Original Article

Genetic Alterations in Preinvasive Lung Synchronous Lesions

Soyeon Ahn, PhD[1]
Jisun Lim, PhD[1]
Soo Young Park, PhD[2]
Hyojin Kim, MD, PhD[2]
Hyun Jung Kwon, MD[2]
Yeon Bi Han, MD[2]
Choon-Taek Lee, MD, PhD[3]
Sukki Cho, MD, PhD[4]
Jin-Haeng Chung, MD, PhD[2]

[1]*Division of Statistics, Medical Research Collaborating Center, Departments of* [2]*Pathology and Translational Medicine,* [3]*Internal Medicine, and* [4]*Thoracic and Cardiovascular Surgery, Seoul National University Bundang Hospital, Seongnam, Korea*

Correspondence: Jin-Haeng Chung, MD, PhD
Department of Pathology, Seoul National
University Bundang Hospital, 82 Gumi-ro
173beon-gil, Bundang-gu, Seongnam 13620, Korea
Tel: 82-31-787-7713
Fax: 82-31-787-4012
E-mail: chungjh@snu.ac.kr

Co-correspondence: Soyeon Ahn, PhD
Division of Statistics, Medical Research
Collaborating Center, Seoul National University
Bundang Hospital, 82 Gumi-ro 173beon-gil,
Bundang-gu, Seongnam 13620, Korea
Tel: 82-31-787-8121
Fax: 82-31-787-4825
E-mail: ahnsoyeon@snubh.org

Received April 14, 2020
Accepted June 4, 2020
Published Online June 5, 2020

Purpose

Despite advances in treatment, lung cancer remains the leading cause of cancer mortality. This study aimed to characterise genome-wide tumorigenesis events and to understand the hypothesis of the multistep carcinogenesis of lung adenocarcinoma (LUAD).

Materials and Methods

We conducted multiregion whole-exome sequencing of LUAD with synchronous atypical adenomatous hyperplasia (AAH), adenocarcinoma *in situ*, or minimally invasive adenocarcinoma of 19 samples from three patients to characterize genome-wide tumorigenesis events and validate the hypothesis of the multistep carcinogenesis of LUAD. We identified potential pathogenic mutations preserved in preinvasive lesions and supplemented the finding by allelic variant level from RNA sequencing.

Results

Overall, independent mutational profiles were observed per patient and between patients. Some shared mutations including epidermal growth factor receptor (*EGFR* , p.L858R) were present across synchronous lesions.

Conclusion

Here, we show that there are driver gene mutations in AAH, and they may exacerbate as a sequence in a histological continuum, supporting the Darwinian evolution model of cancer genome. The intertumoral and intratumoral heterogeneity of synchronous LUAD implies that multi-biomarker strategies might be necessary for appropriate treatment.

Key words

Adenocarcinoma of lung, Atypical adenomatous hyperplasia, Mutation, Clonal evolution, Whole exome sequencing

Introduction

Although molecular targeted therapies and immune checkpoint inhibitors have markedly improved treatment outcomes in lung cancer, it remains the leading cause of cancer deaths [1]. Lung adenocarcinoma (LUAD) is the most common subtype of lung cancer, accounting for approximately 28%-50% of all cases [2,3], and its incidence is continuously increasing. LUAD is often diagnosed at an advanced stage, leading to a poor prognosis. Multiregional sequencing in lung cancer showed high degree of intratumoral heterogeneity [4], highlighting that understanding the processes of LUAD genome evolution by accumulating somatic mutations over time is important for the early diagnosis and prevention of LUAD.

LUAD with ground-glass/lepidic feature is hypothesised

to follow a multistep tumorigenesis starting from atypical adenomatous hyperplasia (AAH), to adenocarcinoma *in situ* (AIS), to minimally invasive adenocarcinoma (MIA), and finally to invasive or lepidic-predominant adenocarcinoma (ADC) as a histologic stepwise continuum [5,6]. AAH is the only reported precursor lesion to LUAD and is occasionally discovered in the surgically resected lung tissue harboring lung cancer [6]. However, studies on AAH are limited due to its rarity and small lesion size.

The first targeted sequencing study on AAH and paired ADC lesions reported increasing mutational abundance of synchronous epidermal growth factor receptor (*EGFR*), *KRAS*, and *TP53* mutations in the tumor [7]. Another targeted sequencing study combined with transcriptome analysis suggested that exclusive pathways in the driver gene *BRAF* or *KRAS* mutate and initiate the progression of precursor lesions to malignancy [8]. Recently, we reported targeted deep sequencing in sequential lesions [9]. The increased proportions of overall mutated lesions in advanced lesions and shared mutations of *EGFR* between synchronous lesions implied a linear stepwise progression of LUAD [10]. However, the interlesional and intralesional heterogeneity reported in previous AAH studies and the analysis being limited to the focused gene list make it difficult to understand the overall genetic alteration events.

This study aimed to characterize genome-wide tumorigenesis events and to elucidate the hypothesis of the multistep carcinogenesis of LUAD. Towards this goal, we conducted multiregion whole-exome sequencing (WES) of LUAD and preinvasive lesions and paired normal tissue samples.

Materials and Methods

1. Sample selection

We used patients' specimens collected in the previous targeted sequencing study [9]. Histologic slides from patients who underwent wedge resection or lobectomy were stained with haematoxylin and eosin for routine pathologic diagnosis. From each formalin-fixed, paraffin-embedded (FFPE) tissue, 10 μm-thick sections were cut for DNA extraction after minimum trimming, and pathologists reviewed cases and microdissected the area containing > 60% neoplastic cells. Three patients (P1, P5, and P8) with complete AAH-AIS-MIA-ADC sequences of synchronous lesions were chosen. For WES, the remaining extracted DNA from the FFPE tissues were prepared. The characteristics of the entire paired samples for WES, targeted sequencing, and RNA sequencing are summarized in S1 Table.

Information regarding patient characteristics and sample collections was detailed in the previously reported paired targeted sequencing study [9].

2. DNA and total RNA preparation

For RNA sequencing (RNA-seq), total RNAs were extracted from additional FFPE tissue sections from the same patients, and cDNA were synthesized according to the manufacturer's protocol. Non-tumorous samples were used as controls for RNA-seq. Genomic DNA was extracted using the Maxwell (R) 16 FFPE Plus LEV DNA Purification Kit (Promega, Mannheim, Germany) and quantified with PicoGreen dsDNA quantitation assay (Molecular Probes, Eugene, OR). All samples passed the in-house quality control criteria of next-generation sequencing library. The library preparation was performed through Agilent SureSect V5 (Agilent Technologies, Santa Clara, CA) and TruSeqProtocol with TruSeq Exome Enrichment (Illumina, San Diego, CA). DNA sequencing and RNA-seq was performed using an Illumina HiSeq 2500 with 100 bp×2 paired-end reads.

3. Somatic DNA variant calling

WES variants were called with three different caller strategies: Genomon2 pipeline, Mutect (ver. 1.1.4), and MuTect2 (S2 Fig.). Genomon2 pipeline with default parameters identified any potential somatic mutation if Fisher exact test resulted in a p-value < 0.05. For Mutect and MuTect2, raw reads were aligned using BWA (ver. 0.7.12) and then preprocessed using GATK (ver. 32.6) per the best practices recommended by the Broad Institute. To maximize sensitivity, MuTect2 was run with a low cut-off (--max_alt_alleles_in_normal_count 10000000 --max_alt_allele_in_normal_fraction 0.10). Single-nucleotide variants (SNVs) were filtered with a minimum read depth of 20 in the lesions and the matched control, and variant allele frequency (VAF) of being greater than 4% in any lesions and less than 1% in the matched control was set as cut-off values. The 4% cut-off was decided based on the paired comparison between the WES samples and the targeted sequencing samples (S2 Fig.). Variants were annotated via ANNOVAR (ver. 2016-02-01) or Cancer Genome Interpreter [11], and only exonic variants were further filtered. Read depth was obtained via bam-readcount (ver. 0.8.0). Insertions and deletions (indels) were identified with the same read-depth and VAF filter criteria and manually reviewed using the Integrative Genomics Viewer.

4. Copy number analysis

Copy number analysis was performed using Sequenza (ver. 2.1.2) and Excavator2. Segmental somatic copy number alterations (CNAs) were defined according to the intersection between Sequenza (filtered by Bayes factor > 0.3) and Excavator2 (filtered by call probability > 0.9).

5. Inferred clonal tree construction

An individual inferred clonal tree was constructed based on the VAF matrix with SNV as rows and samples as columns. Clonal sequences rather than multiregion trees were

used to avoid biased inference of the underlying subcolonal structures [12]. For each patient, two VAF and read-depth matrices (size of [no. of SNVs]×number of lesions) were decomposed into a genotype matrix (size of [no. of SNVs]× no. of clones) and a clone frequency matrix (size of [no. of clone]×no. of lesions) using an Expectation-Maximization algorithm. The algorithm required a fixed clone number, which we set as four. As a result, the genotype matrix determined the clonal membership of each variant, and each clone was linked to lesions according to the estimated clone frequencies. An unrooted phylogenetic tree was drawn from the inferred clones based on the minimum evolution algorithm. The Clomial and ape R packages were used. LUAD driver genes taken from Bailey et al. [13] were depicted in the trees.

6. Pathway analysis/functional annotation

The mutated genes were investigated further via pathway enrichment analysis using REACTOME (ver. 66, https://reactome.org). The pathogenic status of mutations of driver genes were reviewed via the NCBI ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/, accessed on November 18, 2018).

7. Mutational signature analysis

Somatic mutational signatures were generated and compared to the 30 known mutational signatures in the Catalogue of Somatic Mutations in Cancer database (http://cancer.sanger.ac.uk/cosmic/signatures) using deconstructSigs R package. It quantified the linear combination of well-annotated Catalogue of Somatic Mutations in Cancer (COSMIC) signatures from a single sample input. SNVs were annotated with one of 96 trinucleotide-context substitutions, and the prevalent mutation signatures were illustrated as lego plots.

8. Somatic allelic imbalance

Mutational abundance between the genome and the transcriptome was generally consistent, showing that the mutated allele was expressed according to its mutational frequency in the genome. By contrast, somatic allelic imbalance is a deviation of a consistent expression of somatic alleles, and several studies reported that preferentially allelic selections may be associated with the functionality of cancer genomes [14]. We first included genes with a minimum RNA read depth of 20 and fitted a regression model between RNA VAF versus WES VAF. We then defined somatic allelic imbalance using a data-driven binomial model. For each SNV found in WES, unlikely observation of the read number of RNA was calculated given the WES VAF as the expected probability of binomial distribution. Either SOM-L or SOM-E was defined if the adjusted p-value of Hochberg method is < 0.05. The protein-protein interaction graph was drawn using Cytoscape (ver. 3.7.0) with a STRING database (ver. 10.5) plugin.

9. Ethical statement

This study was approved by the Institutional Review Board of Seoul National University Bundang Hospital and all the patients provided written informed consent (IRB No. B-1607/355-301), and all methods were performed in accordance with the relevant guidelines and regulations.

Results

1. Mutational landscape of synchronous LUAD

Initially, 26 samples from three patients were prepared for WES. Of these, four AAH samples that did not pass the quality control and three non-tumorous normal samples were excluded in the analysis. In total, 19 samples (5 AAH, 3 AIS, 4 MIA, 4 ADC, and 3 matched lymph node controls) were included. WES was conducted with an on-target average depth of 217× (range, 185 to 268) per sample. We found an average of 205 exonic mutations (range, 45 to 682) per sample, and 63% (range, 51% to 70%) of these were nonsynonymous somatic mutations (S2 and S3 Figs.). On average, we identified 251 mutations in smokers (P1 and P8) and 66 mutations in a nonsmoker (P5), consistent with previous studies showing that smoking contributed to higher mutational burden [15]. There was no increasing pattern in tumor mutational burden across sequential lesions (75, 341, 196, and 275 variants in AAH, AIS, MIA, and ADC, respectively), but the VAF tended to increase from preinvasive to invasive lesions (6.1%, 9.7%, 9.2%, and 15.2%) (S4 Fig.). A total of 2,198 genes were affected in all patients, and 172 genes were mutated in more than one patient. Mutations were largely private (82%), that is, mutations were observed in a single sample. Among the driver genes in LUAD, *EGFR* had mutated across two patients (P1 and P5). Moreover, other frequently mutated genes in LUAD had recurrently mutated within one patient (*TP53* for P1 and *KRAS* and *BRAF* for P8) (Table 1). Our data also showed that mutations in *KRAS* and *EGFR* genes are mutually exclusive [16], while *TP53* and *EGFR* genes are commutated [17]. However, we observed concomitant *KRAS* and *BRAF* mutations, which were in contrast to previous studies [8,18]. On average, 62%, 61%, and 65% of the mutations were annotated as passenger mutations in P1, P5, and P8, respectively. Similarly, the proportion of passenger mutations or not protein-affecting mutations were 92%, 89%, and 94%, respectively, which implies that most of mutations in smokers were non-driver mutations.

2. Interlesional heterogeneity

To characterize interlesional heterogeneity, multiregional VAF distributions were mapped (Fig. 1). We found some shared mutations that existed in multiple samples, including *EGFR* mutation (encoding p.L858R) in P5 with increasing mutational abundance throughout the sequential histologi-

Table 1. Variant allele frequencies of driver gene mutations

Gene	Amino acid change	Clinical significance (CliniVar ID)	P1								P5				P8			
			AAH1	AAH2	AAH3	AIS	MIA1	MIA2	ADC1	ADC2	AAH	AIS	MIA	ADC	AAH	AIS	MIA	AD
EGFR	p.G719D	-	-	-	-	-	-	-	-	9	-	-	-	-	-	-	-	-
	p.G719C	Pathogenic (45225)	-	-	-	-	-	-	-	-	-	17	-	-	-	-	-	-
	p.L858R	Drug response (16609)	-	-	-	-	-	-	-	-	20	0.5[a]	15	30	-	-	-	-
TP53	c.240_240delinsT		-	-	-	-	13	-	-	-	-	-	-	-	-	-	-	-
	p.Q60X (stopgain)		-	-	-	-	-	-	42	43	-	-	-	-	-	-	-	-
KRAS	p.Q61H	Pathogenic/likely pathogenic (177881)	-	-	-	-	-	-	-	-	-	-	-	-	7	-	-	-
	p.G13C	Pathogenic (45123)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	8
	p.G12V	Pathogenic (12583)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15	-
BRAF	p.G466V	Pathogenic/likely pathogenic (13967)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	15
SETD2	p.G464R	Pathogenic (279992)	-	-	-	-	-	-	-	-	-	-	-	-	-	4	16	-
	p.P2124Q		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	p.R441Q		-	5	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CTNNB1	p.D115V		-	-	-	-	12	11	-	-	-	-	-	-	-	-	-	-
NF1	p.N1388K		-	-	-	-	-	-	-	-	5	-	-	-	-	-	-	-
	p.Y505X		-	-	-	-	-	-	-	-	-	-	8	-	-	-	-	-
RBM10	c.2337_2337delins-G		-	-	-	-	27	13	-	-	-	-	-	-	-	-	-	-
	p.G780V		-	-	-	-	28	14	-	-	-	-	-	-	-	-	-	-
RBM23	p.A355A		-	-	-	-	-	-	-	-	20	19	22	-	-	17	15	19
SMARCA4	p.E1133E		-	-	-	12	-	-	-	-	-	-	-	-	-	-	-	-
U2AF1	p.S34F	Likely pathogenic (376025)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	46
MUC4	p.P1952S		10	7	11	8	8	7	17	-	-	-	-	-	-	-	-	-
	p.N2437D		-	-	-	-	-	-	-	-	-	-	-	-	17	-	16	32
	p.P540P		-	-	-	-	-	-	-	-	-	5	-	-	-	-	-	-
	p.G154R		-	-	-	-	-	-	-	-	-	-	-	-	-	4	-	-
HLA-DQB2	p.G250S		-	-	-	-	-	-	-	-	6	20	11	16	-	-	-	-
	p.R247H		-	-	-	-	-	-	-	-	-	-	-	-	5	20	8	14
ALK	p.H755A		-	-	-	11	-	-	-	-	-	-	-	-	-	-	-	-

Values are presented as percentage. AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma *in situ*; MIA, minimally invasive adenocarcinoma; ADC, adenocarcinoma. [a]Epidermal growth factor receptor (*EGFR*) mutation (p.L858R) in P5 AIS was not called because it failed to pass the cut-off of 4%.
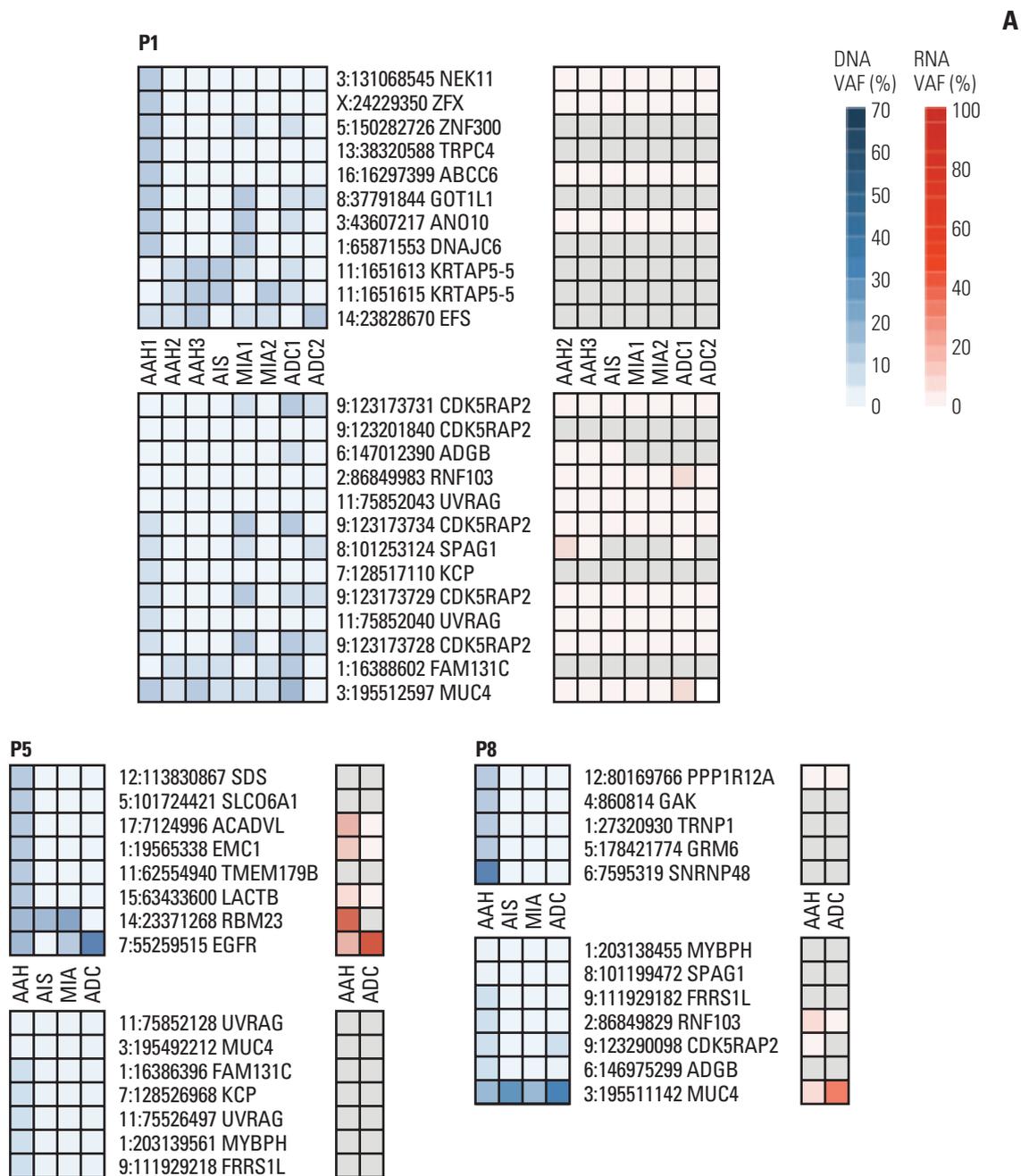
Fig. 1. Multiregional variant allele frequency. Blue and red colours denote the heatmap of variant allele frequency (VAF) of the genome and the transcriptome, respectively. The upper panel shows the private mutations, while the lower panel shows gene mutations found in at least two patients. (A) Atypical adenomatous hyperplasia (AAH). *(Continued to the next page)*

cal continuum. *MUC4* in P1 (p.P1952S) and P8 (p.N2437D), and *HLA-DQB2* in P8 (p.G250S and p.R247H) also showed increase VAF in advanced lesions (Table 1). However, the multiregional heatmaps in the overall population indicated few overlapping mutations (S5 Fig.). Indels of *DNAH7*, *NDUFA10*, and *WDR88* recurrently occurred in more than one patient, and 129 detected indels were mutated within one patient.

3. Intrapatient heterogeneity

To explore intrapatient heterogeneity, a clone relationship was inferred by estimating the hidden clonal status using observed VAFs (Fig. 2). Associated driver genes were portrayed on the inferred clonal trees. In P1, one clone containing *TP53* mutation (c.240_240delinsT-) was detected in AIS, and an *EGFR* mutation (p.G719D) was detected in ADC. Another clone in P1 lacked *TP53* (p.Q60X) in ADC. In P5,

B

P1

| | 7:103417061 RELN |
| 12:22487153 ST8SIA1 |
| X:119211094 RHOXF2 |
| X:3235351 MXRA5 |
| X:101092795 NXF5 |
| 11:66043284 RAB1B |
| X:152818702 ATP2B3 |
| 7:149430956 KRBA1 |
| 5:175111129 HRH2 |
| X:151869952 MAGEA6 |
| 7:93073022 CALCR |
| 3:97851995 OR5H1 |
| 4:41635629 LIMCH1 |
| X:41332850 NYX |
| X:117959939 ZCCHC12 |
| X:23945401 CXorf58 |
| 14:75514332 MLH3 |
| X:34148631 FAM47A |
| 4:155719248 RBM46 |
| X:53115105 TSPYL2 |
| 7:143632726 OR2F2 |
| X:21675566 KLHL34 |
| X:16189465 MAGEB17 |
| 22:26164879 MYO18B |
| 18:39593560 PIK3C3 |
| X:21675565 KLHL34 |
| 7:146818093 CNTNAP2 |
| 1:198697487 PTPRC |
| 1:198697488 PTPRC |
| 7:18674364 HDAC9 |
| 1:198697486 PTPRC |
| 22:43830988 MPPED1 |
| X:153036067 PLXNB3 |
| 7:16566580 LRRC72 |
| 22:26231273 MYO18B |
| X:153297824 MECP2 |
| 7:20825135 SP8 |

AAH1 AAH2 AAH3 AIS MIA1 MIA2 ADC1 ADC2

AAH2 AAH3 AIS MIA1 MIA2 ADC1 ADC2

DNA VAF (%) — 70, 60, 50, 40, 30, 20, 10, 0

RNA VAF (%) — 100, 80, 60, 40, 20, 0

| 1:156917725 ARHGEF11 |
| 12:49434924 KMT2D |
| 10:99337563 ANKRD2 |
| 1:17256526 CROCC |
| 19:31768540 TSHZ3 |
| 3:195512597 MUC4 |
| 21:45713050 AIRE |
| 6:112453982 LAMA4 |
| 8:100829893 VPS13B |
| 14:52507388 NID2 |
| 13:96293918 DZIP1 |
| 10:15614281 ITGA8 |
| 1:17270767 CROCC |
| 18:54398768 WDR7 |
| 17:18022561 MYO15A |
| 1:39789920 MACF1 |
| 1:183083854 LAMC1 |
| 11:47463219 RAPSN |
| 5:175690489 SIMC1 |
| 7:151860195 KMT2C |
| 12:53216827 KRT79 |

Fig. 1. *(Continued from the previous page)* (B) Adenocarcinoma *in situ* (AIS). *(Continued to the next page)*
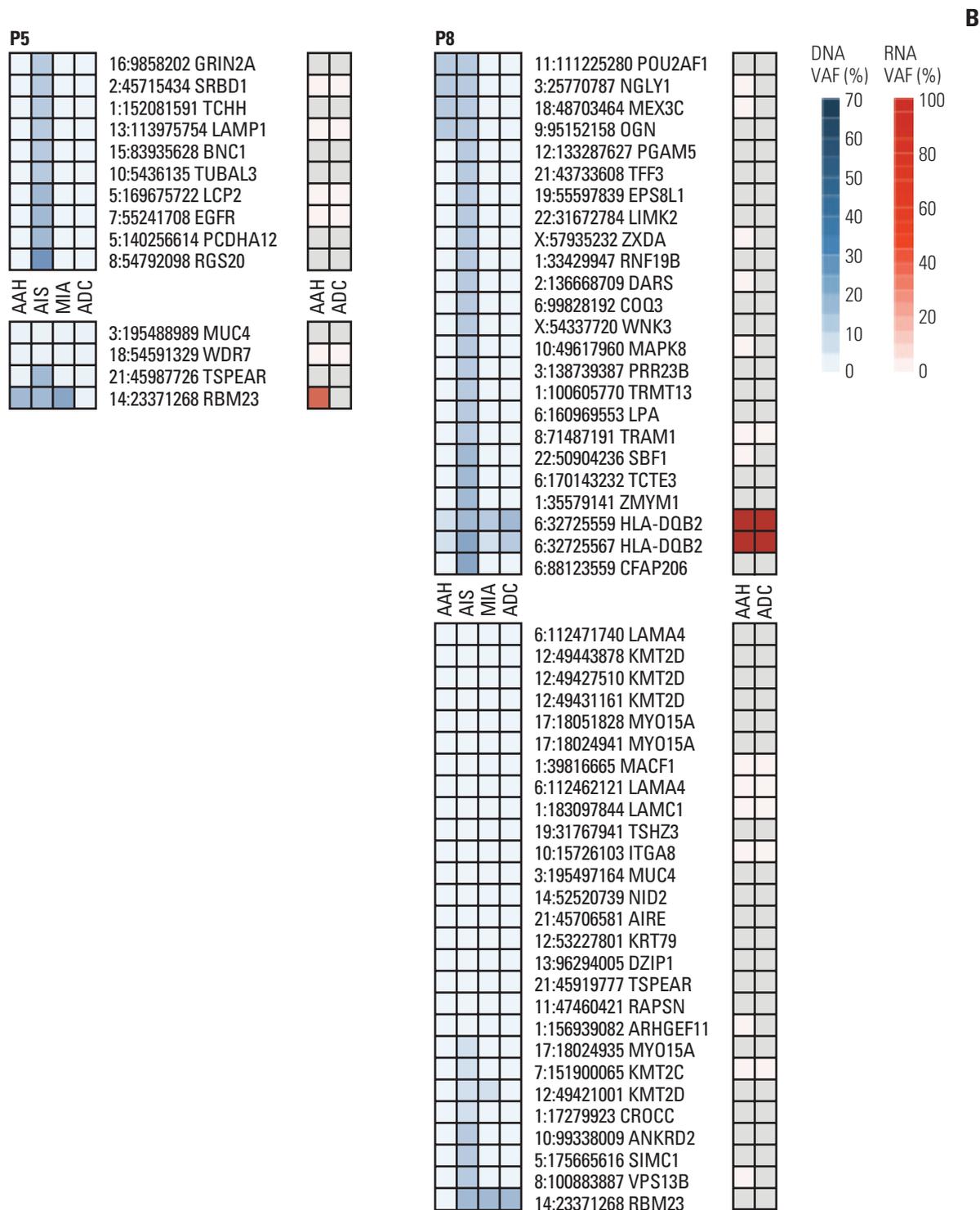
Fig. 1. *(Continued from the previous page) (Continued to the next page)*

EGFR mutation (p.L858R) appeared pre-dominant at the root, and other clones were divergent from each other. Only one clone included *NF1* mutation (p.N1388K) in AAH. In P8, however, multiple clones simultaneously had pathogenic or likely pathogenic mutations. One clone had *KRAS* mutation (p.G12V) coupled with *BRAF* mutation (p.G464R) in MIA, and another clone had *KRAS* (p.G13C) and *BRAF* (p.G466V) in ADC.

CNAs were recurrently observed in 1p13.3, 3q13.33, 4q13.2, 5q35.1, 5q35.3, 8q23.1, 10q21.3, 11p15.5, 14q11.2, 14q13.2, 16p-

Fig. 1. *(Continued from the previous page)* (C) Minimally invasive adenocarcinoma (MIA). *(Continued to the next page)*

13.3, 19q13.42, 20p13, and 22q11.23 (S6 Fig.).

4. Mutation signature analysis

We observed that C:G>A:T transitions were more dominant types in smokers (41% in P1, 51% in P8) than in nonsmokers (30% in P5). Meanwhile, C:G>T:A transitions were more frequent in nonsmokers (38% in P5, 32% in P1, and 27% in P8) [15]. We then further performed the mutation base-

substitution signatures (Fig. 3, S7 Fig.) [19]. Overall, signatures associated with smoking (signature 4 [smoking] and 29 [tobacco] chewing) and aging (signature 1) were highly enriched. Signature 1, composed of C:G>T:A single-base substitutions at CpG sites, was frequently observed in all cancer types. Signatures 4 and 29 were characterized mainly by C:G>A:T mutations with transcriptional strand bias. P1 was a 20 pack-year male current smoker who showed smoking-

Fig. 1. *(Continued from the previous page) (Continued to the next page)*

related signatures, whereas P5 was a never-smoker female and showed abnormalities in DNA maintenance across all lesions (signatures 6 and 15 [DNA mismatch repair (MMR)] and 3 and 20 [defective DNA repair]). P8, a previous smoker male patient with a 30 pack-year history, showed a combination of smoking-related and DNA MMR-related signatures. Exposure to ultraviolet light, which is a mutagenic agent, was common in the AAHs of P1 and P5. The TRACERx study reported a significant relationship between pack-years and smoking-related signature in late clonal mutations, consistent with the enrichment of signature 4 or 29 in our data [20]. Interestingly, differential signature patterns were observed in P1, where signature 4 was dominant in AAH1 and signature 1 was common in AAH2 and AAH3. This suggested that clonal compositions even in preneoplasia lesions could be different, which was also implied in the clonal status analysis (Fig. 2).

5. RNA-seq for allele-specific expression

A total of 14 RNA samples were used in this study. We determined allele-specific expression to compare RNA VAF and WES VAF. A linear regression model showed that RNA VAF was concordant with WES VAF (RNA VAF-1.1×WES VAF; Pearson correlation coefficient, 0.46) (S8 Fig.). Although we did not find similar findings that missense or silent variants in the genome implicated overexpression of certain alleles (S9 Fig.), somatic variant lost in the transcriptome (SOM-L) and somatic variant overexpressed in the transcriptome (SOM-E) variants were discovered in P1, and protein interaction graphs within each lesion type in P1 were derived (S10 Fig.). In AIS, oncogene Yes-associated protein 1 and enzyme ubiquitin-specific peptidase 9, X-linked were found, indicating that preinvasive lesion is related to loss of function in suppressed tumor growth related to the Hippo signaling pathway [21]. In MIA, genes related to mitochondrial outer member, including voltage-dependent anion channel 1 and inner mitochondrial membrane 22, were overexpressed. In ADC, loss alleles of integrin alpha 1 (*ITGA1*) and Versican core protein precursor (*VCAN*) interacted with overexpressed allele of *EGFR* [22]. *ITGA1* and *EGFR* are well-documented prognostic markers, whereas *VCAN* enhances tumor recurrence [23].

In genes in which mutant allelic expression levels were maintained at least 4%, REACTOME analysis revealed that the biological processes of P1, P5, and P8 were regulation of *TP53* expression (*TP53*), fibroblast growth factor receptor 2
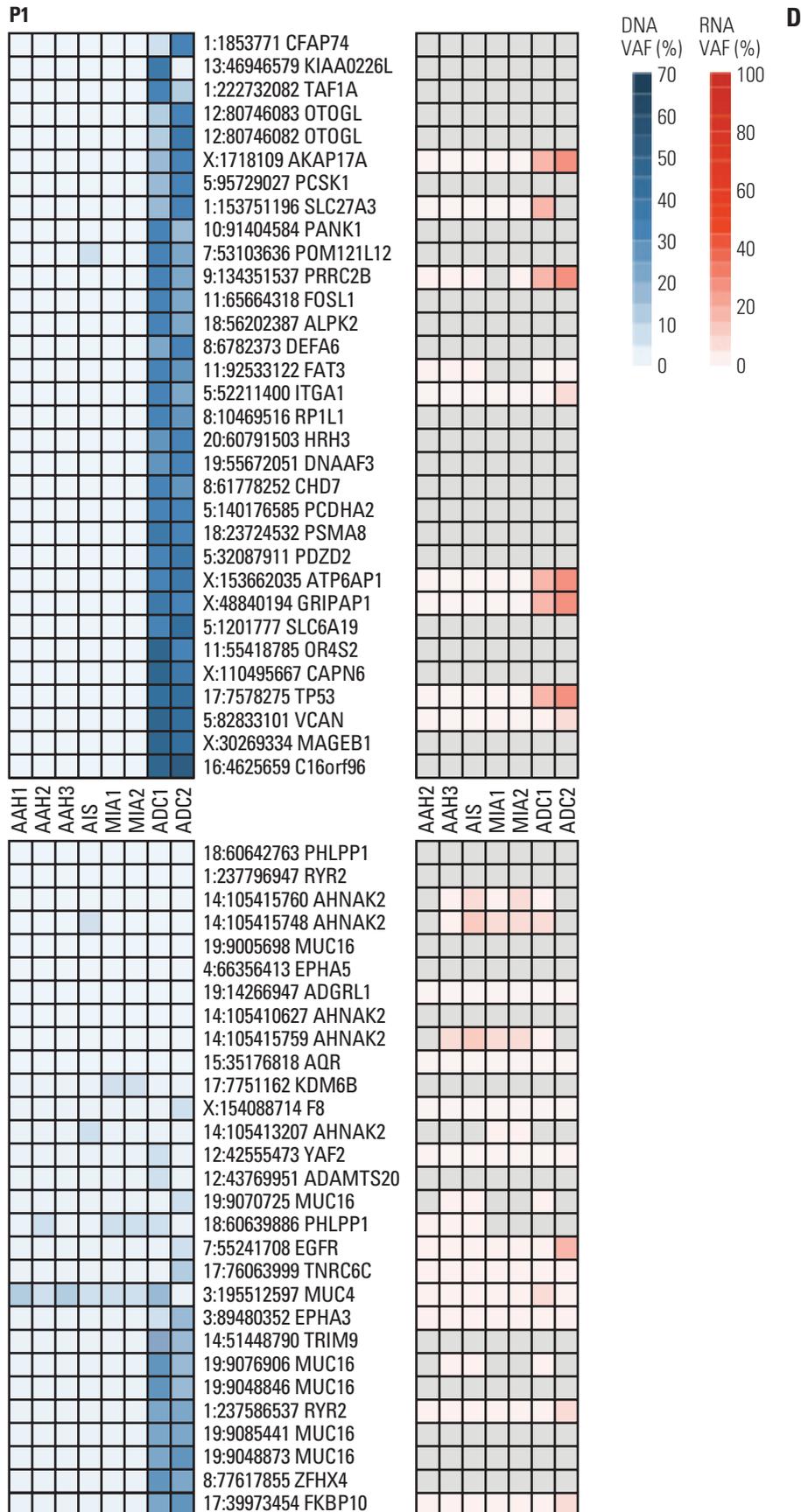
Fig. 1. *(Continued from the previous page)* (D) Adenocarcinoma (ADC). *(Continued to the next page)*
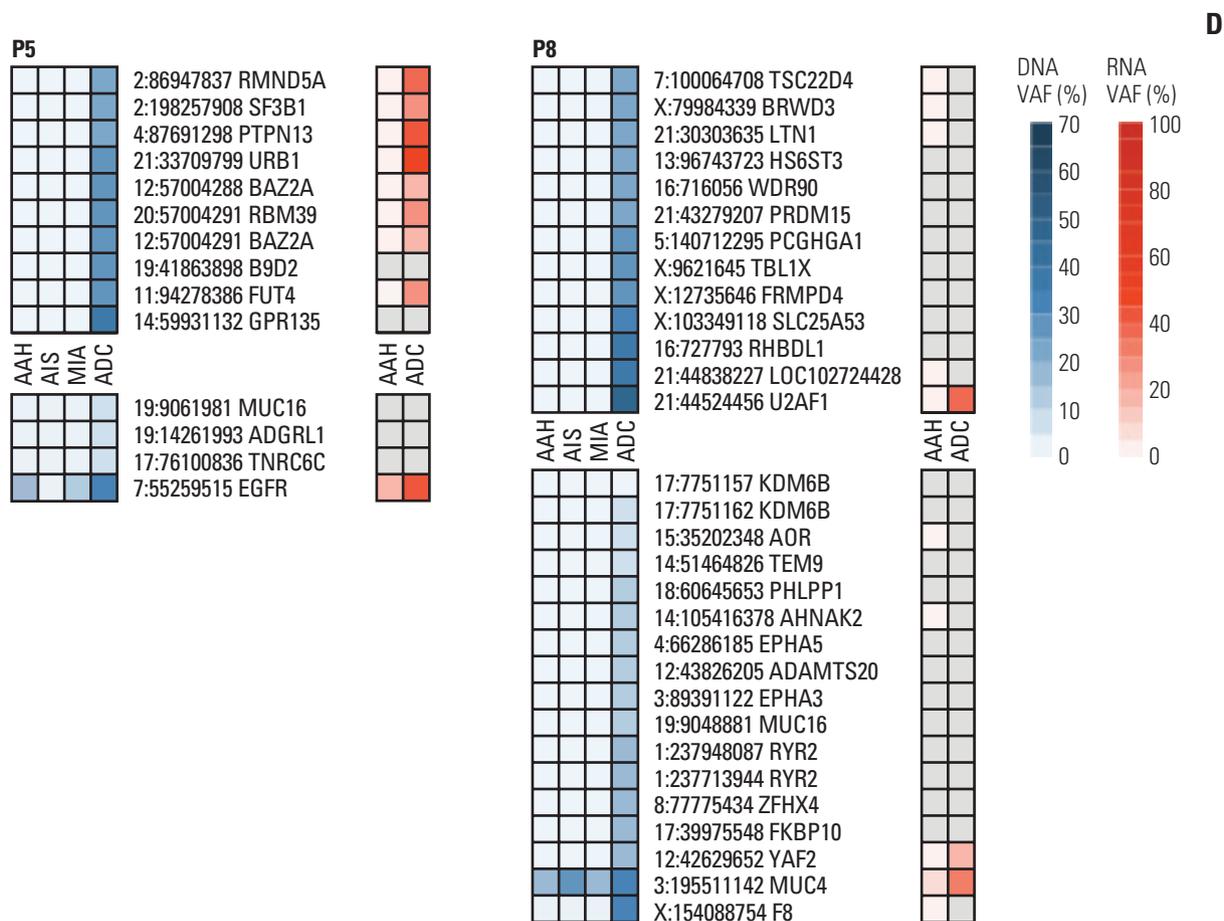
Fig. 1. *(Continued from the previous page)* ADC-N, non-invasive portion of ADC.

Discussion

In this study, we identified the genomic alterations in the precursor lesions of LUAD and inferred clonal evolution in LUAD development through WES supplemented with transcriptome analysis. Shared *EGFR* pathogenic mutation was observed across synchronous lesions, indicating that identical mutations occurred in the early tumorigenesis. Overall increase in VAF but not in tumor burden (mutation number) in invasive lesions indicated that accumulated mutation of certain driver genes is functionally important in cancer development [24]. Furthermore, heterogeneous mutation profiles strongly implied that each lesion underwent largely independent genetic alteration events. Although mutation signature analyses beyond single-gene mutations helped in understanding the combinatory base change mechanism

associated with smoking and aging, the individual pathway was evitable.

The receptor tyrosine kinase/RAS/RAF pathway was frequently mutated and crucial in the development of LUAD [6]. Our results are consistent with those of previous studies on AAH showing the exclusive nature of *EGFR* mutation and *KRAS* mutations, and they were frequently observed in nonsmokers and smokers, respectively. EGFR is a receptor tyrosine kinase belonging to the ERBB family, and its mutations are more widespread in the Asian population and more common in women and nonsmokers. P5, a nonsmoker female Asian patient, harbored the most common pathogenic mutations in *EGFR* (p.L858R) across all lesions, which exemplified the role of *EGFR*-L858R mutation in tumor invasiveness during the early stage of lung cancer. Meanwhile, *KRAS* oncogenes encoding guanosine-5′-triphosphate-binding proteins contribute to invasive lesions when supplementary genomic alternations occur. P8, a male current smoker patient with a 30 pack-year history, had *KRAS* mutation in preinvasive and invasive lesions and *UTAF1* and *BRAF* mutations in invasive lesions. We can further postulate the interaction between *KRAS* and human leukocyte antigen (*HLA*) genes in

P1

AAH2 (3%)
AAH3 (3%)
ADC1 (44%) C3
ADC2 (42%)

CTNNB1 (p.D115V)
RBM10 (p.G780V)

AAH2 (3%)
AAH3 (3%)
MIA1 (33%)
MIA2 (19%)

C2

TP53 (c240_240delinsT-)
SMARCA4 (p.E1133E)
SETD2 (p.R441Q)
EGFR (p.G719D)

TP53 (p.Q60X)

C1

RBM10
(c.2337_2337delins-G)

AAH1 (1%)
AAH2 (3%)
AAH3 (3%)
AIS (13%)
MIA1 (2%)
MIA2 (1%)
ADC1 (2%)
ADC2 (2%)

AAH1 (98%)
AAH2 (90%)
AAH3 (90%)
AIS (86%)
MIA1 (65%)
MIA2 (80%)
ADC1 (54%)
ADC2 (56%)

P5

AIS (3%)
MIA (3%)
ADC (38%) C3

NF1 (p.N1388K)

AAH (14%)
AIS (3%)
MIA (3%)
ADC (1%) C1

EGFR (p.G719C)
RBM10 (p.Y505X)

C2

AIS (3%)
MIA (3%)
ADC (3%)

EGFR (p.L858R)

AAH (86%)
AIS (90%)
MIA (90%)
ADC (58%)

P8

KRAS (p.Q61H)
KRAS (p.G13C)
U2AF1 (p.S34F)
BRAF (p.G466V)

SETD2 (p.P2124Q)

AAH (1%)
AIS (3%)
MIA (3%) C1

C3

AAH (1%)
AIS (3%)
ADC (26%)

C2

AIS (3%)
MIA (30%)

KRAS (p.G12V)
BRAF (p.G464R)

AAH (98%)
AIS (90%)
MIA (66%)
ADC (74%)

Fig. 2. Inferred clone status. For each patient, two variant allele frequency (VAF) and read-depth matrices [(no. of mutations)×(no. lesions)] were decomposed into a genotype matrix [(no. of mutations)×(no. of clones)] and a clone frequency matrix [(no. of clone)×(no. lesions)]. C1-C3 represents inferred clones. Clone frequencies (i.e., the proportion of a clone in each lesion) are shown in parenthesis. Colours matched to the legions where the mutation occurred (orange, atypical adenomatous hyperplasia [AAH]; green, adenocarcinoma *in situ* [AIS]; blue, minimally invasive adenocarcinoma [MIA]; purple, adenocarcinoma [ADC]; red, all lesions). For example, C3 clone in P1 occupied 3% of AAH2 and AAH3, 44% of ADC1, and 42% of ADC2.

this patient because *HLA-DQB2* mutations were prevalent. The specific role of *HLA-DQB2* in cancer genomics was not reported, but immunological function may be affected by mutant neoantigen peptides of hotspot mutations [25].

P1, a 20 pack-year previous smoking male patient, showed

combination pathways related to chromatin remodeling (*SETD2* and *RBM10*) and clinical prognosis (co-mutation of *TP53* and *EGFR*) [26]. We found that recurrent mutations of tubulin tyrosine ligase-like protein 5 (*TTLL5*) appeared in pre-invasive lesions among known co-mutated genes with SET
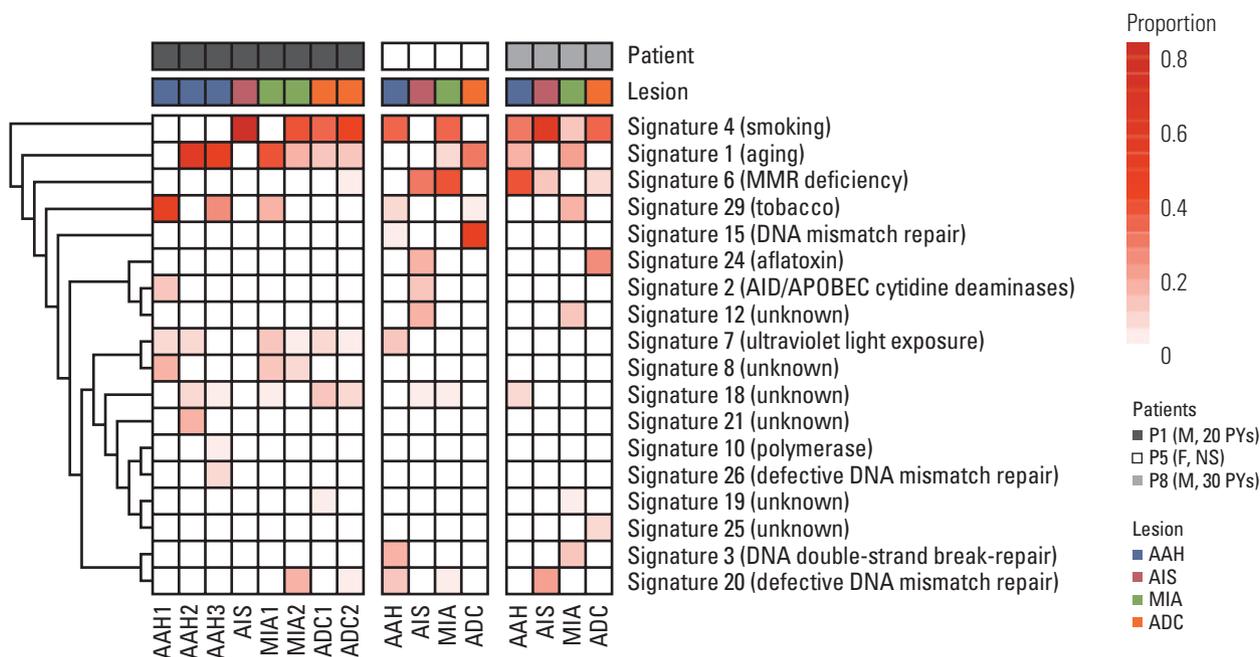
Fig. 3. Mutation signature analysis. Somatic mutational signatures were generated and compared to the 30 known mutational signatures in the Catalogue of Somatic Mutations in Cancer (COSMIC) database. AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma *in situ*; MIA, minimally invasive adenocarcinoma; ADC, adenocarcinoma.

domain containing 2 (*SETD2*), a histone methyltransferase (S12 Table). *SETD2* has been reported to be co-mutated with polybromo 1 (*PBRM1*), particularly in renal cell cancer [27] and in lung cancer [28]. Although WES did not show *PBRM1* mutation in this patient, a previous study using paired targeted sequencing indicated that the patient harbored *PBRM1* mutation (p.K135X) in ADC1 with low frequency (S12 Table, S13 Fig.). We also noted enriched mutations in chromatic-modifying genes (*SMARCA4* [p.E1133E]) and RNA-splicing genes (*RBM10* [c.2337_2337delins-G, and p.G780V]). These findings indicate that accumulated *SETD2* tumor suppressor mutations along with *TTLL5* enzyme mutations in preinvasive lesions exacerbated methylation and chromatic remodeling dysfunction through *SMARCA4*, *RBM10*, and *PBRM1* mutations [29] (S14 Table). The copy number analysis indicated that P1 had aberrant copy number status in invasive lesions (S15 Fig., S16 Table). In a large cohort of 100 early stage lung cancer patients, DNA instability was found to possibly play a key role in cancer malignancy as it can have invasive capacity induced by environmental diversity and can thus be a predictor of clinical outcome [20].

The cancer evolution models are primarily divided into the neutral evolution ('a big bang model') [30] and the Darwinian evolution, depending on the differential clonal selection modes and rates [31]. Darwinian evolution selecting the fittest subclone was further divided into branched, linear, convergent, and parallel evolution. In our study, a linear evolutionary pattern can explain the EGFR-driven selective

sweeps. Our data showed that genes related to epithelial cell proliferation and differentiation (*MUC4*) may be a part of the selective subclonal mechanism. We also conjectured that not only drive specific genomic changes, but also environmental factors contribute to clonal compositions.

The study has limitations. First, because our samples were cross-sectionally obtained at a single time point, the lesions may not be in chronological orders. Second, the patient number was small. Third, our adjusted cut-off value of VAF was extremely lower compared to that in conventional next-generation sequencing studies, although variants with low VAF values are frequently observed in clinical cancer samples [32]. Meanwhile, the major strengths of this study include the series of multiple continuum lesions within a patient and the reproducibility of results through multiple sequencing platforms (WES, targeted, and RNA-seq). We also observed that the mutation signature patterns were consistent to those reported in large-scale studies of early lung cancer patients. Collectively, the variants repeatedly confirmed in our studies may have potential utility for studying neoplastic progression (S17 Table, S18 Fig.).

In summary, we performed a comprehensive analysis of somatic alterations across synchronous lesion mutations and identified the multiple evolutionary trajectories of LUAD rooted in preinvasive lesions toward advanced lesions. We observed few shared somatic mutations and cellular heterogeneity in lung cancer, which implied the independent tumorigenic event within certain genes. The intertumoral

and intratumoral heterogeneity of synchronous LUAD implies that multi-biomarker strategies might be necessary for appropriate treatment decisions. The distinct genetic origin implicates that individualized screening strategies are needed. Our finding implied that genomic variant in *EGFR*, *TP53*, *KRAS*, and *BRAF* could occur early in the process of tumor evolution, and different pathways may involve between a smoker and a non-smoker. The multi-modality tests such as imaging with diagnostic test like cell free DNA testing may lead to identifying certain clonal expansions.

Electronic Supplementary Material

Supplementary materials are available at Cancer Research and Treatment website (https://www.e-crt.org).

Conflicts of Interest

Conflicts of interest relevant to this article was not reported.

References

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2018. CA Cancer J Clin. 2018;68:7-30.
2. Pakkala S, Ramalingam SS. Personalized therapy for lung cancer: striking a moving target. JCI Insight. 2018;3:e120858.
3. Choi CM, Kim HC, Jung CY, Cho DG, Jeon JH, Lee JE, et al. Report of the Korean Association of Lung Cancer Registry (KALC-R), 2014. Cancer Res Treat. 2019;51:1400-10.
4. Zhang J, Fujimoto J, Zhang J, Wedge DC, Song X, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science. 2014;346:256-9.
5. Travis WD, Nicholson AG. WHO classification of tumours of the lung, pleura, thymus and heart. Lyon: International Agency for Research on Cancer; 2015.
6. Inamura K. Clinicopathological characteristics and mutations driving development of early lung adenocarcinoma: tumor initiation and progression. Int J Mol Sci. 2018;19:1259.
7. Izumchenko E, Chang X, Brait M, Fertig E, Kagohara LT, Bedi A, et al. Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. Nat Commun. 2015;6:8258.
8. Sivakumar S, Lucas FA, McDowell TL, Lang W, Xu L, Fujimoto J, et al. Genomic landscape of atypical adenomatous hyperplasia reveals divergent modes to lung adenocarcinoma. Cancer Res. 2017;77:6119-30.
9. Park E, Ahn S, Kim H, Park SY, Lim J, Kwon HJ, et al. Targeted sequencing analysis of pulmonary adenocarcinoma with multiple synchronous ground-glass/lepidic nodules. J Thorac Oncol. 2018;13:1776-83.
10. Noguchi M. Stepwise progression of pulmonary adenocarcinoma: clinical and molecular implications. Cancer Metastasis Rev. 2010;29:15-21.
11. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med. 2018;10:25.
12. Alves JM, Prieto T, Posada D. Multiregional tumor trees are not phylogenies. Trends Cancer. 2017;3:546-50.
13. Bailey MH, Tokheim C, Porta-Pardo E, Sengupta S, Bertrand D, Weerasinghe A, et al. Comprehensive characterization of cancer driver genes and mutations. Cell. 2018;173:371-85.
14. Rhee JK, Lee S, Park WY, Kim YH, Kim TM. Allelic imbalance of somatic mutations in cancer genomes and transcriptomes. Sci Rep. 2017;7:1653.
15. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012;150:1121-34.
16. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014;511: 543-50.
17. Labbe C, Cabanero M, Korpanty GJ, Tomasini P, Doherty MK, Mascaux C, et al. Prognostic and predictive effects of TP53 co-mutation in patients with EGFR-mutated non-small cell lung cancer (NSCLC). Lung Cancer. 2017;111:23-9.
18. Schmid K, Oehl N, Wrba F, Pirker R, Pirker C, Filipits M. EGFR/KRAS/BRAF mutations in primary lung adenocarcinomas and corresponding locoregional lymph node metastases. Clin Cancer Res. 2009;15:4554-60.
19. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. Nature. 2013;500:415-21.
20. Jamal-Hanjani M, Wilson GA, McGranahan N, Birkbak NJ, Watkins TB, Veeriah S, et al. Tracking the evolution of non-small-cell lung cancer. N Engl J Med. 2017;376:2109-21.
21. Valero V 3rd, Pawlik TM, Anders RA. Emerging role of Hpo signaling and YAP in hepatocellular carcinoma. J Hepatocell Carcinoma. 2015;2:69-78.
22. Zheng W, Jiang C, Li R. Integrin and gene network analysis reveals that ITGA5 and ITGB1 are prognostic in non-small-cell lung cancer. Onco Targets Ther. 2016;9:2317-27.
23. Pirinen R, Leinonen T, Bohm J, Johansson R, Ropponen K, Kumpulainen E, et al. Versican in nonsmall cell lung cancer: relation to hyaluronan, clinicopathologic factors, and prognosis. Hum Pathol. 2005;36:44-50.
24. Ma P, Fu Y, Cai MC, Yan Y, Jing Y, Zhang S, et al. Simultaneous evolutionary expansion and constraint of genomic heteroge-

neity in multifocal lung cancer. Nat Commun. 2017;8:823.
25. He XP, Song FJ, Liu XY, Wang Z, Li XX, Liu FY, et al. The relationship between KRAS gene mutations and HLA class I antigen downregulation in the metastasis of non-small cell lung cancer. J Int Med Res. 2013;41:1473-83.
26. Blakely CM, Watkins TB, Wu W, Gini B, Chabon JJ, McCoach CE, et al. Evolution and clinical impact of co-occurring genetic alterations in advanced-stage EGFR-mutant lung cancers. Nat Genet. 2017;49:1693-704.
27. Hakimi AA, Pham CG, Hsieh JJ. A clear picture of renal cell carcinoma. Nat Genet. 2013;45:849-50.
28. Ahn JW, Kim HS, Yoon JK, Jang H, Han SM, Eun S, et al. Identification of somatic mutations in EGFR/KRAS/ALK-nega-

tive lung adenocarcinoma in never-smokers. Genome Med. 2014;6:18.
29. Park IY, Powell RT, Tripathi DN, Dere R, Ho TH, Blasius TL, et al. Dual chromatin and cytoskeletal remodeling by SETD2. Cell. 2016;166:950-62.
30. Sottoriva A, Kang H, Ma Z, Graham TA, Salomon MP, Zhao J, et al. A Big Bang model of human colorectal tumor growth. Nat Genet. 2015;47:209-16.
31. Venkatesan S, Birkbak NJ, Swanton C. Constraints in cancer evolution. Biochem Soc Trans. 2017;45:1-13.
32. Shin HT, Choi YL, Yun JW, Kim NK, Kim SY, Jeon HJ, et al. Prevalence and detection of low-allele-fraction variants in clinical cancer samples. Nat Commun. 2017;8:1377.