49</td>
<td>0</td>
<td>59.7</td>
<td>66</td>
<td>83</td>
<td>45.6</td>
<td>0</td>
</tr>
<tr>
<td>US/Norway</td>
<td>USA (UMD), Norway</td>
<td>67</td>
<td>29</td>
<td>38</td>
<td>0</td>
<td>64.6</td>
<td>37</td>
<td>30</td>
<td>96.9</td>
<td>4</td>
</tr>
<tr>
<td>Directors</td>
<td>USA</td>
<td>276</td>
<td>114</td>
<td>162</td>
<td>0</td>
<td>64.4</td>
<td>131</td>
<td>145</td>
<td>NA</td>
<td>46</td>
</tr>
<tr>
<td>Bhattacharjee</td>
<td>US (Harvard)</td>
<td>76</td>
<td>35</td>
<td>40</td>
<td>1</td>
<td>64.2</td>
<td>32</td>
<td>44</td>
<td>90.8</td>
<td>0</td>
</tr>
<tr>
<td>Tomida</td>
<td>Japan</td>
<td>79</td>
<td>42</td>
<td>37</td>
<td>0</td>
<td>61.4</td>
<td>41</td>
<td>38</td>
<td>50.6</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Seven new cohorts</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tang</td>
<td>USA (MD Anderson)</td>
<td>87</td>
<td>32</td>
<td>55</td>
<td>0</td>
<td>64.1</td>
<td>37</td>
<td>50</td>
<td>NA</td>
<td>22</td>
</tr>
<tr>
<td>Rousseaux</td>
<td>France</td>
<td>81</td>
<td>73</td>
<td>8</td>
<td>0</td>
<td>61.8</td>
<td>65</td>
<td>16</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>Botling</td>
<td>Sweden</td>
<td>70</td>
<td>28</td>
<td>42</td>
<td>0</td>
<td>63.5</td>
<td>31</td>
<td>39</td>
<td>NA</td>
<td>5</td>
</tr>
<tr>
<td>Wilkerson</td>
<td>USA (UNC)</td>
<td>62</td>
<td>31</td>
<td>31</td>
<td>0</td>
<td>66.6</td>
<td>26</td>
<td>36</td>
<td>88.0</td>
<td>0</td>
</tr>
<tr>
<td>Matsuyama</td>
<td>Japan</td>
<td>52</td>
<td>28</td>
<td>24</td>
<td>0</td>
<td>62.3</td>
<td>28</td>
<td>24</td>
<td>46.2</td>
<td>0</td>
</tr>
<tr>
<td>Lee</td>
<td>Korea</td>
<td>36</td>
<td>13</td>
<td>23</td>
<td>0</td>
<td>61.4</td>
<td>16</td>
<td>20</td>
<td>38.9</td>
<td>0</td>
</tr>
<tr>
<td>Bild</td>
<td>USA (Duke)</td>
<td>34</td>
<td>21</td>
<td>9</td>
<td>4</td>
<td>64.8</td>
<td>17</td>
<td>17</td>
<td>NA</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>1069</td>
<td>546</td>
<td>518</td>
<td>5</td>
<td>63.1</td>
<td>527</td>
<td>542</td>
<td>77</td>
<td>524</td>
</tr>
</tbody>
</table>

**NOTE:** Nine cohorts with overall survival information were used in the combined analysis (n=817). Data were obtained from ONCOMINE.

Abbreviation: CT/RT, chemotherapy and/or radiotherapy; NA, not available; RFS, relapse-free survival; CSS, cancer-specific survival; OS, overall survival.
Table 2. Univariable and Multivariable Cox regression of the 4-coding gene classifier in the combined cohort\(^a\) of Stage I, adenocarcinoma patients

<table>
<thead>
<tr>
<th>Variable (n)</th>
<th>Univariable Analysis (^b)</th>
<th>Multivariable Analysis (^b)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HR (95% CI)</td>
<td>P</td>
</tr>
<tr>
<td><strong>TNM Stage I (n=817)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 gene classifier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (276)</td>
<td>Reference</td>
<td>NA</td>
</tr>
<tr>
<td>Medium (271)</td>
<td>1.34 (0.95 ― 1.89)</td>
<td>0.101</td>
</tr>
<tr>
<td>High (270)</td>
<td><strong>2.83 (2.07 ― 3.86)</strong></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Stage(^d)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IB (408) / IA (404)</td>
<td><strong>1.68 (1.29 ― 2.19)</strong></td>
<td>0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>Continous</td>
<td>1.03 (1.02 ― 1.05)</td>
</tr>
<tr>
<td>Gender</td>
<td>Female (409) / Male (408)</td>
<td><strong>0.67 (0.52 ― 0.87)</strong></td>
</tr>
<tr>
<td><strong>TNM Stage IA (n=404)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 gene classifier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (149)</td>
<td>Reference</td>
<td>NA</td>
</tr>
<tr>
<td>Medium (137)</td>
<td>1.47 (0.87 ― 2.49)</td>
<td>0.151</td>
</tr>
<tr>
<td>High (118)</td>
<td><strong>2.69 (1.67 ― 4.34)</strong></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>Continous</td>
<td>1.03 (1.01 ― 1.06)</td>
</tr>
<tr>
<td>Gender</td>
<td>Female (205) / Male (199)</td>
<td><strong>0.61 (0.40 ― 0.91)</strong></td>
</tr>
<tr>
<td><strong>TNM Stage IB (n=408)</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 gene classifier</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Low (125)</td>
<td>Reference</td>
<td>NA</td>
</tr>
<tr>
<td>Medium (132)</td>
<td>1.20 (0.74 ― 1.93)</td>
<td>0.456</td>
</tr>
<tr>
<td>High (151)</td>
<td><strong>2.88 (1.88 ― 4.43)</strong></td>
<td>&lt;0.0001</td>
</tr>
<tr>
<td>Age</td>
<td>Continous</td>
<td>1.04 (1.02 ― 1.05)</td>
</tr>
<tr>
<td>Gender</td>
<td>Female (203) / Male (205)</td>
<td>0.75 (0.54 ― 1.06)</td>
</tr>
</tbody>
</table>

\(^a\) The combined cohort consists of 9 publicly available, independent microarray datasets of stage I patients with overall survival information, including the Directors (276), Bhattacharjee (76), Tomida (79), Botling (70), Tang (87), Rousseaux (81), Matsuyama (52), Wilkerson (62), and Bild (34) cohorts.

\(^b\) The univariable models was adjusted for cohort membership and the multivariable model included the 4 gene classifier, cohort membership, age, gender, and TMN staging.

\(^c\) The 4-coding gene classifier was categorized based on tertiles of Stage I patients for each cohort.

\(^d\) There were a total of 5 stage I cases in the Bhattacharjee (1) and Bild (4) cohorts for which stage IB/IA information is not available. These are included in univariate analyses and excluded in multivariate analyses.
Figure legends.

**Figure 1:** Dataset selection flow chart. A total of 92 datasets from GEO and 10 datasets from ONCOMINE were evaluated. A total of 11 datasets were selected to be included in this meta-analysis.

**Figure 2:** The performance of the 4-coding gene classifier in 12 independent cohorts of stage I lung adenocarcinoma patients. For each cohort, cases were categorized as high, medium or low based on tertiles. P-values were obtained by the log-rank test for trend.

**Figure 3:** Forest plot of the prognostic impact of the 4-coding gene classifier in 12 independent cohorts of stage I lung adenocarcinoma. A) Meta-analysis of all patients with TNM stage IA or IB lung adenocarcinoma. B) Meta-analysis of all patients with TNM stage IA lung adenocarcinoma. C) Meta-analysis of all patients with TNM stage IB lung adenocarcinoma.

**Figure 4:** Nomogram to predict 5-year survival rates for stage I lung adenocarcinoma. Each clinical variable (4-gene score, TNM Stage, sex and age) is assigned a point value. The sum of those points can then be used to estimate probability of survival for 5-years. For example, the sum of the points is 300, a patient has approximately a 30% 5-year survival probability.
Figure 1

Dataset selection flow chart

92 datasets found in initial GEO search

46 datasets were excluded based on the “title”, “summary” and “overall design” in GEO Accession Display
• 23 did not contain clinical samples (cell line or xenograft studies)
• 16 did not contain primary adenocarcinoma samples
• 7 duplicate datasets deposited as superseries composed of corresponding subseries

46 datasets from GEO

10 datasets from ONCOMINE

...5 datasets overlapped between ONCOMINE and GEO

51 datasets for further review

40 datasets were excluded after reviewing the “sample characteristics” in Series Matrix File downloaded from GEO, or the “Dataset Detail” in ONCOMINE. They lacked at least one of the following information.
• TNM stage information
• Histology
• Sufficient survival information
• Sufficient number of stage I adenocarcinoma patients (n≥30)
• Gene expression data for all 4 genes

11 datasets for meta-analysis

7 new datasets
• Botling cohort (GEO)
• Tang cohort (GEO)
• Rousseaux cohort (GEO)
• Wilkerson cohort (GEO)
• Matsuyama cohort (GEO)
• Lee cohort (GEO/ONCOMINE)
• Bild cohort (GEO/ONCOMINE)

4 datasets used in our previous study
• Directors cohort (ONCOMINE)
• Bhattacharjee cohort (ONCOMINE)
• Tomida cohort (GEO)
• Japan cohort (GEO)
Figure 2

Stage I, Adenocarcinoma, 5 original cohorts

- Japan (n=149)
  - qRT-PCR
  - Relapse-free survival
  - Low (n=50)
  - MId (n=50)
  - High (n=49)
  - P<0.001

- US/Norway (n=67)
  - qRT-PCR
  - Cancer-specific mortality
  - Low (n=23)
  - MId (n=23)
  - High (n=21)
  - P<0.001

- Directors (n=276)
  - Affymetrix
  - Overall survival
  - Low (n=92)
  - MId (n=92)
  - High (n=92)
  - P=0.036

- Bhattacharajee (n=76)
  - Affymetrix
  - Overall survival
  - Low (n=26)
  - MId (n=25)
  - High (n=25)
  - P=0.008

- Tomida (n=79)
  - Agilent
  - Overall survival
  - Low (n=27)
  - MId (n=26)
  - High (n=26)
  - P=0.008

Stage I, Adenocarcinoma, 7 new cohorts

- Tang (n=87)
  - Illumina
  - Overall survival
  - Low (29)
  - MId (29)
  - High (20)
  - P=0.046

- Rousseaux (n=81)
  - Affymetrix
  - Overall survival
  - Low (27)
  - MId (27)
  - High (27)
  - P=0.058

- Botling (n=70)
  - Affymetrix
  - Overall survival
  - Low (24)
  - MId (23)
  - High (23)
  - P=0.014

- Wilkerson (n=62)
  - Agilent
  - Overall survival
  - Low (21)
  - MId (21)
  - High (20)
  - P=0.014

- Matsuyama (n=52)
  - Agilent
  - Overall survival
  - Low (18)
  - MId (17)
  - High (17)
  - P=0.028

- Lee (n=36)
  - Affymetrix
  - Overall survival
  - Low (12)
  - MId (12)
  - High (12)
  - P=0.010

- Bild (n=34)
  - Affymetrix
  - Overall survival
  - Low (12)
  - MId (11)
  - High (11)
  - P=0.120
Figure 3

A. Stage I, ADC (9 cohorts, n=817)

B. Stage IA, ADC (9 cohorts, n=404)

C. Stage IB, ADC (9 cohorts, n=408)

Relapse-free survival

Cohorts (n) | Stage I | HR (95% CI) | %Weight
--- | --- | --- | ---
Japan (149) | | 2.19 (1.35, 3.56) | 
Lee (36) | | 2.48 (1.18, 5.20) | 
US/Norway (67) | | 2.71 (1.61, 4.55) | 
Overall survival

Cohorts (n) | Stage IA | HR (95% CI) | %Weight
--- | --- | --- | ---
Japan (100) | | 2.51 (1.17, 5.40) | 
Lee (13) | | 1.20 (0.23, 6.28) | 
US/Norway (29) | | 3.08 (1.41, 6.72) | 
Overall survival

Cohorts (n) | Stage IB | HR (95% CI) | %Weight
--- | --- | --- | ---
Japan (49) | | 1.80 (1.01, 3.23) | 
Lee (23) | | 2.96 (1.24, 7.03) | 
US/Norway (38) | | 2.36 (1.17, 4.76) | 
Overall survival
Figure 4

Nomogram to predict 5-year survival probability for cases with stage I lung cancer

Points

4-gene: low | medium | high
Stage: IA | IB
Sex: Female | Male
Age: 30 | 40 | 50 | 60 | 70 | 80 | 90

Total Points

5-year survival: 95% | 90% | 80% | 70% | 60% | 50% | 40% | 30% | 20%
The expression of four genes as a prognostic classifier for stage I lung adenocarcinoma in 12 independent cohorts

Hirokazu Okayama, Aaron J. Schetter, Teruhide Ishigame, et al.

Cancer Epidemiol Biomarkers Prev  Published OnlineFirst September 21, 2014.

Updated version
Access the most recent version of this article at:
doi:10.1158/1055-9965.EPI-14-0182

Supplementary Material
Access the most recent supplemental material at:
http://cebp.aacrjournals.org/content/suppl/2014/09/23/1055-9965.EPI-14-0182.DC1

Author Manuscript
Author manuscripts have been peer reviewed and accepted for publication but have not yet been edited.

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.