Introduction

Non–small cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancer cases, with adenocarcinoma as the main histologic subtype (1). Groundbreaking discoveries of treatment predictive alterations in the EGFR and ALK genes (2, 3) primarily made in adenocarcinomas, have led to development of targeted treatments for patients harboring these alterations. However, the number of patients eligible for these treatments is limited because of low mutation frequencies in a general patient population (4), leaving chemotherapy and/or radiotherapy as the primary treatment options for the majority of lung cancer patients. Even for NSCLC patients with the best prognosis, resectable stage I disease, the 5-year survival rate after resection is only 58% to 73%, with 3% to 10% additional survival increase if adjuvant chemotherapy is used (5). Clearly, additional tools are needed to improve patient stratification, prognostication, and prediction of response to therapy in primary lung adenocarcinoma.

In 2000, a seminal study by Perou and colleagues (6) defined gene-expression phenotypes (GEPs) in breast cancer, subsequently shown to be associated with prognosis and response to chemotherapy, independent of standard clinicopathologic variables (7, 8). In lung adenocarcinoma, similar studies of GEPs associated with patient outcome and different molecular and clinicopathologic variables have been reported (9–13). In 2006, Hayes and colleagues (11) defined three GEPs (the bronchioid, magnoid, and squamoid subtypes) in lung adenocarcinoma associated with patient outcome and specific clinicopathologic characteristics. These GEPs have subsequently been validated in independent cohorts by a derived single-sample predictor (SSP), comprising a 506-gene nearest centroid classifier (10, 14). This SSP classifies single samples into one of the three molecular subtypes based on similarities in gene-expression pattern between a specific sample and the molecular subtype centroids. Optimally, this type of SSP requires gene-expression data for each tumor to be gene centered against a large and heterogeneous reference set, which makes the assignment of a subtype to a tumor highly dependent on the composition of other tumors in both the investigated cohort and the reference set from which the SSP centroids were derived (15, 16).

The Hayes and colleagues subtypes were recently renamed as the terminal respiratory unit (TRU, formerly bronchioid), proximal-terminating, and proximal-proliferating adenocarcinoma. The Hayes and colleagues subtypes were recently renamed as the terminal respiratory unit (TRU, formerly bronchioid), proximal-terminating, and proximal-proliferating adenocarcinoma. In a large-scale evaluation, we show that GEPs add prognostic value to standard clinicopathologic variables in lung adenocarcinoma. Subject to classification, GEPs have potential to affect the prognostication of adenocarcinoma patients through a molecularly driven disease stratification.
Translational Relevance

Improved molecular characterization and stratification of lung adenocarcinoma is important for identification of new predictors of prognosis and treatment response. In this context, molecular phenotypes identified by gene-expression analysis (GEP) have been associated with distinct molecular and clinicopathologic characteristics and, importantly, also with patient outcome. If GEPs in adenocarcinoma could be validated for clinical benefit and methodologic robustness, they could provide clinically useful and molecularly driven disease stratification. On the basis of a multicohort analysis including 2,395 adenocarcinoma patients, we demonstrate that a commonly used GEP predictor adds independent prognostic information compared with standard clinicopathologic variables in most patients, but is not predictive of response to adjuvant chemotherapy. We also identify weaknesses of this and similar classifiers concerning robustness that must be addressed before such signatures may become robust, single-sample predictors that can be used effectively in a clinical context.

Expression Phenotypes and Prognosis in Lung Adenocarcinoma

proximal-proliferative (PP, magnoid), and proximal-inflammatory (PI, squamoid) subtypes (14). TRU-like tumors have been associated with better patient outcome, never-smoking patient status, EGFR mutations, lower expression of proliferation-related genes, higher expression of surfactant genes, lower tumor stage and grade, and possible sensitivity to anti-EGFR (gefitinib) treatment (10, 14, 17). Patients with PP tumors have been associated with poorer outcome, smoking, KRAS mutations, higher proliferation, and response to adjuvant cisplatin and vinorelbine chemotherapy (10, 14, 17). Finally, PI tumors have been associated with poorer patient outcome, co-mutation of NFI and TP53, higher proliferation, and a solid-type morphologic pattern (10, 14, 17). Taken together, TRU-like adenocarcinomas have the characteristics of a good prognosis group, whereas the PP and PI subtypes represent high-risk poor prognosis groups (10, 14).

In breast cancer, the GEPs originally proposed by Perou and colleagues (6) have now evolved into gene panels that are included in the international guidelines for breast cancer treatment to guide adjuvant chemotherapy treatment (18). In lung adenocarcinoma, considerable challenges remain. First, only a few of the so far proposed GEPs are actually available as SSPs for independent validations, a prerequisite for independent assessment of their potential clinical value. Second, the overlap and consistency between different GEP classifications have not been systematically investigated. In a general setting, it has been suggested that across different GEPs there is a main division of tumors into a TRU subtype enriched for females and never-smokers that express IFIT1 and surfactant proteins, and a second non-TRU subgroup with a poorer patient outcome (19). Moreover, GEPs in lung adenocarcinoma have not been comprehensively investigated for prognostic value, chemotherapy predictive value, or classification robustness in large multicohort analyses. Here, we provide the first such multicohort evaluation of the robustness and clinical benefit of the TRU, PP, and PI subtypes in lung adenocarcinoma with respect to patient outcome and response to adjuvant chemotherapy. On the basis of analysis of 17 cohorts comprising transcriptional profiles from 2,395 patients with lung adenocarcinoma, we demonstrate that classification as the TRU or non-TRU-like subtypes add prognostic information in addition to the standard clinicopathologic variables for the majority of patients. Moreover, we show that expression of proliferation-related genes is the likely causative transcriptional process behind this prognostic association. In contrast, GEPs do not seem to be predictive of response to adjuvant chemotherapy. Finally, we identify the weaknesses of both the current and similar types of GEP classifiers concerning classification robustness, highlighting the need for development of improved molecular subtype classifiers more suitable for a clinical setting. Importantly, the division of lung adenocarcinoma into molecular subgroups associated with prognosis and characteristic genomic, transcriptional, and histopathologic alterations have the potential to provide a molecularly driven disease stratification that may also be clinically useful.

Materials and Methods

Gene-expression cohorts

Published gene-expression profiles from 17 cohorts comprising 2,395 lung adenocarcinomas with available patient outcome data were collected from authors’ websites or public repositories and summarized as previously described (17). Included studies were performed in both western and Asian countries. Overall patient characteristics are summarized in Table 1, and available in detail for each cohort in Supplementary Table S1.

Four cohorts had adjuvant chemotherapy data (treatment/no treatment), including 322 cases from Shedden and colleagues (ref. 20; chemotherapy type not explicitly specified), 133 cases from Sato and colleagues (ref. 21; UT Lung SPORE randomized trial, combination of mainly carboplatin plus taxanes, see also Supplementary Methods), 85 cases from Fouret and colleagues (ref. 22; cisplatin-based chemotherapy), and 28 adenocarcinomas from Zhu and colleagues (ref. 23; JBR.10 randomized trial, combination of cisplatin/vinorelbine chemotherapy). Specific data on treatment cycles, treatment duration, and chemotherapy doses were not available. In total, 562 patients from these cohorts had associated outcome data (overall survival, OS), and 176 of these patients received adjuvant chemotherapy (Table 1).

Gene-expression analyses

Cohorts analyzed by Affymetrix gene-expression microarrays were individually normalized as described in (24), whereas cohorts analyzed by other microarray platforms or by RNA sequencing were processed as described in the Supplementary Methods. On a per cohort basis, tumors were classified as the TRU, PP, or PI subtypes using the 506-gene nearest centroid predictor reported by Wilkerson and colleagues (10). The baseline methodologic classification approach for a cohort/sample included: (i) match of genes common to the 506-gene predictor, (ii) median centering of cohort expression data, (iii) averaging the expression of multiple probes for a specific centroid gene for each sample, (iv) Pearson correlation as the similarity metric between a specific sample’s matched expression vector and the centroids, and (v) the maximum correlation coefficient determining the GEP for a sample. To assess classification robustness, we introduced different alterations to the original classification approach including (i) change of gene-centering method (mean instead of median), (ii) change of correlation method (Spearman instead of Pearson), (iii) classification using different subsets of
Table 1. Clinical characteristics of patients in the total cohort and in the cohort of patients receiving adjuvant chemotherapy

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Total cohorta</th>
<th>Adjuvant chemotherapy cohortb</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>2,395</td>
<td>176</td>
</tr>
<tr>
<td>Number of cohorts</td>
<td>17</td>
<td>4</td>
</tr>
<tr>
<td>Stage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>63%</td>
<td>45%</td>
</tr>
<tr>
<td>II</td>
<td>20%</td>
<td>26%</td>
</tr>
<tr>
<td>III</td>
<td>15%</td>
<td>29%</td>
</tr>
<tr>
<td>IV</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>53%</td>
<td>56%</td>
</tr>
<tr>
<td>Male</td>
<td>47%</td>
<td>44%</td>
</tr>
<tr>
<td>Age (median and range)</td>
<td>64 (21–91)</td>
<td>62 (35–83)</td>
</tr>
<tr>
<td>Smoking status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never-smoker</td>
<td>24%</td>
<td>25%</td>
</tr>
<tr>
<td>Smoker</td>
<td>76%</td>
<td>75%</td>
</tr>
<tr>
<td>Mutation status</td>
<td></td>
<td></td>
</tr>
<tr>
<td>EGFR</td>
<td>29%</td>
<td>39%</td>
</tr>
<tr>
<td>KRAS</td>
<td>22%</td>
<td>27%</td>
</tr>
<tr>
<td>Outcome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>OS, median (y)</td>
<td>2.94</td>
<td>3.76</td>
</tr>
<tr>
<td>OS number of events (% of all patients)</td>
<td>38%</td>
<td>55%</td>
</tr>
<tr>
<td>DMFS, median (y)</td>
<td>3</td>
<td>1.59</td>
</tr>
<tr>
<td>DMFS number of events (% of all cases with DMFS data)</td>
<td>42%</td>
<td>81%</td>
</tr>
<tr>
<td>GEPs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TRU</td>
<td>39%</td>
<td>29%</td>
</tr>
<tr>
<td>PP</td>
<td>30%</td>
<td>37%</td>
</tr>
<tr>
<td>PI</td>
<td>31%</td>
<td>34%</td>
</tr>
</tbody>
</table>

*Characteristics are presented as percentages of all cases with available annotation.

1Only patients with available OS data.

were females) in the total cohort, whereas higher rates of KRAS mutations were found in smokers (29% of smokers carried KRAS mutations; Fisher exact test \( P < 0.05 \) for all individual comparisons). In univariate analysis using OS as endpoint, tumor stage was associated with patient outcome in 10 of 13 cohorts (\( n = 1,867/2,059 \) patients) with available clinical data (\( P < 0.05 \)). Univariate analysis of gender, smoking history, age, and adjuvant chemotherapy (yes/no, irrespective of type) showed association with OS (\( P < 0.05 \)) in only 2 of 14 cohorts (\( n = 306/2,143 \) patients), 1 of 11 cohorts (\( n = 204/1,843 \) patients), 2 of 13 cohorts (\( n = 554/2,059 \) patients), and 2 of 4 cohorts (\( n = 401/562 \) patients) with available clinical available data, respectively. In the subset of patients treated with adjuvant chemotherapy (\( n = 176 \)), there were proportionally less stage I patients and TRU-classified patients compared with the total cohort (Table 1). This difference is consistent with the fact that adjuvant chemotherapy is primarily administered to high-risk patients (higher stage).

Validation of the GEP classification approach

To confirm that our GEP classification approach was accurate, we compared our subtype classifications with the subtypes reported by the authors of the original GEPs (10, 11) in two cohorts: The recently reported set of 230 cases from The Cancer Genome Atlas (TCGA) consortium based on RNA sequencing data (14) and the Wilkerson and colleagues cohort (ref. 10; \( n = 116 \)) based on Agilent microarray data. Here, the 230-sample TCGA cohort represents a subset of the 435-sample TCGA cohort used later in this study. In both cohorts, we observed 100% agreement between our classifications and the previously reported subtypes (10, 14). A striking similarity in the proportions of the different GEPs was observed when the predictor was applied to the 17 cohorts, with approximately 30% to 40% of cases classified as TRU, approximately 25% to 35% of cases classified as PP, and approximately 25% to 35% of cases classified as PI (Fig. 1). Moreover, there were no obvious differences in subtype proportions between
larger (e.g., TCGA and Shedden and colleagues; ref. 20) and smaller sized cohorts (e.g., Hou and colleagues; ref. 30, Lee and colleagues (31), and Zhu and colleagues (23)).

In the total cohort, GEP classification confirmed previous molecular and clinicopathologic associations of the subtypes, including a higher frequency of stage I tumors, EGFR mutations, females, and never-smokers among TRU-classified cases, and a higher frequency of KRA5 mutations in PP-classified cases ($\chi^2$ test $P < 1 \times 10^{-4}$ for all comparisons; refs. 10, 14).

**Prognostic value of the GEP classification**

To assess the prognostic value of the TRU, PP, and PI classification in lung adenocarcinoma, we first performed a univariate analysis using OS or DMFS as the clinical endpoints. The GEP classification (PP and/or PI) was associated with OS in the total cohort (PP: HR, 2.1; 95% confidence interval [CI], 1.2–3.9; and PI: HR, 1.9; 95% CI, 1.2–2.6) and in 5 of 7 individual cohorts with available OS data (data not shown). Consistently, no differences in OS were found between PP and PI cases in any cohort in the univariate analysis (Supplementary Table S2). Only in the total cohort does having high or low metastasis expression based on being above or below the median metastasis expression in its cohort. We excluded the Napsin A/surfactant metagene due to its inverse correlation with the proliferation metagene (Fig. 3A). Next, for each metagene we added this high/low classification to a multivariate model including TRU/non-TRU classification and tumor stage, with OS as the endpoint (Fig. 4A and B). When including a proliferation covariate, the prognostic impact of the GEP classification clearly decreased in the total cohort (non-TRU: HR, 1.4; 95% CI, 1.1–1.7; $P = 0.006$), and remained statistically significant in only one of the 13 cohorts with available clinical data (CLCGP, $n = 91$ patients, $P = 0.04$; Fig. 4A). In contrast, inclusion of other metagene classifications as covariates in the multivariate analysis retained the significance of the GEP classification in more cohorts (five cohorts, with additional borderline nonsignificant results in the Tomida and colleagues and the TCGA cohorts; Fig. 4B). To further confirm the prognostic importance of the high/low proliferation stratification, we performed multivariate analysis using only high/low proliferation status, tumor stage, and patient age as covariates and OS as the endpoint. Here, the proliferation stratification added independent prognostic information in the same cohorts as the GEP classification (Fig. 2C, model B), with one exception (CLCGP cohort, $P = 0.07$; data not shown).

Together, these analyses suggest that expression of proliferation-related genes is one of the main dividing factors between TRU and non-TRU adenocarcinomas, and a key component in the prognostic association of the GEPs.

**Chemotherapy predictive value of the GEPs**

To assess the predictive value of the GEPs in patients receiving adjuvant chemotherapy treatment, we performed univariate analysis in the subset of 176 patients treated with adjuvant chemotherapy (mixed chemotherapy regimens, including cisplatin, carboplatin, taxanes, and vinorelbine; Table 1, Fig. 4C). In these 176 cases, neither univariate analysis of the TRU/PP/PI classification, the PP/PI classification, or the TRU/non-TRU classification showed statistically significant associations with OS ($P > 0.05$ for all comparisons). Similar nonsignificant results were also found in the four individual cohorts comprising the 176 samples, including the Zhu and colleagues and Sato and colleagues randomized trial cohorts (21, 23). The nonsignificant trend in these analyses was that high-risk groups...
Figure 2.
Prognostic association of the TRU, PP, and PI GEPs in lung adenocarcinoma. A, univariate analysis of the association of TRU, PP, or PI classification with OS across 16 cohorts (Lee et al. was excluded, as OS was not available). In the panel, the TRU subtype acts as the reference group. B, univariate analysis of association with OS for the TRU, and non-TRU classification (PP and PI cases combined) across the 16 cohorts. In the panel, the TRU subtype acts as the reference group. C, multivariate analysis including GEP classification (TRU/non-TRU) and different covariates in three different models (A, GEP and tumor stage; B, GEP, tumor stage, and patient age; C, GEP, tumor stage, patient age, smoking status, and gender) using OS as endpoint. Only HRs and 95% CIs for the GEP covariate in the different models are shown. For some cohorts, not all clinicopathologic data were available. Cohorts with significant GEP associations (P < 0.05) are marked with asterisks.
(non-TRU tumors) had higher HRs, and thus worse outcome, than TRU-like tumors (the low-risk group).

In a second analysis, we investigated whether adjuvant chemotherapy was associated with outcome within TRU (n = 223), PP (n = 176), PI (n = 163), or non-TRU (n = 339) tumors from cohorts with treatment data (chemotherapy yes/no) using univariate analysis and OS as the endpoint (Fig. 4D). However, we found no support for adjuvant chemotherapy being associated with OS in the total or individual cohorts for any subgroup (univariate $P > 0.05$ for all comparisons).

Together, these analyses, albeit performed in limited patient materials, do not support that GEP classification is clearly predictive of response to adjuvant chemotherapy in lung adenocarcinoma.

Robustness of the GEP classification

In a clinical setting, classification robustness of individual samples is crucial. To evaluate the robustness of the TRU, PP, and PI classification, we investigated classification variation due to methodologic changes in the classification approach, missing values for the gene-expression centroids, and random or biased shifts in cohort composition. We first analyzed the variability in classification due to methodologic changes in gene-centering method and similarity metric (correlation type) across the 17 cohorts. For each alteration, typically $<10\%$ of cases in each cohort switched GEP (Fig. 5A, top). Next, we performed classification using our standard approach, but randomly varied the number of genes available for classification. The purpose of this analysis was to mimic a reduction in
Figure 4.
Influence of transcriptional metagenes on prognostic association of GEPs and association of subtypes with response to adjuvant chemotherapy. A, multivariate analysis of the association with patient outcome based on OS for a model including TRU/non-TRU classification, tumor stage, and high/low proliferation status across 13 cohorts with available clinical data. In the panel, HRs with 95% CIs are displayed for the non-TRU subgroup only. Cohorts with significant associations (P < 0.05) are marked with asterisks. B, multivariate analysis of the association with OS for (i) a model including TRU/non-TRU, tumor stage, and high/low immune response (IR) status (model A), (ii) a model including TRU/non-TRU, tumor stage, and high/low stroma/ECM status (model B), and (iii) a model including TRU/non-TRU, tumor stage, and high/low basal/squamous status (model C) across the same 13 cohorts. In the panel, HRs are displayed for the non-TRU subgroup only. Cohorts with significant associations (P < 0.05) are marked with asterisks. C, univariate analyses of the association of GEPs with OS in patients treated with adjuvant chemotherapy. x, all univariate P > 0.05 in the three-group analysis, where the TRU group is the reference group; PP P = 0.19, PI P = 0.07. D, univariate analyses of association of chemotherapy treatment with OS in different gene-expression subtypes for patients in the total cohort with chemotherapy data (yes/no, n = 562).
centroid size due to missing expression data in a cohort. Across cohorts, typically <10% of cases switched GEP if the random selection included ≥50% of the original genes (i.e., approximately ≥250 genes) (Fig. 5A, center). These results are consistent with gene signature stability analyses performed by Lauss and colleagues in bladder cancer (34). Next, we evaluated the classification robustness when randomly selecting a subset of samples from a cohort that was subsequently used to re-center the entire cohort before classification. The purpose of this analysis was to investigate the classification stability when...
introducing differences in gene-centering due to random perturbations of the sample composition. Again, GEP classification remained stable with typically <10% of cases switching subtype (Fig. 5A, bottom).

To investigate the effect of nonrandom perturbations of the cohort composition, we performed a similar analysis as recently described by Paquet and Hallet (25) in the 230-sample TCGA cohort (previously used to confirm our classification approach; ref. 14). In a cumulative sample-by-sample exclusion of TRU-classified tumors, in each step followed by gene recentering and reclassification, we observed a growing fraction of samples originally classified as PP or PI switching subtype (Fig. 5B). Samples switching subtype were almost exclusively reclassified as TRU-like. Notably, the subtype proportions before sample removal and after complete removal of all TRU classified samples were highly similar for all three subtypes: 38.7% versus 35.5% for TRU, 27.4% versus 31.2% for PP, and 33.9% versus 33.3% for PI. Similarly, when cumulatively excluding PP or PI tumors, samples switching subtype were consistently reclassified as the excluded subtype (Supplementary Fig. S1). Next, we performed the same analysis in all 17 cohorts, and found that cumulative exclusion of TRU-classified cases caused 30% to 40% of cases originally classified as PP or PI to switch subtype across cohorts. Similar results, albeit with lower fractions of cases switching GEP, were found when performing the same exclusion analysis for the PP and PI subtypes (Supplementary Fig. S1).

Together, these analyses demonstrate that the studied SSP is generally stable to random perturbations, whereas unbalanced shifts in cohort composition introduce greater variability in subtype assignment. Although methodologic factors are easily addressable, classification variance due to missing expression data or shifts in cohort composition may be more difficult to address in the current SSP. Consequently, there is a need for improved, or new, SSPs more suitable for a clinical setting.

Discussion

In this study, we conducted a large-scale multicohort analysis to assess the prognostic value of GEPs in lung adenocarcinoma, showing that the TRU/non-TRU division adds prognostic value to standard clinicopathologic variables in this disease. We also show that expression of genes relating to proliferation drives the prognostic differences between the TRU and non-TRU subtypes. Although GEPs in lung adenocarcinoma have previously been proposed to be associated with distinct molecular and clinicopathologic characteristics, and importantly also patient outcome (10, 11, 14, 19, 26, 33), they have not been validated on such a large scale before. Moreover, based on extensive analysis of classifier robustness, we demonstrate that additional classifier development is required to derive robust, SSPs for a clinical setting.

In this study, we demonstrate that the most prognostically important division of the studied GEP classifier is between TRU and non-TRU cases, where TRU represents a low-risk group and non-TRU a high-risk group. Thus, despite the reported associations of the PP and PI subtypes with different histopathologic, clinicopathologic, molecular, and transcriptional differences (10, 11, 14, 19, 26, 33), the stratification into two proximal subtypes seems to have less prognostic importance. This lack of prognostic difference is presumably due to an equally high level of tumor proliferation in both subtypes. Importantly, the TRU/non-TRU division provides independent prognostic information in multivariate models, including current clinicopathologic prognostic covariates, both overall and in most individual cohorts. These results support that GEP classification, or prognostic gene signatures for that matter, have clinical relevance in lung adenocarcinoma compared with standard prognostic variables (20). Nevertheless, the GEP classification was not a significant factor for prognosis in several cohorts, despite the similarity in GEP proportions across cohorts. Several explanations are conceivable. For some cohorts, the lack of prognostic association could simply be due to small sample sizes (e.g. the Zhu and colleagues cohort) or insufficient patient follow-up time. In other cohorts, a biased sample selection may play a larger role. Finally, intrinsic characteristics of the SSP itself, as discussed further below, may explain why cohorts with more early-stage tumors, which are typically low proliferative with better patient outcome, did not have higher proportions of TRU-classified cases. Importantly, the last two types of bias cannot be properly addressed by the current SSP. Taken together, the prognostic heterogeneity between cohorts is likely to be dependent on both cohort/patient characteristics and methodological factors.

We and others have shown that prognostic high-risk groups in NSCLC benefit more from adjuvant chemotherapy than less proliferative low-risk cases (23, 35, 36). In this study, there was no support for GEP classification providing predictive value for adenocarcinoma patients receiving adjuvant chemotherapy. This observation is consistent with results from Zhu and colleagues (23), whose chemotherapy response signature was not predictive in the adjuvant cisplatin-treated patient arm alone. In contrast with previous studies (23, 35, 36), we found no differences in patient outcome within the high-risk PP, PI, or non-TRU patient subgroups when stratified by chemotherapy treatment. One explanation may be that our pooled cohort analysis only included patients with adenocarcinoma, whereas previous studies included both adenocarcinomas and squamous cell carcinomas (SCC; refs. 23, 35, 36). Moreover, for two of the analyzed cohorts (Shedden and colleagues (20) and Fouret and colleagues (22)), it is unclear whether all or only subsets of the analyzed patients were randomized to adjuvant chemotherapy treatment. This uncertainty represents a major source of potential bias in the analysis, as high-risk patients typically receive adjuvant chemotherapy. Thus, it may be argued that this type of analysis may only be performed in treatment-randomized cohorts. However, a nonsignificant association of GEP classification with chemotherapy response was also found in the two included cohorts based on randomized trials, Zhu and colleagues (ref. 23; JBR.10 trial) and Sato and colleagues (ref. 21; UT Lung SPORE trial). In perspective, the survival benefit of traditional platinum-based adjuvant chemotherapy in patients with NSCLC is only approximately 4% to 5%, with an unclear benefit for stage IA patients and no proven effect of tumor histology (37, 38). Besides stressing the importance of new prediction tools to avoid overtreatment of patients and cytotoxic side effects, this low efficacy suggests that large patient numbers are needed to derive such tools, and also to determine interactions and efficacy of specific chemotherapy combinations with predictive signatures. Consequently, we acknowledge that analyses of larger randomized adjuvant...
Expression Phenotypes and Prognosis in Lung Adenocarcinoma

chemotherapy trials may provide more definite conclusions about the chemotherapy predictive value of GEPs in lung adenocarcinoma. Unfortunately, at this point, current chemotherapy-predictive gene signatures have not provided insights into whether results are related to the histologic composition of the investigated cohorts.

This study also dissects the transcriptional programs defining the GEPs, showing that, for example, TRU-classified tumors may be defined by lower proliferation and higher expression of Napsin A and different surfactant genes. These characteristics are in excellent agreement with the proposed existence of a TRU-like adenocarcinoma subgroup based on positive TTF-1 immunostaining showing morphologic similarity to type II pneumocytes, Clara cells, and nonciliated bronchioles (19). Moreover, these analyses also identified expression of proliferation-related genes as an important carrier of prognostic information in the GEP classification and in lung adenocarcinoma in general, consistent with previous studies (12, 20, 35, 36).

A crucial characteristic of a clinically useful predictor is its robustness. In our analyses, the investigated SSP shows reasonable stability (low percentage of cases switching GEP) when introducing modest alterations in the methodologic classification approach, decreased centroid size, or random perturbations of cohort composition. The intrinsic stability problems with SSPs based on nearest centroid predictors, caused by their dependence on gene centering, become evident when classification is reperformed after removing increasing fractions of a specific GEP, causing large number of cases to switch subtype (Fig. 5). Specifically, the gene-centering dependence of the SSP makes it biased toward keeping the proportions of the subtypes similar in each cohort, more or less irrespectively of the sample composition. The classification variance caused by unbalanced shifts in cohort composition may be further illustrated by the simultaneous shifts introduced in the transcriptional metagenes, for example, the proliferation metagene, during the cumulative GEP-specific sample removal. For the proliferation metagene, exclusion of increasing numbers of tumors originally classified as TRU-like (typically low-proliferative) followed by gene-recentering shifts the center of proliferation (mean/median) toward more proliferative cases. The latter cases now appear as less proliferative and become reclassified as TRU-like by the SSP (due to the strong dependence on expression of proliferation-related genes in the classification).

Notably, the classification variance introduced when selectively removing one of the high-proliferative groups (PP or PI) was lower than selectively removing TRU-like cases. This is because cumulative removal of only one of the two high-proliferative groups retains the median proliferation closer to the original, as one high-proliferative group remains in the gene-centering step. Although it could hypothetically be argued that all 17 investigated cohorts in this study are equally representative of primary lung adenocarcinoma, the strikingly similar subtype fractions across cohorts (Fig. 1) support some type of classification bias inherent to the used SSP. In further support of this hypothesis, we did not observe any clear trends of higher proportions of TRU-classified cases in cohorts with more early-stage tumors, or vice versa higher proportions of cases classified as PP or PI in cohorts with more higher-stage tumors (based on data presented in Supplementary Table S1). Importantly, these findings are in line with the results from the recent study by Paquet and Hallet (25) regarding GEPs in breast cancer. Furthermore, our findings are not limited to GEPs in lung adenocarcinoma, but should also apply to other proposed GEPs in lung cancer, for example, in SCC (39). In addition, our observations may also explain why different reported prognostic gene-expression signatures and SSPs in lung cancer often are less successful when applied to independent cohorts.

Although slightly different SSP implementations, or differences in cohort compositions between studies, may have limited impact on the broad characteristics of the GEPs in terms of, for example, patient outcome, it is not satisfactory in a clinical setting that classifications of individual patients are not robust. Thus, there is an imminent need of robust, standardized methods for molecular subtype classification of individual lung adenocarcinomas independent of data from other tumors. Although various modifications of existing SSPs based on nearest centroid classification are conceivable, Paquet and Hallet recently proposed a new SSP for molecular subtyping of breast cancer (25). This new predictor relates raw expression measurements of subtype-specific genes to the levels of other genes within each tumor sample, omitting the gene-centering step and thus representing a more true SSP. Not only would such a classifier likely be more robust across different adenocarcinoma cohorts, but also to highly selected cohorts, such as high-risk patients from neoadjuvant trials.

Provided that a more robust SSP can be developed, adaptation of the TRU/non-TRU classification for clinical use may require deriving mixed prognostic models that include GEP classification, standard clinicopathologic variables, and possibly other gene-expression covariates (e.g., metagene scores) to obtain the best prognostic performance. Mixed prognostic models have already been reported in lung adenocarcinoma and breast cancer, including both strictly prognostic signatures as well as GEP classifications (breast cancer) (8, 20, 40). Importantly, such a risk classifier should also be applicable to formalin-fixed paraffin-embedded tissue, as this is the most commonly available clinical tissue type today, especially for advanced lung cancer. Here, the reduction of large multigene centroids, like the studied 506-gene Wilkerson and colleagues (10) centroids, into smaller sets of hub genes/metagenes may be important for the actual clinical implementation of gene expression–based assays, especially if more focused multigene platforms are required for analysis of challenging routine clinical specimens (8, 36).

In summary, our study demonstrates that GEPs in lung adenocarcinoma can have clinical value. Importantly, we also identify several key issues that need to be addressed before such signatures can become clinically relevant, robust, SSPs. Given appropriate classifier development, our results suggest that GEPs in lung adenocarcinoma have the potential to add prognostic information beyond the currently used clinicopathologic variables, based on a molecularly driven disease stratification.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

Authors’ Contributions

Conception and design: J. Staaf

Development of methodology: M. Ringnér, G. Jönsson, J. Staaf

www.aacjournals.org Clin Cancer Res; 22(1) January 1, 2016 227

Downloaded from clincancerres.aacrjournals.org on November 11, 2017. © 2016 American Association for Cancer Research.
Acquisition of data (provided animals, acquired and managed patients, provided facilities, etc.): J. Staaf

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): M. Ringnér, G. Jonsson, J. Staaf

Writing, review, and/or revision of the manuscript: M. Ringnér, G. Jonsson, J. Staaf

Study supervision: J. Staaf

Grant Support

This study was financially supported by the Swedish Cancer Society, the Knut and Alice Wallenberg Foundation, the Foundation for Strategic Research through the Lund Centre for Translational Cancer Research (CREATE Health), the Mrs Berta Kamprad Foundation, the Gunnar Nilsson Cancer Foundation, the Swedish Research Council, and the Gustav V’s Jubilee Foundation.

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 March 6, 2015; revised July 9, 2015; accepted August 3, 2015; published OnlineFirst August 11, 2015.

References


Prognostic and Chemotherapy Predictive Value of Gene-Expression Phenotypes in Primary Lung Adenocarcinoma

Markus Ringnér, Göran Jönsson and Johan Staaf


Updated version Access the most recent version of this article at: doi:10.1158/1078-0432.CCR-15-0529

Supplementary Material Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2015/08/13/1078-0432.CCR-15-0529.DC1

Cited articles This article cites 38 articles, 20 of which you can access for free at: http://clincancerres.aacrjournals.org/content/22/1/218.full#ref-list-1

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, use this link http://clincancerres.aacrjournals.org/content/22/1/218. Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.