Prognostic and chemotherapy predictive value of gene expression phenotypes in primary lung adenocarcinoma

Markus Ringnér\textsuperscript{1,2}, Göran Jönsson\textsuperscript{1,2}, Johan Staaf\textsuperscript{1,2,*}

\textsuperscript{1}Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, Medicon Village, SE 22381, Lund, Sweden
\textsuperscript{2}CREATE Health Strategic Center for Translational Cancer Research, Lund University, Medicon Village, SE 22381, Lund, Sweden

*Corresponding author

Running title: Expression phenotypes and prognosis in lung adenocarcinoma

Key words: lung cancer, adenocarcinoma, molecular subtype, prognosis, gene expression, chemotherapy response

Corresponding author:

Johan Staaf, e-mail: johan.staaf@med.lu.se, Address: Division of Oncology and Pathology, Department of Clinical Sciences Lund, Lund University, Medicon Village, SE 22381 Lund, Sweden, Fax: +4646147327

Competing interests

Authors declare that they have no competing interests.

Word count: 4999 Total number of figures and tables: 6
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 (GEPs) have been associated with distinct molecular and clinicopathological characteristics and, importantly, also with patient outcome. If GEPs in adenocarcinoma could be validated for clinical benefit and methodological robustness, they could provide clinically useful and molecularly driven disease stratification. Based on a multicohort analysis including 2395 adenocarcinoma patients, we demonstrate that a commonly used GEP predictor adds independent prognostic information compared with standard clinicopathological 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.
Abstract

Purpose

Primary lung adenocarcinoma remains a deadly disease. Gene expression phenotypes (GEPs) in adenocarcinoma have potential to provide clinically relevant disease stratification for improved prognosis and treatment prediction, given appropriate clinical and methodological validation.

Experimental Design

2395 transcriptional adenocarcinoma profiles were assembled from 17 public cohorts and classified by a nearest centroid GEP classifier into three subtypes: terminal respiratory unit (TRU), proximal-proliferative, and proximal-inflammatory, and additionally scored by five transcriptional metagenes representing different biological processes, including proliferation. Prognostic and chemotherapy predictive associations of the subtypes were analyzed by univariate and multivariate analysis using overall survival or distant metastasis-free survival as endpoints.

Results

Overall, GEPs were associated with patient outcome in both univariate and multivariate analyses, although not in all individual cohorts. The prognostically relevant division was between TRU and non-TRU classified cases, with expression of proliferation-associated genes as a key prognostic component. In contrast, GEP classification was not predictive of adjuvant chemotherapy response. GEP classification showed stability to random perturbations of genes or samples and alterations to classification procedures (typically <10% of cases per cohort switching
subtype). High classification variability (>20% of cases switching subtype) was observed when removing larger or entire fractions of a single subtype, due to gene-centering shifts not addressable by the classifier.

**Conclusions**

In a large-scale evaluation we show that GEPs add prognostic value to standard clinicopathological variables in lung adenocarcinoma. Subject to classifier refinement and confirmation in prospective cohorts, GEPs have potential to impact the prognostication of adenocarcinoma patients through a molecularly driven disease stratification.
Introduction

Non-small cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancer cases, with adenocarcinoma as the main histological 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 due to 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 58-73%, with 3-10% 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 et al. (6) defined gene expression phenotypes (GEPs) in breast cancer, subsequently shown to be associated with prognosis and response to chemotherapy, independent of standard clinicopathological variables (7, 8). In lung adenocarcinoma, similar studies of GEPs associated with patient outcome and different molecular and clinicopathological variables have been reported (9-13). In 2006, Hayes et al. defined three GEPs (the bronchioid, magnoid, and squamoid subtypes) in lung adenocarcinoma associated with patient outcome and specific clinicopathological characteristics (11). These GEPs have subsequently been validated in independent cohorts by a derived single sample predictor (SSP), comprising of 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.
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 et al. subtypes were recently renamed as the terminal respiratory unit (TRU, formerly bronchioid), 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, \textit{EGFR} mutations, lower expression of proliferation-related genes, higher expression of surfactant genes, lower tumor stage and grade, and possible sensitivity to anti-\textit{EGFR} (gefitinib) treatment (10, 14, 17). Patients with proximal-proliferative tumors have been associated with poorer outcome, smoking, \textit{KRAS} mutations, higher proliferation, and response to adjuvant cisplatin and vinorelbine chemotherapy (10, 14, 17). Finally, proximal-inflammatory tumors have been associated with poorer patient outcome, co-mutation of \textit{NF1} and \textit{TP53}, higher proliferation, and a solid-type morphological pattern (10, 14, 17). Taken together, TRU-like adenocarcinomas have the characteristics of a good prognosis group, while the proximal-proliferative and proximal-inflammatory subtypes represent high-risk poor prognosis groups (10, 14).

In breast cancer, the GEPs originally proposed by Perou et al. (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. Firstly, 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. Secondly, 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 TITF-1 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, proximal-proliferative, and proximal-inflammatory subtypes in lung adenocarcinoma with respect to patient outcome and response to adjuvant chemotherapy. Based on analysis of 17 cohorts comprising transcriptional profiles from 2395 patients with lung adenocarcinoma we demonstrate that classification as the TRU or non-TRU-like subtypes add prognostic information in addition to the standard clinicopathological 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 histopathological 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 2395 lung adenocarcinomas with available patient outcome data were collected from authors’ web-sites 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 et al. (20) (chemotherapy type not explicitly specified), 133 cases from Sato et al. (21) (UT Lung SPORE randomized trial, combination of mainly carboplatin plus taxanes, see also Supplementary Methods), 85 cases from Fouret et al. (22) (cisplatin-based chemotherapy), and 28 adenocarcinomas from Zhu et al. (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), 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 reference (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, proximal-proliferative, or proximal-inflammatory subtypes using the 506-gene nearest centroid predictor reported by Wilkerson et al. (10). The baseline methodological 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 centroid genes (mimicking missing data values), and iv) classification after random or biased perturbation of the cohort composition. Specifically, we introduced shifts in the gene-centering in a cohort by centering on random subset of samples, or biased shifts by exclusion of samples of a specific GEP similar to the study by Paquet and Hallet (25).

In addition, tumors were scored according to five expression metagenes previously identified in lung cancer (26); proliferation, immune response, basal / squamous, stroma / extra cellular matrix, and expression of Napsin A / surfactants. Explicit details on cohorts, normalization, scoring procedures, and classification procedures are available in the Supplementary Methods.

Survival analyses

Univariate and multivariate Cox analyses of overall survival (OS) or distant metastasis-free survival (DMFS) data were performed in R (27) using the survival
package (see also Supplementary Methods). Due to differences in patient follow-up time between cohorts, a five-year censoring was used in all survival analyses.

Results

Cohort demographics

In total, we analyzed 2395 transcriptional adenocarcinoma profiles from 17 cohorts (Table 1, and Supplementary Table S1). Consistent with the literature (28, 29), we observed higher rates of \textit{EGFR} mutations and female gender in never-smokers (54% of never-smokers carried \textit{EGFR} mutations and 81% of never-smokers were females) in the total cohort, whereas higher rates of \textit{KRAS} mutations were found in smokers (29% of smokers carried \textit{KRAS} mutations) (Fisher’s exact test p<0.05 for all individual comparisons). In univariate analysis using OS as endpoint, tumor stage was associated with patient outcome in 10/13 cohorts (n=1867 of 2059 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 2/14 cohorts (n=306/2143 patients), 1/11 cohorts (n=204/1843 patients), 2/13 cohorts (n=554/2059 patients), and 2/4 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 consortium based on RNA sequencing data (14) and the Wilkerson et al. cohort (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 ~30-40% of cases classified as TRU, ~25-35% of cases classified as proximal-proliferative, and ~25-35% of cases classified as proximal-inflammatory (Figure 1). Moreover, there were no obvious differences in subtype proportions between larger (e.g., TCGA and Shedden et al. (20)) and smaller sized cohorts (e.g., Hou et al. (30), Lee et al. (31), and Zhu et al. (23)).

In the total cohort, GEP classification confirmed previous molecular and clinicopathological 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 KRAS mutations in proximal-proliferative classified cases (Chi-square p<1e-4 for all comparisons) (10, 14).

**Prognostic value of the GEP classification**

To assess the prognostic value of the TRU, proximal-proliferative, proximal-inflammatory classification in lung adenocarcinoma, we first performed a univariate analysis using OS or DMFS as the clinical endpoints. The GEP classification (proximal-proliferative and/or proximal-inflammatory) was associated with OS in the
total cohort (proximal-proliferative: hazard ratio (HR) = 2, 95% confidence interval (CI) = 1.7-2.4, and proximal-inflammatory: HR=2, 95% CI=1.6-2.4) and in 10 out of the 16 individual cohorts with available OS data (Lee et al. (31) excluded as no OS data available) (Figure 2A and Supplementary Table S2). For DMFS, GEP classification was also significant in the total cohort (proximal-proliferative: HR=1.6, CI=1.3-2.1, and proximal-inflammatory: HR=1.9, 95% CI=1.5-2.4) and in five out of seven individual cohorts with available DMFS data (p<0.05, Supplementary Table S3). While TRU-classified cases consistently showed the best patient outcome in Kaplan-Meier analyses, which of the proximal-proliferative and proximal-inflammatory subtypes that had the worst outcome varied across cohorts (data not shown). Consistently, no differences in OS were found between proximal-proliferative and proximal-inflammatory cases in any cohort in the univariate analyses (Supplementary Table S2). Only in one cohort (CLCGP consortium (32)), was a weak difference in DMFS observed between proximal-proliferative and proximal-inflammatory cases (Supplementary Table S3). These results suggest that the division of cases into TRU and non-TRU (proximal-proliferative and proximal-inflammatory combined) subtypes is the most relevant from a prognostic standpoint. Indeed, this TRU/non-TRU division was associated with outcome in 9/16 and 3/7 cohorts for OS and DMFS, respectively, and in the total cohort for OS (non-TRU: HR=2, CI=1.7-2.3) and DMFS (non-TRU: HR=1.8, CI=1.5-2.2) (Figure 2B and Supplementary Tables S2 and S3). Moreover, we found an association of TRU/non-TRU classification with OS in stage I tumors by univariate analysis in the total cohort (non-TRU: HR=2, CI=1.5-2.5, n=1318 patients) and in five out of the nine cohorts in which the TRU/non-TRU classification was significant in the total cohort (see Figure
2B), with additional borderline non-significance in the Shedden et al. cohort (p=0.051) (data not shown).

Multivariate analysis was performed to investigate the clinical benefit of the TRU/non-TRU classification compared with conventional prognostic variables. Three different multivariate models were considered that included different covariates: i) GEP and tumor stage, ii) GEP, tumor stage and patient age, and iii) GEP, tumor stage, patient age, smoking status, and gender. In the total cohort, the TRU/non-TRU classification added independent prognostic information for all three models, and also in five out of eight individual cohorts showing univariate significance, with two additional cohorts (Tomida et al. (33) and TCGA) being borderline non-significant (Figure 2C and Supplementary Table S2).

Together, these analyses demonstrate that the TRU/non-TRU-like classification add independent prognostic information for most patients in addition to the standard clinicopathological variables in lung adenocarcinoma.

*Transcriptional processes behind the prognostic value of the GEPs*

The origin of the TRU, proximal-proliferative, and proximal-inflammatory subtypes is based on unsupervised gene expression analysis (10, 11). To identify biological processes associated with the GEPs, which may also account for their prognostic association, we analyzed the subtypes with respect to five lung cancer derived transcriptional metagenes representing different biological processes: proliferation, Napsin A / surfactant expression, immune response, stroma / extracellular matrix, and basal / squamous expression (26). Notably, the three GEPs appeared readily definable by different combinations of these metagenes (Figures 3A and B). However, in the total cohort the expression of these metagenes all mimicked unimodal distributions of
continuous data values (Figure 3C). Importantly, these unimodal distributions imply that associations of GEPs with different metagenes will be dependent on gene-centering and arbitrary cut-offs.

To assess the prognostic influence of these metagenes in relation to the GEP classification, we first classified each patient as having high or low metagene expression based on being above or below the median metagene expression in its cohort. We excluded the Napsin A / surfactant metagene due to its inverse correlation with the proliferation metagene (Figure 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 (Figures 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, 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) (Figure 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 non-significant results in the Tomida et al. and the TCGA cohorts; Figure 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 (Figure 2B, 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, Figure 4C). In these 176 cases, neither univariate analysis of the TRU/proximal-proliferative/proximal-inflammatory classification, the proximal-proliferative/proximal-inflammatory classification, or the TRU/non-TRU classification showed statistically significant associations with OS (p>0.05 for all comparisons). Similar non-significant results were also found in the four individual cohorts comprising the 176 samples, including the Zhu et al. and Sato et al. randomized trial cohorts (21, 23). The non-significant trend in these analyses was that high-risk groups (non-TRU tumors) had higher hazard ratios, 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), proximal-proliferative (n=176), proximal-inflammatory (n=163), or non-TRU (n=339) tumors from cohorts with treatment data (chemotherapy yes/no) using univariate analysis and OS as the endpoint (Figure 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, proximal-proliferative, and proximal-inflammatory classification, we investigated classification variation due to methodological 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 methodological 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 (Figure 5A, top panel). 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 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) (Figure 5A, center panel). These results are consistent with gene signature stability analyses performed by Lauss et al. 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 prior to 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 (Figure 5A, bottom panel).

To investigate the effect of non-random 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) (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 proximal-proliferative or proximal-inflammatory switching subtype (Figure 5B). Samples switching subtype were almost exclusively reclassified as TRU-like. Notably, the subtype proportions prior to sample removal and after complete removal of all TRU classified samples were highly similar for all three subtypes: 38.7% vs 35.5% for TRU, 27.4% vs 31.2% for proximal-proliferative, and 33.9% vs 33.3% for proximal-inflammatory. Similarly, when cumulatively excluding proximal-proliferative or proximal-inflammatory tumors, samples switching subtype were consistently reclassified as the excluded subtype (Supplementary Figure S1). Next, we performed the same analysis in all 17 cohorts, and found that cumulative exclusion of TRU-classified cases led to a continually increasing fraction of cases switching GEP across all cohorts (Figure 5C). In the end, complete removal of TRU-classified cases caused 30-40% of cases originally classified as proximal-proliferative or proximal-inflammatory 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 proximal-proliferative and proximal-inflammatory subtypes (Supplementary Figure S1).

Together, these analyses demonstrate that the studied SSP is generally stable to random perturbations, while unbalanced shifts in cohort composition introduce...
greater variability in subtype assignment. While methodological 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 the current 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 clinicopathological 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 clinicopathological 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 the current 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 proximal-proliferative and proximal-inflammatory subtypes with different histopathological, clinicopathological, 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 clinicopathological 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 to 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 et al. 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 the current 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 et al. (23), whose chemotherapy response signature was not predictive in the adjuvant cisplatin-treated patient arm alone. In contrast to previous studies (23, 35, 36), we found no differences in patient outcome within the high-risk proximal-proliferative, proximal-inflammatory, 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 (23, 35, 36). Moreover, for two of the analyzed cohorts (Shedden et al. (20) and Fouret et al. (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 non-significant association of GEP classification with chemotherapy response was also found in the two included cohorts based on randomized trials, Zhu et al. (23) (JBR.10 trial) and Sato et al. (21) (UT Lung SPORE trial). In perspective, the survival benefit of traditional platinum-based adjuvant chemotherapy in patients with NSCLC is only ~4-5%, 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 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 histological composition of the investigated cohorts.

The current study also dissects the transcriptional programs defining the GEPs, showing that, e.g., 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 morphological 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 methodological 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 re-performed after removing increasing fractions of a specific GEP, causing large number of cases to switch subtype (Figure 5). Specifically, the gene-centering dependence of the SSP makes it biased towards 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, e.g., 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) towards 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 (proximal-proliferative or proximal-inflammatory) 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 the current study are equally representative of primary lung adenocarcinoma, the strikingly similar subtype fractions across cohorts (Figure 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 proximal-proliferative or proximal-inflammatory in cohorts with 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, e.g., in squamous cell carcinoma (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.

While 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, e.g., 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. While 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 be applicable to highly selected cohorts, such as high-risk patients from neoadjuvant trials.

Provided that a more robust SSP can be developed, adaption of the TRU/non-TRU classification for clinical use may require deriving mixed prognostic models that include GEP classification, standard clinicopathological 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 506-gene Wilkerson et al. (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 clinicopathological variables, based on a molecularly driven disease stratification.

Funding

Financial support for this study was provided by the Swedish Cancer Society, the Knut & 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.

References

5. Crino L, Weder W, van Meerbeeck J, Felip E. Early stage and locally advanced (non-metastatic) non-small-cell lung cancer: ESMO Clinical Practice


### Table 1. Clinical characteristics of patients in the total cohort and in the cohort of patients receiving adjuvant chemotherapy.

<table>
<thead>
<tr>
<th></th>
<th>Total cohort*</th>
<th>Adjuvant chemotherapy cohort*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of patients</td>
<td>2395</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 &amp; 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>23%</td>
<td>27%</td>
</tr>
<tr>
<td>Outcome</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Overall survival (OS), median (years)</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>Distant metastasis-free survival (DMFS), median (years)</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>Gene expression phenotypes</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Terminal respiratory unit (TRU)</td>
<td>39%</td>
<td>29%</td>
</tr>
<tr>
<td>Proximal-proliferative (PP)</td>
<td>30%</td>
<td>37%</td>
</tr>
<tr>
<td>Proximal-inflammatory (PI)</td>
<td>31%</td>
<td>34%</td>
</tr>
</tbody>
</table>

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

** Only patients with available overall survival data.
Figure legends

Figure 1. Terminal respiratory unit (TRU), proximal-proliferative, and proximal-inflammatory classification of lung adenocarcinoma. Frequency of the TRU, proximal-proliferative, and proximal-inflammatory subtypes in 2395 tumors across 17 cohorts according to the baseline classification approach.

Figure 2. Prognostic association of the terminal respiratory unit (TRU), proximal-proliferative, and proximal-inflammatory gene expression phenotypes in lung adenocarcinoma. (A) Univariate analysis of the association of TRU, proximal-proliferative or proximal-inflammatory classification with overall survival (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 (proximal-proliferative and proximal-inflammatory cases combined) across the 16 cohorts. In the panel, the TRU subtype acts as the reference group. (C) Multivariate analysis including gene expression phenotype 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 hazard ratios (HR) and 95% confidence intervals (CI) for the GEP covariate in the different models are shown. For some cohorts, not all clinicopathological data were available. Cohorts with significant GEP associations (p<0.05) are marked with *.
Figure 3. Transcriptional metagenes defining terminal respiratory unit, proximal-proliferative an proximal-inflammatory gene expression phenotypes in lung adenocarcinoma.

(A) Expression of five lung cancer transcriptional metagenes representing different biological processes across 2395 adenocarcinomas in 17 cohorts ordered by terminal respiratory unit (TRU), proximal-proliferative, and proximal-inflammatory subtype classification. For each cohort, metagene expression was transformed to a Z-score prior to pooling all cohorts. (B) Schematic representation of the adenocarcinoma gene expression phenotypes by expression of five lung cancer transcriptional metagenes based on panel A, using terms of relative expression for each metagene. (C) Distribution of Z-score values from panel A for all 2395 samples for each metagene. In all panels, the red vertical line indicates metagene expression 0. ECM = extracellular matrix.

Figure 4. Influence of transcriptional metagenes on prognostic association of gene expression phenotypes and association of subtypes with response to adjuvant chemotherapy. (A) Multivariate analysis of the association with patient outcome based on overall survival (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, hazard ratios (HR) with 95% confidence intervals (CI) are displayed for the non-TRU subgroup only. Cohorts with significant associations (p<0.05) are marked with *. (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, hazard ratios are displayed for the non-TRU subgroup only. Cohorts with significant associations (p<0.05) are marked with *. (C) Univariate analyses of the association of gene expression phenotypes with OS in patients treated with adjuvant chemotherapy. § = all univariate p>0.05 in the three-group analysis, where the TRU group is the reference group; proximal-proliferative (PP) p=0.19, proximal-inflammatory (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). ECM= extra cellular matrix.

**Figure 5. Robustness of terminal respiratory unit (TRU), proximal-proliferative, and proximal-inflammatory classification.** (A) Classification variability due to methodological alterations or random perturbations of genes or samples. The top panel shows the percentage of cases switching gene expression phenotype (GEP) when using Spearman correlation instead of Pearson correlation as the similarity metric, or mean centering instead of median gene-centering across all cohorts. For the Shedden et al. cohort the latter analysis was not performed due to reported site-specific expression differences. The center panel shows the percentages of cases switching GEP after random selection of different subsets of genes from the 506-gene centroid followed by classification, mimicking the effect of missing data values. The bottom panel shows the percentages of cases switching GEP per cohort after random selection of different subsets of cases followed by gene recentring and classification according to the standard approach. In the center and bottom panel the presented data value for each cohort represents the mean of 20 permutations. (B) Detailed analysis of the variability in GEP classification caused by proportional changes of a specific GEP
in the 230-sample TCGA cohort (14). In a step-by-step process, one TRU-classified tumor (n=89 originally) was cumulatively removed from the cohort followed by gene recentering and reclassification of the remaining cases. The panel shows how the GEP classification for an individual sample (rows) changes with increasing fractions of originally TRU-classified cases excluded (columns). (C) The same analysis as in panel B is shown, but across all 17 cohorts. Each line corresponds to a unique cohort. Variance is binned in 5% bins (x-axis). A box plot for each bin is superimposed in red.
Figure 1

- **TRU**
- **Proximal-proliferative**
- **Proximal-inflammatory**

Frequency of subtypes (%)
A) Coorts

All : PP *
Roepman : PP
Wilkerson : PP
Okayama : PP *
Der : PP *
Hou : PP
Shedden : PP *
Bild : PP *
CLCGP : PP *
Sato : PP *
Chitale : PP *
Fouret : PP
Tomida : PP *
Beer : PP *
TCGA : PP *
Botling : PP
Zhu : PP

B) Coorts

All *
Roepman
Wilkerson
Okayama *
Der *
Hou
Shedden *
Bild *
CLCGP *
Sato *
Chitale *
Fouret
Tomida *
Beer
TCGA *
Botling
Zhu

HR, 95% CI

C) Coorts

All : GEP+Stage (A) *
GEP+Stage+Age (B) *
GEP+Stage+Age+Sex+Smoking (C) *
Roepman : A
B
C
Wilkerson : A
B
C
Okayama : A *
B *
C *
Der: A *
B *
C *
Shedden : A *
B *
C *
CLCGP : A *
B *
C
Sato : A *
B *
C *
Chitale : A *
B *
C *
Fouret : A *
B *
C
Tomida : A
B
C
TCGA : A
B
C
Botling : A
B
C
Zhu : A
B
C

HR, 95% CI
Cohorts

A) HR, 95% CI

B) Cohorts

C) TRU / PP / PI, p > 0.05

D) TRU, n = 223

Author Manuscript Published OnlineFirst on August 11, 2015; DOI: 10.1158/1078-0432.CCR-15-0529
Prognostic and chemotherapy predictive value of gene expression phenotypes in primary lung adenocarcinoma

Markus Ringnér, Göran Jönsson and Johan Staaf

Clin Cancer Res  Published OnlineFirst August 11, 2015.

Updated version
Access the most recent version of this article at:

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