



Comprehensive molecular analysis of genomic profiles and PD-L1 expression in lung adenocarcinoma with a high-grade fetal adenocarcinoma component

Masaki Suzuki¹, Rika Kasajima^{2,3}, Tomoyuki Yokose¹, Hiroyuki Ito⁴, Eigo Shimizu³, Seira Hatakeyama⁵, Kazuaki Yokoyama⁶, Rui Yamaguchi^{3,7,8}, Yoichi Furukawa⁵, Satoru Miyano^{3,9}, Seiya Imoto³, Emi Yoshioka¹, Kota Washimi¹, Yoichiro Okubo¹, Kae Kawachi¹, Shinya Sato², Yohei Miyagi²

¹Department of Pathology, Kanagawa Cancer Center, Yokohama, Japan; ²Molecular Pathology and Genetics Division, Kanagawa Cancer Center Research Institute, Yokohama, Japan; ³Division of Health Medical Intelligence, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; ⁴Department of Thoracic Surgery, Kanagawa Cancer Center, Yokohama, Japan; ⁵Division of Clinical Genome Research, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; ⁶Department of Hematology/Oncology, Research Hospital, Institute of Medical Science, The University of Tokyo, Tokyo, Japan; ⁷Division of Cancer Systems Biology, Aichi Cancer Center Research Institute, Nagoya, Japan; ⁸Division of Cancer Informatics, Nagoya University Graduate School of Medicine, Nagoya, Japan; ⁹Department of Integrated Data Science, Medical and Dental Data Science Center, Tokyo Medical and Dental University, Tokyo, Japan

Contributions: (I) Conception and design: M Suzuki, R Kasajima, Y Miyagi; (II) Administrative support: T Yokose; (III) Provision of study materials or patients: M Suzuki, T Yokose, H Ito; (IV) Collection and assembly of data: M Suzuki, R Kasajima, H Ito; (V) Data analysis and interpretation: M Suzuki, R Kasajima, E Shimizu, S Hatakeyama, K Yokoyama, R Yamaguchi, Y Furukawa, S Miyano, S Imoto, Y Miyagi; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

Correspondence to: Yohei Miyagi, MD. Molecular Pathology and Genetics Division, Kanagawa Cancer Center Research Institute, 2-3-2 Nakao, Asahi-ku, Yokohama-city, Kanagawa 241-8515, Japan. Email: miyagi@gancen.asahi.yokohama.jp.

Background: Fetal adenocarcinoma of the lung is a rare variant of lung adenocarcinoma and is subcategorized into low-grade and high-grade (H-FLAC) fetal adenocarcinoma. We previously reported poor prognosis in pulmonary adenocarcinomas with an H-FLAC component; however, the genetic abnormalities involved in H-FLAC remain unclear. Therefore, this study aimed to elucidate molecular abnormalities as potential therapeutic targets for H-FLACs.

Methods: We performed immunohistochemical analysis and comprehensive genetic analyses using whole-exome sequencing in 16 lung cancer samples with an H-FLAC component. DNA was extracted from formalin-fixed paraffin-embedded tissues after macrodissection of the H-FLAC component.

Results: Cancer-related mutations were identified in *TP53* (7/16 cases), *KMT2C* (6/16 cases), *KRAS* (4/16 cases), *NF1* (3/16 cases), *STK11* (3/16 cases), *CTNNB1* (2/16 cases), and *EGFR* (1/16 cases). A high tumor mutation burden of ≥ 10 mutations per megabase was observed in 3/16 cases. A high microsatellite instability was not detected in any case. Based on the cosine similarity with the Catalogue of Somatic Mutations in Cancer mutational signatures, H-FLACs were hierarchically clustered into three types: common adenocarcinoma-like (five cases), surfactant-deficient (ten cases), and signatures 2 and 13-related (one case). All common adenocarcinoma-like cases presented thyroid transcription factor-1 (TTF-1) expression, whereas surfactant-deficient cases often presented loss of TTF-1 and surfactant protein expression and included cases with mutations in the surfactant system genes *NKX2-1* and *SFTPC*. H-FLACs displayed low programmed death ligand-1 (PD-L1) expression (1–49% of tumor cells) in 5/16 cases, and no case displayed high PD-L1 expression ($\geq 50\%$ of tumor cells).

Conclusions: This study indicates that lung cancers with an H-FLAC component rarely harbor currently targetable driver gene mutations for lung cancer but display a high frequency of *KMT2C* mutations. The microsatellite instability, tumor mutation burden, and PD-L1 expression status suggest a poor response to immune checkpoint therapy.

Keywords: High-grade fetal adenocarcinoma; whole-exome sequencing (WES); *CTNNB1*; *KMT2C*; programmed death ligand-1 (PD-L1)

Submitted Oct 31, 2020. Accepted for publication Jan 29, 2021.

doi: 10.21037/tlcr-20-1158

View this article at: <http://dx.doi.org/10.21037/tlcr-20-1158>

Introduction

Lung cancer is a leading cause of cancer-related mortality globally. Significant advancements in lung cancer treatment are based on the identification of the histologic subtypes and targeted agents, which are important factors that guide treatment-related decisions. Lung adenocarcinoma (LUAD) is the most common type of lung cancer, and invasive adenocarcinoma is classified into major histological patterns including lepidic, acinar, papillary, solid, and micropapillary, and variants such as mucinous, colloid, enteric, and fetal (1,2). Fetal adenocarcinomas are further divided into low-grade fetal adenocarcinomas (L-FLACs), and high-grade fetal adenocarcinomas (H-FLACs) (2,3). The association between prognosis and major histological patterns has been previously reported and indicate an association between solid/micropapillary patterns and poor prognosis (4-8). Few studies have focused on the prognosis of rare variants. We recently examined the prognostic influence of an H-FLAC component and revealed that LUADs with H-FLAC components had a poor prognosis comparable to that of micropapillary adenocarcinoma (9).

L-FLACs and H-FLACs display characteristic glandular structures resembling the airway epithelium during the pseudoglandular stage of fetal lung development and exhibit different clinicopathological and biological behaviors. L-FLACs reportedly constitute 0.3% of primary LUADs, usually appearing in the early stages in young-to-middle-aged women with a history of smoking, and display histological features including mild nuclear atypia and morula formation (3,10,11). H-FLACs constitute 0.5–1.4% of primary LUADs, frequently develop at advanced stages in elderly men with a history of heavy smoking, and display histological features including prominent nuclear atypia and the absence of morula formation (3,9,11-13). L-FLACs exhibit aberrant nuclear β -catenin localization with frequent *CTNNB1* (β -catenin gene) mutations, whereas most H-FLACs do not display β -catenin aberrations (9,11-15). H-FLACs are therefore speculated to originate through a tumorigenesis process distinct from that of L-FLACs. However, the tumorigenesis process of H-FLACs remains

unclear as only few systematic studies have focused on their genetic and molecular abnormalities, which may be considered as candidate molecular therapeutic targets.

Therefore, in this study, we performed whole-exome sequencing (WES) of 16 lung cancer cases with an H-FLAC component, to elucidate the genetic features of H-FLAC. Furthermore, based on the profile obtained from genetic alterations, an immunohistochemical (IHC) analysis of programmed death ligand-1 (PD-L1) expression was performed to predict the efficacy of immune checkpoint inhibitors for lung cancers with an H-FLAC component. To our knowledge, this is the first study to comprehensively assess H-FLAC. We present the following article in accordance with the MDAR reporting checklist (available at <http://dx.doi.org/10.21037/tlcr-20-1158>).

Methods

Case selection and clinicopathological evaluation

The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the ethics committee of the Kanagawa Cancer Center (approval no. H28-Res60) and individual consent for this retrospective analysis was waived. Sixteen samples of surgically resected primary lung cancers with an H-FLAC component present in various proportions (at least 5% of the total tumor volume) were obtained. All resected tissues were fixed in 10% formalin, embedded in paraffin, and stained with hematoxylin and eosin (HE). All cases were previously reported, and H-FLAC was histologically defined as previously described (9). The proportion of the H-FLAC component was evaluated as a percentage of the total tumor volume in HE-stained sections. Additionally, other coexisting components were assessed and classified in accordance with the World Health Organization classification of lung tumors (2). Clinical data including patient age, sex, smoking history, and tumor size and stage were retrieved from the Thoracic Surgical Database and medical records. Although the Union for International Cancer Control (UICC) published the 8th edition of

the UICC TNM classification of malignant tumors in 2017 (16), the lung cancer stages in the study were evaluated in accordance with the 7th edition of the UICC TNM classification due to the difficulty in retrospective measurement of accurate invasive sizes of old lung cancer cases (17).

IHC analysis

IHC staining was performed for formalin-fixed paraffin-embedded (FFPE) tissue sections using an automated immunostainer (Histostainer 48A; Nichirei, Tokyo, Japan) with the following primary antibodies: mouse monoclonal anti-thyroid transcription factor-1 (TTF-1) (Agilent Cat# IR056, RRID:AB_2755006, dilution 1:200), anti-surfactant protein (SP)-A (IBL, Cat# 10375, RRID:AB_10679116, dilution 1:40), anti-SP-B (Santa Cruz Biotechnology, Cat# sc-17827, RRID:AB_2275110, dilution 1:400), anti- β -catenin (Leica, Cat# PA0088, RRID:AB_10555989, dilution 1:50), anti-p53 (Agilent Cat# M7001, RRID:AB_2206626, dilution 1:100), rabbit polyclonal anti-SP-C (Proteintech, Cat# 10774-1-AP, RRID:AB_2185497, dilution 1:200), and rabbit monoclonal anti-PD-L1 (Cell Signaling Technology, Cat# 13684, RRID:AB_2687655, dilution 1:200) antibodies. TTF-1, SP-A, SP-B, SP-C, β -catenin, and p53 expression was detected using the Histofine Simple Stain MAX-PO kit (Nichirei, Tokyo, Japan). PD-L1 expression was detected using Novolink Polymer Detection Systems (Leica Biosystems, Newcastle, United Kingdom). Appropriate positive and negative control tissues were used in each case. IHC staining for β -catenin, TTF-1, SP-A, SP-B, and SP-C was considered positive if these molecules were detected in >10% of H-FLAC cells. Upon IHC analysis of p53, staining of >50% of H-FLAC cells was considered as positive. The H-FLAC component was considered positive for PD-L1 expression if \geq 1% PD-L1 staining was detected and was estimated in increments of 5% except for 1% positivity of H-FLAC cells.

WES

Multiple thinly sliced sections were prepared in succession from each FFPE tissue block, and the first section was stained with HE. Based on the results of HE staining, tissues from the H-FLAC component were manually obtained from the next 10 consecutive thinly sliced sections through macrodissection. DNA was extracted from the H-FLAC component, using the GeneRead DNA FFPE Kit (Qiagen,

Hilden, Germany), in accordance with the manufacturer's instructions, to reduce artefactual C>T|G>A transitions. Furthermore, DNA was extracted from non-neoplastic lung tissue in each case, using the same procedure to obtain the reference germline sequence. Double-stranded DNA was quantified using Qubit 2 (Thermo Fisher Scientific, Waltham, MA, USA). To assess DNA purity, the A_{260}/A_{280} and A_{260}/A_{230} optical density ratios were evaluated using NanoPhotometer (Implen, Munich, Germany).

In total, 200 ng of extracted DNA was used to generate panel libraries for each specimen. Libraries were amplified using SureSelect XT Target Enrichment System V6 (Agilent, Santa Clara, CA, USA) and sequenced using a HiSeq2500 system (Illumina, San Diego, CA, USA) with a standard 125-bp paired-end read protocol. We used the Genomon 2 DNA analysis pipeline (<https://github.com/Genomon-Project>) to detect somatic mutations and structural variations (SVs) for each case. Mutations such as single nucleotide variations (SNVs), short insertions and deletions (indels) with minimum depth \geq 8, base quality \geq 15, variant read \geq 4, $P < 0.01$ (Fisher's exact test), and variant allele frequency ≥ 0.02 in the tumor, were selected as mutation candidates. Analysis of the candidates included predicting changes in amino acids, determining if the variants were previously reported as somatic changes and if the variants are potentially pathogenic. SNVs were further evaluated using the sequences obtained as queries against the Catalogue of Somatic Mutations in Cancer (COSMIC; <https://cancer.sanger.ac.uk/cosmic>) database, the 1000 Genomes Project (<https://www.internationalgenome.org/>) (frequency of variants < 0.01), and dbSNP (NCBI, NIH; <https://www.ncbi.nlm.nih.gov/snp/>) [minor allele frequency (MAF) < 0.01]. Furthermore, literature searches in combination with database searches were performed to exclude potential single nucleotide polymorphisms. SVs such as large insertions, deletions, tandem duplications, and inversions, with tumor allele frequency ≥ 0.07 , depth threshold ≥ 10 , control depth ≥ 10 , removal of repeat sequence and inversion size threshold $> 1,000$ in the tumor, were considered as SV candidates.

Analyses of tumor mutation burden (TMB), microsatellite instability (MSI), and copy number variations (CNVs)

To determine the TMB as numbers of somatic mutations per DNA megabase, SNVs and splice-site variants were considered. The MSI status was determined using MicroSatellite Instability Classifier (MOSAIC;

Table 1 Clinicopathological features of lung cancers with a high-grade fetal adenocarcinoma component (H-FLAC)

Variables	Value
Total number (n)	16
Age (year)	
Median	67
Range	52–81
Sex (n)	
Female	3
Male	13
Smoking (n)	
Never	2
Ever	14
Tumor size (mm)	
Median	36
Range	21–122
Stage (n)	
I	8
II	5
III	3
Background lung tissue (n)	
Usual interstitial pneumonia	3
Severe emphysema	2

<https://github.com/ronaldhause/mosaic>) (18). CNVs were identified as genomic regions with an abnormal copy number, using DNACopy version 1.56.0. (<https://bioconductor.org/packages/release/bioc/html/DNACopy.html>), an R/Bioconductor package.

Evaluation of mutation patterns and signatures

MutationalPatterns version 2.0.0. (<https://bioconductor.org/packages/release/bioc/html/MutationalPatterns.html>), an R/Bioconductor-based package developed by Blokzijl *et al.* (19), was used to characterize the broad range of patterns in base substitutions and mutational signatures identified via WES, to analyze the underlying mechanisms involved in the development and acquisition of the malignant characteristics of H-FLAC. Furthermore, data of the same analyses performed using a WES dataset comprising 514 LUADs

obtained from The Cancer Genome Atlas (TCGA; <https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga>) and the results were compared with those obtained for the H-FLACs in this study.

Statistical analysis

Mann-Whitney U test was performed to compare the TMB obtained from H-FLACs and the TMB obtained from LUADs in TCGA database. Cosine similarity was used to determine the relative contribution of the 30 COSMIC mutational signatures (Mutational Signatures version 2, https://cancer.sanger.ac.uk/cosmic/signatures_v2) in each sample.

Results

Clinicopathological features

The clinicopathological findings for the 16 lung cancer cases with an H-FLAC component are summarized in *Table 1*. Lung cancers with an H-FLAC component were predominant among elderly men with a smoking history. Eight patients had stage I, five patients had stage II, and three patients had stage III disease. These cases displayed H-FLAC histology in various proportions: 5–9%, one case; 10–19%, two cases; 50–59%, one case; 60–69%, four cases; 70–79%, three cases; 80–89%, one case; 90–99%, two cases; and 100%, two cases. The H-FLAC component displayed complex glandular, cribriform, and papillary structures composed of columnar clear cells with pseudostratified atypical nuclei, resembling “piano keyboards” (*Figure 1*). Only two cases presented pure H-FLAC histology, whereas the others presented H-FLAC histology along with diverse components: conventional adenocarcinoma, 12 cases; invasive mucinous adenocarcinoma, three cases; enteric adenocarcinoma, one case; squamous cell carcinoma, one case; and sarcomatoid component, two cases. Pulmonary fibrosis with the usual interstitial pneumonia (UIP) pattern and severe emphysematous changes were observed in the background lung tissue in three (cases 4, 15, and 16) and two cases (cases 3 and 6), respectively.

Somatic gene mutations identified via WES

The means of the reads obtained and the depth were 107,849,503 and 165.34, respectively. Each depth is indicated in *Table S1*. We found that 99.9% of the total reads were accurately aligned with the hg19 human

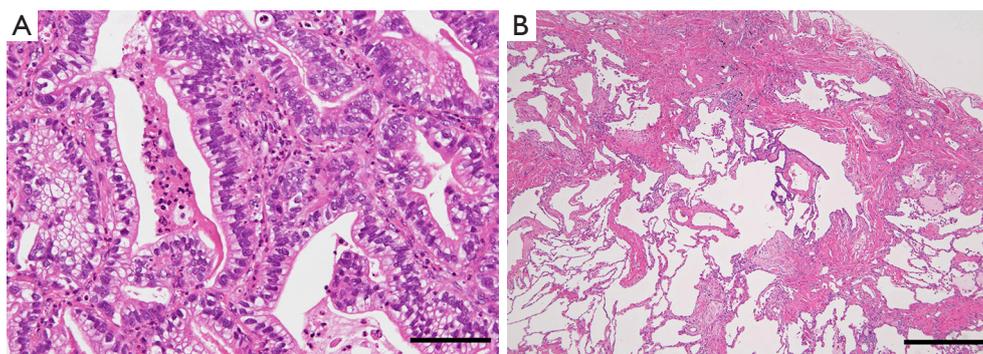


Figure 1 Microscopic characteristics (hematoxylin and eosin stain). (A) High-grade fetal adenocarcinoma (H-FLAC) formed complex glandular structures composed of columnar clear cells with pseudostratified nuclei, thereby displaying a “piano keyboard-like appearance”. Magnification: $\times 100$. Scale bar: 100 μm . (B) Some cases presented pulmonary fibrosis with the usual interstitial pneumonia pattern in the background lung tissue. Magnification: $\times 40$. Scale bar: 500 μm .

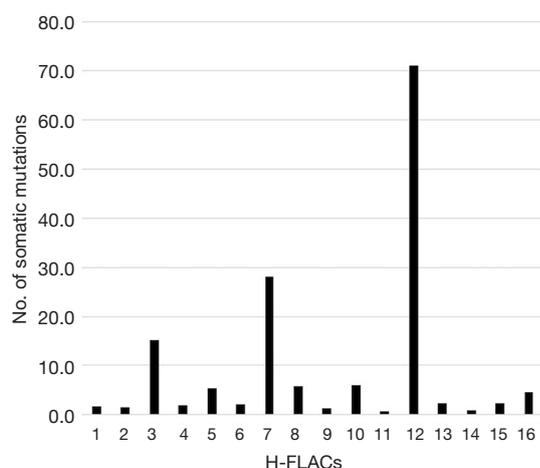


Figure 2 Tumor mutation burden (TMB) of the mutational landscape of high-grade fetal adenocarcinomas (H-FLACs). The number of mutations per megabase is denoted as the TMB. Each vertical bar corresponds to each H-FLAC for case IDs 1–16.

genome reference sequence, thus indicating the accuracy and reliability of WES. The total number of somatic mutations satisfying the criterion set ($P < 0.01$ and variant allele frequency ≥ 0.02) in each H-FLAC ranged from 26 to 2,402, and the TMB was 0.7–71.1 (median, 2.3; *Figure 2*). Through the same analysis, we obtained 12.5 as the TMB for the 514 TCGA LUADs, and the median TMB value of our cohort was significantly lower than that of the TCGA data ($P < 0.01$). A high TMB of ≥ 10 mutations per Mb was observed in 3/16 cases (19%). All the nonsynonymous somatic mutations that led to peptide sequence alterations

(including splice site mutations) identified in this study are included in table online (<https://cdn.amegroups.com/static/public/tlcr-20-1158-1.xlsx>) along with the allele frequency.

Nonsynonymous somatic mutations in H-FLACs were compared with a panel of 37 mutated genes in LUADs, as these genes displayed significant or near-significant mutation frequencies in 405 LUAD cases, and the data for this gene panel is available at TumorPortal (http://www.tumorportal.org/tumor_types?ttype=LUAD), hosted by the Broad Institute of MIT and Harvard (Cambridge, MA, USA). The most frequently mutated gene in our study was *TP53* (7/16 cases, 44%), followed by *KMT2C* (also known as *MLL3*) (6/16 cases, 38%), *KRAS* (4/16 cases, 25%), *NF1* (3/16 cases, 19%), and *STK11* (3/16, 19%). *SMAD4*, *RB1*, *MET*, *CTNNB1*, *CDKN2A*, and *BRAF* mutations were reported in two cases each (2/16, 13% each). The other 14 mutated genes in LUAD, including *EGFR* and *KEAP1*, were observed in one case each. The H-FLACs shared 25 of the 37 (68%) genes mutated in LUAD; this may reflect the profile of common nonspecific LUADs.

We then analyzed the list of somatic hotspot mutations occurring in major cancers (20), as they may be the probable cancer driver genes even though they may not have been confirmed as driver genes through *in vitro* or *in vivo* analyses. Along with genes mutated in LUAD cases on TumorPortal, this approach identified *PTEN*, *ACVR1B*, *COBL*, *KLF5*, *SHROOM1*, and *TPSD1*; however, these mutations were observed in only one case each.

Furthermore, we analyzed the genes that are not well characterized as cancer-related genes but were repeatedly detected during the current WES analysis. We focused on

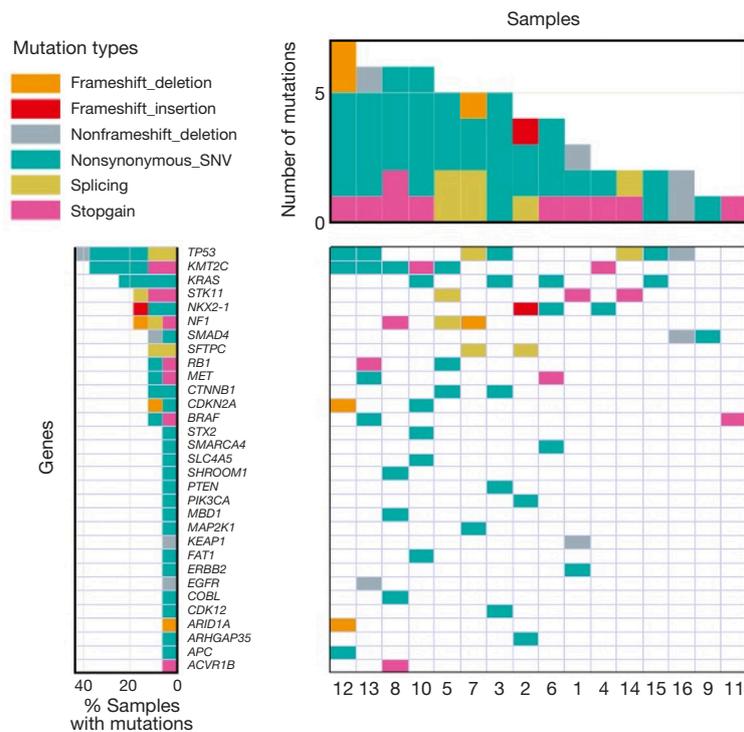


Figure 3 A schematic representation of the mutational landscape of high-grade fetal adenocarcinomas (H-FLACs). H-FLACs were characterized by somatic nonsynonymous mutations in this study. The types of mutations are denoted by different colors on the upper left side. The H-FLACs in this figure have been sorted in accordance with the number of mutations.

two genes, *NKX2-1* and *SFTPC*, involved in the pulmonary surfactant system. Three cases harbored *NKX2-1* mutations, including two missense mutations and one frameshift deletion, and two cases harbored *SFTPC* mutations that were splice-site mutations. These results are summarized in [Table S2](#) and schematically presented as the mutational landscape of H-FLAC in [Figure 3](#). All *KMT2C* mutations were detected in lung cancers that contained $\geq 60\%$ of the H-FLAC component. However, there was no appreciable difference in other gene mutations according to the proportion of H-FLAC component.

MSI status and CNVs

The MSI status was determined using the MicroSatellite Instability Classifier (MOSAIC), and a high MSI was not detected in any H-FLAC case. We analyzed genes harboring CNVs for the functional interpretation of large lists of genes, using DAVID (<https://david.ncifcrf.gov/>); however, none of the common CNVs were strongly associated with H-FLACs examined in this study (data not shown).

IHC analysis

The results of IHC analysis are displayed in [Figure 4](#). The H-FLAC component displayed nuclear β -catenin localization in three cases, including two cases harboring a *CTNNB1* mutation. One of the cases was a pure H-FLAC, and the other cases presented a combination with a lepidic adenocarcinoma component, which, in contrast, displayed membranous β -catenin staining. Diffuse p53 expression was detected in six cases, all presenting a *TP53* mutation. Considering the results of WES and mutational signature analysis displayed below, TTF-1 and SPs including SP-A, SP-B, and SP-C were often not expressed in H-FLAC components.

IHC analysis was performed for PD-L1 due to the significantly lower TMB in our H-FLAC cohort compared with the TCGA data of LUADs. Five cases displayed low PD-L1 expression (1–49% of H-FLAC cells), whereas the other cases were negative for PD-L1. In all cases wherein PD-L1 was not expressed in H-FLAC cells, it was not expressed in the other combined components as well.

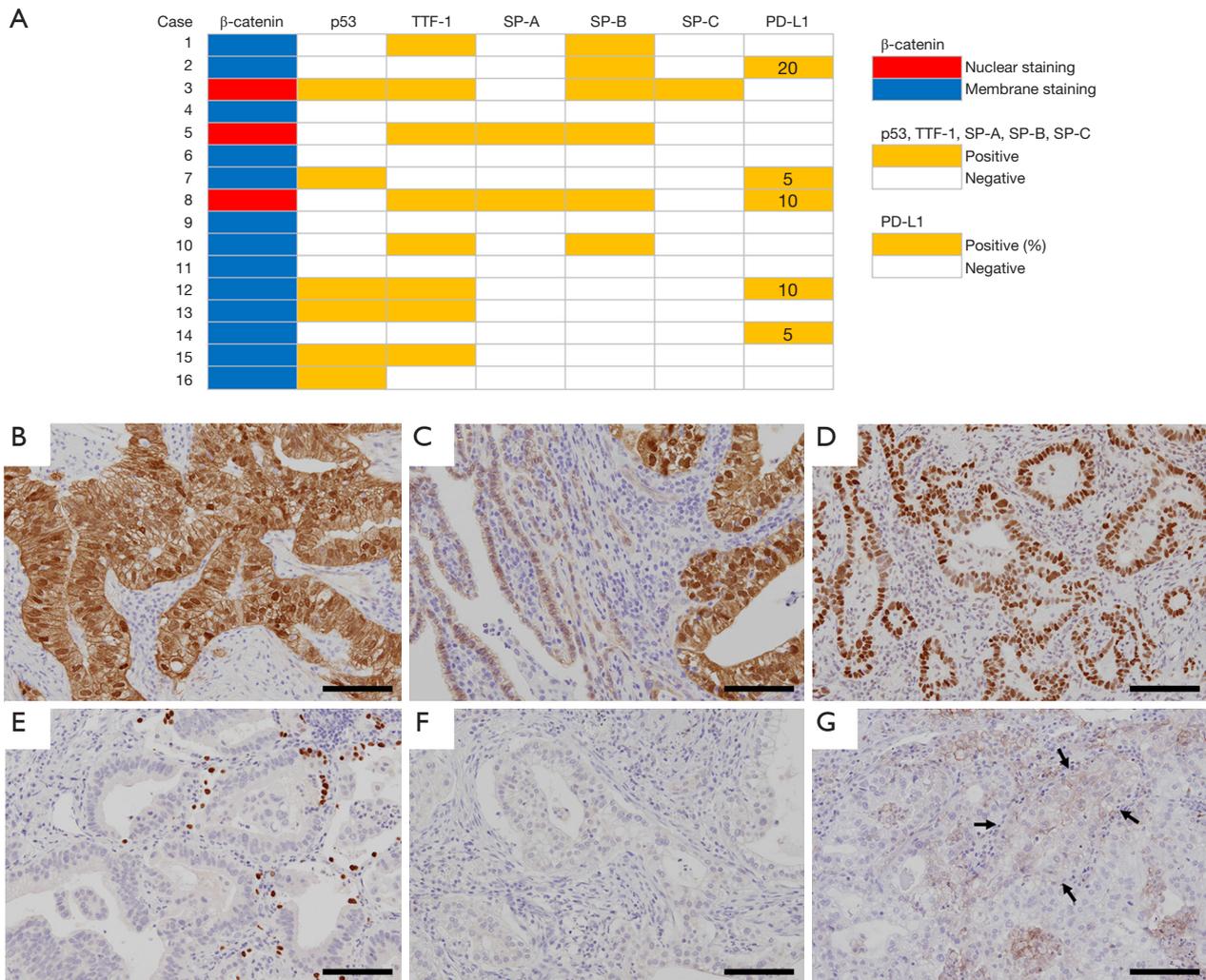


Figure 4 Immunohistochemical characteristics of high-grade fetal adenocarcinomas (H-FLACs). (A) Summary of the results of immunohistochemical analysis. (B) Nuclear β-catenin localization was detected. (C) H-FLAC component with nuclear β-catenin localization was seen on the right, and the coexisting lepidic adenocarcinoma component displayed membranous staining of β-catenin on the left. (D) Diffuse p53 expression was detected. (E,F) TTF-1 (E) and SP-C (F) expression was often not observed. (G) H-FLAC displayed low PD-L1 expression levels in some cases (arrows: tumor cells with weak PD-L1 expression). Scale bars: 100 μm.

Analyses of mutational patterns and mutational signatures

A plot of the six base substitution type spectra of all the 16 H-FLAC samples was obtained (Figure S1). Overall, the most common base change identified in H-FLACs was C>T; in contrast, the most common base change in the 514 cases of TCGA LUADs was C>A. Among the cases with the C>T base change, C>T was predominant in the non-CpG dinucleotide context, and this tendency was the same as that noted in the TCGA LUADs.

Thereafter, we obtained the mutation matrix of each

of the 16 H-FLACs, which displayed a mutation pattern in terms of the relative contribution of the identified mutation numbers for all 96 trinucleotide changes (Figure 5A). Two common mutational signatures, designated as signatures A and B, were extracted *de novo* from the mutation matrices of 96 trinucleotide changes through non-negative matrix fractionation with the method provided in the MutationalPatterns package (Figure 5B). The same analysis was performed for the TCGA LUAD data, and the signatures extracted from them were markedly different

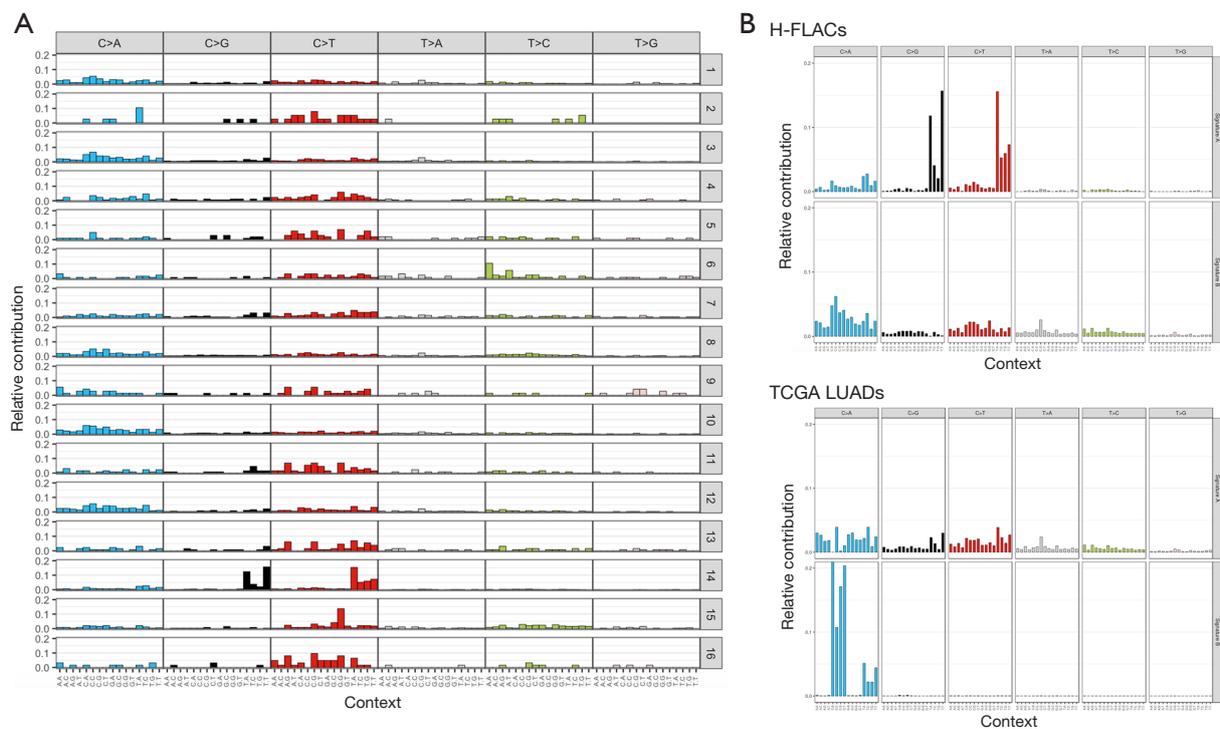


Figure 5 Mutational pattern of each high-grade fetal adenocarcinoma (H-FLAC) and the extracted common signatures in the context of trinucleotide changes. (A) The mutational pattern for each of the 16 H-FLACs is indicated in terms of the relative contribution of the identified mutation numbers for all 96 trinucleotide changes. (B) The extracted common signatures A and B for 16 H-FLACs (upper panel) and for 514 TCGA lung adenocarcinomas (LUADs; lower panel) are indicated. TCGA, The Cancer Genome Atlas.

(Figure 5B). The relative contribution of the well-known 30 COSMIC mutational signatures to the mutational profile of each H-FLAC was further evaluated by determining the cosine similarity through a *MutationalPatterns* package. H-FLACs were hierarchically clustered into three types in accordance with the cosine similarity to the COSMIC signatures. These clusters were well correlated with the immunostaining status of TTF-1 and the SPs SP-A, SP-B, and SP-C and to the status of surfactant gene mutations, designated as common adenocarcinoma-like (five cases), surfactant-deficient (ten cases), and signatures 2 and 13-related (one case) (Figure 6). All H-FLACs in the common adenocarcinoma-like cluster were positive for TTF-1, and all cases, except for one, displayed positive immunostaining for at least one of the three SPs. In contrast, H-FLACs in the surfactant-deficient cluster were negative for TTF-1 upon immunostaining, with three exceptions, and negative for the SPs, with two exceptions, and included all four cases of H-FLACs with mutations in the surfactant genes *NKX2-1* or *SFTPC*. Additionally, all three cases with a mucinous adenocarcinoma component

and one case with an enteric adenocarcinoma component were included in the surfactant-deficient cluster. Other coexistent histological components were not apparently associated with the mutational signature clusters. The common adenocarcinoma-like H-FLACs were strongly associated with signature 4 and, specifically, in comparison with the surfactant-deficient type, associated with signatures 3, 8, 24, and 29. The surfactant-deficient type was associated with signatures 1 and 6. The same analysis was performed for the 514 TCGA LUADs. Most TCGA LUADs were clustered into a group whose pattern in the cosine similarity heat map was quite similar to that of the common adenocarcinoma-like H-FLACs, revealing the appropriateness of our cluster nomenclature (Figure S2). Furthermore, a mutational signature pattern of the surfactant-deficient type of H-FLACs was similar to that of a minor cluster of TCGA LUADs.

Discussion

Molecular targeted therapy has been the standard systemic

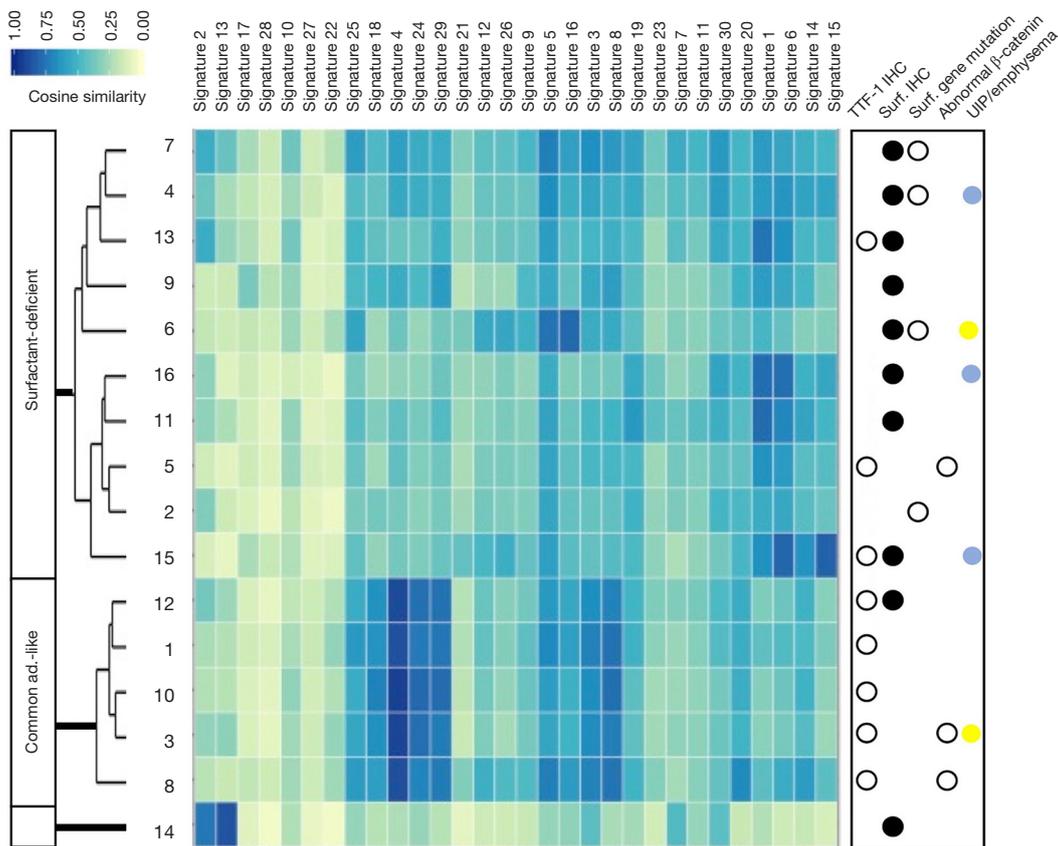


Figure 6 Heatmap of the cosine similarity to the 30 mutational signatures and relationship with information regarding immunohistochemistry, gene mutation, and clinical background for each high-grade fetal adenocarcinoma (H-FLAC). Each H-FLAC has been vertically aligned, and 30 Catalogue of Somatic Mutations in Cancer (COSMIC) mutational signatures have been aligned on the horizontal line. Cosine similarity to each mutational signature was noted from 1.0 (identical) to 0 (completely independent) using the colors included in the upper-left corner. The features of each case are indicated on the right side of the heatmap. Positive immunostaining for TTF-1 or nuclear β -catenin and the presence of mutations in the surfactant system genes *NKX2-1* and *SFTPC* are indicated with open circles, and negative immunostaining for any of the surfactant proteins is indicated with closed circles. Background usual interstitial pneumonia (UIP) is indicated with blue circles and severe emphysema with yellow circles. The types of H-FLAC clusters are indicated on the left side of the panel. Surf.: surfactant.

treatment method for patients with oncogene-driven advanced-stage non-small cell lung cancer (NSCLC); however, the currently targeted driver gene mutations for lung cancer were not detected in H-FLAC in this study, except for an *EGFR* mutation. We are currently analyzing RNA-sequencing data for RNAs extracted from the sample sets used for WES; currently no targetable gene fusion has been identified (data not shown). *KRAS* G12C inhibitors are being developed for treating cancers harboring this mutation (21,22); however, the G12C mutation was detected in only one case in this study. Six of the 16 cases (38%) harbored eight mutations in *KTM2C*, which was

identified as the most frequently mutated gene in the present WES analysis. *KTM2C* is a key histone H3K4 monomethyltransferase along with *KTM2D* at enhancers, primarily implementing H3K4 monomethylation (23). Missense and nonsense somatic mutations in *KTM2C* have been identified in various malignancies including NSCLC, and *KTM2C* is inactivated through these mutations (24,25). According to the data in TumorPortal, 14% of LUADs harbored a somatic mutation in *KTM2C*, which was the third most frequently mutated gene along with *KEAP1* (14%), following the most frequently mutated *TP53* (51%) and *KRAS* (26%). The status of *TP53* and *KRAS* mutations

was similar between H-FLACs in this study and the LUADs from TumorPortal. Therefore, the high frequency of *KTM2C* mutations can be considered as a characteristic of H-FLACs. As H3K4 monomethylation is a histone mark of enhancer activation (24), perturbation of the epigenetic regulation of gene expression is potentially associated with the biological features of H-FLACs. Comparative genetic, epigenetic, and/or transcriptomic analyses of H-FLAC and non-H-FLAC components in the same tumors are required to further elucidate the characteristics of H-FLAC.

Although only one H-FLAC harbored an *EGFR* mutation, a markedly higher *EGFR* mutation frequency (46–64%) has been reported in conventional LUADs, especially among East Asian patients (7,26–29). The *EGFR* mutation rate in H-FLACs is reportedly low (6–20%), as also observed in our study (11–13). *EGFR* mutations are more prevalent among women and never-smokers among East Asian LUADs cases (26,27,29), whereas, H-FLAC reportedly occurs in elderly men with a smoking history (3,9,11–13). The gap in the *EGFR* mutation status potentially reflects differences in the etiological factors among lung cancer patients. The *KRAS* mutation associated with a smoking history and mucinous histology (7,30) was correspondingly detected in 25% of H-FLACs in this study. No significant differences between the cases displaying fetal-type histological morphology with or without driver mutations were noted.

Cellular genome integrity is always endogenously and exogenously challenged by DNA damage, and accumulated mutations can lead to diseases including cancer. Therefore, identification of the mutation signatures is important for determining the etiology of cancer (31). Based on the similarity of mutational patterns in H-FLACs with 30 well-established mutation signatures provided in COSMIC, we could classify H-FLACs into two major types based on the status of TTF-1 and surfactant expression, surfactant system gene mutations, and their association with UIP/severe emphysema. Although one-third of H-FLACs were classified as common adenocarcinoma-like H-FLACs, the remaining two-thirds could be considered as the surfactant-deficient type, with one exception that was strongly associated with signatures 2 and 13. Honda *et al.* performed WES for numerous LUAD cases and reported associations among mutations in pulmonary surfactant system genes (*NKX2-1*, *SFTPA1*, *SFTPA2*, *SFTPB*, and *SFTPC*), smoking, and a background of UIP (32). In this study, all three UIP cases and 1 of the 2 severe emphysema cases were included in the surfactant-deficient type of H-FLACs.

Smoking or tobacco-chewing habits are associated with signatures 4 and 29 and were associated with the TTF-1-positive common adenocarcinoma-like type of H-FLACs in this study. Although clusters displaying mutational signature patterns similar to those of H-FLACs were observed in the cosine similarity heat map of TCGA LUADs, the proportions of clusters were markedly different between H-FLACs and TCGA LUADs. Because the association between mutational signatures and the predicted etiology remains unclear, precise clinicopathological studies with larger numbers of H-FLACs are required.

Owing to the heterogeneous mutational profiles and diverse combined non-fetal type histology, a previous study reported that H-FLAC is not a true entity but a morphological pattern (12), as revealed through the detection of various cancer-related mutations. Some mutational signature clusters identified in this H-FLAC cohort were shared with the LUAD datasets in TCGA. However, in rare cases, H-FLAC could occur in its pure form and can be confused with L-FLAC; therefore, H-FLAC should be recognized as a rare variant. Furthermore, regardless of the various genetic backgrounds, H-FLACs had characteristic clinicopathological features, unique fetal lung-like histology, and a potentially common oncogenic mechanism not detectable in this genomic study; hence, it should be considered as an independent entity of LUAD with fetal morphological appearance and high-grade behavior.

In this study, three cases presented aberrant nuclear β -catenin accumulation, and two of them harbored *CTNNB1* hotspot mutations D32Y and D32V. All the samples in our previous study displayed membrane β -catenin localization (9); however, nuclear β -catenin localization was detected through IHC analysis in this study via another clone of the monoclonal antibody. No case of H-FLAC harboring a *CTNNB1* mutation has been reported previously (11–13,15), unlike L-FLAC cases, which are characterized by *CTNNB1* mutations and nuclear β -catenin localization (11,14,15). Therefore, H-FLAC is considered more similar to conventional LUAD than to L-FLAC. In this study, two cases with nuclear β -catenin localization were combined with the lepidic adenocarcinoma component, and the lepidic components did not display nuclear β -catenin localization. No L-FLAC component was observed in tumors in all cases. Therefore, it might be speculated that the H-FLAC component developed from a pre-existing lepidic component after the acquisition of the *CTNNB1* mutation, although DNA sequencing of the

lepidic components was not performed. However, Nakatani *et al.* reported that nuclear β -catenin localization was immunohistochemically detected in two of eight H-FLAC cases and that one of the two cases presented histologic features intermediate between L-FLAC and H-FLAC (15). L-FLAC may rarely progress to a high-grade form. Aberrant β -catenin localization is not considered sufficient for the differential diagnosis of L-FLAC and H-FLAC. The present results indicate that old age, prominent nuclear atypia, no morula formation, and the coexistence of a conventional adenocarcinoma component are important for distinguishing H-FLAC from L-FLAC.

In advanced NSCLC cases with high PD-L1 expression ($\geq 50\%$ of tumor cells), pembrolizumab, which is used for first-line immune checkpoint inhibitor monotherapy, reportedly improves progression-free and overall survival (33). In previous studies using the 22C3 PD-L1 IHC assay, which was approved as a companion diagnostic test for pembrolizumab by the Food and Drug Administration of the United States, positive PD-L1 staining ($\geq 1\%$ of tumor cells) was detected in 34–66% and high PD-L1 expression ($\geq 50\%$ of tumor cells) in 11–30% of NSCLC cases (33–37). In this study, positive PD-L1 staining ($\geq 1\%$ of H-FLAC cells) was observed in 5/16 (31%) cases; however, no sample displayed high PD-L1 expression ($\geq 50\%$ of H-FLAC cells). The E1L3N clone was used for the IHC analysis of PD-L1 in this study, owing to its reportedly higher sensitivity than the 22C3 clone (38). Only two a few studies have performed IHC analysis for PD-L1 in H-FLAC cases, and both studies reported cases negative for PD-L1 (39,40). Furthermore, the findings for MSI and TMB, which are surrogate markers for pembrolizumab application, did not indicate H-FLACs as a target for pembrolizumab treatment. Mutation signature 6, known to be associated with defective mismatch repair, and signature 1, known to be associated with an endogenous mutation process initiated by 5-methylcytosine deamination, are considered characteristics of the surfactant-deficient type. However, MSI analysis using MOSAIC in this study did not reveal any high MSI case, and the median TMB value in this cohort is significantly lower than that of TCGA LUADs. H-FLACs tend to display no/low PD-L1 expression and a comparatively low TMB/MSI, suggesting resistance to immunotherapy in these patients.

In conclusion, H-FLACs rarely harbor driver mutations that are currently targeted for lung cancers; however, they harbor highly frequent *KMT2C* mutations, implying the epigenetic regulation of the biological features of H-FLACs.

Mutational signature analysis helped classify H-FLACs into two major types; however, further studies are required to clarify the underlying etiological factors. MSI, TMB, and the PD-L1 expression status suggest a poor response of H-FLACs to PD-L1 immune checkpoint therapy. As H-FLAC has a high malignant potential, further studies are required to identify candidate therapeutic target molecules.

Acknowledgments

We are grateful to the technical staff at Kanagawa Cancer Center for their expert technical assistance and the members of the Clinical Sequencing Team at Human Genome Center, Institute of Medical Science, The University of Tokyo, for their expert technical assistance and intellectual discussions. The super-computing resource was provided by the Human Genome Center, Institute of Medical Science, The University of Tokyo (<http://sc.hgc.jp/shirokane.html>). Genomon 2 was provided by Kenichi Chiba, Ai Okada, and Yuichi Shiraishi.

Funding: This work was supported in part by the Japan Society for the Promotion of Science (JP18K15111 to MS, JP16H06276 to YM).

Footnote

Reporting Checklist: The authors have completed the MDAR reporting checklist. Available at <http://dx.doi.org/10.21037/tlcr-20-1158>

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at <http://dx.doi.org/10.21037/tlcr-20-1158>). Dr. SM reports other from Fujitsu Laboratories, Ltd., other from Liquid Mine, Ltd., outside the submitted work. The other authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013). The study was approved by the ethics committee of the Kanagawa Cancer Center (approval no. H28-Res60) and individual consent for this retrospective analysis was waived.

Open Access Statement: This is an Open Access article

distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

References

1. Travis WD, Brambilla E, Noguchi M, et al. International Association for the Study of Lung Cancer/American Thoracic Society/ European Respiratory Society International Multidisciplinary Classification of Lung Adenocarcinoma. *J Thorac Oncol* 2011;6:244-85.
2. Travis WD, Brambilla E, Burke AP, et al. WHO classification of tumours of the lung, pleura, thymus and heart. Lyon: IARC Press, 2015.
3. Nakatani Y, Kitamura H, Inayama Y, et al. Pulmonary adenocarcinomas of the fetal lung type: a clinicopathologic study indicating differences in histology, epidemiology, and natural history of low-grade and high-grade forms. *Am J Surg Pathol* 1998;22:399-411.
4. Yoshizawa A, Motoi N, Riely GJ, et al. Impact of proposed IASLC/ATS/ERS classification of lung adenocarcinoma: prognostic subgroups and implications for further revision of staging based on analysis of 514 stage I cases. *Mod Pathol* 2011;24:653-64.
5. Warth A, Muley T, Meister M, et al. The novel histologic International Association for the Study of Lung Cancer/ American Thoracic Society/European Respiratory Society classification system of lung adenocarcinoma is a stage-independent predictor of survival. *J Clin Oncol* 2012;30:1438-46.
6. Tsuta K, Kawago M, Inoue E, et al. The utility of the proposed IASLC/ATS/ERS lung adenocarcinoma subtypes for disease prognosis and correlation of driver gene alterations. *Lung Cancer* 2013;81:371-6.
7. Yoshizawa A, Sumiyoshi S, Sonobe M, et al. Validation of the IASLC/ATS/ERS lung adenocarcinoma classification for prognosis and association with EGFR and KRAS gene mutations: analysis of 440 Japanese patients. *J Thorac Oncol* 2013;8:52-61.
8. Yanagawa N, Shiono S, Abiko M, et al. The Clinical Impact of Solid and Micropapillary Patterns in Resected Lung Adenocarcinoma. *J Thorac Oncol* 2016;11:1976-83.
9. Suzuki M, Nakatani Y, Ito H, et al. Pulmonary adenocarcinoma with high-grade fetal adenocarcinoma component has a poor prognosis, comparable to that of micropapillary adenocarcinoma. *Mod Pathol* 2018;31:1404-17.
10. Sato S, Koike T, Yamato Y, et al. Resected well-differentiated fetal pulmonary adenocarcinoma and summary of 25 cases reported in Japan. *Jpn J Thorac Cardiovasc Surg* 2006;54:539-42.
11. Zhang J, Sun J, Liang XL, et al. Differences between low and high grade fetal adenocarcinoma of the lung: a clinicopathological and molecular study. *J Thorac Dis* 2017;9:2071-8.
12. Morita S, Yoshida A, Goto A, et al. High-grade lung adenocarcinoma with fetal lung-like morphology: clinicopathologic, immunohistochemical, and molecular analyses of 17 cases. *Am J Surg Pathol* 2013;37:924-32.
13. Suzuki M, Yazawa T, Ota S, et al. High-grade fetal adenocarcinoma of the lung is a tumour with a fetal phenotype that shows diverse differentiation, including high-grade neuroendocrine carcinoma: a clinicopathological, immunohistochemical and mutational study of 20 cases. *Histopathology* 2015;67:806-16.
14. Proctor L, Folpe AL, Esper A, et al. Well-differentiated fetal adenocarcinoma of the lung: cytomorphic features on fine-needle aspiration with emphasis on use of β -catenin as a useful diagnostic marker. *Diagn Cytopathol* 2007;35:39-42.
15. Nakatani Y, Masudo K, Miyagi Y, et al. Aberrant nuclear localization and gene mutation of β -catenin in low-grade adenocarcinoma of fetal lung type: up-regulation of the Wnt signaling pathway may be a common denominator for the development of tumors that form morules. *Mod Pathol* 2002;15:617-24.
16. James DB, Mary KG, Christian W (editors). TNM classification of malignant tumors. 8th edition. Hoboken: Wiley-Blackwell, 2017.
17. Leslie HS, Mary KG, Christian W. TNM classification of malignant tumors. 7th edition. Hoboken: Wiley-Blackwell, 2009.
18. Hause RJ, Pritchard CC, Shendure J, et al. Classification and characterization of microsatellite instability across 18 cancer types. *Nat Med* 2016;22:1342-50.
19. Blokzijl F, Janssen R, van Boxtel R, et al. MutationalPatterns: comprehensive genome-wide analysis of mutational processes. *Genome Med* 2018;10:33.
20. Chang MT, Asthana S, Gao SP, et al. Identifying recurrent mutations in cancer reveals widespread lineage diversity and mutational specificity. *Nat Biotechnol* 2016;34:155-63.

21. Canon J, Rex K, Saiki AY, et al. The clinical KRAS(G12C) inhibitor AMG 510 drives anti-tumour immunity. *Nature* 2019;575:217-23.
22. Hallin J, Engstrom LD, Hargis L, et al. The KRAS(G12C) Inhibitor MRTX849 Provides Insight toward Therapeutic Susceptibility of KRAS-Mutant Cancers in Mouse Models and Patients. *Cancer Discov* 2020;10:54-71.
23. Hu D, Gao X, Morgan MA, et al. The MLL3/MLL4 branches of the COMPASS family function as major histone H3K4 monomethylases at enhancers. *Mol Cell Biol* 2013;33:4745-54.
24. Sze CC, Shilatifard A. MLL3/MLL4/COMPASS Family on Epigenetic Regulation of Enhancer Function and Cancer. *Cold Spring Harb Perspect Med* 2016;6:a026427.
25. Liu P, Morrison C, Wang L, et al. Identification of somatic mutations in non-small cell lung carcinomas using whole-exome sequencing. *Carcinogenesis* 2012;33:1270-6.
26. Song Z, Zhu H, Guo Z, et al. Correlation of EGFR mutation and predominant histologic subtype according to the new lung adenocarcinoma classification in Chinese patients. *Med Oncol* 2013;30:645.
27. Kosaka T, Yatabe Y, Onozato R, et al. Prognostic implication of EGFR, KRAS, and TP53 gene mutations in a large cohort of Japanese patients with surgically treated lung adenocarcinoma. *J Thorac Oncol* 2009;4:22-9.
28. Haneda H, Sasaki H, Lindeman N, et al. A correlation between EGFR gene mutation status and bronchioloalveolar carcinoma features in Japanese patients with adenocarcinoma. *Jpn J Clin Oncol* 2006;36:69-75.
29. Mitsudomi T, Kosaka T, Endoh H, et al. Mutations of the epidermal growth factor receptor gene predict prolonged survival after gefitinib treatment in patients with non-small-cell lung cancer with postoperative recurrence. *J Clin Oncol* 2005;23:2513-20.
30. Li H, Pan Y, Li Y, et al. Frequency of well-identified oncogenic driver mutations in lung adenocarcinoma of smokers varies with histological subtypes and graduated smoking dose. *Lung Cancer* 2013;79:8-13.
31. Alexandrov LB, Nik-Zainal S, Wedge DC, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500:415-21. Erratum in: *Nature*. 2013 Oct 10;502(7470):258. Imielinski, Marcin [corrected to Imielinski, Marcin].
32. Honda T, Sakashita H, Masai K, et al. Deleterious Pulmonary Surfactant System Gene Mutations in Lung Adenocarcinomas Associated With Usual Interstitial Pneumonia. *JCO Precis Oncol* 2018. doi: 10.1200/PO.17.00301
33. Reck M, Rodriguez-Abreu D, Robinson AG, et al. Pembrolizumab versus Chemotherapy for PD-L1-Positive Non-Small-Cell Lung Cancer. *N Engl J Med* 2016;375:1823-33.
34. Garon EB, Rizvi NA, Hui R, et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N Engl J Med* 2015;372:2018-28.
35. Herbst RS, Baas P, Kim DW, et al. Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *Lancet* 2016;387:1540-50.
36. Chen Q, Fu YY, Yue QN, et al. Distribution of PD-L1 expression and its relationship with clinicopathological variables: an audit from 1071 cases of surgically resected non-small cell lung cancer. *Int J Clin Exp Pathol* 2019;12:774-86.
37. Skov BG, Rorvig SB, Jensen THL, Skov T. The prevalence of programmed death ligand-1 (PD-L1) expression in non-small cell lung cancer in an unselected, consecutive population. *Mod Pathol* 2020;33:109-17.
38. Munari E, Zamboni G, Lunardi G, et al. PD-L1 expression in non-small cell lung cancer: evaluation of the diagnostic accuracy of a laboratory-developed test using clone E1L3N in comparison with 22C3 and SP263 assays. *Hum Pathol* 2019;90:54-9.
39. Kubouchi Y, Matsuoka Y, Araki K, et al. Pure high-grade fetal adenocarcinoma of the lung: a case report. *Surg Case Rep* 2018;4:70.
40. Ishida H, Yasuda M, Nitanda H, et al. Pulmonary high-grade fetal adenocarcinoma associated with cystic airspace: A case report. *Thorac Cancer* 2020;11:1703-7.

Cite this article as: Suzuki M, Kasajima R, Yokose T, Ito H, Shimizu E, Hatakeyama S, Yokoyama K, Yamaguchi R, Furukawa Y, Miyano S, Imoto S, Yoshioka E, Washimi K, Okubo Y, Kawachi K, Sato S, Miyagi Y. Comprehensive molecular analysis of genomic profiles and PD-L1 expression in lung adenocarcinoma with a high-grade fetal adenocarcinoma component. *Transl Lung Cancer Res* 2021;10(3):1292-1304. doi: 10.21037/tlcr-20-1158

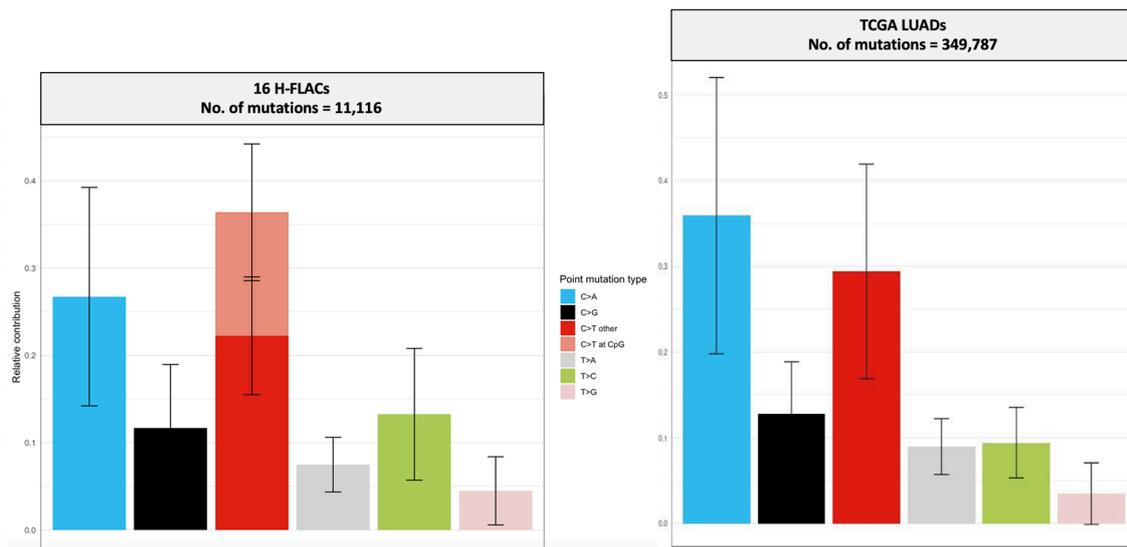


Figure S1 Six base substitution type spectra of mutations in 16 high-grade fetal adenocarcinomas (H-FLACs) and The Cancer Genome Atlas (TCGA) lung adenocarcinomas (LUADs). Six base substitution type spectra of mutations of H-FLACs were compared with those of the TCGA LUADs.

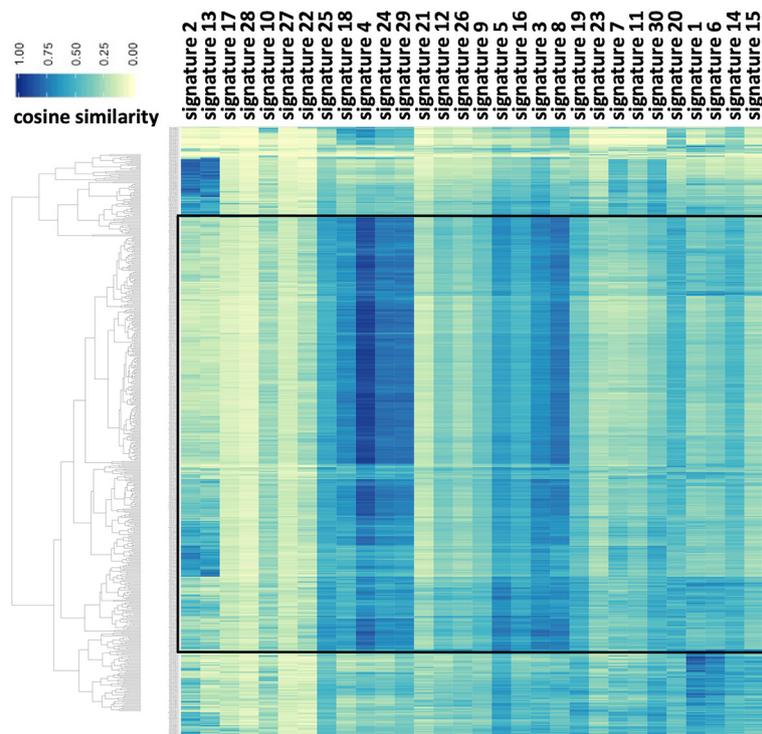


Figure S2 Heatmap of cosine similarity with the 30 mutational signatures of the 514 The Cancer Genome Atlas (TCGA) lung adenocarcinomas. The black-boxed cases resemble the common adenocarcinoma-like cluster of high-grade fetal adenocarcinomas.

Table S1 Depth of whole-exome sequencing in this study

Case	Tumor/normal	Mapped bases	Total reads	Mapped reads	Average depth
1	Normal	14,207,250,876	145,588,534	145,567,663	173.4
	Tumor	10,549,991,287	108,989,542	108,978,966	120.3
2	Normal	10,298,302,219	107,095,830	106,331,339	118.8
	Tumor	9,516,642,628	98,964,910	98,264,021	125.8
3	Normal	12,682,724,915	130,077,950	130,044,402	176.2
	Tumor	17,464,278,431	179,188,518	179,147,506	249.4
4	Normal	20,852,408,890	213,940,710	213,918,518	236.6
	Tumor	16,681,722,588	170,880,258	170,866,769	176.0
5	Normal	16,012,910,291	163,063,046	163,047,551	197.3
	Tumor	9,201,544,823	93,860,074	93,851,652	129.6
6	Normal	9,406,739,344	96,408,052	96,386,977	132.5
	Tumor	14,296,561,905	146,616,016	146,577,527	207.3
7	Normal	6,794,952,150	69,226,006	69,220,129	196.2
	Tumor	7,765,555,835	79,415,868	79,409,065	109.3
8	Normal	7,030,664,437	72,716,628	72,222,492	227.3
	Tumor	10,286,952,074	106,927,174	106,153,003	129.1
9	Normal	5,273,558,726	53,633,592	53,628,956	176.0
	Tumor	3,922,298,094	39,910,676	39,907,336	162.4
10	Normal	22,138,805,708	227,217,152	227,164,600	309.9
	Tumor	14,965,894,854	153,125,194	153,087,820	209.0
11	Normal	3,741,774,558	38,021,092	38,018,030	140.8
	Tumor	7,912,218,220	80,486,498	80,481,374	169.2
12	Normal	10,561,042,636	107,402,422	107,387,426	147.4
	Tumor	10,745,806,494	109,531,452	109,516,062	154.2
13	Normal	5,732,584,430	58,236,288	58,231,333	149.1
	Tumor	8,346,102,076	84,717,568	84,706,190	116.9
14	Normal	7,562,510,568	78,990,740	78,982,023	100.5
	Tumor	10,458,370,047	106,369,746	106,360,625	120.8
15	Normal	7,334,071,274	74,404,062	74,399,424	190.9
	Tumor	10,418,159,370	106,539,070	106,533,342	133.8
16	Normal	8,269,812,792	84,597,748	84,592,882	149.5
	Tumor	6,502,322,511	66,481,680	66,478,559	155.3

Table S2 Summary of nonsynonymous somatic mutations of high-grade fetal adenocarcinomas

Gene	Patient ID															
	1	2	3 [#]	4 ^â	5	6 [#]	7	8	9	10	11	12	13	14	15 ^â	16 ^â
<i>KMT2C</i>				C391X	Y512C			K339N		C391X		Multiple*	D2703E			
<i>TP53</i>			S241F				Splice site					L130F	L194R	Splice site	P250L	240_243del
<i>KRAS</i>			G12D			G12C				G12D					G12D	
<i>STK11</i>	Q37X				Splice site									E256X		
<i>SMAD4</i>								D351Y								528_530del
<i>RB1</i>					R320P								Y813X			
<i>NF1</i>					Splice site		Multiple*	L2089X								
<i>MET</i>						C180X							V444G			
<i>CTNNB1</i>			D32Y		D32V											
<i>CDKN2A</i>										G110C		A76fX				
<i>BRAF</i>											E715X		L711F			
<i>STX2</i>										R150C						
<i>SMARCA4</i>						R1135L										
<i>SLC4A5</i>										Q1100H						
<i>PIK3CA</i>		W498L														
<i>MBD1</i>								Q403H								
<i>MAP2K1</i>							E233K									
<i>KEAP1</i>	515_519del															
<i>FAT1</i>										P546L						
<i>ERBB2</i>	R978H															
<i>EGFR</i>													746_750del*			
<i>CDK12</i>			S288R													
<i>ARID1A</i>													P1592fX			
<i>ARHGAP35</i>		P1038H														
<i>APC</i>													P421T			
<i>PTEN</i>			K13E													
<i>COBL</i>								E480D								
<i>ACVR1B</i>								R485X								
<i>SHROOM1</i>								P300H								
<i>NKX2-1</i>		A154fX		P232L		V235G										
<i>SFTPC</i>		Splice site					Splice site									

[#], patients with severe emphysema; ^â, patients with UIP. *KMT2C* multiple*: E5X, G568C and S2350Y; *NF1* multiple*: S521C and A1453fX. UIP, usual interstitial pneumonia.