Research Article in Cancer Research

Development of Lung Adenocarcinomas with Exclusive Dependence on Oncogene Fusions

Motonobu Saito¹², Yoko Shimada¹, Kouya Shiraishi¹, Hiromi Sakamoto³, Koji Tsuta⁴, Hirohiko Totsuka⁵, Suenori Chiku⁶, Hitoshi Ichikawa³, Mamoru Kato⁷, Shun-ichi Watanabe⁸, Teruhiko Yoshida³, Jun Yokota¹⁹, Takashi Kohno¹

¹Division of Genome Biology, ³Division of Genetics and ⁷Department of Bioinformatics, National Cancer Center Research Institute, Tokyo, Japan
²Department of Organ Regulatory Surgery, Fukushima Medical University School of Medicine, Fukushima, Japan
⁴Division of Pathology and Clinical Laboratories and ⁸Division of Thoracic Surgery, National Cancer Center Hospital, Tokyo, Japan
⁵Bioinfomatics Group, Research and Development Center, Solution Division 4, Hitachi Government and Public Corporation System Engineering Ltd, Tokyo, Japan
⁶Science Solutions Division, Mizuho Information and Research Institute Inc., Tokyo, Japan
⁹Cancer Genome Biology, Institute of Predictive and Personalized Medicine of Cancer, Barcelona, Spain

Running title: Oncogenic fusion-positive lung adenocarcinoma development
Keywords: lung adenocarcinoma, oncogenic fusion, oncogene mutation, oncogene amplification, tyrosine kinase inhibitor.

Financial support: This work was supported by the Advanced Research for Medical Products Mining Program of the National Institute of Biomedical Innovation (NIBIO); Grants-in-Aid from the Japan Agency for Medical Research and Development (AMED) for the Health and Labor Sciences Research Expenses for Commission (the Practical Research for Innovative Cancer Control: H26-practical-general-007); a Grant-in-Aid from the Japanese Society for the Promotion of Science (JSPS) and Scientific Research (B), and the National Cancer Center Research and Development Fund (26A-1).

Corresponding author: Takashi Kohno, Ph.D., Division of Genome Biology, National Cancer Center Research Institute, 1-1 Tsukiji 5-chome, Chuo-ku, Tokyo 104-0045, Japan.
Tel: +81-3-3542-2511; Fax: +81-3-3542-0807; Email: tkkohn@ncc.go.jp

Disclosure of potential conflicts of interest: The authors have no potential conflicts of interest to disclose.

Word of text (<5000): 3515

Total number of figures and tables (<7): 5

References (<50): 47
ABSTRACT (<250 words)

This report delivers a comprehensive genetic alteration profile of lung adenocarcinomas (LADCs) driven by ALK, RET and ROS1 oncogene fusions. These tumors are difficult to study because of their rarity: each drives only a low percentage of LADCs. Whole exome sequencing and copy number variation analyses were performed on a Japanese LADC cohort (n = 200) enriched in patients with fusions (n = 31, 15.5%), followed by deep re-sequencing for validation. The driver fusion cases showed a distinct profile with smaller numbers of non-synonymous mutations in cancer-related genes or truncating mutations in SWI/SNF chromatin remodeling complex genes than in other LADCs (P < 0.0001). This lower mutation rate was independent of age, gender, smoking status, pathological stage, and tumor differentiation (P < 0.0001) and was validated in nine fusion-positive cases from a US LADCs cohort (n = 230). In conclusion, our findings indicate that LADCs with ALK, RET, and ROS1 fusions develop exclusively via their dependence on these oncogene fusions. The presence of such few alterations beyond the fusions supports the use of monotherapy with tyrosine kinase inhibitors targeting the fusion products in fusion-positive LADCs. (185 words)
INTRODUCTION

Lung adenocarcinoma (LADC) is the most frequent histological type of lung cancer and its incidence is rising in Asian and Western countries. Oncogenic fusions of the protein tyrosine kinase genes \textit{ALK}, \textit{RET} and \textit{ROS1}, identified by us and others, are believed to drive the development of a subset (3–4\%, 1–2\% and 1–2\%, respectively) of LADCs (1-4). These fusion-positive LADCs often, but not always, show mucinous-cribriform patterns (2, 5-7). In addition, fusion-positive LADCs tend to occur in young and non/light-smoking individuals (2, 4, 8-10) and show a high therapeutic response to tyrosine kinase inhibitors (TKIs) that suppress the kinase activity of the fusion products (11-13). These results indicate that fusion-positive LADCs are a distinct LADC molecular entity. Despite recent large-scale genome sequencing studies in LADCs (14-16), the genetic profile of fusion-positive LADCs remains unknown due to the rarity of these tumors. Better genetic characterization of fusion-positive LADCs is required to improve therapeutic strategies. If other genetic abnormalities are detected, agents targeting these defects could be used in combination with TKIs to improve efficacy and outcome (11, 17, 18).

LADCs carrying \textit{ALK}, \textit{ROS1} and \textit{RET} fusions have already been shown to lack activating mutations in other oncogenes, such as \textit{EGFR}, \textit{KRAS}, \textit{BRAF} and \textit{HER2/ERBB2} (2-4, 14-16); however, the mutational status of other genes frequently mutated in lung and other cancers, including those identified in the cancer gene census (CGC) (19) or those identified as significantly mutated genes (SMGs) in 12 common cancers (20), is unknown. These gene sets include tumor suppressor genes, such as \textit{TP53}, \textit{CDKN2A}, \textit{KEAP1} and \textit{STK11/LKB1}, and chromatin remodeling/modifying genes, such as \textit{ARID1A} and \textit{SMARCA4},
which are the targets of genetic loss-of-function aberrations in cancer cells (20). Notably, recent studies suggested that these deleterious aberrations are therapeutically targetable; drugs restoring the function of mutant p53 proteins are being developed (21, 22), and synthetic lethality-based therapies have been considered by us and others to treat cancers with TP53, LKB1, ARID1A and SMARCA4 deficiencies (23-28).

We performed the comparative genetic aberration profiling of oncogenic fusion-positive and -negative LADCs. Two hundred cases of snap-frozen surgical LADC tissues were subjected to whole exome sequencing using a next generation sequencer and to copy number variation analysis using a DNA chip. The selected cases were enriched in oncogenic ALK, RET or ROS1 fusions (n = 31, 15.5%) and included 96 cases (48.0%) of activating mutations in EGFR, KRAS, HER2, BRAF and HRAS oncogenes and 73 cases (35.5%) without any such aberrations. The study revealed that fusion-positive LADCs have a unique genetic profile that includes fewer genetic aberrations than other LADCs.

PATIENTS AND METHODS

Patients

A total of 200 LADC cases (the study cohort) were selected from 608 consecutive cases (the NCC cohort) (NCC Biobank, Tokyo, Japan) who underwent surgical resection between 1997 and 2008 at the National Cancer Center Hospital, Tokyo, and for whom snap-frozen cancerous and non-cancerous lung tissues were available (Fig. 1A-B). All of the 608 cases were screened for EGFR, KRAS, BRAF and HER2 hot spot mutations by the HRM method, and for EML4- and KIF5B-ALK, KIF5B- and CCDC6-RET, and CD74-, EZR- and
SLC34A2-ROS1 fusions by RT-PCR, as described (4, 7) (Supplementary Table S2). In addition, a case with a novel type of RET fusion, KIAA1468-RET, which was detected by whole RNA sequencing, was included in this cohort.

Driver fusion study subjects (n = 31), i.e., those with ALK (n = 11), RET (n = 11) or ROS1 (n = 9) fusions were selected from all 50 fusion-positive cases in the NCC original cohort, i.e., those with ALK fusions (n = 23), RET (n = 13) or ROS1 (n = 14) fusions based on the criterion that sufficient amounts of genomic DNA for whole exome sequencing were available (Fig. 1A). EGFR-positive (n = 72), other driver mutation (n = 23), and pan-negative cases (n = 74), with sufficient amounts of genomic DNA, were randomly selected from the NCC original cohort together with EGFR-positive (n = 282), other driver mutation (n = 79), and pan-negative cases (n = 197) to obtain a EGFR-positive case:pan-negative case ratio of approximately 1:1 and to make the total number of samples 200. The exome sequencing analysis revealed an activating HRAS mutation (Q61L), so this case was classified as driver mutation. Thus, the study cohort included 73 pan-negative cases and a HRAS mutation-positive case. The selection resulted in a cohort that was more enriched in driver fusion cases than the original cohort (Fig. 1B). The study subjects were diagnosed according to the 7th TNM classification of malignant tumors (29, 30). The study was approved by the Institutional Review Boards of the NCC.

**Genome copy analysis and tumor content estimation**

Genome copy number and allelic status were assessed in all 200 study cases by Illumina OMNI 2.5M array analysis using both cancerous and non-cancerous lung DNA. Tumor cell
content in each tumor sample and copy numbers for each gene were deduced using the Global Parameter Hidden Markov Model (GPHMM) method (31).

**Exome sequencing**

Exome sequencing was conducted from 2.5 μg of cancerous or non-cancerous DNA isolated from snap-frozen tissues. Exome capture was performed using the Agilent SureSelect Human All Exon 50-Mb, V4 or V5 according to the manufacturer’s instructions. Exome sequencing was performed on the Illumina HiSeq 2000 platform using 75 bp paired-end reads (Illumina). Basic alignment and sequence quality control were conducted using the Picard and Firehose pipelines. The reads were aligned against the reference human genome from UCSC human genome 19 (Hg19) using the Burrows Wheeler Aligner Multi-Vision software package. Since duplicate reads were generated during the PCR amplification process, paired-end reads that aligned to the same genomic positions were removed using SAMtools.

Somatic SNVs were called by the MuTect program, which applies a Bayesian classifier to allow the detection of somatic mutations with a low allele frequency (32). Somatic InDel mutations were called by the GATK Somatic Indel Detector (33). SNV and InDel detection was supported by visual examination using the Integrative Genomics Viewer software (34).

**Verification of somatic mutations by deep re-sequencing**
Mutations in all coding exons of the following 28 genes were examined by targeted genome capture and massively parallel sequencing using an Illumina HiSeq 2000 system and the Haloplex Custom Enrichment Kit (Agilent Technologies): ten representative cancer census genes (19), AKT1, APC, CTNNB1, KEAP1, MAP2K1, MET, NRAS, PIK3CA, STK11 and TP53, and 18 SWI/SNF chromatin remodeling genes (35) whose mutations were detected in one or more tumors by exome sequencing, ACTL6B, ARID1A, ARID1B, ARID2, BPTF, DPF1, EP400, HLTF, PBRM1, RAD54L2, SHPRH, SMARCA2, SMARCA4, SMARCA1, SMARCAL1, SMARCB1, SMARCC1 and SRCAP. Average read depths were approximately 1,000.

**Cancer Gene Census**

A list of somatic mutations from the CGC was downloaded from the most recently released COSMIC V70 (36).

**Validation in a U.S. cohort**

Validation analysis was performed using TCGA LADC study data (16). The TCGA cohort of 230 cases was selected from 678 patients with previously untreated lung adenocarcinoma based on tumor percentage, availability of clinical data, and availability of sufficient amounts of nucleic acid (37). Driver gene fusions and mutations were evaluated in all cases; thus, the 230 cases were subgrouped according to driver gene type: nine driver fusion cases (3.9%) with ALK (n = 3), RET (n = 2), and ROS1 (n = 2) fusions, 121 driver mutation cases (52.6%) with hot spot mutations in EGFR (n = 26), KRAS (n = 74), HER2 (n
= 4), \textit{BRAF} (n = 16) and \textit{HRAS} (n = 1), and 100 pan-negative cases (43.5%) (Supplementary Fig. S4A). Among the cancer-related and SWI/SNF chromatin remodeling genes investigated in Figure 2, information on \textit{CDKN2A}, \textit{RBM10}, \textit{RB1}, \textit{NF1}, \textit{KEAP1}, \textit{MET}, \textit{MGA}, \textit{U2AF1}, \textit{PIK3CA}, \textit{STK11}, \textit{TP53}, \textit{SMARCA4} and \textit{ARID1A} was available and therefore was used in the analysis.

\textit{Statistics}

Statistical analyses of differences in genetic alterations, clinical and pathological factors between the driver aberration groups, or smoking status were assessed by using the Kruskal-Wallis test, two-sided Mann-Whitney test, two-sided Fisher’s exact test or Chi-square test, and two-sided Spearman r test in GraphPad Prism 5 software (GraphPad Software). Multivariate regression analysis including the number of non-synonymous mutations per Mb and clinicopathological factors (age, gender, smoking status, pathological stage, and tumor differentiation) was conducted using JMP 10 software (SAS Institute). \( P < 0.05 \) was considered significant.

\textbf{RESULTS}

\textit{Study cohort}

The study cohort of 200 LADC cases included 31 cases (15.5\%) with \textit{ALK}, \textit{RET} or \textit{ROS1} fusions, 96 cases (48.0\%) with hot spot mutations in \textit{EGFR}, \textit{KRAS}, \textit{HER2}, \textit{BRAF} or \textit{HRAS}, and 73 cases (36.5\%) without any of these driver gene aberrations (Table 1). The study cohort samples were selected from the original National Cancer Center (NCC) cohort
samples (n = 608) to enrich fusion-positive cases (Fig. 1A-B and Supplementary Table S1). The driver gene aberrations detected were mutually exclusive, as predicted (14, 16, 38). There was no case of NTRK1 fusion, a recently reported oncogene fusion (39), in this cohort. The 31, 96 and 73 cases were classified into three groups designated the “driver fusion”, “driver mutation” and “pan-negative” groups, respectively (Fig. 1A-B).

Patient characteristics are summarized in Table 1. The driver fusion cases included a significantly higher frequency of young, female and never-smoker cases, and the pan-negative cases included a significantly higher frequency of older, male and heavy smokers (P = 0.024 by Kruskal-Wallis test, 0.001 by Chi-square test, and < 0.0001 by Chi-square test, respectively) than the other two groups, representing the characteristics of the original NCC cohort (Supplementary Table S2). In addition, the pan-negative cases showed a significant predominance of poor differentiation (P = 0.0007 by Chi-square test). These features are consistent with previous reports (2, 4-6, 8-10) indicating the authenticity and suitability of our study cohort to establish the genetic profile of LADCs driven by specific aberrations.

**Genome-wide mutation profiling**

Genomic DNA from cancerous and non-cancerous lung tissues was subjected to whole exome sequencing and SNP chip analyses. The average sequencing depth of the driver fusion, driver mutation and pan-negative groups was similar; median: 106 (range: 84–218), 98 (82–216) and 104 (82–145), respectively (P ≥ 0.05 by Kruskal-Wallis test) (Supplementary Fig. S1A). The tumor contents deduced from genome-wide copy number
and allelic imbalance data obtained by SNP chip analysis were also similar \( (P \geq 0.05 \text{ by Kruskal-Wallis test}) \) (Supplementary Fig. S1B). Therefore, having ruled out differential sensitivity between the groups, the sample set was found suitable to compare the genetic aberration profiles of the three groups. In addition, the SMGs deduced by the MutSigCV analysis were consistent with recent large-scale sequencing studies (14-16), supporting the authenticity of the present sample set (Supplementary Table S3).

**Infrequent gene mutations in fusion-positive LADCs**

Non-synonymous mutations (missense, nonsense, indel and splicing site mutations) based on driver genes are depicted in Fig. 2, focusing on CGC genes, SMGs, and SWI/SNF chromatin remodeling complex genes (20, 40). The driver fusion cases appeared to harbor fewer mutated genes. In fact, the median number of non-synonymous mutations per Mb was lowest in driver fusion cases (0.37, range: 0–1.5) compared to driver mutation cases (0.65, 0–6.0) and pan-negative cases (0.87, 0–24.8) \( (P < 0.0001 \text{ by Kruskal-Wallis test}) \) (Fig. 3A and Supplementary Fig. S2A). The median number of mutated CGC genes was smaller in driver fusion cases (1.0, range: 0–4) than in driver mutation cases (3.0, range: 1–12) and pan-negative cases (3.0, range: 0–46) \( (P < 0.0001 \text{ by Kruskal-Wallis test}) \). The median number of mutated SMGs was also smaller in driver fusion cases (1.0, range: 0–4) than in driver mutation cases (3.0, range: 1–7) and pan-negative cases (2.0, range: 0–25) \( (P < 0.0001 \text{ by Kruskal-Wallis test}) \). A positive correlation was observed between the number of non-synonymous mutations in all genes and that in CGC genes or SMGs \( (P < 0.0001 \text{ by Spearman’s test}) \) (Supplementary Fig. S2B). Driver fusion was associated with fewer
mutations regardless of age, gender, smoking status, pathological stage or tumor differentiation \((P < 0.0001\) by multivariate regression model analysis) (Supplementary Table S4). Thus, driver fusion-positive LADCs seem to develop from the accumulation of a significantly smaller number of gene mutations than other types of LADCs.

**Infrequent mutation of lung cancer-related genes in fusion-positive LADCs**

Mutations in several genes known to contribute to lung carcinogenesis were examined. The \(TP53\) gene is a representative tumor suppressor gene, included both in the CGC genes and SMGs, whose mutation occurs during the progression of LADCs following \(EGFR\) and \(KRAS\) mutations (41). \(TP53\) was the most frequently mutated gene in our study cohort (Table 2) and in the US cases (20), with a mutation frequency of \(41/96\) (42.7\%) in the driver mutation cases and \(29/73\) (39.7\%) in the pan-negative cases. Notably, \(TP53\) mutations were significantly less frequent in the driver fusion group (5/31 or 16.1\%, \(P = 0.026\) by Chi-square test) than in the other two groups (Fig. 3B and Supplementary Fig. S3A). Truncating mutations in SWI/SNF chromatin remodeling complex genes, which are frequently observed in a variety of cancers (40), were not detected either (Fig. 2 and 3B, and Supplementary Fig. S3B). The SMGs included 20 cellular process genes, most of which were less frequently mutated in fusion-positive cases than in mutation-positive and pan-negative cases (Supplementary Table S5). Two WNT signaling genes, \(APC\) and \(CTNNB1/\beta\)-catenin, and a PI(3)K signaling gene, \(PIK3CA\), which activates PI(3)K signaling, are representative signaling genes known to be mutated in a small subset of
LADCs (14, 16). Notably, these three genes were mutated irrespective of driver gene status (Fig. 3B, Table 2, Supplementary Table S5 and Supplementary Fig. S3C-D).

All 31 driver fusion cases and five driver mutation cases were subjected to deep re-sequencing of 28 genes with a mean depth of 1,000, using the same DNA samples. The 28 genes included ten cancer-related genes and 18 SWI/SNF chromatin remodeling genes. All 19 non-synonymous mutations detected by exome sequencing in these 36 cases were confirmed. In addition, an ARID1A missense mutation, which was not detected due to low depth by exome sequencing (i.e., no mutant reads among 27 reads), was detected in a driver fusion case. This re-sequencing verified that LADCs with oncogenic fusions develop through a pathway that involves only a small number of gene mutations.

Similar findings were also observed in a LADC cohort from the USA consisting of 230 cases (16). This cohort included nine fusion-positive cases consisting of three ALK fusions, two RET fusions and four ROS1 fusions. Mutation distributions according to driver gene status are summarized in Supplementary Table S6. In this cohort, driver fusion cases carried fewer non-synonymous mutations ($P = 0.046$ by Kruskal-Wallis test) and cancer-related gene mutations, including TP53 ($P = 0.003$ by Chi-square test) mutations, than other cases (Supplementary Fig. S4A-C).

**Copy number gain in oncogenes**

Genomic copy number gains in nine oncogenes recently defined as amplified in histological types of lung cancer (15) were examined to assess whether genomic aberrations other than mutations contribute to the development of fusion-positive LADCs (Supplementary Table
According to DNA chip analysis, these aberrations were infrequent regardless of the driver fusion.

Frequent gene mutations in pan-negative LADCs

The mutational characteristics of pan-negative cases were also investigated. These cases were associated with higher proportion of individuals of male gender, smokers, and individuals with poorly differentiated tumors (Table 1), and had the highest median number of non-synonymous mutations per Mb among the three groups (Fig. 3A and Supplementary Fig. S2A). Furthermore, they had more frequent mutations in CGC genes ($P < 0.0001$ by Kruskal-Wallis test) (Fig. 3A) than the other cases, and most of the SMGs were more highly mutated in these cases than in the other groups (Supplementary Table S5). Among 127 SMGs, TSHZ3, SETBP1, EPHA3 and NAV3 were preferentially mutated in pan-negative cases (Table 2). In addition, truncating mutations in SWI/SNF chromatin remodeling genes, such as ARID1A and SMARCA4/BRG1, which have been reported to be mutated in LADC (14, 16), and PBRM, another SWI/SNF gene frequently mutated in renal cell carcinoma (ccRCC) (42), were significantly predominant in pan-negative cases ($P < 0.0001$ by Kruskal-Wallis test) (Fig. 3B, Supplementary Fig. S3B, and Supplementary Table S5). On the other hand, KEAP1, NF1 and RIT1 mutations, which were previously shown to occur frequently in pan-negative tumors in the TCGA cohort (16), were not frequent in our cases (Supplementary Table S6). Therefore, there might be a difference in the carcinogenic pathways of pan-negative cases between Asians and Europeans/Americans.
Some pan-negative cases did not show mutations in cancer related genes (Fig. 2). A small number of mutations were associated with cases with low tumor content, suggesting that the failure to detect mutations in some pan-negative cases was due to the low mutation detection power in tumors with low purity (Supplementary Fig. S5). To further address this point, we subjected 29 pan-negative cases, in which few or no cancer-related gene mutations were detected, to deep re-sequencing, in the same way as we did for fusion-positive cases. In addition to the non-synonymous mutations already detected by exome sequencing, we additionally detected several mutations, including those with a lower mutation allele frequency than expected from tumor content (Supplementary Table S8). Therefore, some of the pan-negative tumors most likely had intra-tumor heterogeneity that hampered the detection of mutations by exome sequencing.

DISCUSSION
This study compared gene aberrations based on driver gene status in a cohort enriched in LADCs with oncogenic ALK, RET and ROS1 fusions. The fusion-positive cases showed significantly fewer mutations than the other cases in all genes and in known cancer-related genes represented by the CGC genes and SMGs. The lower mutation rate was independent of age, gender, smoking status, pathological stage, and tumor differentiation and was validated in a LADC cohort consisting of 230 selected US patients (16) (Supplementary Fig. S4). The rate of non-synonymous mutation in driver fusion LADCs was similar to that in ovarian, breast, brain, kidney and hematopoietic tumors, for which mutation rates are low (20, 43, 44). Notably, fusion-positive LADCs had a lower frequency of C>A
transversion, which is predominant in tumors of ever-smokers and is a signature of cigarette
smoke exposure, than others (20, 43, 44) (Supplementary Fig S6A). Thus, LADCs with
oncogene fusions develop through a distinct pathway that includes fewer gene aberrations
than other LADCs. Interestingly, similar differential genomic profiles were observed in
PAX fusion-positive and -negative rhabdomyosarcomas (45). This might be due to the
strong ability of oncogene fusions to drive carcinogenesis; therefore, tumor cells with such
fusions might not need many other genetic alterations to progress. The fact that LADC
patients with oncogene fusions are younger than those without fusions in this and other
populations (Table 1) might also reflect the oncogenic robustness of these fusions (2, 8-10,
38).

By contrast, pan-negative LADCs appear to develop from a large number of gene
mutations, including mutations in TP53 and other cancer-related genes. In addition,
pan-negative cases carry truncating mutations in several SWI/SNF chromatin remodeling
genes more frequently than other cases. Notably, truncating mutations in PBRM1 were
detected only in pan-negative cases. Truncating mutations in PBRM1 and other SWI/SNF
genes are frequent in ccRCC (42); therefore, some pan-negative LADC and ccRCC might
develop through a common carcinogenic pathway. Smoking is linked to high numbers of
mutations (Supplementary Fig. S6B and C), suggesting that the large numbers of gene
mutations caused by exposure to tobacco carcinogens can lead to tumor development
without any known driver oncogene alteration. In addition, a subset of pan-negative tumors
were found to have high intra-tumor heterogeneity that hampered mutation detection by the
exome sequencing method. Thus, the mutation profiles of pan-negative tumors require
further investigation by more sensitive methods, such as deep-whole exome sequencing, to understand how these tumors develop.

The present study has implications for therapeutic approaches to fusion-positive LADCs. Personalized therapy targeting fusion products with TKIs has become the first-line therapeutic method in advanced and/or recurrent tumors (12, 46, 47). This study indicates that targeting fusion products is the best approach since only a small number of mutations have occurred in other genes. Notably, driver fusion cases (and also driver mutation cases) lack SWI/SNF chromatin remodeling gene aberrations. Recent studies demonstrated that deficiencies in *SMARCA4* and *ARID1A* make cancer cells treatable by inhibiting the activity of their paralogs, *SMARCA2* and *ARID1B*, based on synthetic lethality (23, 24, 28); however, LADC patients with driver fusions will not benefit from such therapies because their tumors have a low frequency of aberrations in *SMARCA4* and *ARID1A* genes. Notably, a small subset of fusion-positive LADCs carry gene mutations in genes, such as *PIK3CA*, *APC* and *CTNNB1*, that could affect signal transduction, although how prevalent or specific these mutations are to fusion-positive LADCs remains unclear due to a small number of study subjects. This issue should be further investigated to unravel the carcinogenic pathways that lead to fusion-positive LADC and find efficient therapeutic targets for this type of cancer.
Author contributions

**Conception and design:** Motonobu Saito, Takashi Kohno

**Administrative support:** Teruhiko Yoshida

**Provision of study materials or patients:** Kouya Shiraishi, Koji Tsuta, Shun-ichi Watanabe

**Collection and assembly of data:** Motonobu Saito, Kouya Shiraishi, Yoko Shimada, Hiromi Sakamoto, Teruhiko Yoshida, Hitoshi Ichikawa, Shun-ichi Watanabe

**Data analysis and interpretation:** Motonobu Saito, Kouya Shiraishi, Suenori Chiku, Hirohiko Totsuka

**Manuscript writing:** Motonobu Saito, Takashi Kohno

**Final approval of manuscript:** All authors

Acknowledgments

We thank Sumiko Ohnami, Yoko Odaka and Misuzu Okuyama for technical assistance.
REFERENCES


Table 1. Clinical and pathological characteristics of the 200 lung adenocarcinomas of the study cohort

<table>
<thead>
<tr>
<th>Variable</th>
<th>Driver fusion</th>
<th>Driver mutation</th>
<th>Pan-negative</th>
<th>P</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All</td>
<td>Total ALK RET</td>
<td>ROS1</td>
<td>Total EGFR KRAS HÉR2 BRAF HRAS</td>
</tr>
<tr>
<td>Cases-no. (%)</td>
<td>200</td>
<td>31 (15.5)</td>
<td>11 11 9</td>
<td>96 (48.0) 72 17 4 2 1</td>
</tr>
<tr>
<td>Age (y)</td>
<td></td>
<td>58.8</td>
<td>57 52 57 59</td>
<td>60 59 62 59 57 65</td>
</tr>
<tr>
<td>Range</td>
<td>28-82</td>
<td>28-78 30-68</td>
<td>28-78 40-68</td>
<td>31-82 31-69 49-82 50-67 53-61</td>
</tr>
<tr>
<td>Gender-no. (%)</td>
<td></td>
<td>0.024</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>108 (54.0)</td>
<td>11 (35.5)</td>
<td>4 6 1</td>
<td>46 (47.9) 28 12 3 2 1</td>
</tr>
<tr>
<td>Female</td>
<td>92 (46.0)</td>
<td>20 (64.5)</td>
<td>7 5 8</td>
<td>50 (52.1) 44 5 1 0 0</td>
</tr>
<tr>
<td>Smoking (pack year)-no. (%)</td>
<td></td>
<td>&lt;0.0001</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Never</td>
<td>88 (44.0)</td>
<td>22 (71.0)</td>
<td>5 9 8</td>
<td>46 (47.9) 38 6 2 0 0</td>
</tr>
<tr>
<td>&lt;20</td>
<td>26 (13.0)</td>
<td>3 (9.7)</td>
<td>3 0 0</td>
<td>18 (18.8) 18 0 0 0 0</td>
</tr>
<tr>
<td>≥20</td>
<td>86 (43.0)</td>
<td>6 (19.3)</td>
<td>3 2 1</td>
<td>32 (33.3) 16 11 2 2 1</td>
</tr>
<tr>
<td>TNM stage-no. (%)</td>
<td></td>
<td>0.009</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I</td>
<td>76 (38.0)</td>
<td>14 (45.2)</td>
<td>2 8 4</td>
<td>32 (33.3) 24 5 2 1 0</td>
</tr>
<tr>
<td>II</td>
<td>59 (29.5)</td>
<td>9 (29.0)</td>
<td>6 2 1</td>
<td>21 (21.9) 12 8 0 0 1</td>
</tr>
<tr>
<td>III</td>
<td>56 (28.0)</td>
<td>8 (25.8)</td>
<td>3 1 4</td>
<td>35 (36.5) 29 3 2 1 0</td>
</tr>
<tr>
<td>IV</td>
<td>9 (4.5)</td>
<td>0 0 0 0 0 0 8</td>
<td>8 (8.3)</td>
<td>7 1 0 0 0 0 1 (1.4)</td>
</tr>
<tr>
<td>Differentiation-no. (%)</td>
<td></td>
<td>0.0007</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Well</td>
<td>90 (45.0)</td>
<td>14 (45.2)</td>
<td>2 7 5</td>
<td>46 (47.9) 34 11 1 0 0</td>
</tr>
<tr>
<td>Mod</td>
<td>74 (37.0)</td>
<td>14 (45.2)</td>
<td>7 4 3</td>
<td>41 (42.7) 32 4 3 1 1</td>
</tr>
<tr>
<td>Poor</td>
<td>34 (17.0)</td>
<td>2 (6.5)</td>
<td>1 0 1</td>
<td>8 (8.3) 5 2 0 1 0</td>
</tr>
<tr>
<td>Unknown</td>
<td>2 (1.0)</td>
<td>1 (3.1)</td>
<td>1 0 0</td>
<td>1 (1.0) 1 0 0 0 0</td>
</tr>
</tbody>
</table>

*P* value was derived from the comparison between driver fusion, driver mutation, and pan-negative cases by Kruskal-Wallis test or two-sided Chi-square test, where appropriate.
<table>
<thead>
<tr>
<th>Genes-no. ( % )</th>
<th>All cases ( n=200 )</th>
<th>Fusion ( n=31 )</th>
<th>Mutation ( n=96 )</th>
<th>Pan-negative ( n=73 )</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TP53</strong></td>
<td>75 (37.5)</td>
<td>5 (16.1)</td>
<td>41 (42.7)</td>
<td>29 (39.7)</td>
</tr>
<tr>
<td><strong>EGFR</strong></td>
<td>72 (36.0)</td>
<td>0</td>
<td>72 (75.0)</td>
<td>0</td>
</tr>
<tr>
<td><strong>KRAS</strong></td>
<td>17 (8.5)</td>
<td>0</td>
<td>17 (17.7)</td>
<td>0</td>
</tr>
<tr>
<td><strong>NAV3</strong></td>
<td>14 (7.0)</td>
<td>0</td>
<td>5 (5.2)</td>
<td>9 (12.3)</td>
</tr>
<tr>
<td><strong>APC</strong></td>
<td>13 (6.5)</td>
<td>2 (6.5)</td>
<td>8 (8.3)</td>
<td>1 (1.4)</td>
</tr>
<tr>
<td><strong>CDKN2A</strong></td>
<td>13 (6.5)</td>
<td>0</td>
<td>3 (3.1)</td>
<td>10 (13.7)</td>
</tr>
<tr>
<td><strong>STK11</strong></td>
<td>12 (6.0)</td>
<td>0</td>
<td>4 (4.2)</td>
<td>8 (11.0)</td>
</tr>
<tr>
<td><strong>CTNNB1</strong></td>
<td>11 (5.5)</td>
<td>2 (6.5)</td>
<td>8 (8.3)</td>
<td>1 (1.4)</td>
</tr>
<tr>
<td><strong>EPHA3</strong></td>
<td>11 (5.5)</td>
<td>0</td>
<td>3 (3.1)</td>
<td>8 (11.0)</td>
</tr>
<tr>
<td><strong>TSHZ2</strong></td>
<td>11 (5.5)</td>
<td>0</td>
<td>5 (5.2)</td>
<td>6 (8.2)</td>
</tr>
<tr>
<td><strong>TSHZ3</strong></td>
<td>11 (5.5)</td>
<td>0</td>
<td>2 (2.1)</td>
<td>9 (12.3)</td>
</tr>
<tr>
<td><strong>MLL2</strong></td>
<td>10 (5.0)</td>
<td>2 (6.5)</td>
<td>2 (2.1)</td>
<td>6 (8.2)</td>
</tr>
<tr>
<td><strong>SETBP1</strong></td>
<td>10 (5.0)</td>
<td>0</td>
<td>3 (3.1)</td>
<td>7 (9.6)</td>
</tr>
<tr>
<td><strong>MLL3</strong></td>
<td>9 (4.5)</td>
<td>0</td>
<td>4 (4.2)</td>
<td>5 (6.8)</td>
</tr>
<tr>
<td><strong>SETD2</strong></td>
<td>9 (4.5)</td>
<td>3 (9.7)</td>
<td>2 (2.1)</td>
<td>4 (5.5)</td>
</tr>
<tr>
<td><strong>ARID1A</strong></td>
<td>8 (4.0)</td>
<td>0</td>
<td>0</td>
<td>8 (11.0)</td>
</tr>
<tr>
<td><strong>ATM</strong></td>
<td>8 (4.0)</td>
<td>0</td>
<td>2 (2.1)</td>
<td>6 (8.2)</td>
</tr>
<tr>
<td><strong>ATRX</strong></td>
<td>8 (4.0)</td>
<td>0</td>
<td>3 (3.1)</td>
<td>5 (6.8)</td>
</tr>
<tr>
<td><strong>EPPK1</strong></td>
<td>8 (4.0)</td>
<td>1 (3.2)</td>
<td>0</td>
<td>7 (9.6)</td>
</tr>
<tr>
<td><strong>PDGFRA</strong></td>
<td>8 (4.0)</td>
<td>0</td>
<td>5 (5.2)</td>
<td>3 (4.1)</td>
</tr>
<tr>
<td><strong>POLQ</strong></td>
<td>8 (4.0)</td>
<td>0</td>
<td>3 (3.1)</td>
<td>5 (6.8)</td>
</tr>
</tbody>
</table>
Figure legends

Fig. 1. Patient selection. A. Two hundred surgically resected LADC cases (the study cohort) were selected from 608 consecutive cases (the NCC cohort). Patients were classified into three groups; “driver fusion”, “driver mutation”, and “pan-negative”. The original cohort consists of 608 cases screened for driver mutations in EGFR, KRAS, HER2 and BRAF and for driver fusions involving ALK, RET and ROS1. Other cases were classified as pan-negative cases. B. The study cohort, consisting of 200 cases enriched in ALK, RET and ROS1 fusion-positive cases (n = 31), was subjected to whole exome sequencing. The study cohort includes 96 driver mutation cases, including activating mutations in EGFR, KRAS, BRAF and HER2, and a case with a HRAS mutation detected in the present exome sequence.

Fig. 2. Genetic aberration profile of lung adenocarcinoma. Genetic aberrations in LADCs according to driver aberration. Driver gene aberrations, numbers of aberrant CGC genes, SMGs and SWI/SNF chromatin remodeling genes, and aberrations in representative cancer-related and SWI/SNF chromatin remodeling genes are shown for each case with clinical characters. Bar chart (right) indicates fractions of cases with gene aberrations. The numbers of all non-synonymous mutations are indicated with vertical bars at the bottom.

Fig. 3. Gene mutations by driver aberration. A. The numbers of non-synonymous mutations per Mb in all genes and the numbers of CGC genes and SMGs with non-synonymous mutations are shown. Whiskers represent the 5–95 percentiles and dots indicate the outliers.
P value was assessed by Kruskal-Wallis test. B. Frequency of mutation of the TP53 gene, SWI/SNF chromatin remodeling genes, APC or CTNNB1 genes, and PIK3CA gene. P value was assessed by Chi-square test. N.S.: not significant.
Fig. 1

A

The NCC cohort (n=608)

- ALK, RET, ROS1 fusion (n=50)
- EGFR, KRAS, HER2, BRAF mutation (n=361)
- None of ALK, RET, ROS1 fusion or EGFR, KRAS, HER2, BRAF, HRAS mutation (n=197)

Exclude cases with insufficient amounts of DNAs for exome sequencing

EGFR-positive and pan-negative cases were randomly selected to make the ratio approximately 1:1. Other driver mutation cases were also randomly selected.

Driver gene fusion (n=31)
Driver gene mutation (n=96): EGFR+ (n=72) & others (n=24)
Pan-negative (None) (n=73)

Eligible for the study cohort (n=200)

B

NCC cohort (n=608)

- None 32.4%
- ROS1 fusion 2.3%
- RET fusion 2.1%
- ALK fusion 3.8%
- BRAF mut 1.0%
- HER2 mut 2.0%
- EGFR mut 46.4%

Study cohort (n=200)

- None 36.5%
- ROS1 fusion 4.5%
- RET fusion 5.5%
- ALK fusion 5.5%
- HRAS mut 0.5%
- HER2 mut 2.0%
- BRAF mut 1.0%
- KRAS mut 8.5%
- Driver Mutation (35.3%)
- Fusion (15.6%)
- None (36.5%)
Fig. 3

A

**All genes**

Number of non-synonymous mutations per Mb

- **CGC**
  - Number of mutated genes
  - $P < 0.0001$

- **SMGs**
  - Number of mutated genes
  - $P < 0.0001$

B

**TP53**

Frequency of mutation

- Fusion
- Mutation
- Pan-negative

$P = 0.026$

**SWI/SNF genes**

Frequency of truncating mutation

- Fusion
- Mutation
- Pan-negative

$P < 0.0001$

**APC/CTNNB1**

Frequency of mutation

- Fusion
- Mutation
- Pan-negative

N.S.

**PIK3CA**

Frequency of mutation

- Fusion
- Mutation
- Pan-negative

N.S.
Development of Lung Adenocarcinomas with Exclusive Dependence on Oncogene Fusions

Motonobu Saito, Yoko Shimada, Kouya Shiraishi, et al.

Cancer Res Published OnlineFirst April 8, 2015.

Updated version
Access the most recent version of this article at: doi:10.1158/0008-5472.CAN-14-3282

Supplementary Material
Access the most recent supplemental material at: http://cancerres.aacrjournals.org/content/suppl/2015/04/10/0008-5472.CAN-14-3282.DC1

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