Genomic and Transcriptional Alterations in Lung Adenocarcinoma in Relation to EGFR and KRAS Mutation Status.

Planck, Maria; Edlund, Karolina; Botling, Johan; Micke, Patrick; Isaksson, Sofi; Staaf, Johan

Published in:
PLoS ONE

DOI:
10.1371/journal.pone.0078614

2013

Citation for published version (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal

Take down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
Genomic and Transcriptional Alterations in Lung Adenocarcinoma in Relation to EGFR and KRAS Mutation Status

Maria Planck, Karolina Edlund, Johan Botling, Patrick Micke, Sofi Isaksson, Johan Staaf

1 Department of Oncology, Clinical Sciences, Lund University and Skåne University Hospital, Medicon Village, Lund, Sweden
2 Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden
3 CREATE Health Strategic Center for Translational Cancer Research, Lund University, Medicon Village, Lund, Sweden

Abstract

Introduction: In lung adenocarcinoma, the mutational spectrum is dominated by EGFR and KRAS mutations. Improved knowledge about genomic and transcriptional alterations in and between mutation-defined subgroups may identify genes involved in disease development or progression.

Methods: Genomic profiles from 457 adenocarcinomas, including 113 EGFR-mutated, 134 KRAS-mutated and 210 EGFR and KRAS-wild type tumors (EGFRwt/KRASwt), and gene expression profiles from 914 adenocarcinomas, including 309 EGFR-mutated, 192 KRAS-mutated, and 413 EGFRwt/KRASwt tumors, were assembled from different repositories. Genomic and transcriptional differences between the three mutational groups were analyzed by both supervised and unsupervised methods.

Results: EGFR-mutated adenocarcinomas displayed a larger number of copy number alterations and recurrent amplifications, a higher fraction of total loss-of-heterozygosity, higher genomic complexity, and a more distinct expression pattern than EGFR-wild type adenocarcinomas. Several of these differences were also consistent when the three mutational groups were stratified by stage, gender and smoking status. Specific copy number alterations were associated with mutation status, predominantly including regions of gain with the highest frequency in EGFR-mutated tumors. Differential regions included both large and small regions of gain on 1p, 5q34-q35.3, 7p, 7q11.21, 12p12.1, 16p, and 21q, and losses on 6q16.3-q21, 8p, and 9p, with 20-40% frequency differences between the mutational groups. Supervised gene expression analyses identified 96 consistently differentially expressed genes between the mutational groups, and together with unsupervised analyses these analyses highlighted the difficulty in broadly resolving the three mutational groups into distinct transcriptional entities.

Conclusions: We provide a comprehensive overview of the genomic and transcriptional landscape in lung adenocarcinoma stratified by EGFR and KRAS mutations. Our analyses suggest that the overall genomic and transcriptional landscape of lung adenocarcinoma is affected, but only to a minor extent, by EGFR and KRAS mutation status.

Introduction

Lung cancer is a heterogeneous malignancy with poor survival due to diagnosis at an often advanced stage [1]. Lung cancer is broadly divided into small cell lung cancer (~15% of all lung cancers) and non-small cell lung cancer with adenocarcinoma as the most frequent histological type [2]. In adenocarcinoma, the mutational spectrum is dominated by EGFR and KRAS mutations, where the former is an established predictor of response to EGFR inhibitors [3,4]. EGFR and KRAS mutations are nearly always mutually exclusive and associated with differences in patient gender and smoking history [5]. Together, this suggests that these genetic alterations may be drivers of pathogenesis for specific
Genomic Alterations in EGFR/KRAS/wt Adenocarcinoma

Materials and Methods

Tumor material

Genomic profiles from 1272 adenocarcinoma tumors and cell lines were obtained from a previous study (n=1210) [21], with addition of adenocarcinomas from Wilkerson et al. [20] (n=62, GSE36363). All genomic profiles were analyzed in an unmatched fashion and sample uniqueness was assured as described [21]. All included tumors represented primary driver events from original studies. EGFR and KRAS mutational status was available for 457 adenocarcinoma, including 113 EGFR-mutated, 134 KRAS-mutated, and 210 EGFRwt/KRASwt cases.

Gene expression profiles from 914 adenocarcinoma tumors, including 309 EGFR-mutated, 192 KRAS-mutated and 413 EGFRwt/KRASwt cases, were collected from eight studies analyzed by different microarray platforms [8,10,20,22-26]. Samples from Chitale et al. [10] were further divided into two cohorts according to their different Affymetrix platforms (U133A and U133 2plus).

EGFR and KRAS mutation analysis

EGFR and KRAS mutation status was determined as described in either File S1 (for GSE37745 [25] and GSE28572 [27]) or in each of the original articles.

Genomic analyses

Normalized copy number and B allele frequency estimates for Affymetrix microarrays and Illumina SNP beadchips, and normalized copy number estimates for Agilent 44K, Agilent 244K and ROMA 85K cohorts were generated and/or assembled as described in Staaf et al. [21] and File S1. Probe annotations for all array platforms were updated to the hg18/NCBI36 genome build. Genomic profiles were partitioned, centralized, and merged to a common probe set as described [21] and File S1.

A modified version of Genomic Identification of Significant Targets in Cancer (GISTIC) [28], referred to as mGISTIC herein, was used for identification of focal copy number alterations and recurrent amplifications from the 1272-sample cohort (see [21] and File S1). Robustness of identified regions was assessed by permutation analysis (Figure S1 and File S1). A genome-wide screen of differential copy number gain and loss between the three EGFR and KRAS defined mutation groups was performed by division of genomic profiles into 12,698 sequential segments of ~200 Kbp size, excluding reported regions of copy number variation. Each segment was subsequently tested for differences in frequency of copy number gain or loss individually. Fisher’s exact test or the Chi-square test was used to identify genomic regions and recurrent amplifications with different frequency between mutation groups.

For tumors analyzed by SNP microarrays (n=141), B allele frequency estimates were partitioned [29], integrated with copy number data, and subjected to Genome Alteration Print (GAP) [30] analysis for estimation of allele-specific copy numbers and in silico tumor ploidy (referred to as GAP-ploidy herein) as described [21]. Loss-of-heterozygosity (LOH), copy-neutral LOH, and copy-neutral allelic imbalance were estimated from GAP results as described [21]. The fractions of the genome altered by copy number alterations, LOH, copy-neutral LOH, and copy-neutral allelic imbalance were calculated as described [21]. Data processing steps are further described in File S1 and [21].

Gene expression analyses

Affymetrix cohorts were individually normalized using GC Robust Multi-array Averaging (GCRMA) [31]. For non-Affymetrix cohorts, normalized expression data were obtained from Gene Expression Omnibus [32]. In total, nine cohorts were analyzed individually for transcriptional differences.
between mutation groups as either discovery (n=5) or validation cohorts (n=4) (Table 2). Differentially expressed genes between EGFR-mutated, KRAS-mutated and EGFRwt/KRASwt tumors were identified by ANOVA with false discovery rate adjustment using a 5% threshold for statistical significance. Hierarchical clustering was performed using Pearson correlation and complete linkage. Data processing steps are further described in File S1.

Results

Copy number alterations in lung adenocarcinoma

To identify copy number alterations (CNAs) of general importance in lung adenocarcinoma, which may serve as basis for supervised comparisons between EGFR-mutated, KRAS-mutated and EGFRwt/KRASwt tumors, we analyzed 1272 tumors and cell lines profiled by SNP or aCGH microarrays (Figure 2A, Table 1). To pinpoint recurrent CNAs in lung...
Genomic Alterations in *EGFR/KRAS/wt* Adenocarcinoma

**Table 1. Characteristics of individual aCGH and SNP genomic adenocarcinoma cohorts stratified by microarray platform.**

<table>
<thead>
<tr>
<th>Genomic cohort</th>
<th>No. of AC^A/Microarray platform</th>
<th>Included in CN / GAP analysis^B</th>
<th>No. of tumors / cell lines</th>
<th>No. of <em>EGFR</em>-mutated / Kras^C</th>
<th>No. of stage III/IV tumors^C</th>
<th>Gender</th>
<th>Smoking status NS/Smoker</th>
<th>Smoking status NS/Smoker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zhao [49]</td>
<td>36 Affymetrix 100K</td>
<td>Yes/No</td>
<td>3/0</td>
<td>12/0</td>
<td></td>
<td>2/1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE18252 [50]</td>
<td>4 Affymetrix 100K</td>
<td>Yes/Yes</td>
<td>4/0</td>
<td>0/4/0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weir [51]</td>
<td>112 Affymetrix 250K Syl</td>
<td>Yes/No</td>
<td>112/0</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Weir [51]</td>
<td>196 Affymetrix 250K Syl</td>
<td>Yes/No</td>
<td>196/0</td>
<td>15/49/51</td>
<td>41/14/15/3</td>
<td>110/83</td>
<td>17/116</td>
<td></td>
</tr>
<tr>
<td>GSE19399 [52]</td>
<td>19 Affymetrix 250K Syl</td>
<td>Yes/Yes</td>
<td>19/0</td>
<td>1/1/8</td>
<td>19/1/4/0</td>
<td>10/9</td>
<td>2/16</td>
<td></td>
</tr>
<tr>
<td>GSE17247 [53]</td>
<td>49 Affymetrix 250K Syl</td>
<td>Yes/No</td>
<td>0/49</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE28872 [27]</td>
<td>44 Affymetrix 250K Syl</td>
<td>Yes/Yes</td>
<td>44/0</td>
<td>9/18/17</td>
<td>21/10/8/2</td>
<td>27/17</td>
<td>4/37</td>
<td></td>
</tr>
<tr>
<td>GSE34140 [54]</td>
<td>141 Affymetrix 250K Syl</td>
<td>Yes/No</td>
<td>141/0</td>
<td>-</td>
<td>79/26/9/7</td>
<td>-</td>
<td>10/107</td>
<td></td>
</tr>
<tr>
<td>GSK [55]</td>
<td>10 Affymetrix 250K Syl</td>
<td>Yes/No</td>
<td>0/10</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE19804 [56]</td>
<td>12 Affymetrix 6.0</td>
<td>Yes/No</td>
<td>12/0</td>
<td>-</td>
<td>6/3/2/1</td>
<td>12/0</td>
<td>12/0</td>
<td></td>
</tr>
<tr>
<td>GSE25016 [34]</td>
<td>58 Affymetrix 6.0</td>
<td>Yes/No</td>
<td>58/0</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSE33848 [9]</td>
<td>146 Affymetrix 6.0</td>
<td>Yes/No</td>
<td>146/0</td>
<td>-</td>
<td>95/0/0/0</td>
<td>-</td>
<td></td>
<td></td>
</tr>
<tr>
<td>TCGA-AC [57]</td>
<td>135 Affymetrix 6.0</td>
<td>Yes/No</td>
<td>135/0</td>
<td>-</td>
<td>73/23/27/9</td>
<td>78/57</td>
<td>19/109</td>
<td></td>
</tr>
<tr>
<td>Chitale [10]</td>
<td>184 Agilent 44K</td>
<td>Yes/No</td>
<td>184/0</td>
<td>42/45/97</td>
<td>121/26/32/5</td>
<td>107/77</td>
<td>39/145</td>
<td></td>
</tr>
<tr>
<td>GSE20393 [58]</td>
<td>2 Agilent 244K</td>
<td>Yes/No</td>
<td>2/0</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E-TABM-926 [59]</td>
<td>17 Agilent 244K</td>
<td>Yes/No</td>
<td>17/0</td>
<td>13/1/3</td>
<td>8/1/8/0</td>
<td>14/3</td>
<td>17/0</td>
<td></td>
</tr>
<tr>
<td>E-TABM-1169 [59]</td>
<td>40 Illumina 370K</td>
<td>Yes/Yes</td>
<td>40/0</td>
<td>28/2/10</td>
<td>22/5/13/0</td>
<td>36/4</td>
<td>40/0</td>
<td></td>
</tr>
<tr>
<td>GSE31586 5</td>
<td>ROMA 85K</td>
<td>Yes/No</td>
<td>0/5</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TOTAL</td>
<td>1272</td>
<td>1272/733</td>
<td>120/64</td>
<td>113/134/210</td>
<td>515/117/129/27</td>
<td>433/276</td>
<td>165/586</td>
<td></td>
</tr>
</tbody>
</table>

A: Number of used adenocarcinoma cases per cohort. B: Included in overall GISTIC analysis (CN) and/or GAP-analysis for *EGFR/KRAS* mutation groups. C: For tumors only.

D: NS = never-smoker, S = smoker

The table presents characteristics of individual aCGH and SNP genomic adenocarcinoma cohorts stratified by microarray platform, including the number of cases, microarray platforms used, and various genomic alterations. The table also highlights the relationship between smoking status and the presence of genomic alterations.

Adenocarcinoma, we performed an mGISTIC analysis of the entire 1272-sample set identifying 59 gains and 31 losses distributed across all autosomes (Figure 2A, Table S1). Several of the identified mGISTIC regions harbored known or putative adenocarcinoma driver candidates, such as **EGFR**, **MM2**, **KRAS**, **MYC**, **MYCL**, **TERM**, **MET**, **CCND1**, **NXX2-1/TITF1**, **CDK4**, **ERBB2**, **ID1**, **RB1**, **CDKN2A**, and **PTEN**.

Copy number alterations in *EGFR/KRAS* mutation groups

Stratification of the 457 adenocarcinoma tumors with known **EGFR** and **KRAS** mutation status into **EGFR**-mutated (n=113), **KRAS**-mutated (n=134) and **EGFR**/**KRAS** (n=210) tumors revealed both common alterations across mutation groups, such as gains of chromosome 1q and 8q, and loss of 3p, and regions with apparently different prevalence between mutation groups, including gains on chromosome 7p (**EGFR**-mutated) and 16p (**EGFR**-mutated), and losses on 6q (**KRAS**-mutated) (Figures 2B-D). In general, **EGFR**-mutated tumors displayed more copy number alterations (estimated by the fraction of the genome altered by CNA, CN-FGA) than non-**EGFR**-mutated tumors (Figure 3A). This pattern was consistent also in five out of six individual cohorts that included both **EGFR**-mutated and non-**EGFR**-mutated tumors. When the three mutation groups were stratified by clinicopathological variables, **EGFR**-mutated tumors continued to display higher CN-FGA fractions in stage I tumors, female patients, and never-smokers (Figure 3A).

Analysis of the 90 focal mGISTIC regions (derived from analysis of the 1272 adenocarcinoma sample cohort) identified 17 regions discriminating between the three mutation groups. 15 of these 17 regions showed the highest alteration frequency in **EGFR**-mutated tumors, while the remaining two regions showed highest frequency in **KRAS**-mutated tumors (Bonferroni adjusted Fisher’s exact test p<0.05 and frequency difference >20%, Figure 3B and Table 3). Specifically, **EGFR**-mutated tumors showed higher frequencies of copy number gain on chromosomes 1q34.2 (including **MYCL**), 5q15.33, 5q35.1, 7p22.3-22.2, 7p21.1, 7p11.2 (including **EGFR**), 7q11.21, 14q21.2, and 16p13.13, and copy number loss in regions at 8p (including **DUSP4**), 9p (including **CDKN2A**), and 10q23.2-q33.1 (**PTEN**). **KRAS**-mutated tumors showed higher frequencies of gain on 12p12.1 (**KRAS**) and loss at 6q16.3-q21.

A genome-wide analysis of differences in copy number frequency between the three mutation groups identified nine large coherent genomic regions (seven gains and two losses), all with higher alteration frequency in **EGFR**-mutated tumors. Regions were located on 1p, 5q, 7p, 7q, 8q, 16p and 21q, and involved 8% (7% gain, 1% loss) of the analyzed genome (Hochberg adjusted Fisher’s exact test p<0.01 and minimum frequency difference >20%, Tables 3 and S2).
Similar to copy number gain and loss in general, *EGFR*-mutated tumors also displayed more recurrent amplifications in the 59 mGISTIC regions of gain compared with the non-*EGFR*-mutated tumors (p=0.004, Chi-square test). This finding was consistent also in patients with stage I disease (p=0.05, Fisher’s exact test), but not in tumors of higher stages (≥II), or tumors stratified by gender. In contrast, no significant differences in copy-neutral LOH or copy-neutral allelic imbalance fractions were observed between mutation groups overall or when stratified by stage or gender. The highest frequencies of total LOH (>50%) were most often found in regions of copy number loss, while copy-neutral LOH and copy-neutral allelic imbalance showed an overall lower prevalence across chromosomes in all mutation groups (generally ≤10-15% frequency for copy-neutral LOH, and <25% for copy-neutral allelic imbalance across chromosomes) (Table 3 and Figure S2).

Taken together, these results suggest a higher genomic complexity in *EGFR*-mutated adenocarcinomas compared with *KRAS*-mutated and *EGFRwt/KRASwt* tumors.

Patterns of tumor ploidy and allelic imbalance in *EGFR/ KRAS* mutation groups

Patterns of tumor ploidy and allelic imbalances between mutational groups were evaluated by GAP [30] analysis of 141 tumors (n=43 *EGFR*-mutated, 39 *KRAS*-mutated, and 59 *EGFRwt/KRASwt*) profiled by SNP microarrays. Primarily, no differences in distribution of tumor ploidy (estimated by GAP-ploidy) were observed between mutation groups (p=0.96, ANOVA, Figure 3C). Secondly, *EGFR*-mutated adenocarcinomas were weakly associated with higher fractions of total LOH compared with *KRAS*-mutated and *EGFRwt/KRASwt* tumors both overall and in stage I disease (p=0.05, ANOVA), but not in tumors of higher stages (≥II), or tumors stratified by gender. In contrast, no significant differences in copy-neutral LOH or copy-neutral allelic imbalance fractions were observed between mutation groups overall or when stratified by stage or gender. The highest frequencies of total LOH (>50%) were most often found in regions of copy number loss, while copy-neutral LOH and copy-neutral allelic imbalance showed an overall lower prevalence across chromosomes in all mutation groups (generally ≤10-15% frequency for copy-neutral LOH, and <25% for copy-neutral allelic imbalance across chromosomes) (Table 3 and Figure S2).

Taken together, this implies that the weak associations of differences in allelic imbalances between the mutation groups are predominantly related to LOH caused by copy number loss in *EGFR*-mutated tumors.

**Supervised and unsupervised analysis of transcriptional differences between *EGFR/KRAS* mutation groups**

To identify a robust set of differentially expressed genes between the three mutation groups we performed supervised analysis of five adenocarcinoma Affymetrix cohorts (n=624 tumors, discovery cohorts, Table 2). 96 genes showed consistent differential expression in ≥4 cohorts, while only 21
genes were differentially expressed across all five cohorts (Tables 4 and S3). We validated the 96 identified genes in four independent adenocarcinoma cohorts analyzed by different microarray platforms (n=290 tumors, Table 2). In the

Figure 2. Copy number alterations in lung adenocarcinoma. Frequency of copy number gain (red) and loss (green) for adenocarcinoma stratified by EGFR and KRAS mutational status using \( \log_2 \) ratio ± 0.12 as threshold for identification of copy number gain and loss. Probes matched to known copy number variations are excluded. Black regions indicate genomic position of significant mGISTIC regions, which were identified from analysis of the entire 1272-sample cohort across chromosomes. Arrows indicate genomic regions with apparently different copy number alteration frequency between EGFR/KRAS mutation groups (6q, 7p, and 16p). (A) All 1272 adenocarcinomas. (B) 113 EGFR-mutated adenocarcinoma tumors. (C) 134 KRAS-mutated adenocarcinoma tumors. (D) 210 EGFRwt/KRASwt adenocarcinoma tumors.

doi: 10.1371/journal.pone.0078614.g002
Figure 3. Copy number alterations and tumor ploidy in EGFR/KRAS mutation groups. (A) Pattern of gross copy number alterations measured as fraction of the genome altered by copy number gain or loss in adenocarcinoma tumors stratified by EGFR and KRAS mutation status (EGFR:red, KRAS:light blue, EGFRwt/KRASwt:gray), stage, gender and patient smoking status. Copy number alterations were called using log_{2} ratio ± 0.12 as thresholds for identification of copy number gain and loss. P-values were calculated using ANOVA for indicated groups, ***: P< 0.001, **: P< 0.01, *: P< 0.05. Top axis indicates number of cases per group. (B) mGISTIC regions discriminating between EGFR-mutated (red), KRAS-mutated (light blue) and EGFRwt/KRASwt (gray) adenocarcinoma tumors. mGISTIC regions identified by Fisher’s exact test (Bonferroni adjusted p-value < 0.05) with an additional requirement of > 20% frequency difference between the lowest and highest groups. The y-axis describes the frequency of copy number gain or loss in respective group. (C) Distribution of GAP-ploidy across the adenocarcinoma EGFR/KRAS mutation groups for 141 tumors analyzed by GAP. A GAP-ploidy of two equals a diploid, three a triploid genome and four a tetraploid genome. Curves were generated by an Epanechnikov smoothing kernel with 0.1 smoothing bandwidth.

doi: 10.1371/journal.pone.0078614.g003
Table 3. Differences and similarities in genomic alterations and allelic imbalances between EGFR/KRAS mutation groups.

<table>
<thead>
<tr>
<th>Investigated property</th>
<th>EGFR-mutated</th>
<th>KRAS-mutated</th>
<th>EGFRwt/KRASwt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fraction of the genome altered by copy number gain and loss</td>
<td>More a</td>
<td>Less</td>
<td>Less</td>
</tr>
<tr>
<td>Fraction of the genome altered by LOH</td>
<td>More</td>
<td>Less</td>
<td>Less</td>
</tr>
<tr>
<td>Fraction of the genome altered by copy number neutral LOH or copy-neutral allelic imbalance</td>
<td>Less</td>
<td>Less</td>
<td>Less</td>
</tr>
<tr>
<td>Overall frequency of recurrent amplifications and genomic complexity</td>
<td>More</td>
<td>Less</td>
<td>Less</td>
</tr>
<tr>
<td>Amplifications differing between mutation groups (mGISTIC regions)</td>
<td>7p11.2 (EGFR), 12q14.2-q14.3, 12q15 (MDM2)</td>
<td>8p12 (FGFR1)</td>
<td></td>
</tr>
<tr>
<td>Copy number alterations differing between mutation groups (mGISTIC regions)</td>
<td>+1p34.2, +5p15.33, +5q35.1, +7p22.3-p22.2,</td>
<td>+12p12.1 (KRAS),</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+7p21.1, +7p11.2, +7q11.21, +14q21.2, +16p13.13,</td>
<td>+6q16.3-q21</td>
<td></td>
</tr>
<tr>
<td></td>
<td>-8p23.2-p23.1, -6p21.2-p12, -9p24.3, -9p23,</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>-9p21.3, -10q23.2-q23.31</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Copy number alterations differing between mutation groups (genome-wide screen)</td>
<td>+1p36.33-p31.1, +5q34-q35.3, +7p22.3-p11.1,</td>
<td>+7q11.21, +16p13.3-p12.1, +16p11.2-q11.1,</td>
<td></td>
</tr>
<tr>
<td></td>
<td>+7q11.21, +16p13.3-p12.1, +16p11.2-q11.1,</td>
<td>+21q22.1-q22.3, -8p22-p11.21, -8q11.23</td>
<td></td>
</tr>
<tr>
<td>Characteristic total LOH regions (≥50% frequency)b</td>
<td>8p, 9, 13q, 17p</td>
<td>19p</td>
<td></td>
</tr>
<tr>
<td>Predominant tumor ploidy (GAP-ploidy)</td>
<td>2N (highest), 3N</td>
<td>2N (highest), 3N</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>2N (highest), 3N</td>
<td></td>
</tr>
</tbody>
</table>

a Less indicates relatively lower estimates or frequencies between groups. More indicates relatively higher estimates or frequencies.
b Includes LOH caused by copy number loss as well as copy-neutral LOH.

doi: 10.1371/journal.pone.0078614.0003

independent cohorts, 41-96% of the 96 genes were present and thus available for further comparisons. Of the available genes 46-67% showed differential expression between the mutation groups in the independent cohorts (p<0.05 ANOVA, Table S3).

A three-group centroid classifier was used to explore the predictive power of the 96 genes in calling true mutation status (File S1 and Table S3). Classification of the four independent cohorts showed an overall accuracy of 40-90% in classification across a range of classification cut-offs (Figure S3A).

Sensitivity was highest in classification of EGFR-mutated tumors across the different cohorts (80-100%), followed by KRAS-mutated tumors (Figure 4A). However, specificities of the 96-gene classifier were lower (60-90%) for the EGFR and KRAS-mutated groups (Figure 4A). In contrast, for EGFRwt/KRASwt tumors, sensitivity was poor (10-60%) but specificity higher (80-100%).

To further analyze the transcriptional patterns between the three mutation groups we performed unsupervised hierarchical clustering of four Affymetrix discovery cohorts (Table 2). Each cohort was individually clustered using a) a signature of genes overexpressed in EGFRwt/KRASwt tumors [26], b) a KRAS dependency gene signature [33], and c) probe sets derived from three different expression variance filters reflecting at different stringency the variation in expression across all tumors in a cohort (Affymetrix probe set range n=1356-24052).

In none of these analyses did clustering resolve the three mutation groups into discrete transcriptional groups without notable inclusion of tumors from other mutation groups (Figure S4). However, supportive of results from the supervised analyses we found that EGFR-mutated adenocarcinomas in general appeared to display a more distinctive expression pattern with enrichment of EGFR-mutated tumors (~60% of all mutations) in specific clusters. In contrast, KRAS-mutated and EGFRwt/KRASwt tumors often appeared more intermixed, even when clustered using the KRAS dependency gene signature [33] (Figure S4).

Taken together, results from the supervised and unsupervised gene expression analyses suggest that mutation status is not translated into a clearly distinctive and prominent expression signature.

Discussion

In the current study we delineate genomic and transcriptional alterations in lung adenocarcinoma stratified by EGFR and KRAS mutation status. We show that a few specific copy number and transcriptional alterations exist between the three mutational groups, but also a considerable similarity caused by high intra-group heterogeneity and/or less distinctive inter-group differences. Together, this suggests that the overall genomic and transcriptional landscape of adenocarcinoma is affected, but only to a minor extent, by the mutational status of EGFR and KRAS.

Stratification of genomic profiles from 457 tumors with available EGFR and KRAS mutation status into three mutation groups revealed differences in the overall pattern of CNAs, amplifications and genomic architecture, as well as specific regions and amplifications differing in frequency between the groups (summarized in Table 3). Overall, EGFR-mutated tumors displayed more CNAs, more amplifications, and higher genomic complexity than non-EGFR-mutated tumors consistent with previous reports [11,15,18]. Specific patterns of recurrent amplifications in between the mutation groups, such as 8p12 (harboring FGFR1) in EGFRwt/KRASwt and 12q amplifications (including the p53 repressor MDM2) in EGFR-mutated tumors were observed. FGFR1 mutations are rarely observed in NSCLC, while FGFR1 amplification is frequent in, e.g., squamous cell lung carcinoma and associated with increased protein levels and a FGFR1 proliferation dependency.
Several of the regions have been reported previously, but as presence of specific genomic circuits acting as driving forces in groups. These regions predominantly include regions of copy larger and less defined regions [9, 11-13, 17, 18], while others

Table 4. Differentially expressed genes between EGFR/KRAS mutation groups in ≥4 of five Affymetrix adenocarcinoma cohorts.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Name</th>
<th>Gene</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACSF2</td>
<td>acyl-CoA synthetase family member 2</td>
<td>ISG20</td>
<td>interferon stimulated exonuclease gene 20kDa</td>
</tr>
<tr>
<td>ADCY9</td>
<td>adenylate cyclase 9</td>
<td>ITPR3</td>
<td>inositol 1,4,5-triphosphate receptor, type 3</td>
</tr>
<tr>
<td>AGFG1</td>
<td>AFGAP with FG repeats 1</td>
<td>KCN5</td>
<td>potassium channel, subfamily K, member 5</td>
</tr>
<tr>
<td>AHR</td>
<td>aryl hydrocarbon receptor</td>
<td>KIAA0319L</td>
<td>polycystic kidney disease 1-like</td>
</tr>
<tr>
<td>APOH</td>
<td>apolipoprotein H precursor</td>
<td>KIAA0494</td>
<td>hypothetical protein LOC9613</td>
</tr>
<tr>
<td>ARMCX6</td>
<td>armadillo repeat containing, X-linked 6</td>
<td>KIAA0495</td>
<td>hypothetical protein LOC57212</td>
</tr>
<tr>
<td>ARSD</td>
<td>arylsulfatase D</td>
<td>KIAA1033</td>
<td>hypothetical protein LOC23252</td>
</tr>
<tr>
<td>BAG1</td>
<td>BCL2-associated athanogene 1</td>
<td>KRAS</td>
<td>c-K-ras2 protein isoform b</td>
</tr>
<tr>
<td>BLVRA</td>
<td>biliverdin reductase A</td>
<td>LDLRAP1</td>
<td>low density lipoprotein receptor adaptor protein</td>
</tr>
<tr>
<td>C16orf58</td>
<td>hypothetical protein LOC64755</td>
<td>LRRRC3</td>
<td>leucine rich repeat containing 31</td>
</tr>
<tr>
<td>C7orf23</td>
<td>chromosome 7 open reading frame 23</td>
<td>MANBA</td>
<td>mannosidase, beta A, lysosomal</td>
</tr>
<tr>
<td>CAPD52</td>
<td>Ca2+-dependent activator protein for secretion 2</td>
<td>MEAF6</td>
<td>MYST/Esa1-associated factor 6</td>
</tr>
<tr>
<td>CAMTA1</td>
<td>calmodulin-binding transcription activator 1</td>
<td>MMP15</td>
<td>matrix metalloproteinase 15 preproprotein</td>
</tr>
<tr>
<td>CLDN10</td>
<td>claudin 10</td>
<td>MTPAP</td>
<td>mitochondrial poly(A) polymerase</td>
</tr>
<tr>
<td>COL21A1</td>
<td>collagen, type XXI, alpha 1 precursor</td>
<td>MYST1</td>
<td>MYST histone acetyltransferase 1</td>
</tr>
<tr>
<td>CTNNBIP1</td>
<td>catenin, beta interacting protein 1</td>
<td>NAT5</td>
<td>N(alpha)-acetyltransferase 60, NatF catalytic subunit</td>
</tr>
<tr>
<td>DDAH1</td>
<td>dimethylarginine dimethylaminohydrolase 1</td>
<td>NBPFF10</td>
<td>hypothetical protein LOC440673</td>
</tr>
<tr>
<td>DDX21</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 21</td>
<td>NFYC</td>
<td>nuclear transcription factor Y, gamma</td>
</tr>
<tr>
<td>DNAJC9</td>
<td>DnaJ homolog, subfamily C, member 9</td>
<td>NIPAL3</td>
<td>NIPA-like domain containing 3</td>
</tr>
<tr>
<td>DUSP4</td>
<td>dual specificity phosphatase 4</td>
<td>PDSS1</td>
<td>prenyl diphosphate synthase, subunit 1</td>
</tr>
<tr>
<td>EFHC2</td>
<td>EF-hand domain (C-terminal) containing 2</td>
<td>PEF1</td>
<td>penta-EF-hand domain containing 1</td>
</tr>
<tr>
<td>EGFR</td>
<td>epidermal growth factor receptor</td>
<td>PER3</td>
<td>period 3</td>
</tr>
<tr>
<td>ELN</td>
<td>elastin</td>
<td>PIGV</td>
<td>phosphatidylinositol glycan class V</td>
</tr>
<tr>
<td>ENC1</td>
<td>ectodermal-neural cortex (with BTB-like domain)</td>
<td>PIK3IP1</td>
<td>HGFL protein</td>
</tr>
<tr>
<td>ENTPD4</td>
<td>ectonucleoside triphosphate diphosphohydrolase</td>
<td>PPCS</td>
<td>phosphoanthenololysylcysteine synthetase isofrom</td>
</tr>
<tr>
<td>ETV5</td>
<td>ets variant gene 5 (ets-related molecule)</td>
<td>PPFIB2P</td>
<td>PTPRF interacting protein, binding protein 2</td>
</tr>
<tr>
<td>FAAH</td>
<td>fatty acid amide hydrolase</td>
<td>PPIF</td>
<td>peptidylprolyl isomerase F precursor</td>
</tr>
<tr>
<td>FAM184A</td>
<td>family with sequence similarity 184, member A</td>
<td>PRDM4</td>
<td>PR domain containing 4</td>
</tr>
<tr>
<td>FGF13</td>
<td>fibroblast growth factor 13</td>
<td>PYROXD1</td>
<td>pyridine nucleotide-disulphide oxidoreductase domain 1</td>
</tr>
<tr>
<td>FGG</td>
<td>fibrinogen, gamma chain</td>
<td>RAPGEF5</td>
<td>Rap guanine nucleotide exchange factor (GEF) 5</td>
</tr>
<tr>
<td>FGGY</td>
<td>FGFY carboxykinase kinase domain containing</td>
<td>RFK</td>
<td>riboflavin kinase</td>
</tr>
</tbody>
</table>

doi: 10.1371/journal.pone.0078614.t004

[34]. Moreover, FGF-FGFR pathway activation has been suggested to be one mediator of resistance to EGFR inhibitors, together with, e.g., MET amplification (see [35] for review and [36]). In the current study, FGFR1 and MET amplifications were restricted to the EGFRwt/KRASwt tumor group, and were mutually exclusive (MET amplification was borderline non-significant for difference in frequency between mutation groups, p=0.09, Fisher’s exact test). Together, this could indicate presence of specific genomic circuits acting as driving forces in pathogenesis in the different mutation groups.

Taken together, the analyses of differential genomic regions point to only a few, variably sized, regions with moderate frequency differences (20-40%) between the mutation groups. These regions predominantly include regions of copy number gain with higher frequency in EGFR-mutated tumors. Several of the regions have been reported previously, but as larger and less defined regions [9, 11-13, 17, 18], while others such as 5q34-35.3, appear novel (see Table 4 for literature comparison of 34 previously reported regions from five independent studies [9, 11-13, 18]). For instance, 17 of our mGISTIC regions were present in 34 previously reported regions differing between EGFR-mutated and EGFR-wild type tumors, or KRAS-mutated tumors and KRAS-wild type tumors (Table S4). Although 24 of the 34 reported regions showed statistical significance for the original comparisons in our cohort, only 16 of these 24 regions also showed >20% frequency difference between the three groups. The absolute majority of these regions (88%) were located on chromosome 1p, 7p, and 16p (gains) and 8p (losses). Together, this emphasizes the need for adequately sized cohorts in order to draw reproducible conclusions when only moderate differences exist between investigated groups.

Few genome-wide analyses of differential allelic imbalance between EGFR-mutated, KRAS-mutated, and EGFRwt/ KRASwt tumors exist in the literature. Blons et al. reported that EGFR-mutated tumors in general display more fractional allelic loss than KRAS-mutated tumors [18], consistent with our observation of higher fractions of total LOH in EGFR-mutated tumors. Moreover, Nakamichi et al. reported that two regions, 4q13 and 4q22, differ in allelic imbalance between the mutation...
groups [37]. However, in the current study we could not verify this finding using total LOH, copy-neutral LOH or copy-neutral allelic imbalance as measurements. Instead, we found that the frequency of total LOH was strongly correlated with regions of copy number loss. In contrast, the less frequent copy-neutral LOH and copy-neutral allelic imbalance events were overall more evenly distributed across chromosomes in the mutation groups. These findings are consistent with results for lung cancer histology groups in general [21], and also recent reports from breast cancer [38,39], suggesting that LOH is predominantly caused by copy number loss in these tumor types. Moreover, the similarity in the observed tumor ploidy patterns appears consistent with that the mutation groups do not exhibit gross differences in either CNAs or allelic imbalances. However, it should be noted that the analyses of allelic imbalances in the current study are based on a smaller subset of samples (n=141), which could be a source of variability.

Several studies have reported KRAS mutant signatures or differentially expressed genes between adenocarcinomas with EGFR and/or KRAS mutations and respective wild type cases [8,10,12,14,16,26,33,40]. However, the overlap between these public signatures is very low when directly compared (Figures S3B and C). We identified 96 differentially expressed genes by supervised gene expression analyses between mutation groups across multiple discovery cohorts, of which several could be validated in independent cohorts. Reasons for the lower number of significant genes in the independent cohorts likely include smaller sample sizes and different microarray platforms compared with the discovery cohorts. The low number of differentially expressed genes between the mutation groups (only 21 genes consistently differentially expressed in all five discovery cohorts) is similar to results from other studies [8,12,14,16]. This low number of differentially expressed genes argues against that the mutational subgroups represent distinct transcriptional groups. Moreover, the overlaps between our 96 genes and previous studies [8,14,16,26,33,40] were poor (1-5% individual overlap between signatures). These results underline the need for a multicohort approach for identification of robust transcriptional differences between the mutation groups. Notably, our 96 genes mapped to a higher extent (43% of genes) to genomic regions showing differences in frequency of copy number gain or loss between the mutation groups compared to gene signatures from five reported studies [8,12,14,16,26] (6-15% of reported genes). However, the influence of the modest differences in CNA frequency (20-40%) between the mutation groups on transcriptional levels is difficult to assess. In addition to EGFR and KRAS, differentially expressed genes between the mutation groups included several other genes reported to be involved in tumorigenesis (DUSP4, RPS6KA1, ID1, TNFRSF10B, CAMTA1) [10,41], and, consistent with the enrichment of never-smokers in the EGFR-mutated patient group, genes reported as deregulated by smoking (AHR, CLDN10, FGG, GGA2, GUSB, TXNRD1) [42-45].

In supervised classification, the 96 differentially expressed genes identified EGFR-mutated adenocarcinomas with high sensitivity, but poorer specificity, while opposite results was found for EGFRwt/KRASwt tumors. Together with the results from unsupervised hierarchical clustering of multiple gene expression cohorts using different gene or probe sets these analyses demonstrate the difficulty in separating the mutation groups, especially KRAS-mutated and EGFRwt/KRASwt tumors, into more discrete transcriptional entities. Chitale et al. [10] proposed that the more distinctive expression pattern of EGFR-mutated tumors compared to KRAS-mutated tumors may depend on either a less prominent effect of KRAS mutations on expression, a biological or etiological heterogeneity among KRAS-mutated tumors, or that EGFR mutations arise in a more homogeneous and restricted cell type. Our results may be interpreted as support for potentially all three hypotheses, given the differences observed between

Figure 4. Supervised classification of adenocarcinoma gene expression cohorts with respect to EGFR and KRAS mutation status. Sensitivity (solid line) and specificity (dashed line) by EGFR/KRAS mutation group for classification of four independent validation cohorts using a 96-gene centroid classifier. The x-axis shows Pearson correlation cut-off for assigning a sample to the centroid with the highest correlation. Increasing correlation cut-offs correspond to increased stringency in classification, but introduces growing numbers of unclassified samples excluded in the calculation of sensitivity and specificity. doi: 10.1371/journal.pone.0078614.g004
and within mutation groups. Together, the results from our supervised and unsupervised gene expression analyses suggest that only modest, reproducible, transcriptional differences exist between the mutation groups. This conclusion appears consistent with the somewhat mixed inclusion of EGFR-mutated, KRAS-mutated and EGFRwt/KRASwt adenocarcinomas in different reported molecular subtypes of adenocarcinomas [16,19,20]. Although the bronchiod molecular subtype [19] has been strongly associated with EGFR-mutated tumors, this subtype also includes notable fractions of KRAS-mutated and EGFRwt/KRASwt tumors (see, e.g., [19,20]). Moreover, ~30% or more of EGFR-mutated have been classified as non-bronchiod (magnoid or squamoid) in discovery cohorts in previous studies [19,20,46]. In the absence of bronchiod classified tumors we found no significant association between the magnoid and squamoid subtypes and EGFR/KRAS mutation status in any of the five discovery cohorts in the current study (data not shown). These findings appear consistent with our unsupervised analysis showing a more distinct expression pattern of a subset of EGFR-mutated tumors across multiple cohorts, while the KRAS-mutated and EGFRwt/KRASwt groups are more intermixed (Figure S4). These results also suggest that EGFR-mutated tumors could be divided into additional subgroups, which we have recently demonstrated [46]. Taken together, EGFR and KRAS mutational status do not appear to be translated into a clearly distinctive and prominent expression signature in lung adenocarcinoma.

To further delineate the observed heterogeneous patterns of CNAs, allelic imbalances and gene expression patterns in the three mutational groups identification and/or definition of new molecular subgroups within the EGFR-mutated, KRAS-mutated and EGFRwt/KRASwt tumor groups are needed. For instance, although EGFRwt/KRASwt adenocarcinomas with ALK rearrangements are reported to display distinct expression profiles compared with ALK-negative tumors [26], it remains unclear whether this is also true for CNAs and allelic imbalances. Recent studies of lung adenocarcinoma have suggested that molecular profiling could be of value in future clinical decision making by providing clues about, e.g., treatment response to EGFR inhibitors [17,20,47,48]. For instance, Yuan et al. recently reported that clustered CNAs (copy number gains) on chromosome 7p were associated with poorer survival and less favorable response to EGFR tyrosine kinase inhibitors in EGFR-mutated adenocarcinomas specifically [17]. In support of Yuan et al., we recently identified a gene signature associated with poorer survival for patients with EGFR-mutated adenocarcinomas, where the high-risk patient group showed more copy number gains and amplifications on chromosome 7p [46]. As regions on chromosome 7p display some of the largest frequency differences between the mutational groups (~40%) these findings highlight the need for a more detailed characterization of this chromosome arm. The growing number of detected tyrosine kinase fusions in predominantly EGFRwt/KRASwt adenocarcinomas (including ALK, RET, and ROS1) are also becoming increasingly important in the therapeutic setting, as these alterations are/may become targets for specialized molecular agents. However, it remains to be investigated whether there exist similar regions and/or gene signatures associated with treatment response also for these adenocarcinoma subgroups. Clearly, further molecular stratification within the EGFR and KRAS mutation-defined lung adenocarcinoma groups has the potential to reveal new targets for synergistic treatment and provide insights into resistance mechanisms.

In summary, our multicohort analyses of genomic and transcriptional alterations demonstrate both differences and strong similarities between the EGFR and KRAS mutation defined adenocarcinoma groups. Moreover, our results suggest that the overall genomic and transcriptional landscape of adenocarcinoma is only to a minor extent affected by the mutational status of EGFR and KRAS.

Supporting Information

Figure S1. Permutation analysis of mGISTIC regions. Close support from permutation analysis for an mGISTIC region (n=90) is defined as the % of times the region was enclosed or overlapped by a permuted region based on a 75% sample subset of the 1272 samples (n=100 permutations). (A) Cumulative fraction of regions (all, gain, loss) stratified into bins of 10% close support. (B) Hexagonal binning of mGISTIC regions (all, gain, loss) for % of close support versus -log_{10}(p-value) of detected regions. A general trend of higher p-values connected to lower % close support is observed. Colors of bins indicate number of regions. Taken together, regions showing the lowest permutation detection rates also showed the lowest g-scores [28] and p-values. This is consistent with that these regions are present in only a small subset of the 1272 cases, which makes the identification and delineation of these regions to sensitive to sample composition. (PDF)

Figure S2. Pattern of CNAs, LOH, CNN-LOH, and CNN-AI in EGFR/KRAS mutation groups. Panels show in decreasing order from the top pattern (frequency) of copy number gain (red) and loss (green) relative to GAP-ploidy with mGISTIC regions identified from the 1272 sample cohort indicated by blue dots, LOH, copy-neutral LOH (CN-LOH), copy-neutral allelic imbalance (CN-AI), and variation of FGA values versus GAP-ploidy for copy number (black), CNN-LOH (blue), and CNN-AI (light blue) in the bottom panel. For the bottom panel GAP-ploidy estimates were binned in bins of size 0.3, which is represented by tick marks on the x-axis. For each bin the median FGA value of the included samples is plotted (points) for copy number, LOH, CN-LOH and CN-AI. Bins contain different numbers of samples (top axis). The 141 tumors with mutation status analyzed by GAP were stratified into (A) EGFR-mutated (n=43), (B) KRAS-mutated (n=39), and (C) EGFRwt/KRASwt (n=59) tumors. (PDF)

Figure S3. Comparison of public EGFR/KRAS signatures and classification by a set of genes differentially expressed between EGFR/KRAS mutation groups across
multiple cohorts. (A) Overall accuracy for classification of four independent adenocarcinoma cohorts using a 96-gene centroid classifier. The number of genes in the centroid matching to the different cohorts varies. The x-axis shows Pearson correlation cut-off for assigning a sample to the centroid with the highest correlation. Increasing correlation cut-offs introduces growing numbers of unclassified samples, which are excluded in calculation of accuracy. (B) Venn-diagram of the gene overlap between four reported gene lists of differentially expressed genes between $EGFR$-mutated and $EGFR$-wild type adenocarcinoma tumors. (C) Venn-diagram of the gene overlap between two reported gene lists of differentially expressed genes between $KRAS$-mutated and $KRAS$-wild type adenocarcinoma tumors, and two reported $KRAS$ mutant signatures [33,40].

Figure S4. Unsupervised analyses of four Affymetrix adenocarcinoma gene expression cohorts using different probe sets. Unsupervised hierarchical clustering was performed using Pearson correlation and complete linkage analysis. Dendrograms for each cluster tree were cut into the top two or three clusters, and the number of probe sets used in the clustering is shown for each cohort. For each cohort the distribution of $EGFR$-mutated (red), $KRAS$-mutated (blue), and $EGFR$wt/$KRAS$wt (black) tumors are shown across clusters as bars. Percentages in bar plots correspond to, e.g., how many $EGFR$-mutated tumors of the total number of $EGFR$-mutated cases that reside in a particular cluster. (A) Clustering based on probe sets from a list of 190 probe sets reported to be upregulated in $EGFR$wt/$KRAS$wt adenocarcinomas [26]. (B) Clustering based on probe sets with log2ratio standard deviation $>0.3$ across tumors in a cohort. (C) Clustering based on probe sets with log2ratio standard deviation $>0.5$ across tumors in a cohort. (D) Clustering based on probe sets with log2ratio standard deviation $>1$ across tumors in a cohort. (E) Clustering based on matching genes from the list of top 250 genes reported by Singh et al. [33]. Each dendrogram is cut into the top two clusters. Division of dendrograms into three groups did not identify $KRAS$-mutants as a single group without notable inclusion of $EGFR$wt/$KRAS$wt tumors in all cohorts.

File S1. Document with details concerning used analysis methods.

(DOC)

Table S1. Genomic mGISTIC regions identified from analysis of 1272 lung adenocarcinomas.

(XLSX)

Table S2. Differential regions of copy number gain and loss between mutation groups obtained from genome-wide analysis.

(XLSX)

Table S3. Differentially expressed genes between mutation groups across at least four gene expression cohorts.

(XLSX)

Table S4. Analysis of genomic regions reported in the literature to stratify mutation groups in the current cohort.

(DOC)

Author Contributions

Conceived and designed the experiments: JS MP SI. Performed the experiments: KE. Analyzed the data: JS. Contributed reagents/materials/analysis tools: JS. Wrote the manuscript: JS MP. Contributed patient data: KE PM JB.

References

(2008) An oncogenic KRAS2 expression signature identified by cross-
expression profiling reveals reproducible human lung adenocarcinoma
subtypes in multiple independent patient cohorts. J Clin Oncol 24:

(2011) GISTIC2.0 facilitates sensitive and confident localization of the
cancer genomic profiles obtained by SNP arrays. Genome Biol 10:

(2008) A comparative and integrative approach identifies ATPase family, AAA
domain containing 2 as a likely driver of cell proliferation in lung

(2005) Activation of the FGF2-GFGR1 Autocrine Pathway: A Novel Mechanism
of Acquired Resistance to Gefitinib in NSCLC. Mol Cancer Res 11:

(2009) Whole genome comparison of allelic imbalance between
EGFR-mutated and EGFR/KRAS-wild type lung adenocarcinoma
patients with dismal prognosis. J Clin Oncol 27: 2793-2799. doi:

(2012) Identification of genes upregulated in ALK-positive and EGFR/KRAS/
ALK-negative lung adenocarcinomas. Cancer Res 72: 100-111. doi:

(2011) Identification of transcriptional subgroups in EGFR-mutated and EGFR/KRAS-wild type
lung adenocarcinoma reveals gene signatures associated with patient

(2005) Epithelial versus mesenchymal phenotype determines in vitro
sensitivity and predicts clinical activity of erlotinib in lung cancer
patients. Clin Cancer Res 11: 8866-8869. doi:

(2013) An epithelial-mesenchymal transition gene signature predicts resistance to
EGFR and PI3K inhibitors and identifies Axi as a therapeutic target for

(2005) Homozygous deletions and chromosome amplifications in human lung
carcinomas revealed by single nucleotide polymorphism array analysis.