

## Identification of Transcriptional Subgroups in *EGFR*-Mutated and *EGFR/KRAS* Wild-Type Lung Adenocarcinoma Reveals Gene Signatures Associated with Patient Outcome

Maria Planck<sup>1</sup>, Sofi Isaksson<sup>1</sup>, Srinivas Veerla<sup>1</sup>, and Johan Staaf<sup>1,2</sup>

### Abstract

**Purpose:** In lung adenocarcinoma, *EGFR* and *KRAS* mutations dominate the mutational spectrum and have clear therapeutic implications. We sought to determine whether transcriptional subgroups of clinical relevance exist within *EGFR*-mutated, *KRAS*-mutated, or *EGFR* and *KRAS* wild-type (*EGFRwt/KRASwt*) adenocarcinomas.

**Experimental Design:** Gene expression profiles from 1,186 adenocarcinomas, including 215 *EGFR*-mutated, 84 *KRAS*-mutated, and 219 *EGFRwt/KRASwt* tumors, were assembled and divided into four discovery ( $n = 522$ ) and four validation cohorts ( $n = 664$ ). Subgroups within the mutation groups were identified by unsupervised consensus clustering, significance analysis of microarrays (SAM) analysis, and centroid classification across discovery cohorts. Genomic alterations in identified mutation subgroups were assessed by integration of genomic profiles for 158 cases with concurrent data. Multicohort expression subgroup predictors were built for each mutation group using the discovery cohorts, and validated in the four validation cohorts.

**Results:** Consensus clustering within the mutation groups identified reproducible transcriptional subgroups in *EGFR*-mutated and *EGFRwt/KRASwt* tumors, but not in *KRAS*-mutated tumors. Subgroups displayed differences in genomic alterations, clinicopathologic characteristics, and overall survival. Multicohort gene signatures derived from the mutation subgroups added independent prognostic information in their respective mutation group, for adenocarcinoma in general and stage I tumors specifically, irrespective of mutation status, when applied to the validation cohorts. Consistent with their worse clinical outcome, high-risk subgroups showed higher expression of proliferation-related genes, higher frequency of copy number alterations/amplifications, and association with a poorly differentiated tumor phenotype.

**Conclusions:** We identified transcriptional subgroups in *EGFR*-mutated and *EGFRwt/KRASwt* adenocarcinomas with significant differences in clinicopathologic characteristics and patient outcome, not limited to a mutation-specific setting. *Clin Cancer Res*; 19(18); 5116–26. ©2013 AACR.

### Introduction

Lung cancer is the leading cause of cancer-related mortality worldwide (1). The disease is heterogeneous but may be broadly divided into small cell lung cancer and non-small cell lung carcinoma (NSCLC). NSCLC accounts for approximately 85% of all diagnosed cases, with adenocarcinoma as the most frequent histologic type (2). The *EGF* receptor gene, *EGFR*, located at 7p11.2 and the *V-Ki-ras2*

Kirsten rat sarcoma viral oncogene homolog gene, *KRAS*, located at 12p12.1 represent the two most frequently mutated oncogenes in adenocarcinoma (3). *EGFR* and *KRAS* mutations are essentially mutually exclusive in these tumors and are associated with differences in, for example, patient gender and smoking history, suggesting that these molecular defects may be drivers of pathogenesis for specific subgroups (ref. 3; and references therein). In line with this, *EGFR* mutations have been associated with improved overall survival, whereas *KRAS* mutations may predict shorter survival for patients with advanced lung adenocarcinoma (4). Furthermore, the occurrence of *EGFR* mutations predicts an improved response to *EGFR* tyrosine kinase inhibitors and is therefore routinely assessed in clinical practice (5).

*EGFR* and *KRAS* wild-type (*EGFRwt/KRASwt*) adenocarcinomas represent a still unclear group, with different potential driver mutations as well as mutually exclusive genomic rearrangements of the *ALK*, *RET*, and *ROS1* genes (3, 6, 7). Interestingly, a subgroup of *EGFRwt/KRASwt*

**Authors' Affiliations:** <sup>1</sup>Department of Oncology, Clinical Sciences, Skåne University Hospital; and <sup>2</sup>CREATE Health Strategic Center for Translational Cancer Research, Lund University, Lund, Sweden

**Note:** Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

**Corresponding Author:** Johan Staaf, Department of Oncology, Lund University, Medicion Village, SE22381 Lund, Sweden. Phone: 46-462221444; Fax: 46-46147327; E-mail: johan.staaf@med.lu.se

doi: 10.1158/1078-0432.CCR-13-0928

©2013 American Association for Cancer Research.

### Translational Relevance

*EGFR* and *KRAS* mutations dominate the mutational spectrum of lung adenocarcinoma. *EGFR* mutation is an established predictive marker for response to targeted therapy, although primary or acquired resistance impairs the result of treatment with EGF receptor (EGFR) inhibitors. Further characterization of mutation groups may identify new molecular subgroups, additional targets for synergistic treatment, and provide new insights into resistance mechanisms and molecular pathogenesis. On the basis of a multicohort discovery and validation strategy, we identified transcriptional subgroups in *EGFR*-mutated and *EGFR/KRAS* wild-type adenocarcinomas with significant differences in genomic alterations, clinicopathologic characteristics, and prognosis. Moreover, these subgroup gene signatures also added independent prognostic information in adenocarcinomas in general and in stage I disease specifically, irrespective of mutation status. Further investigations on the predictive value of these gene signatures in the setting of targeted treatment may provide a future basis for refined diagnosis and treatment of lung adenocarcinoma.

adenocarcinomas seems to benefit from EGFR inhibitors, although there are currently no predictive markers to identify these patients (8, 9). Consequently, increased knowledge about the molecular background of tumor subgroups defined by *EGFR/KRAS* mutation status may identify novel genomic signatures and additional targets for synergistic treatment, and also provide new insights into resistance mechanisms and molecular pathogenesis. Numerous single biomarkers (e.g., reviewed in refs. 10, 11) and microarray-based gene signatures (12–16) associated with clinical outcome in NSCLC/adenocarcinoma have been reported to date. In addition, different molecular subgroups in lung adenocarcinoma have also been reported (17, 18), of which the bronchioid, squamoid, and magnoid subtypes originally defined by Hayes and colleagues (17) have been reproduced in multiple cohorts (17, 19). Bronchioid tumors are generally of lower grade, have a higher proportion of *EGFR* mutations and higher expression of excretion, asthma, and surfactant genes, occur predominantly in women and never-smokers, and have better overall survival (17, 19). In contrast, magnoid and squamoid tumors harbor more *KRAS* mutations, seem to be more closely related in gene expression, occur more often in men and smokers, and have poorer overall survival (17, 19). Despite these findings, the clinical usefulness of many of these markers/signatures remains debated (11, 20), and the prognostic performance in adenocarcinoma subgroups defined by *EGFR* and *KRAS* mutational status remains relatively unclear. Several studies have also illustrated the difficulties in separating *EGFR*-mutated, *KRAS*-mutated, and *EGFRwt/KRASwt* tumors into distinct transcriptional entities (16, 21–23). Moreover, it has not been systematically

investigated whether clinically relevant transcriptional subgroups exist within the mutation groups.

Herein, we sought to determine whether transcriptional subgroups of clinical relevance exist within *EGFR*-mutated, *KRAS*-mutated, and/or *EGFRwt/KRASwt* adenocarcinomas. Through a multicohort discovery and validation strategy, we identified reproducible gene expression subgroups within *EGFR*-mutated tumors and *EGFRwt/KRASwt* tumors associated with different genomic alterations, clinicopathologic characteristics, and patient outcomes. In addition, these subgroup gene signatures also added independent prognostic information in adenocarcinomas in general and in stage I disease specifically, irrespective of mutation status.

### Materials and Methods

#### Tumor cohorts

Gene expression profiles from 522 lung adenocarcinomas, including 215 *EGFR*-mutated, 84 *KRAS*-mutated, and 219 *EGFRwt/KRASwt* cases, were obtained from GSE31210 (16), E-MTAB-923 (21), and Chitale and colleagues (22) and were used as discovery cohorts. Samples from Chitale and colleagues (22) were divided into two cohorts based on the different Affymetrix platforms used in this study (U133A and U133 2plus) creating four final discovery cohorts (Table 1). Prognostic associations of the derived gene signatures were validated in 117 tumors with known mutation status from GSE13213 (ref. 15;  $n = 45$  *EGFR*-mutated, 15 *KRAS*-mutated, and 57 *EGFRwt/KRASwt*), and 547 additional adenocarcinomas with unknown mutation status from Shedden and colleagues (ref. 14;  $n = 356$ ), GSE3141 (ref. 24;  $n = 58$ ), and the University of Texas Lung Specialized Program of Research Excellence cohort (ref. 13; UT Lung SPORE, GSE42127,  $n = 133$  adenocarcinomas). Matched and analyzed genomic profiles from 158 tumors (53 *EGFR*-mutated and 105 *EGFRwt/KRASwt*) belonging to the E-MTAB-923 and Chitale and colleagues cohorts were extracted from Staaf and colleagues (25).

Response to adjuvant chemotherapy (ACT) for patients with NSCLC was investigated in the complete UT Lung SPORE cohort (GSE42127;  $n = 176$ ) including patients treated mainly with carboplatin plus taxanes, and in the National Cancer Institute of Canada Clinical Trials Group JBR.10 clinical trial cohort ( $n = 90$ ; GSE14814; ref. 12) including patients treated mainly with vinorelbine plus cisplatin (Supplementary Table S1). For both of these cohorts, gene expression profiling was conducted before therapy. Response to sorafenib, a drug that targets different tyrosine and Raf kinases, treatment of patients with NSCLC with advanced chemorefractory metastatic disease was explored in the GSE33072 cohort (ref. 26; Supplementary Table S1). The biopsy samples in this study were taken from the lung, liver, lymph node, bone/soft tissue, and adrenal glands of patients enrolled in the Biomarker-integrated Approaches of Targeted Therapy for Lung Cancer Elimination (BATTLE) trial (27).

Explicit information on patient ethnicity or specific mutation type was not available for the majority of included

**Table 1.** Clinical characteristics of adenocarcinoma patients in the prognostic gene expression cohorts

|                         | Discovery cohorts        |                       |                            |                          | Validation cohorts |                    |                          |                     |
|-------------------------|--------------------------|-----------------------|----------------------------|--------------------------|--------------------|--------------------|--------------------------|---------------------|
|                         | GSE31210<br>(16)         | Chitale<br>U133A (22) | Chitale U133<br>2plus (22) | E-MTAB-923<br>(21)       | GSE13213<br>(15)   | GSE42127<br>(13)   | GSE3141<br>(24)          | Shedden<br>(14)     |
| Total no. of patients   | 226                      | 91                    | 102                        | 103                      | 117                | 133                | 58                       | 356                 |
| Gender                  |                          |                       |                            |                          |                    |                    |                          |                     |
| Male                    | 105                      | 41                    | 42                         | 16                       | 60                 | 68                 | —                        | 189                 |
| Female                  | 121                      | 50                    | 60                         | 87                       | 57                 | 65                 | —                        | 166                 |
| Smoking status          |                          |                       |                            |                          |                    |                    |                          |                     |
| Never-smokers           | 115                      | 17                    | 19                         | 63                       | 56                 | —                  | —                        | 33                  |
| Smokers                 | 111                      | 73                    | 83                         | 40                       | 61                 | —                  | —                        | 229                 |
| Mutation status         |                          |                       |                            |                          |                    |                    |                          |                     |
| EGFR-mutated            | 127                      | 15                    | 24                         | 49                       | 45                 | —                  | —                        | —                   |
| KRAS-mutated            | 20                       | 11                    | 36                         | 17                       | 15                 | —                  | —                        | —                   |
| EGFRwt/KRASwt           | 79                       | 65                    | 42                         | 33                       | 57                 | —                  | —                        | —                   |
| Stage                   |                          |                       |                            |                          |                    |                    |                          |                     |
| I                       | 168                      | 53                    | 70                         | 60                       | 79                 | 89                 | —                        | 224                 |
| II                      | 58                       | 20                    | 10                         | 10                       | 13                 | 22                 | —                        | 77                  |
| III                     | 0                        | 18                    | 17                         | 33                       | 25                 | 20                 | —                        | 51                  |
| IV                      | 0                        | 0                     | 5                          | 0                        | 0                  | 1                  | —                        | 0                   |
| ACT <sup>a</sup>        | 226/0                    | Unknown               | Unknown                    | 52/33                    | 117/0              | 94/39              | Unknown                  | 172/62              |
| Median follow-up, years | 5                        | 3.5                   | 1.5                        | 3.6                      | 5.6                | 3.8                | 2.5                      | 4                   |
| Platform                | Affymetrix<br>U133 2plus | Affymetrix<br>U133A   | Affymetrix<br>U133 2plus   | Affymetrix<br>U133 2plus | Agilent<br>44K     | Illumina<br>WG6 V3 | Affymetrix<br>U133 2plus | Affymetrix<br>U133A |

<sup>a</sup>ACT, number of untreated/reated patients.

patients, and was therefore omitted from analyses. Included studies were carried out in both western and Asian countries. Patient and tumor characteristics are summarized in Table 1 and Supplementary Table S1.

### Gene expression analyses

Affymetrix cohorts were normalized using GC robust multi-array averaging (GCRMA; ref. 28), except for GSE3141 (24) and GSE33072 (26) for which normalized expression data were obtained from Gene Expression Omnibus (29). Normalized expression data were obtained from Gene Expression Omnibus for non-Affymetrix cohorts. Unsupervised subgroup discovery within mutation groups was carried out using consensus clustering through the ConsensusClusterPlus R package (30) on the four discovery cohorts individually. Significance analysis of microarrays (SAM; ref. 31) from the siggenes R package (32) was used to identify differentially expressed probe sets between consensus clusters for the discovery cohorts. Probe sets with false discovery rate less than 5% were considered statistically significant. Nearest-centroid predictors and multicohort centroids were created for each cohort or mutation group as described (Supplementary Data). Proliferation differences between samples were assessed by an expression metagene based on the proliferation/chromosome instability (CIN70) signature (ref. 33; referred to as the CIN70 metagene hereon), which includes numerous proliferation/cell cycle-related genes (Supplementary

Data). Data processing steps are further described in Supplementary Data.

### Survival analysis

Survival analyses were conducted in R using the Survival package with overall survival as endpoint. Survival curves were compared using Kaplan–Meier estimates and the log-rank test. Follow-up time for overall survival was censored at 5 years for all cohorts.

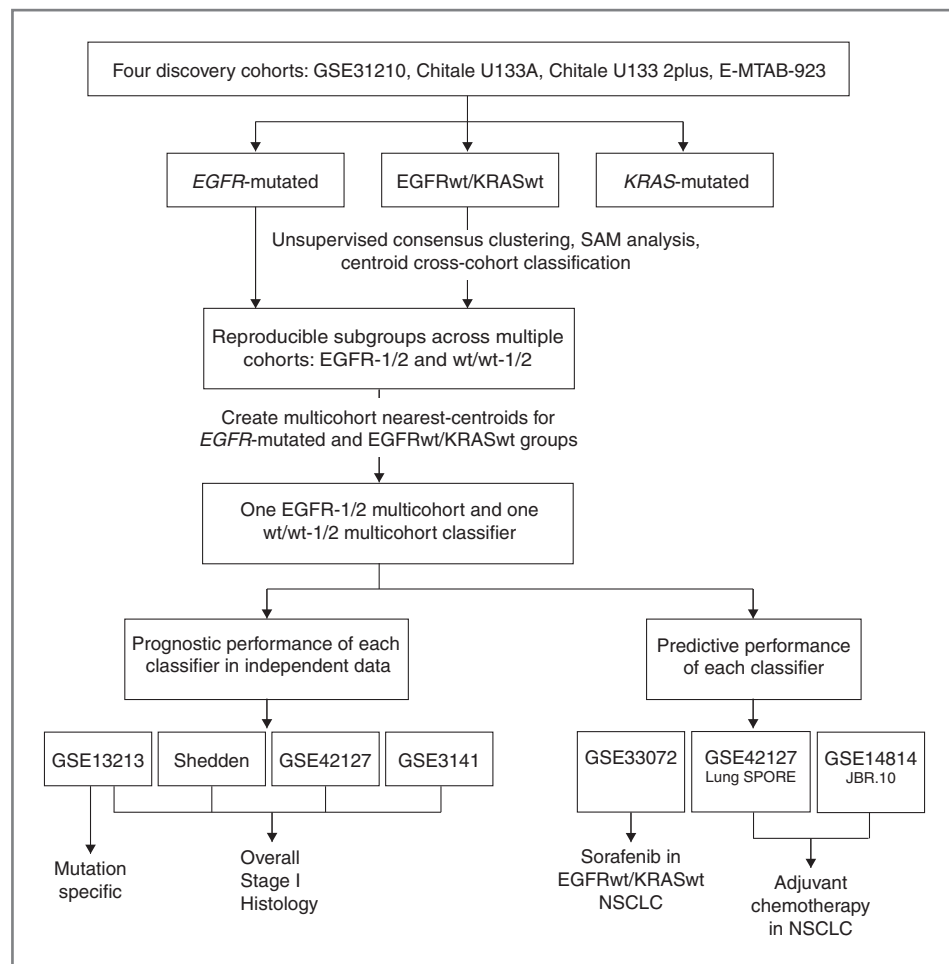
### Results

#### Unsupervised discovery of reproducible transcriptional subgroups in EGFR-mutated and EGFRwt/KRASwt tumors

Figure 1 shows a schematic of the steps taken to identify and validate transcriptional subgroups within the three mutation groups (see also Supplementary Data). For each cohort and mutation group, consensus clustering was used to identify two sample clusters (Fig. 2) from which a centroid classifier was constructed on the basis of differentially expressed genes from a SAM analysis. Next, each mutation and cohort-specific classifier was used to classify tumors of similar mutation status in the remaining three discovery cohorts, and the overlap between the predicted groups and the original consensus clusters was compared for each classifier and for each cohort.

This cross-cohort approach identified two reproducible subgroups in EGFR-mutated tumors, termed EGFR-1 and

Figure 1. Schematic of the multicohort discovery and validation design.

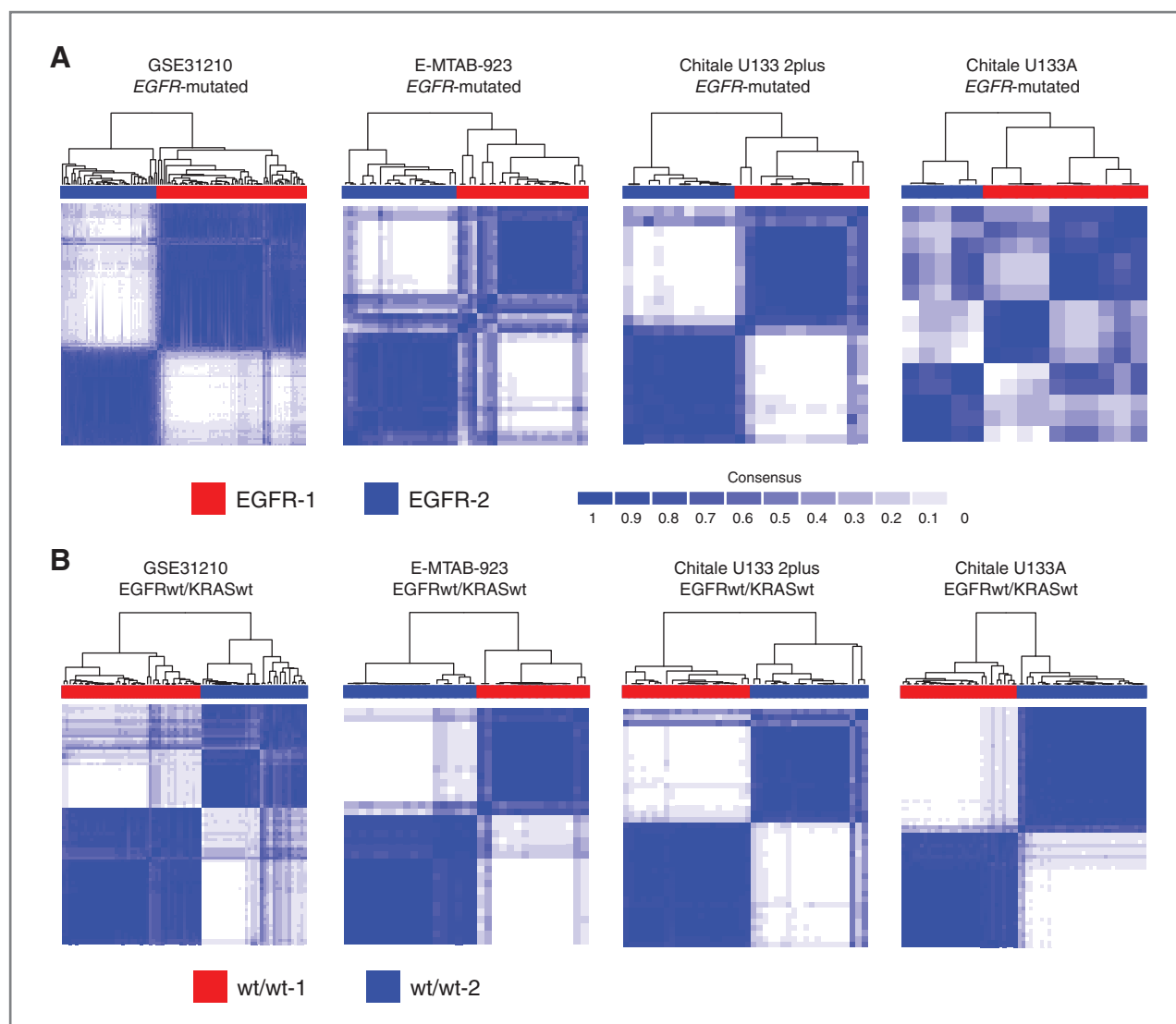


EGFR-2 herein, based on consistency between predicted and original consensus clusters in three of four discovery cohorts (Figs. 2A and Supplementary Fig. S1A). Similarly, two subgroups, termed wt/wt-1 and wt/wt-2 hereon, were identified in EGFRwt/KRASwt tumors based on all four discovery cohorts (Figs. 2B and Supplementary Fig. S1B). In contrast, no robust subgroups (across at least three discovery cohorts) were identified in KRAS-mutated tumors (data not shown). For EGFRwt/KRASwt tumors, the wt/wt-1 and wt/wt-2 subgroups in GSE31210 corresponded strongly to the clusters identified by Okayama and colleagues (16); 95% of the tumors were similarly grouped (Fisher's exact test  $P = 2 \times 10^{-17}$ ). Notably, for both EGFR-1/2 and wt/wt-1/2, large transcriptional differences between groups were observed in the SAM analysis, comprising thousands of probe sets (Supplementary Fig. S1A and S1B).

#### EGFR-mutated and EGFRwt/KRASwt transcriptional subgroups are associated with clinicopathologic differences, molecular subtypes, and patient outcome

Across the discovery cohorts, EGFR-mutated and EGFRwt/KRASwt transcriptional subgroups were associated with differences in (i) adenocarcinoma molecular subtype (19) patterns, (ii) expression of a proliferation

metagene (CIN70; ref. 33), (iii) clinicopathologic characteristics, and (iv) overall survival. In EGFR-mutated tumors, the EGFR-1 subgroup included 75% to 85% of all bronchioid but few magnoid or squamoid-classified tumors (19), displayed lower expression of the CIN70 metagene, included older patients, and displayed better overall survival compared with EGFR-2 (Fig. 3A and Supplementary Fig. S1C). No associations with tumor stage ( $P < 0.05$  in GSE31210; Fisher's exact test), patient gender ( $P > 0.05$  all cohorts; Fisher's exact test), or smoking status ( $P < 0.05$  in GSE31210; Fisher's exact test) were observed in more than one discovery cohort for EGFR-1 or -2, potentially due to the small group sizes for certain cohorts. In EGFRwt/KRASwt tumors, the wt/wt-2 subgroup included 88% to 100% of the bronchioid tumors, but few squamoid/magnoid tumors, displayed lower expression of the CIN70 metagene, included more never-smokers, and displayed better overall survival (with the exception of E-MTAB-923) compared with wt/wt-1 (Figs. 3B and Supplementary Fig. S1D). No associations with patient gender, age, or tumor stage were observed for wt/wt-1 or wt/wt-2 in the discovery cohorts, potentially due to the small group sizes for certain cohorts ( $P > 0.05$  all cohorts; Fisher's exact test or Wilcoxon test).



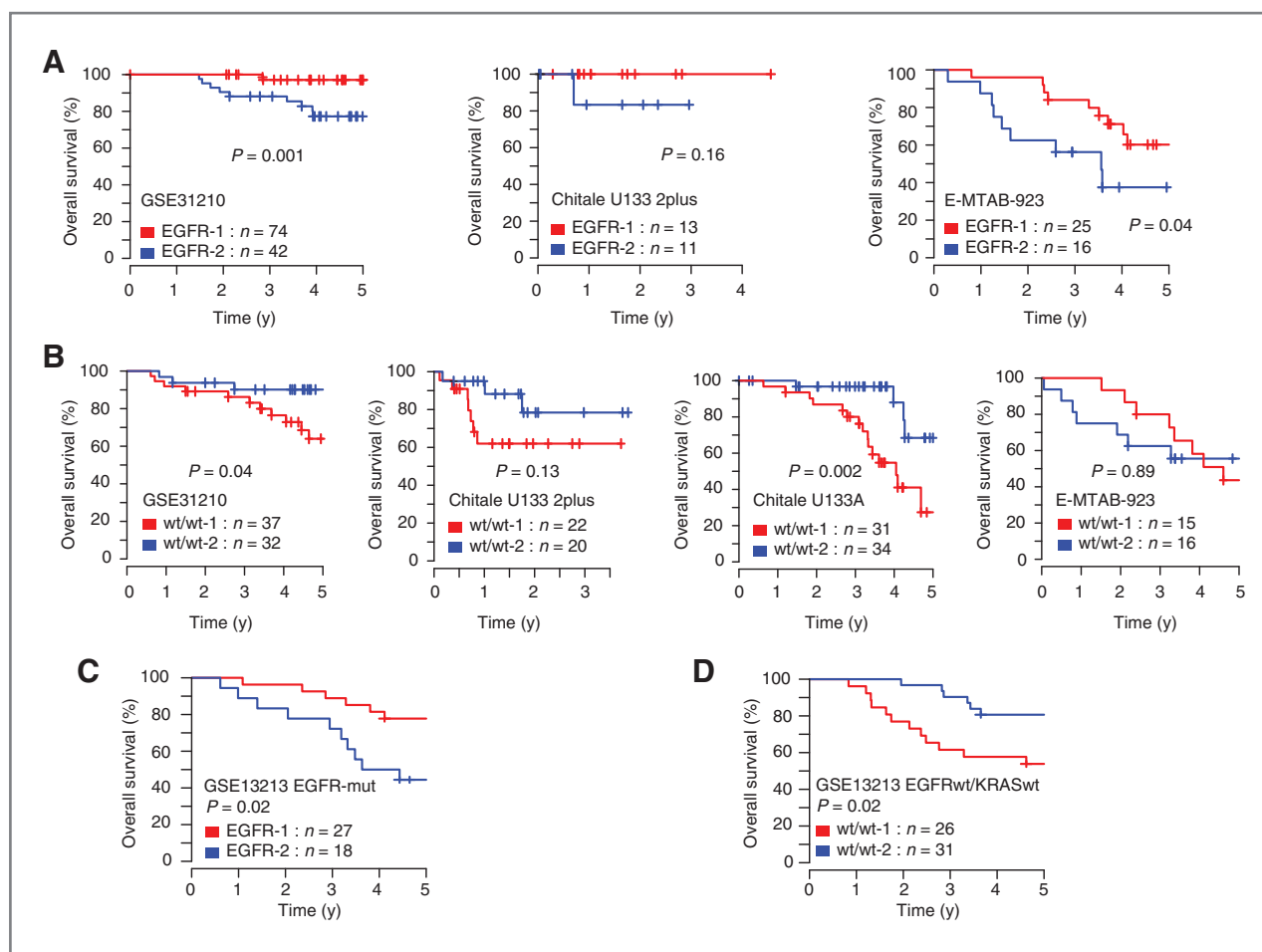
**Figure 2.** Consensus clustering of discovery cohorts for *EGFR*-mutated and *EGFR*wt/*KRAS*wt tumors. Consensus clustering was conducted on normalized, mean-centered, and variance-filtered gene expression data using  $k = 2$  groups as described (Supplementary Data). Heatmaps display consensus values between pairs of tumors by blue shading. High consensus corresponds to samples that always occur in the same cluster and is shaded dark blue. *KRAS*-mutated tumors were omitted as no reproducible subgroups from clustering were found in later cross-cohort analyses. A, clustering of the four discovery cohorts for *EGFR*-mutated adenocarcinomas. Note that the Chitale U133A cohort does not show distinct clusters, and later cross-cohort analyses could not reproduce subgroups from the other cohorts in this cohort. B, clustering of the four discovery cohorts for *EGFR*wt/*KRAS*wt adenocarcinomas.

### ***EGFR*-mutated and *EGFR*wt/*KRAS*wt transcriptional subgroups display different genomic alterations**

In 158 tumors with matched genomic and transcriptional profiles, both the high-risk subgroups, *EGFR*-2 and wt/wt-1, were associated with overall more copy number alterations in their respective mutation group (measured as the fraction of the genome altered by copy number alterations; see Supplementary Data; and ref. 25) compared with their corresponding low-risk subgroups ( $P = 0.04$  and  $0.0007$ , respectively; Wilcoxon test). *EGFR*-mutated *EGFR*-2 tumors displayed more copy number gain in regions on 7p (including *EGFR*; ~85% of cases) and 3q, and more frequent losses on 4q, 9p, 15q, and 16q compared with *EGFR*-mutated

*EGFR*-1 tumors (Supplementary Fig. S2A; >25% frequency difference). Moreover, 80% of the *EGFR* amplifications, 89% of all 7p amplifications, and all *NKX2-1/TITF* amplifications were found in the *EGFR*-2 subgroup compared with *EGFR*-1 in the analyzed *EGFR*-mutated tumors.

In *EGFR*wt/*KRAS*wt tumors, the high-risk wt/wt-1 subgroup displayed more copy number gains in regions on 1q, 3q, 5p, 7q, and 12p and more frequent losses on 4q, 5q, 15q, and 22q compared with the wt/wt-2 subgroup (Supplementary Fig. S2B; >25% frequency difference). Overall, wt/wt-1 tumors also displayed more amplifications compared with wt/wt-2 ( $P = 0.0003$ ; Fisher's exact test). Two recurrently amplified regions were found to



**Figure 3.** Association with overall survival for *EGFR*-mutated and *EGFR*wt/*KRAS*wt transcriptional subgroups. **A**, *EGFR*-mutated subgroups (*EGFR*-1 and *EGFR*-2) display difference in overall survival. **B**, *EGFR*wt/*KRAS*wt subgroups (wt/wt-1 and wt/wt-2) display difference in overall survival, with the exception of E-MTAB-923. **C**, multicohort *EGFR*-1/2 centroid classification is associated with overall survival in *EGFR*-mutated tumors in GSE13213. **D**, multicohort wt/wt-1/2 centroid classification is associated with overall survival in *EGFR*wt/*KRAS*wt tumors in GSE13213.

differ significantly between the wt/wt-1 and wt/wt-2 subgroups: 8p12 (*FGFR1*) amplifications in wt/wt-1, and 12q15 (*MDM2*) amplifications in wt/wt-2 ( $P = 0.02$  and  $0.03$ , respectively; Fisher's exact test).

### Independent validation of *EGFR*-mutated and *EGFR*wt/*KRAS*wt transcriptional subgroups

To validate the *EGFR*-mutated and *EGFR*wt/*KRAS*wt transcriptional subgroups in independent cohorts, we first created a single *EGFR*-mutated (*EGFR*-1/2) and a single *EGFR*wt/*KRAS*wt (wt/wt-1/2) multicohort centroid classifier from the individual discovery cohort centroids for respective mutation group (Fig. 1; Supplementary Data and Supplementary Table S2). Notably, 18% to 20% of the genes in the *EGFR*-1/2 and wt/wt-1/2 multicohort classifiers matched reported lists of potential therapeutic targets and modulators of chemotherapy drugs' effects in lung cancer cells (refs. 34, 35; Supplementary Table S2), whereas 24% to 26% of the genes in the two signatures overlapped with the bronchioid, magnoïd, and

squamous subtype centroids reported by Wilkerson and colleagues (19).

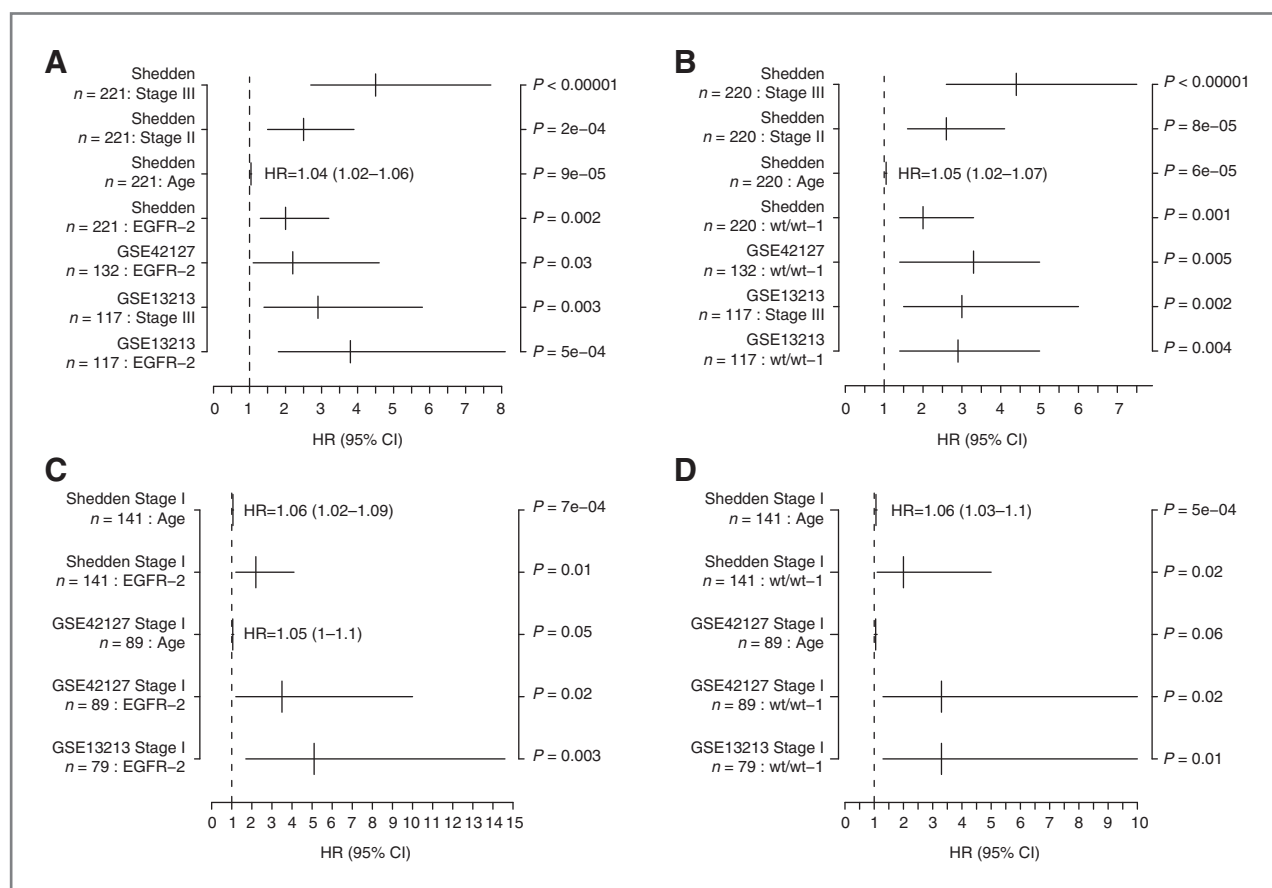
The two multicohort centroid classifiers were next applied to their respective mutation group in the GSE13213 (15) cohort. Consistent patterns of molecular subtype distribution, CIN70 metagene expression, patient age, smoking status, and overall survival, were observed for the predicted subgroups in GSE13213 compared with the discovery cohorts, although not always reaching statistical significance due to small sample sizes (Fig. 3C and D and Supplementary Fig. S1E and S1F). Consistent with the largest *EGFR*-mutated discovery cohort (GSE31210), the low-risk *EGFR*-1 subgroup in GSE13213 contained nearly twice as many never-smokers as the *EGFR*-2 subgroup (Supplementary Fig. S1E and S1F). Classification of all 117 samples in GSE13213 showed that both multicohort signatures were associated with overall survival (log-rank test;  $P = 2 \times 10^{-5}$  for *EGFR*-1/2 and  $P = 2 \times 10^{-4}$  for wt/wt-1/2), and that there was a high consistency between *EGFR*-1/2 and wt/wt-1/2 classifications ( $P = 1 \times 10^{-21}$ ; Fisher's

exact test). Both classifications also added independent prognostic information in multivariate analysis including tumor stage, patient age, mutation status, and smoking status as covariates, and overall survival as the endpoint (Fig. 4A and B).

To further validate the general prognostic association of the two multicohort signatures, we classified 547 additional independent adenocarcinomas with unknown mutation status from Shedden and colleagues (14), GSE3141 (24), and GSE42127 (ref. 13; Table 1). Consistently, centroid classifications were associated with overall survival in Shedden and colleagues (log-rank test;  $P = 4 \times 10^{-5}$  for EGFR-1/2 and  $P = 9 \times 10^{-5}$  for wt/wt-1/2), GSE3141 ( $P = 0.01$  and 0.006, respectively), and GSE42127 ( $P = 0.008$  and 0.0006, respectively). In Shedden and colleagues, the high-risk EGFR-2 and wt/wt-1 groups were both strongly enriched with poorly differentiated tumors, whereas the corresponding low-risk EGFR-1 and wt/wt-2 groups included more than 90% of the well-differentiated tumors in this cohort ( $P = 6 \times 10^{-20}$  and  $4 \times 10^{-19}$ , respectively; Fisher's exact test). In multivariate analysis, both multicohort classifica-

tions added independent prognostic information in GSE42127 with tumor stage, patient age, and ACT as covariates, and in Shedden and colleagues when including tumor stage, patient age, ACT, and smoking status as covariates (Fig. 4A and B). Furthermore, both multicohort classifications also added independent prognostic information for patients with stage I disease in Shedden and colleagues, GSE42127, and GSE13213 (Fig. 4C and D).

Patients with bronchioid-classified tumors have repeatedly been shown to have superior outcomes to patients with the magnoid or squamoid subtypes (17, 19). Given the enrichment of bronchioid-classified tumors in the low-risk groups, we investigated whether the multicohort classifiers added independent prognostic information when including the molecular subtypes in the previous multivariate models for Shedden and colleagues, GSE13213 and GSE42127. In GSE42127, neither the EGFR-1/2 ( $P = 0.37$ ), the wt/wt-1/2 ( $P = 0.07$ ), nor the molecular subtype classification were significant in the multivariate analysis. In the Shedden and colleagues cohort, both multicohort classifiers added independent prognostic



**Figure 4.** Multivariate Cox regression analyses of the EGFR-1/2 and wt/wt-1/2 signatures in independent adenocarcinoma cohorts. Multivariate analyses of EGFR-1/2 and wt/wt-1/2 were conducted using overall survival as endpoint. For GSE13213 tumor stage, patient age, mutation status, and smoking status were included as covariates in the multivariate analyses. For GSE42127 tumor stage, patient age, and ACT were included as covariates. For Shedden and colleagues tumor stage, patient age, ACT, and smoking status were included as covariates. Only hazard ratios (HR) and  $P$  values for significant or borderline nonsignificant covariates are displayed. A, EGFR-1/2 analysis for all cases for respective cohort. B, wt/wt-1/2 analysis for all cases for respective cohort. C, EGFR-1/2 analysis in stage I tumors for respective cohort. D, wt/wt-1/2 analysis in stage I tumors for respective cohort.

information in multivariate analysis [ $n = 221$ , EGFR-2:  $P = 0.01$ ; HR, 2.2; 95% confidence interval (CI), 1.2–4, and  $n = 220$ , wt/wt-1:  $P = 0.03$ ; HR, 1.9; 95% CI, 1.1–3.3]. Similarly, in GSE13213, both multicohort classifiers also added independent prognostic information after inclusion of molecular subtype in the multivariate model ( $n = 117$ , EGFR-2:  $P = 0.0002$ ; HR, 10; 95% CI, 3–34, and  $n = 117$ , wt/wt-1:  $P = 0.045$ ; HR, 4.3; 95% CI, 1.03–20). Similar multicohort centroid classification of squamous cell carcinoma tumors in GSE3141 ( $n = 53$ ) or GSE42127 ( $n = 43$ ) was not associated with overall survival (log-rank test;  $P > 0.05$ ; data not shown), suggesting that the prognostic association of the multicohort signatures is adenocarcinoma specific.

#### EGFR-mutated and EGFRwt/KRASwt gene signatures are associated with response to treatment in NSCLC

To assess whether the EGFR-1/2 and wt/wt-1/2 signatures were associated with response to treatment, we first analyzed gene expression profiles from tumor biopsies taken from patients with NSCLC with advanced chemorefractory metastatic disease enrolled in the BATTLE trial (refs. 26, 27; GSE33072, Supplementary Table S1). We restricted the analysis to the 30 patients with EGFRwt/KRASwt NSCLC treated with sorafenib due to otherwise small sample groups and poor sample annotations. For the sorafenib-treated cohort, the high-risk EGFR-2 and wt/wt-1 groups included the majority of patients who did not meet the primary endpoint of 8-week disease control (85% and 83%, respectively). Moreover, the wt/wt-1 classification was borderline nonsignificant for worse progression-free survival compared with wt/wt-2 (log-rank test;  $P = 0.07$ ).

Second, we investigated whether the multicohort classifiers predicted survival benefits from ACT in NSCLC based on the GSE14814 (JBR.10 clinical trial cohort;  $n = 90$ ; ref. 12) and GSE42127 (UT Lung SPORE;  $n = 176$ ; ref. 13) cohorts (Fig. 1 and Supplementary Table S1). In GSE14814, ACT-treated NSCLC patients showed significantly better overall survival than those without treatment in the high-risk EGFR-2 ( $n = 58$ ; HR, 0.39; 95% CI, 0.16–0.97;  $P = 0.04$ ) and wt/wt-1 ( $n = 56$ ; HR, 0.39; 95% CI, 0.15–0.99;  $P = 0.05$ ) groups, whereas in the low-risk groups ACT had no significant survival benefits. Similar results were found for the GSE42127 cohort, with patients with NSCLC in the high-risk groups benefiting from ACT (EGFR-2,  $n = 98$ : HR, 0.42; 95% CI, 0.17–1;  $P = 0.05$  and trend-like for wt/wt-1,  $n = 83$ : HR, 0.46; 95% CI, 0.18–1.2;  $P = 0.11$ ), whereas patients in the low-risk groups did not benefit from ACT.

#### Discussion

In this study, we investigated whether reproducible transcriptional subgroups could be identified within EGFR-mutated, KRAS-mutated, and/or EGFRwt/KRASwt lung adenocarcinomas based on a multicohort discovery and validation approach. Unsupervised subgroup discovery within individual mutation groups identified transcriptional subgroups in EGFR-mutated and EGFRwt/KRASwt tumors associated with differences in genomic alterations, clinico-

pathologic characteristics, and patient outcome. The failure to identify reproducible subgroups within KRAS-mutated tumors could be due to the lower number of available tumors and/or a potential biologic or etiologic heterogeneity among these tumors (36). For EGFRwt/KRASwt tumors, our results extend recent findings by Okayama and colleagues (16), by showing that apparently similar subgroups exist in other cohorts using a different analysis approach. Subgroup signatures included a spectrum of genes involved in lung carcinogenesis (*NKX2-1*, *HOPX*, *LOXL2*, several matrix metalloproteinases), DNA-repair (*BRCA1*, *RAD51*, *MSH6*, and *MSH2*), cell proliferation, and cell-cycle control, as well as genes coding for secretory proteins and collagens (see Supplementary Table S2 and Supplementary Fig. S3).

Previous studies on resected lung adenocarcinoma have suggested a large number of gene signatures, single biomarkers (including *ERCC1*, *RRM1*, and different cell-cycle regulators), and molecular subtype signatures to be associated with survival (see e.g., refs. 10–18). With respect to these, our study adds a demonstration of how unsupervised analysis of transcriptional patterns in straightforwardly defined, and clinically relevant adenocarcinoma mutation groups can stratify patients into better or worse prognosis in both a mutation-specific and a general setting. Importantly, cohorts included in both the discovery and validation phases were analyzed by different microarray platforms and conducted in both western and Asian countries. To develop a more clinically practical molecular assay, the multicohort gene signatures would need to be reduced to a smaller set of marker genes, for example, by gene network analyses as recently described (13, 37). Such a reduced gene set may be measured by other techniques with potential application also to formalin-fixed paraffin-embedded tissue.

The prognostic associations of the multicohort EGFR-1/2 and wt/wt-1/2 gene signatures were adenocarcinoma specific, but not limited to a mutation-specific context, as both signatures added independent prognostic information irrespective of mutation status in four independent adenocarcinoma validation cohorts. The association with patient outcome in adenocarcinomas for the two signatures is presumably due to the presence of a strong proliferative component, with elevated cell proliferation and loss of cell-cycle control associated with poor outcome (14, 38). The prognostic importance of the proliferative component in the signatures was supported by the observation that overlapping probe sets between the EGFR-1/2 and the wt/wt-1/2 multicohort signatures were strongly enriched for cell cycle-related genes. In addition, a centroid classifier based on these overlapping probe sets alone yielded nearly identical prognostic results as the original classifiers (data not shown).

The EGFR-1 and wt/wt-2 low-risk groups were notably enriched for bronchioid-classified tumors, whereas the high-risk groups included the majority of magnoid and squamoid tumors. Bronchioid-classified tumors have been repeatedly associated with EGFR alterations (17, 19), although approximately 30% or more of the EGFR-mutated



tumors are classified as nonbronchioid in discovery cohorts from both previous studies (17, 19) and the current study. In the absence of bronchioid tumors, we found no significant association between the magnoid and squamoid subtypes and *EGFR/KRAS* mutation status in our four discovery cohorts and GSE13213. Although a detailed analysis of the characteristics of the molecular subtypes with respect to *EGFR/KRAS* mutation-defined subgroups have not been reported, proliferation differences appear as a possible explanation for the differences in distribution of the bronchioid, magnoid, and squamoid subtypes between our mutation subgroups. This seems likely as bronchioid tumors overall displayed significantly lower expression of the CIN70 metagene compared with magnoid and squamoid tumors in all discovery cohorts irrespective of whether tumors were stratified by mutation status or not (data not shown). In contrast, no significant difference was seen in expression of the CIN70 metagene between magnoid and squamoid tumors (data not shown). In line with this, increasing the number of evaluated clusters in the consensus clustering from two to three for the largest cohorts of *EGFR*-mutated tumors (GSE31210,  $n = 127$  and E-MTAB-923,  $n = 49$ ) or *EGFR*wt/*KRAS*wt tumors (GSE31210,  $n = 79$  and Chitale U133A,  $n = 65$ ) did not resolve magnoid and squamoid tumors as separate clusters (data not shown). In addition, proliferation differences between mutation subgroups also appear as the likely explanation for the enrichment of never-smokers in the low-risk wt/wt-2 subgroup and the low-risk *EGFR*-1 group in GSE31210 and GSE13213 (39, 40).

In recent years, ACT has improved the overall survival for patients with surgically treated NSCLC (41), and there are a few reports of gene signatures that predict response to adjuvant treatment (12, 13, 42). Gene expression-based classification of NSCLC cases in the GSE14814 (12) and GSE42127 (13) cohorts by the *EGFR*-1/2 and wt/wt-1/2 signatures in our study suggested that ACT only benefits patients in the high-risk *EGFR*-2 and wt/wt-1 groups. Notably, these results are consistent with previous studies reporting predictive gene signatures based on the same cohorts (12, 13, 42). However, these results should be interpreted with care given the small cohort sizes and the increasing evidence that tumor histology needs to be considered in clinical decision making for treatment of NSCLC (43), as exemplified by the superior effect of pemetrexed in non-squamous NSCLC (44). For instance, in both the GSE14814 and the GSE42127 cohorts, ACT seemed more beneficial in a 5-year follow-up perspective for patients with squamous cell carcinomas ( $P = 0.06$  and  $0.10$ , respectively; log-rank test) compared with patients with adenocarcinomas ( $P = 0.21$  and  $0.32$ , respectively). Moreover, in both our study and the recent study by Tang and colleagues (13), reporting a 12-gene signature predictive of ACT response, the identified gene signatures were not prognostic in squamous cell carcinomas, which comprise a large part of both the untreated and treated groups in the GSE14814 and GSE42127 cohorts. Furthermore, squamous cell carcinomas generally displayed higher expression of the CIN70 metagene than

adenocarcinomas in all NSCLC cohorts included in the current study, leading to the enrichment of these tumors in high-proliferative risk groups when analyzed together with adenocarcinomas. Together, this highlights the need for adequately sized studies to identify and/or evaluate the clinical value of predictive gene signatures in a histology-specific setting.

*EGFR* tyrosine kinase and *ALK* inhibitors have improved clinical outcome in advanced NSCLC (5, 45), although only a smaller fraction of tumors harbor *EGFR* mutations or *ALK* rearrangements and thus fulfill criteria for such treatment. Moreover, these patients usually relapse because of primary or acquired resistance. Clearly, additional tools are needed to further guide diagnosis and treatment with targeted therapies in patients with lung cancer. Response to *EGFR* inhibitors has been associated with retention of an epithelial phenotype in NSCLC cell lines and tumors by epithelial-mesenchymal transition (EMT) gene signatures derived from cell line experiments (26, 46). However, the performance of these EMT signatures in resected tumor tissue remains to be clarified in larger patient cohorts. We found only a limited overlap (5–6 genes) between the two multicohort signatures and one such EMT gene signature (26). Wilkerson and colleagues reported that bronchioid-classified *EGFR* wild-type tumors displayed higher average gefitinib sensitivity scores, based on a cell line expression signature, than nonbronchioid *EGFR* wild-type tumors (19). This suggests that *EGFR* wild-type tumors responding to gefitinib would be of the bronchioid subtype (19). Given that *KRAS*-mutant adenocarcinomas are resistant to *EGFR* inhibitors (5), this subset of tumors would then primarily correspond to our low-risk wt/wt-2 subgroup. In contrast, no significant difference in gefitinib sensitivity scores was observed between *EGFR*-mutated tumors stratified by molecular subtype (19). Yuan and colleagues recently reported that clustered genomic alterations (copy number gains) on chromosome 7p predicted clinical outcome and response to *EGFR* inhibitors in *EGFR*-mutated, but not in *EGFR* wild-type adenocarcinomas (47). Consistent with Yuan and colleagues (47), the high-risk *EGFR*-2 group showed more copy number alterations and more amplifications on chromosome 7p than the low-risk *EGFR*-1 group in *EGFR*-mutated tumors. Moreover, one of the representative genes from Yuan and colleagues, *VOPPI*, is included in the *EGFR*-1/2 multicohort classifier with highest expression in *EGFR*-2. In addition, *EGFR* itself was significantly upregulated in *EGFR*-mutated *EGFR*-2 tumors in two of three cohorts defining the *EGFR*-1/2 classifier compared with *EGFR*-1 tumors, potentially due to a higher frequency of *EGFR* copy number gain or amplification in this subgroup. To further explore the association of our derived multicohort signatures with response to targeted treatment, we classified tumor biopsies from patients with *EGFR*wt/*KRAS*wt NSCLC with advanced chemorefractory metastatic disease treated with sorafenib enrolled in the BATTLE trial (26). We show that the high-risk *EGFR*-2 and the wt/wt-1 groups included more cases without disease control, and the wt/wt-1 group was borderline nonsignificant for worse

progression-free survival. These results seem to be consistent with a more aggressive phenotype for the EGFR-2 and wt/wt-1 high-risk groups, as suggested by their overall higher CIN70 metagene expression, higher number of copy number alterations/amplifications, and association with a poorly differentiated tumor phenotype. In addition, our results, combined with previous reports (26, 48, 49), show the potential of applying prognostic/predictive gene expression signatures to small biopsy specimens from patients with nonoperable disease, provided that enough tissue material could be sampled.

Together, these findings suggest that a connection between the identified transcriptional subgroups and response to different targeted treatments is possible, and that mutation status and molecular subtyping together could potentially predict therapy response better than mutation status alone. Patients with *EGFR*-mutated adenocarcinomas treated with targeted tyrosine kinase inhibitors represent the most important case. In addition, the growing number of detected tyrosine kinase fusions (including *ALK*, *RET*, and *ROS1*) in predominantly lung adenocarcinoma are also becoming increasingly important as these alterations are/may become targets for specialized molecular agents and comprise a notable fraction of EGFRwt/KRASwt adenocarcinomas. With the exception of 11 confirmed *ALK*-positive cases in GSE31210, patient-specific information about *ALK*, *RET*, and *ROS1* rearrangements were not available for included cohorts in the current study. Marked overexpression of these genes, which could indicate the presence of potential rearrangements, was only observed in a small number of cases across the different cohorts (data not shown). This was especially evident for *ALK* when compared with the expression levels of the known *ALK*-positive cases in GSE31210. Together, this precluded a detailed analysis of the multicohort signatures in these subgroups.

In summary, we identified transcriptional subgroups in *EGFR*-mutated and EGFRwt/KRASwt adenocarcinomas

with clinical and genomic differences based on a multicohort discovery and validation strategy. The identified gene signatures also added independent prognostic information in a general lung adenocarcinoma context irrespective of mutation status, and showed promising associations with response to different treatments. Further analyses in larger well-characterized cohorts with available treatment response data for EGFR inhibitors or other therapeutic agents are required to determine the predictive values of the identified gene signatures in a mutation-specific and general context.

#### Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed.

#### Authors' Contributions

**Conception and design:** M. Planck, S. Isaksson, J. Staaf

**Development of methodology:** J. Staaf

**Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis):** S. Veerla, J. Staaf

**Writing, review, and/or revision of the manuscript:** M. Planck, S. Isaksson, J. Staaf

**Study supervision:** J. Staaf

#### Acknowledgments

The authors thank the editors at Elevate Scientific for helpful comments on the article.

#### Grant Support

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, the Lund University Hospital Research Funds, the Gustav V Jubilee Foundation, and the IngaBritt and Arne Lundberg 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 April 2, 2013; revised June 20, 2013; accepted July 17, 2013; published OnlineFirst August 12, 2013.

#### References

- Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011;61:69–90.
- World Health Organization classification of tumours. In: Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, editors. Pathology and genetics of tumours of the lung, pleura, thymus and heart. Lyon, France: IARC Press; 2004.
- Kadara H, Kabbout M, Wistuba II. Pulmonary adenocarcinoma: a renewed entity in 2011. *Respirology* 2011;17:50–65.
- Johnson ML, Sima CS, Chaft J, Paik PK, Pao W, Kris MG, et al. Association of KRAS and EGFR mutations with survival in patients with advanced lung adenocarcinomas. *Cancer* 2013;119:356–62.
- Pao W, Chmielecki J. Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. *Nat Rev Cancer* 2010;10:760–74.
- Seo JS, Ju YS, Lee WC, Shin JY, Lee JK, Bleazard T, et al. The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res* 2012;22:2109–19.
- Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, et al. RET, ROS1 and ALK fusions in lung cancer. *Nat Med* 2012;18:378–81.
- Bell DW, Lynch TJ, Haserlat SM, Harris PL, Okimoto RA, Brannigan BW, et al. Epidermal growth factor receptor mutations and gene amplification in non-small-cell lung cancer: molecular analysis of the IDEAL/INTACT gefitinib trials. *J Clin Oncol* 2005;23:8081–92.
- Zhu CQ, da Cunha Santos G, Ding K, Sakurada A, Cutz JC, Liu N, et al. Role of KRAS and EGFR as biomarkers of response to erlotinib in National Cancer Institute of Canada Clinical Trials Group Study BR.21. *J Clin Oncol* 2008;26:4268–75.
- Coate LE, John T, Tsao MS, Shepherd FA. Molecular predictive and prognostic markers in non-small-cell lung cancer. *Lancet Oncol* 2009;10:1001–10.
- Ellis PM, Blais N, Soulieres D, Ionescu DN, Kashyap M, Liu G, et al. A systematic review and Canadian consensus recommendations on the use of biomarkers in the treatment of non-small cell lung cancer. *J Thorac Oncol* 2011;6:1379–91.
- Zhu CQ, Ding K, Strumpf D, Weir BA, Meyerson M, Pennell N, et al. Prognostic and predictive gene signature for adjuvant chemotherapy in resected non-small-cell lung cancer. *J Clin Oncol* 2010;28:4417–24.
- Tang H, Xiao G, Behrens C, Schiller J, Allen J, Chow CW, et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in

- non-small-cell lung cancer patients. *Clin Cancer Res* 2013;19:1577–86.
14. Shedden K, Taylor JM, Enkemann SA, Tsao MS, Yeatman TJ, Gerald WL, et al. Gene expression-based survival prediction in lung adenocarcinoma: a multi-site, blinded validation study. *Nat Med* 2008;14:822–7.
  15. Tomida S, Takeuchi T, Shimada Y, Arima C, Matsuo K, Mitsudomi T, et al. Relapse-related molecular signature in lung adenocarcinomas identifies patients with dismal prognosis. *J Clin Oncol* 2009;27:2793–9.
  16. Okayama H, Kohno T, Ishii Y, Shimada Y, Shiraishi K, Iwakawa R, et al. Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* 2012;72:100–11.
  17. Hayes DN, Monti S, Parmigiani G, Gilks CB, Naoki K, Bhattacharjee A, et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J Clin Oncol* 2006;24:5079–90.
  18. Bhattacharjee A, Richards WG, Staunton J, Li C, Monti S, Vasa P, et al. Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc Natl Acad Sci U S A* 2001;98:13790–5.
  19. Wilkerson MD, Yin X, Walter V, Zhao N, Cabanski CR, Hayward MC, et al. Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PLoS ONE* 2012;7:e36530.
  20. Subramanian J, Simon R. Gene expression-based prognostic signatures in lung cancer: ready for clinical use? *J Natl Cancer Inst* 2010;102:464–74.
  21. Fouret R, Laffaire J, Hofman P, Beau-Faller M, Mazieres J, Validire P, et al. A comparative and integrative approach identifies ATPase family, AAA domain containing 2 as a likely driver of cell proliferation in lung adenocarcinoma. *Clin Cancer Res* 2012;18:5606–16.
  22. Chitale D, Gong Y, Taylor BS, Broderick S, Brennan C, Somwar R, et al. An integrated genomic analysis of lung cancer reveals loss of DUSP4 in EGFR-mutant tumors. *Oncogene* 2009;28:2773–83.
  23. Shibata T, Hanada S, Kokubu A, Matsuno Y, Asamura H, Ohta T, et al. Gene expression profiling of epidermal growth factor receptor/KRAS pathway activation in lung adenocarcinoma. *Cancer Sci* 2007;98:985–91.
  24. Bild AH, Yao G, Chang JT, Wang Q, Potti A, Chasse D, et al. Oncogenic pathway signatures in human cancers as a guide to targeted therapies. *Nature* 2006;439:353–7.
  25. Staaf J, Isaksson S, Karlsson A, Jonsson M, Johansson L, Jonsson P, et al. Landscape of somatic allelic imbalances and copy number alterations in human lung carcinoma. *Int J Cancer* 2012;1:2020–31.
  26. Byers LA, Diao L, Wang J, Saintigny P, Girard L, Peyton M, et al. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clin Cancer Res* 2013;19:279–90.
  27. Kim ES, Herbst RS, Wistuba II, Lee JJ, Blumenschein GR Jr, Tsao A, et al. The BATTLE trial: personalizing therapy for lung cancer. *Cancer Discov* 2011;1:44–53.
  28. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 2003;19:185–93.
  29. Gene Expression Omnibus. Available from: <http://www.ncbi.nlm.nih.gov/geo/>.
  30. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* 2010;26:1572–3.
  31. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21.
  32. BioConductor. Available from: <http://www.bioconductor.org>.
  33. Carter SL, Eklund AC, Kohane IS, Harris LN, Szallasi Z. A signature of chromosomal instability inferred from gene expression profiles predicts clinical outcome in multiple human cancers. *Nat Genet* 2006;38:1043–8.
  34. Hammerman PS, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, et al. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 2012;489:519–25.
  35. Whitehurst AW, Bodemann BO, Cardenas J, Ferguson D, Girard L, Peyton M, et al. Synthetic lethal screen identification of chemosensitizer loci in cancer cells. *Nature* 2007;446:815–9.
  36. Rieley GJ, Kris MG, Rosenbaum D, Marks J, Li A, Chitale DA, et al. Frequency and distinctive spectrum of KRAS mutations in never smokers with lung adenocarcinoma. *Clin Cancer Res* 2008;14:5731–4.
  37. Fredlund E, Staaf J, Rantala JK, Kallioniemi O, Borg A, Ringner M. The gene expression landscape of breast cancer is shaped by tumor protein p53 status and epithelial-mesenchymal transition. *Breast Cancer Res* 2012;14:R113.
  38. Beer DG, Kardias SL, Huang CC, Giordano TJ, Levin AM, Misek DE, et al. Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nat Med* 2002;8:816–24.
  39. Staaf J, Jonsson G, Jonsson M, Karlsson A, Isaksson S, Salomonsson A, et al. Relation between smoking history and gene expression profiles in lung adenocarcinomas. *BMC Med Genomics* 2012;5:22.
  40. Landi MT, Dracheva T, Rotunno M, Figueroa JD, Liu H, Dasgupta A, et al. Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS ONE* 2008;3:e1651.
  41. Crino L, Weder W, van Meerbeeck J, Felip E. Early stage and locally advanced (non-metastatic) non-small-cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2010;21(Suppl 5):v103–15.
  42. Chen DT, Hsu YL, Fulp WJ, Coppola D, Haura EB, Yeatman TJ, et al. Prognostic and predictive value of a malignancy-risk gene signature in early-stage non-small cell lung cancer. *J Natl Cancer Inst* 2011;103:1859–70.
  43. Langer CJ, Besse B, Gualberto A, Brambilla E, Soria JC. The evolving role of histology in the management of advanced non-small-cell lung cancer. *J Clin Oncol* 2010;28:5311–20.
  44. Standfield L, Weston AR, Barraclough H, Van Kooten M, Pavlakis N. Histology as a treatment effect modifier in advanced non-small cell lung cancer: a systematic review of the evidence. *Respirology* 2011;16:1210–20.
  45. Kwak EL, Bang YJ, Camidge DR, Shaw AT, Solomon B, Maki RG, et al. Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 2010;363:1693–703.
  46. Yauch RL, Januario T, Eberhard DA, Cavet G, Zhu W, Fu L, et al. Epithelial versus mesenchymal phenotype determines *in vitro* sensitivity and predicts clinical activity of erlotinib in lung cancer patients. *Clin Cancer Res* 2005;11:8686–98.
  47. Yuan S, Yu SL, Chen HY, Hsu YC, Su KY, Chen HW, et al. Clustered genomic alterations in chromosome 7p dictate outcomes and targeted treatment responses of lung adenocarcinoma with EGFR-activating mutations. *J Clin Oncol* 2011;29:3435–42.
  48. Baty F, Facompre M, Kaiser S, Schumacher M, Pless M, Bubendorf L, et al. Gene profiling of clinical routine biopsies and prediction of survival in non-small cell lung cancer. *Am J Respir Crit Care Med* 2010;181:181–8.
  49. Suwinski R, Klusek A, Tyszkiewicz T, Kowalska M, Szczesniak-Klusek B, Gawkowska-Suwinska M, et al. Gene expression from bronchoscopy obtained tumour samples as a predictor of outcome in advanced inoperable lung cancer. *PLoS ONE* 2012;7:e41379.

# Clinical Cancer Research

## Identification of Transcriptional Subgroups in *EGFR*-Mutated and *EGFR/KRAS* Wild-Type Lung Adenocarcinoma Reveals Gene Signatures Associated with Patient Outcome

Maria Planck, Sofi Isaksson, Srinivas Veerla, et al.

*Clin Cancer Res* 2013;19:5116-5126. Published OnlineFirst August 12, 2013.

**Updated version** Access the most recent version of this article at:  
doi:[10.1158/1078-0432.CCR-13-0928](https://doi.org/10.1158/1078-0432.CCR-13-0928)

**Supplementary Material** Access the most recent supplemental material at:  
<http://clincancerres.aacrjournals.org/content/suppl/2013/08/12/1078-0432.CCR-13-0928.DC1>

**Cited articles** This article cites 46 articles, 17 of which you can access for free at:  
<http://clincancerres.aacrjournals.org/content/19/18/5116.full#ref-list-1>

**Citing articles** This article has been cited by 2 HighWire-hosted articles. Access the articles at:  
<http://clincancerres.aacrjournals.org/content/19/18/5116.full#related-urls>

**E-mail alerts** [Sign up to receive free email-alerts](#) related to this article or journal.

**Reprints and Subscriptions** To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at [pubs@aacr.org](mailto:pubs@aacr.org).

**Permissions** To request permission to re-use all or part of this article, use this link  
<http://clincancerres.aacrjournals.org/content/19/18/5116>.  
Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.