



Multomics Analysis Reveals Molecular Changes during Early Progression of Precancerous Lesions to Lung Adenocarcinoma in Never-Smokers

Yun-Ching Chen¹, Chia-Lang Hsu², Hui-Min Wang¹, Shang-Gin Wu³, Yih-Leong Chang⁴, Jin-Shing Chen^{5,6}, Yu-Ching Wu², Yen-Ting Lin³, Ching-Yao Yang⁷, Mong-Wei Lin⁶, Jang-Ming Lee⁶, Shuenn-Wen Kuo⁶, Ke-Cheng Chen⁶, Hsao-Hsun Hsu⁸, Pei-Ming Huang⁶, Yen-Lin Huang⁴, Chong-Jen Yu^{7,9}, Mehdi Pirooznia¹, Bevan E. Huang¹, Rob Yang¹, Jin-Yuan Shih⁷, and Pan-Chyr Yang^{7,10}

ABSTRACT

Lung cancer is the most common cause of cancer mortality globally, and the prevalence of lung adenocarcinoma, the most common lung cancer subtype, has increased sharply in East Asia. Early diagnosis leads to better survival rates, but this requires an improved understanding of the molecular changes during early tumorigenesis, particularly in nonsmokers. In this study, we performed whole-exome sequencing and RNA sequencing of samples from 94 East Asian patients with precancerous lesions [25 with atypical adenomatous hyperplasia (AAH); 69 with adenocarcinoma *in situ* (AIS)] and 73 patients with early invasive lesions [minimally invasive adenocarcinoma (MIA)]. Cellular analysis revealed that the activities of endothelial and stromal cells could be used to categorize tumors into molecular subtypes within pathologically defined types of lesions. The subtypes were linked with the radiologically defined type of lesions and corresponded to immune cell infiltration throughout the early

progression of lung adenocarcinoma. Spatial transcriptomic analysis revealed the distribution of epithelial cells, endothelial cells, fibroblasts, and plasma cells within MIA samples. Characterization of the molecular lesion subtypes identified positively selected mutational patterns and suggested that angiogenesis in the late-stage AIS type potentially contributes to tissue invasion of the MIA type. This study offers a resource that may help improve early diagnosis and patient prognosis, and the findings suggest possible approaches for early disease interception.

Significance: Integrative analysis of multiomics data revealed coordination between immune and nonimmune cells during early progression of precancerous lesions to lung adenocarcinomas and shed light on the molecular characteristics of clinically defined subtypes.

¹Interventional Oncology, Johnson & Johnson Enterprise Innovation, Inc., Boston, Massachusetts. ²Department of Medical Research, National Taiwan University Hospital, Taipei, Taiwan. ³Department of Medicine, National Taiwan University Cancer Center and College of Medicine, National Taiwan University, Taipei, Taiwan. ⁴Department of Pathology, National Taiwan University Cancer Center and College of Medicine, National Taiwan University, Taipei, Taiwan. ⁵Department of Surgery Oncology, National Taiwan University Cancer Center, Taipei, Taiwan. ⁶Department of Surgery, National Taiwan University Hospital, Taipei, Taiwan. ⁷Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, National Taiwan University, Taipei, Taiwan. ⁸Department of Surgery, National Taiwan University Hospital and College of Medicine, National Taiwan University, Taipei, Taiwan. ⁹Department of Internal Medicine, National Taiwan University Hospital Hsin-Chu Branch, Hsinchu, Taiwan. ¹⁰Institute of Biomedical Sciences, Academia Sinica, Taipei, Taiwan.

Y.-C. Chen, C.-L. Hsu, H.-M. Wang, and S.-G. Wu contributed equally to this article and share first authorship.

Corresponding Authors: Rob Yang, Interventional Oncology at Johnson & Johnson, 10th Floor 255 Main Street, Cambridge, Boston, MA 02142. E-mail: yang.robert@gmail.com; Jin-Yuan Shih, Department of Internal Medicine, National Taiwan University Hospital and College of Medicine, National Taiwan University, No. 7, Zhongshan South Road, Zhongzheng District, Taipei 10002, Taiwan. E-mail: jyshih@ntu.edu.tw; and Pan-Chyr Yang, pcyang@ntu.edu.tw

Cancer Res 2025;85:602-17

doi: 10.1158/0008-5472.CAN-24-0821

This open access article is distributed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International (CC BY-NC-ND 4.0) license.

©2024 The Authors; Published by the American Association for Cancer Research

Introduction

Lung cancer is one of the most complex and heterogeneous malignancies and is the leading cause of cancer mortality worldwide (1). Lung adenocarcinoma is the most common pathologic subtype, and the average 5-year overall survival rate of patients with lung adenocarcinoma is 19% (2). However, the 5-year survival of patients with lung adenocarcinoma depends on the stage at diagnosis; it is 100% for those with minimal invasion, about 70% for those with stage I, about 30% for those with locally advanced disease, and about 5% for those with distant metastasis (2–4). This underscores the urgent need to study the early progression of lung adenocarcinoma because the findings of these studies could help improve the early diagnosis and treatment of this cancer.

However, the sensitive detection of early-stage lung cancer is difficult because the pulmonary nodules that are seen on low-dose CT are small masses of tissue that are usually benign and only a small fraction of them are precancerous or cancerous. The radiological appearance of nodules on CT allows their classification as pure ground-glass opacity (pGGO) nodules, partially solid nodules, or solid nodules. Pathologically, the progression of lung adenocarcinoma from a precancerous lesion to an invasive tumor seems to follow a linear multistep process. In particular, this cancer progresses from atypical adenomatous hyperplasia (AAH) to adenocarcinoma *in situ* (AIS) and then to invasive adenocarcinoma, such as minimally invasive adenocarcinoma (MIA) or invasive adenocarcinoma (5, 6). Clinically, AAH and AIS often manifest as pGGO

nodules on CT, and more advanced adenocarcinomas may include larger solid components within the ground-glass opacity (GGO) region (7–9).

Several studies have investigated the molecular profiles of precancerous lung lesions using different types of molecular data, such as targeted sequencing (10–12), whole-exome sequencing (WES; refs. 13, 14), bulk RNA sequencing (RNA-seq; refs. 10, 14), DNA methylation analysis (15), and single-cell RNA-seq (scRNA-seq; refs. 16, 17). Although these studies have greatly expanded our knowledge of the molecular characteristics of lesions during lung adenocarcinoma progression, certain important questions remain. First, the percentage of lung cancers among never-smokers is higher in East Asia (18); these patients are frequently diagnosed with lung adenocarcinoma, and their clinical characteristics are different from smokers with lung adenocarcinoma (18–20). In addition, pGGO lesions are common in lung cancers from nonsmoking patients, and nearly one third of them have confirmed invasive lung adenocarcinoma (21). Many previous studies of different cohorts included mixtures of smokers and nonsmokers (13–15, 22), and there is less information about the etiology of lung adenocarcinoma progression in nonsmokers. Second, many previous studies had small sample sizes or inadequate representation of the different radiological and pathologic types of this cancer (11, 12, 14, 16, 17). A large cohort study that examines AAH, AIS, and MIA and performs an in-depth analysis of clinical phenotypes is needed to provide a more comprehensive understanding of the early progression of lung adenocarcinoma. Finally, although tumor invasion is a hallmark of malignancy, previous studies only provided limited insights on the molecular basis of tissue invasion during early lung adenocarcinoma progression, essentially the transition from AAH and AIS (precancerous lesions) to MIA (cancerous lesions).

In this study, we examined somatic mutations using WES and examined transcriptome profiles using bulk RNA-seq of AAH, AIS, and MIA nodules from a cohort of East Asian patients, most of whom were nonsmokers. We then conducted an integrative analysis by examination of published lung adenocarcinoma scRNA-seq data to characterize the molecular properties of the tumor microenvironment (TME). This approach allowed identification of subtypes within AIS and MIA nodules that had distinct molecular characteristics and were associated with specific radiological types. These findings will provide a better understanding of the molecular changes that occur during lung adenocarcinoma progression from a pGGO lesion to a non-pGGO lesion and may also help determine the mechanism of lung adenocarcinoma progression from a precancerous lesion to an invasive tumor.

Materials and Methods

Patients and samples

From 2017 to 2021, lung adenocarcinoma samples from the data bank of the Department of Pathology, National Taiwan University Hospital, were retrospectively examined. Diagnosis and histologic classification were made according to the World Health Organization (6). Identification of the AAH, AIS, and MIA types were reconfirmed by two experienced pathologists (YLC and CTW). This reconfirmation process included the delineation of cancerous lesions and their differentiation from normal tissues. Moreover, the noninvasive and invasive components of MIA lesions were also identified and distinguished. MIA was defined by limited stromal invasion for less than 5 mm, an invasive component that was not the predominant component in the tumor, and a whole tumor with the

longest axis of 30 mm or less. The noninvasive area was defined by tumor tissue with a lepidic pattern, and the invasive component was defined by other histologic subtypes (i.e., acinar, papillary, etc.) or tumor cells infiltrating the myofibroblast stroma. These criteria align with the World Health Organization Classification of Thoracic Tumors (fifth edition; ref. 6). Only samples with clearly defined noninvasive or invasive segments were selected for further analysis; samples with ambiguous borders between invasive and noninvasive components were excluded. In addition, we only used tissue samples if there was sufficient tissue for subsequent pathologic dissection and genetic testing. The majority of AAH lesions were radiologically classified as pGGO. Surgical excision was only considered for AAH lesions that exceeded 0.8 cm in size or had rapid growth on serial follow-up imaging. The follow-up intervals were based on the Fleischner Society Guidelines and the clinical physician's assessment (23). The average follow-up interval was 6 to 12 months, and treatment decisions were made jointly by the physician and the patient.

Samples from 167 patients were ultimately included. The clinical and demographic characteristics and the imaging reports of these patients were analyzed, including cancer stage, smoking history, chest CT images, surgical interventions, and prognosis. Among the enrolled patients, 56 presented with multifocal lesions and received surgical excision of these lesions either concurrently or as a subsequent procedure. Notably, only one patient harbored an AAH lesion (AAH-013) and an AIS lesion (AIS-016), and each lesion was analyzed independently. This study was approved by the Institutional Review Board of the National Taiwan University Hospital Research Ethics Committee (REC number 201912136RIFC). All enrolled patients signed an informed consent document about the use of tissues for molecular analyses.

Whole-exome sequencing and somatic variant calling

Genomic DNA and total RNA were isolated from formalin-fixed, paraffin-embedded (FFPE) tumor and adjacent normal tissue sections using the truXTRAC FFPE Tissue Extraction Kit (Covaris). DNA quality was assessed using the Agilent NGS FFPE Kit and the Qubit dsDNA HS Assay Kit. RNA quality was assessed using the BioOptic Qsep100 RNA Analyzer and the Qubit RNA HS Assay Kit. The WES library was constructed using the Agilent SureSelectXT-low input reagent. DNA was fragmented using the SureSelect Enzymatic Fragmentation Kit and then ligated using sequencing adapters. The Agilent SureSelect Human All Exon V7 was used for exome capture. Sequencing was performed at Welgene Biotech using the Illumina NovaSeq 6000 platform. Data from sequencing reads were checked for quality and adapter sequence contamination using FastQC (version 0.11.9). Adapter sequences and unqualified bases were trimmed using Cutadapt (version 3.5, RRID: SCR_022192). After trimming, reads with fewer than 50 bases were removed. Qualified reads were aligned to a human reference genome (GRCh38) using BWA-mem (version 0.7.17). The median depth was 173.4 (range: 121.8–404.5). Picard Tools (embedded in GATK 4.2.3.0) were used to sort BAM files and to mark duplicate reads. To prevent interpatient sample swaps, HLA-HD (version 1.5.0) was used to identify alleles with four-digit human leukocyte antigen class I genes in all WES samples; all matched pairs of tumor–normal samples had identical alleles. Somatic mutations were detected using Mutect2 (embedded in GATK 4.2.3.0) and annotated using ANNOVAR (release 20210819, RRID: SCR_012821). To reduce false-positive variant calls, each analyzed variant had a read depth of 50 or more, 5 or more supporting reads,

an allele frequency greater than 0.05, and previous documentation in the COSMIC database (RRID: SCR_002260). Sequenza (RRID: SCR_016662) was used to call copy-number alterations and estimate cell ploidy. Gene-level amplification was called by a mean gene copy number of two or more $\times$ ploidy + one copy, and deletion was called by a mean gene copy number below 0.5.

Transcriptome sequencing and read counts

The RNA-seq library was constructed using the Takara SMARTer-Seq Stranded Kit. Library quality and quantity were assessed using the BioOptic QSep100 DNA Analyzer, the Qubit dsDNA HS Assay Kit, and the KAPA Library Quantification Kit. Sequencing was performed at Welgene Biotech using the Illumina NovaSeq 6000 platform. Data from sequencing reads were checked for quality and adapter sequence contamination using FastQC (version 0.11.9). Adapter sequences and unqualified bases were trimmed using Cutadapt (version 3.5, RRID: SCR_022192). After trimming, reads with fewer than 50 bases were removed. Qualified reads were aligned to a human reference genome (GRCh38) using STAR (version 0.7.17), and gene-level expression was calculated based on annotation from GENCODE (version 35). Only genes that were annotated as protein coding or lncRNA and were expressed (read count >0) in 80% or more of the samples were included. The trimmed mean of M values algorithm from the R package edgeR was used to adjust differences in sequencing throughput among samples, and the batch effect was removed using the removeBatchEffect function in the R package limma (RRID: SCR_010943). A total of 25,227 genes with log₂-transformed transcripts per million were used as the gene expression unit for downstream analyses. ESTIMATE was used to predict tumor purity (24).

Spatial transcriptomics

Spatial transcriptomics data were generated using the 10 $\times$ Visium CytAssist Spatial Gene Expression for the FFPE platform (11 $\times$ 11 mm, Human, PN-1000522). Prior to this analysis, RNA quality was assessed, and only samples with a DV200 greater than 30% were selected for analysis. Additionally, all slides were reviewed by pathologists to ensure that there was sufficient tumor volume. Following the manufacturer's instructions, FFPE slides were mounted onto Visium slides using CytAssist, and libraries were then constructed using the supplied reagents. The libraries were sequenced on the NovaSeq 6000 platform (Illumina), and the data were analyzed using Space Ranger (version 2.0.0). The spatial transcriptomic data were then analyzed using the Seurat package (version 5.1.0). Spots containing fewer than 150 detected genes were excluded from the analysis. For each sample, the gene-spot matrix was normalized using the SCTransform package (25), and dimensionality reduction was performed by principal component analysis (PCA). The first 15 principal components (PC) were used to construct nearest-neighbor plots, and the Louvain algorithm was used to identify clusters with a resolution of 1. The xCell package (26) was used to calculate the enrichment score of cell types in each spot.

Analysis of differentially expressed genes

The analysis of differentially expressed genes (DEG) was performed using the lmFit function and the eBayes function (trend = TRUE) in R software (version 4.0) using limma (version 3.46; RRID: SCR_010943) by fitting to the following linear model:

$$Gene \sim PathType + GGO + EGFR + Covariates \quad (A)$$

in which *Gene* is the gene expression in log₂(transcripts per million), *PathType* is a categorical predictor indicating tissue-based pathologic type [adjacent normal (baseline), AAH, AIS, the noninvasive part of a MIA lesion (MIA_nonInv), or the invasive part of a MIA lesion (MIA_Inv)], GGO is a binary predictor (0 for pGGO and 1 for non-pGGO), *EGFR* is a binary predictor (0 for wild type and 1 for any mutation), and the covariates were age, smoking status, gender, tumor size (TS), and surrogate variable(s). The surrogate variables were estimated using the R package sva (version 3.38; ref. 27).

Genes that had increasing expression among pathologic types satisfied the following two criteria:

Nominally significant difference from normal:

$$P_{AAH} < 0.05 \text{ and } P_{AIS} < 0.05 \text{ and } P_{MIA_nonInv} < 0.05 \text{ and } P_{MIA_Inv} < 0.05 \quad (B1)$$

Increased expression from normal to AAH, to AIS, and to MIA (MIA_nonInv and/or MIA_Inv):

$$(\beta_{MIA_Inv} \text{ and/or } \beta_{MIA_nonInv}) > \beta_{AIS} > \beta_{AAH} > 0 \quad (B2)$$

in which *P* is the *P* value and β is the β -coefficient in the linear model for a gene.

Genes that had decreasing expression among pathologic types satisfied the following two criteria:

Nominally significant difference from normal:

$$P_{AAH} < 0.05 \text{ and } P_{AIS} < 0.05 \text{ and } P_{MIA_nonInv} < 0.05 \text{ and } P_{MIA_Inv} < 0.05 \quad (C1)$$

Decreased expression from normal to AAH, to AIS, and to MIA (MIA_nonInv and/or MIA_Inv):

$$(\beta_{MIA_Inv} \text{ and/or } \beta_{MIA_nonInv}) < \beta_{AIS} < \beta_{AAH} < 0 \quad (C2)$$

in which *P* is the *P* value and β is the β -coefficient in the linear model for a gene.

The following additional linear model was fitted by adding interactions between pathologic types and radiological types (*PathType* $\times$ *GGO*) to examine gene associations with radiological types within each pathologic type. Specifically, the *P* value distribution of *PathType* $\times$ *GGO* was plotted for the AIS, MIA_nonInv, and MIA_Inv types:

$$Gene \sim PathType + GGO + EGFR + PathType * GGO + PathType * EGFR + Covariates \quad (D)$$

The paired differential analysis between MIA_Inv and MIA_nonInv was performed by identification of gene expression differences between matched MIA_Inv and MIA_nonInv samples using the following linear model. In total, 96 genes were upregulated in MIA_Inv compared with MIA_nonInv (intercept > 0, nominal *P* < 0.005).

$$Gene_{diff} \sim GGO + EGFR + Covariates \quad (E)$$

Identification of markers of different cell types

The markers of different types of cells were computed from two public lung cancer scRNA-seq datasets: Sinjab and colleagues (28) and Salcher and colleagues (29). For analysis of data from Sinjab and colleagues, all cells were used because these data were only for cells from patients with lung adenocarcinoma (including cells from

adjacent normal tissues). For analysis of data from Salcher and colleagues, only cells from lung adenocarcinoma or normal tissues were used, and a small fraction of cells with ambiguous cell types (annotated as “other” in cell_type_major) were excluded. To assure that the cell types were consistent between the two datasets, the original cell-type annotations were converted to the same annotations at granularities of L1 and L2 (Supplementary Table S1). The markers of different cell types were the top 50 DEGs of each L2 cell type determined using the function FindAllMarkers in the R package Seurat (version 4.1). Supplementary Table S2 provides the full list of these markers.

Confirmation of specificity of cell-type markers in precancerous scRNA-seq data

To confirm specificity of cell-type markers derived from lung adenocarcinoma scRNA-seq data in the precancerous environment, the scRNA-seq data from cells that were collected from three AIS nodules and three MIA nodules were downloaded (17). Then low-quality cells were filtered out using the following four criteria: (i) read count ≤ 500 ; (ii) gene count < 200 ; (iii) gene count $> 10,000$; and (iv) the percentage of mitochondrial genes ≥ 10 . Because cell annotations were unavailable from the original authors, transfer learning (the functions FindTransferAnchors and TransferData in the R package Seurat version 4.1, with default parameters) was used to annotate the L1 cell types, with the data from Sinjab and colleagues (28) used as the reference. Subsequently, cells from AIS and MIA samples were visualized, and the mean levels of expression of the top 50 DEGs of each L1 cell type identified in the Sinjab and colleagues dataset were inspected on Uniform Manifold Approximation and Projection (UMAP). This analysis revealed that cells from the AIS and MIA samples and from the lung adenocarcinoma samples were generally well-aligned in terms of L1 cell types and that marker genes identified in lung adenocarcinoma samples remained specific to each cell type in the AIS and MIA samples.

Gene set scoring for bulk RNA-seq and scRNA-seq

To compute the gene set score for a bulk RNA-seq sample, patient-specific effects were first removed. Random effects for each patient were regressed out using the function removeBatchEffect in the R limma package (version 3.46; RRID: SCR_010943). For each bulk sample, the gene set score was computed using single-sample gene set enrichment analysis (ssGSEA) that was implemented in the R package GSVA (30). The gene set scores for different cell types were computed using the top 50 markers of each of the L2 cell types from the scRNA-seq datasets. The gene set scores for the L1 cell types were computed as the means of scores of the L2 cell types that were in each L1 cell type. The gene set score for individual cells in the scRNA-seq data [e.g., tip endothelial cell (EC) score for ECs] was computed using AUCell (31).

Clustering of molecular subtypes

The top 20 markers of the nonimmune L2 cell types were used to compute the PCs, and the top 15 PCs were used for subtype clustering. The molecular subtypes were obtained by performing spectrum clustering using the R package Spectrum (version 1.1), and the best number of clusters was determined by this algorithm.

Deconvolution analysis of different cell types

Deconvolution of cell types was performed using TIMER (32) and CLIMB (bioRxiv 2023.05.10.540140). TIMER is a user-friendly web application (<http://timer.cistrome.org>). CLIMB is a recently

published deconvolution method that can deconvolute bulk RNA-seq samples using scRNA-seq data as the reference and outperforms other similar methods. Many cell types analyzed in this study were from scRNA-seq data and were not estimated in TIMER; CLIMB was used to estimate the composition of these cell types.

SCISSOR analysis

The SCISSOR (Single-Cell Identification of Subpopulations with Bulk Sample Phenotype Correlation) analysis (33) focused on ECs and stromal cells. The scRNA-seq data were obtained from Salcher and colleagues (29) because the number of ECs and stromal cells were greater in these data than those in the data of Sinjab and colleagues (28). The ECs included vein ECs, artery ECs, capillary ECs, and lymphatic ECs; the stromal cells included fibroblasts, pericytes, and smooth muscle cells. First, a set of bulk RNA samples was defined as group 0 (e.g., AAH), and another set of bulk RNA samples was defined as group 1 (e.g., AIS3). Then SCISSOR was used to identify associations of the binarized samples with cells in a specific cell population (e.g., ECs) to obtain a set of group 0-associated cells and a set of group 1-associated cells. To test whether the associated cells were enriched for a particular subtype (e.g., tip ECs), all cells were first scored using subtype markers with AUCell, as previously described (31). Then a one-sided Kolmogorov-Smirnov test was applied for group 0- (or group 1-) associated cells against the union of these associated cells. A significant enrichment or depletion was based on a P value below 0.05. Previously reported congruent tip EC markers (34) and the cancer-associated fibroblast (CAF) subtype markers (35) were used in this analysis.

Statistical analysis

Unless specified otherwise, the Wilcoxon rank-sum test was used for analysis of the association between a binary variable and a continuous variable, and the hypergeometric test was used to test for enrichment in analysis of two binary conditions. A P value less than 0.05 was used to define statistical significance. The Pearson correlation coefficient was used to calculate the correlations of continuous variables.

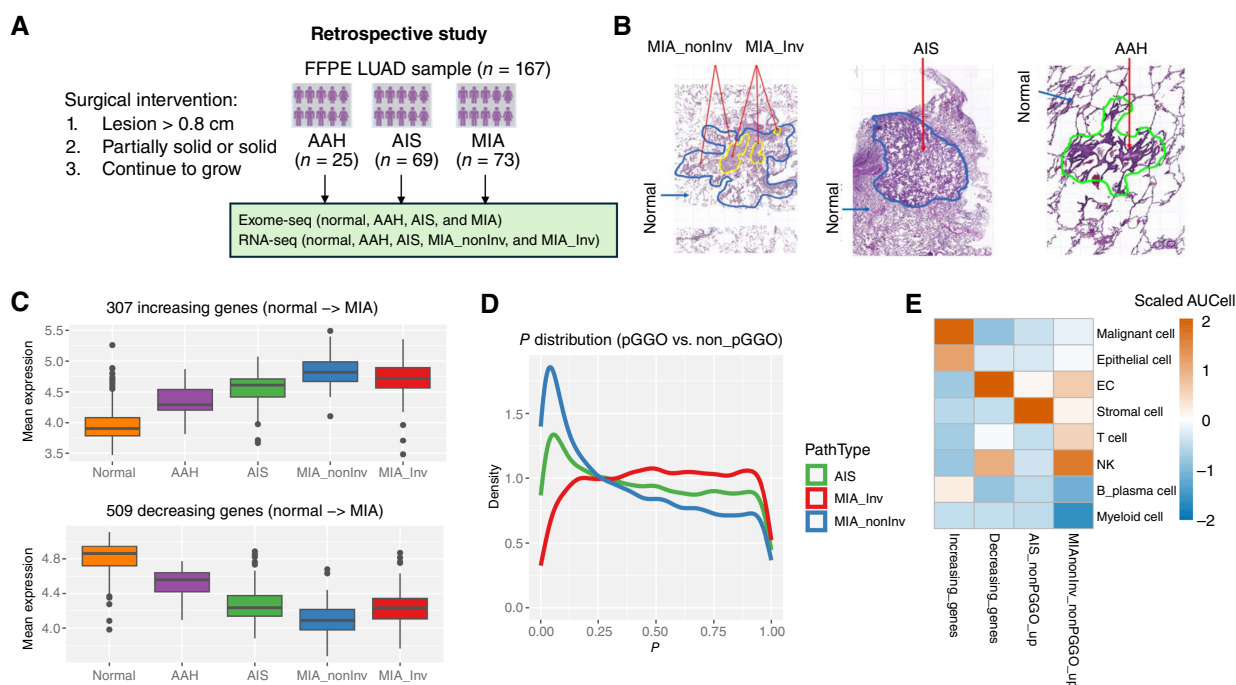
Data availability

The WES and RNA-seq data generated in this study are publicly available at the European Genome-phenome Archive (<https://ega-archive.org/>) with the accession number EGAD50000000637. The Visium spatial transcriptomic data are available at the Gene Expression Omnibus with the accession number GSE277206. The external mutation data analyzed in this study were obtained from the UCSC Xena browser (<https://xenabrowser.net/hub/>) at The Cancer Genome Atlas (TCGA) Hub and the supplementary table provided in ref. 19. All other raw data are available upon request from the corresponding author.

Results

Clinical characteristics of patients

We analyzed FFPE samples from surgically resected tumor tissues and adjacent normal tissues that were collected from 167 patients from Taiwan who had early-stage (0 or I) lung adenocarcinoma (Fig. 1A). These tumors were resected if they met any of the three following criteria: (i) lesion size larger than 0.8 cm on CT, (ii) presence of any solid component (i.e., non-pGGO), or (iii) enlargement based on serial follow-up images. There were 115 females (69%), the median age was 57 years, and 154 patients (92%) were

**Figure 1.**

Overview of study methods and analysis of differential expression of genes from bulk RNA-seq samples. **A**, Overview of the cohort and methods of analysis. **B**, Morphology-based clinicopathologic staging and tissue-based pathologic staging. LUAD, lung adenocarcinoma. **C**, Mean expression of genes in which expression increased (top) or decreased (bottom) during lung adenocarcinoma progression. **D**, Distributions of P values for all genes in the comparison of pGGO with non-pGGO for the AIS, MIA_nonInv, and MIA_Inv types. **E**, Mean expression scores (AUCell scores) of differentially expressed genes (from **C** and **D**) for each major cell type in the scRNA-seq data from Sinjab and colleagues (28).

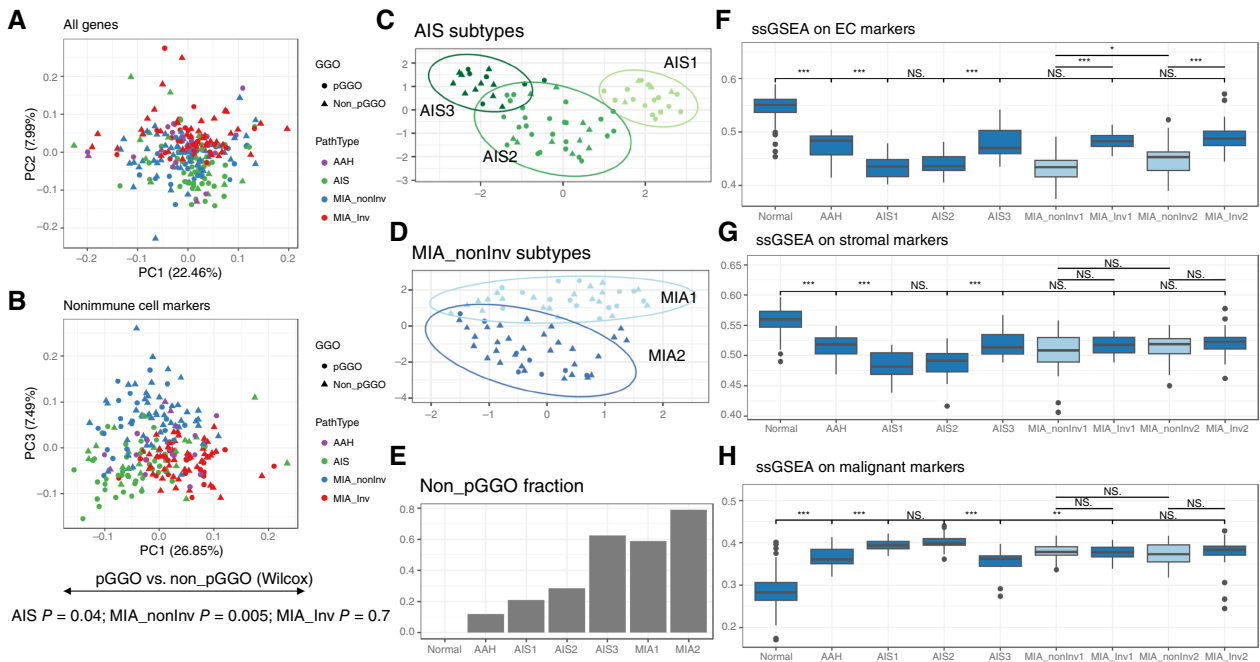
nonsmokers. According to the pathologic reports, we classified the nodules as preneoplasia or AAH ($n = 25$), AIS ($n = 69$), or MIA ($n = 73$). Patients with these three pathologic types had no significant differences in sex ($P = 0.648$), smoking status ($P = 0.566$), or past cancer history ($P = 0.915$). We examined all patients by WES and bulk RNA-seq.

Analysis of the imaging patterns showed that there were 90 nodules (54%) with pGGO, 76 nodules (46%) with partially solid GGO, and one solid nodule (Supplementary Table S3). Because only one patient had a solid nodule, we included this patient in the group with partially solid nodules (non-pGGO group) for subsequent analysis. There were pGGO nodules in 88% (22 of 25) of the patients with AAH, 65% (45 of 69) of the patients with AIS, and 32% (23 of 73) of the patients with MIA ($P < 0.001$). The pGGO and non-pGGO groups had no significant differences in sex ($P = 0.615$), median age ($P = 0.237$), smoking status ($P = 1.000$), TS ($P = 0.969$), or past cancer history ($P = 0.423$). As of December 31, 2023, the median follow-up duration was 57.8 months (range: 46.8–94.3 months). There were no instances of tumor recurrence, and all patients remained alive.

Genes with differential expression in comparisons of pathologic and radiological types have higher expression in nonimmune cells

We performed WES and bulk RNA-seq in all patients. For the patients with MIA, we performed RNA-seq on samples from the noninvasive tissues (MIA_nonInv) and the invasive tissues (MIA_Inv; Fig. 1B). The tumor purity scores were similar for the

AIS and MIA samples but were slightly lower for AAH samples (Supplementary Fig. S1). Although not universally acknowledged, previous research identified linear changes in several molecular characteristics, such as tumor mutational burden (TMB; ref. 13), ratio of CD4/CD8 T-cell infiltration (22), and number of hypomethylation sites (15), during early lung adenocarcinoma progression from normal tissue to AAH, to AIS, to MIA, and then to lung adenocarcinoma. Because there are dramatic changes in cell composition during cancer progression, we hypothesized that linear changes in the transcriptome would occur in parallel with the changes in cell composition (36, 37). We therefore applied a linear model (see Materials and Methods) to identify genes in which the mean expression (compared with normal tissue) increased or decreased during progression from normal to AAH, to AIS, and then to MIA (MIA_nonInv or MIA_Inv). **This analysis led to the identification of 307 genes with increasing expression and 509 genes with decreasing expression based on nominal P values ($P < 0.05$; Fig. 1C).** However, there were no genes with increasing or decreasing expression when the adjusted P values were applied, possibly raising doubt about our hypothesis that there are linear changes in the variation of the transcriptome. Moreover, although the genes with increasing and decreasing expression based on the nominal P value were chosen to demonstrate progressive up- and downregulation during early lung adenocarcinoma progression, these changes seemed to plateau after AIS. Altogether, these results suggest that changes in the transcriptome do not follow a simple linear progression during disease progression. We also identified DEGs in comparisons of pGGO and non-pGGO nodules for the

**Figure 2.**

Use of nonimmune cell markers to explain pathologic types and classify samples into GGO-associated subtypes. **A** and **B**, PCA of all genes (**A**) and nonimmune cell markers (**B**). The two-sided Wilcoxon rank-sum test was used to test the association between PC1 and the radiological types for AIS, MIA_nonInv, and MIA_Inv in **B**. **C**, AIS subtypes identified by clustering using nonimmune cell markers. **D**, MIA_nonInv subtypes identified by clustering using nonimmune cell markers. **E**, Fraction of non-pGGO samples within each subtype. **F–H**, Distribution of EC ssGSEA scores based on EC markers, stromal markers, and malignancy markers according to subtype. Statistical significance was determined from a two-sided Wilcoxon rank-sum test. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$; NS, not significant.

AIS, MIA_nonInv, and MIA_Inv groups (with exclusion of AAH because it only had three non-pGGO nodules). There were only 24 significant DEGs (FDR < 0.05) for MIA_nonInv and no significant DEGs for AIS and MIA_Inv. Interestingly, the distribution of P values showed an enrichment of genes with small P values for AIS and MIA_nonInv, indicating statistical differences in gene expression. In contrast, MIA_Inv displayed a more uniform distribution, indicating no significant differences in gene expression (Fig. 1D). These results suggest that the solid components detected on CT could be related to transcriptome changes in AIS and MIA_nonInv but not in MIA_Inv (subtly invaded tissues in MIA). Among the top 100 DEGs with the smallest P values, 97 were upregulated in non-pGGO nodules in AIS and 99 were upregulated in non-pGGO nodules in MIA_nonInv.

We then investigated the relationships between these DEGs and the composition of different cell types by examination of gene expression in the major cell types using two lung adenocarcinoma scRNA-seq datasets (Supplementary Fig. S2; refs. 28, 29). Our hypothesis was that if these DEGs paralleled the changes in cell types in the TME, then these DEGs would likely contain markers of the different cell types that accounted for increased or decreased fractions. Because the results from these two scRNA-seq datasets were similar, we described the results of Sinjab and colleagues (28) in the main text and the results of Salcher and colleagues (29) in the supplement. The results showed that the DEGs had higher expression in nonimmune cells, such as malignant cells, epithelial cells (a combination of AT1, AT2, basal, ciliated, and club cells), ECs (a combination of artery, vein, capillary, and lymphatic cells), and

stromal cells (a combination of fibroblasts, pericytes, and smooth muscle cells; Fig. 1E; Supplementary Fig. S3). In particular, the increasing (or decreasing) DEGs were highly expressed in malignant cells (or ECs and NK cells), and genes upregulated in non-pGGO nodules had higher expression in stromal cells (AIS) and in ECs and NK cells (MIA_nonInv).

Nonimmune cell markers can distinguish pathologic and radiological types

Because the DEGs identified above by imposing linear changes during disease progression displayed differential expression patterns in nonimmune cell types in the scRNA-seq data, we used an unsupervised approach to further examine this phenomenon. PCA of whole transcriptomes demonstrated a clear separation between normal tissues and lesions (Supplementary Fig. S4), but there was no clear separation among the different pathologic types within tumors (AAH, AIS, MIA_nonInv, and MIA_Inv; Fig. 2A). Because the DEGs had higher expression in nonimmune cells in the previous analysis, we hypothesized that changes in the composition of nonimmune cells might drive changes in the transcriptome during the early progression of lung adenocarcinoma. We therefore performed another PCA that only used nonimmune cell markers (top 20 DEGs of each nonimmune cell type in the scRNA-seq data). These results showed that the pathologic types (except AAH, which will be discussed later) were better separated by PC1 and PC3 and that PC1 was associated with radiological types (pGGO vs. non-pGGO) for AIS and MIA_nonInv but not MIA_Inv (Fig. 2B). The better separation of pathologic types using nonimmune cell markers in

this PCA supports our hypothesis that variations in the composition of nonimmune cells can partially explain changes in the transcriptome in the different pathologic types. In addition, the association of PC1 with GGO is consistent with our previous analysis of DEGs that different radiological types were associated with changes in the transcriptome in AIS and MIA_nonInv but not in MIA_Inv (Fig. 1D).

The association of PC1 with GGO led to our classification of AIS and MIA_nonInv nodules into GGO-associated molecular subtypes by clustering samples within AIS and MIA_nonInv, each using nonimmune cell markers (Fig. 2C and D). This analysis identified three subtypes of AIS (AIS1, AIS2, and AIS3) and two subtypes of MIA_nonInv (MIA_nonInv1 and MIA_nonInv2), and the subtypes within each pathologic type could be characterized by the fraction of non-pGGO nodules (Fig. 2E). Note that the two MIA_nonInv subtypes also classified the paired MIA_Inv samples (and patients with MIA) into two subtypes: MIA_Inv1 and MIA_Inv2 (and MIA1 and MIA2). This classification provided a more detailed characterization of the early progression of lung adenocarcinoma, from three pathologic types to six subtypes. Because AAH is considered the earliest lesion and a non-pGGO nodule on CT is often more clinically aggressive than a pGGO nodule (38), we provisionally ordered the seven subtypes (including normal tissue) according to disease progression first by pathology and then by radiological type within each pathologic type as follows: normal, AAH, AIS1, AIS2, AIS3, MIA1, and MIA2 (Supplementary Table S4). Although the order of these seven subtypes presumably corresponds to lung adenocarcinoma progression, we cannot be certain that all tumors actually pass through all of these subtypes during the progression from a pulmonary precancerous lesion to lung adenocarcinoma.

Gene set enrichment analysis using markers of ECs, stromal cells, and malignant cells revealed nonlinear changes in gene expression during lung adenocarcinoma progression

To examine how each nonimmune cell type varied during cancer progression, we computed ssGSEA scores for samples within each subtype using the top 50 markers for each nonimmune cell type to estimate relative abundance in the TME. Surprisingly, there was evidence of a nonlinear pattern (i.e., not consistently increasing or consistently decreasing, as imposed in the initial analysis of DEGs). The EC and stromal ssGSEA scores had similar patterns, in that they decreased from normal to AAH and to AIS1, remained low in AIS2, increased in AIS3, and remained relatively constant in MIA_Inv1 and MIA_Inv2 (Fig. 2F and G). The only differences occurred between MIA_nonInv and MIA_Inv. The stromal ssGSEA scores were comparable for MIA_nonInv and MIA_Inv, but the EC ssGSEA scores were higher in MIA_Inv. In addition, genes upregulated in MIA_Inv compared with MIA_nonInv based on a paired differential analysis ($P < 0.005$) were highly expressed in ECs, suggesting that ECs were likely to be the source of the transcriptomic differences (Supplementary Fig. S5). These trends generally persisted when considering MIA samples as a whole, in which paired noninvasive and invasive components were combined (Supplementary Fig. S6). This confirmed our observation of nonlinear changes from normal to AAH, to AIS, and then to MIA. However, the malignancy ssGSEA score had the opposite pattern during progression from normal to AIS3; more specifically, this score increased from normal to AAH and to AIS1, remained high in AIS2, decreased in AIS3, and then increased again in MIA_Inv (Fig. 2H). The changes in ssGSEA scores were generally in line with our analysis of DEGs (Supplementary Text). In addition,

these patterns were robust because they did not depend on the number of cell-type markers used (30, 50, or 70) or the source of the scRNA-seq data (Supplementary Figs. S7 and S8), and expression of the top cell-type markers exhibited trends consistent with the ssGSEA scores (Supplementary Fig. S9). To ensure that the markers identified from lung adenocarcinoma scRNA-seq data were comparable in the precancerous environment, we validated their specificity to distinct cell types using scRNA-seq data from patients with AIS and MIA lesions (Supplementary Fig. S10; ref. 17). Moreover, there were similar patterns when bulk RNA-seq samples were deconvolved using the scRNA-seq data as the reference (Supplementary Fig. S11). These results suggest that changes in the ssGSEA scores reflected changes in cell-type composition.

To validate the observed patterns of ssGSEA scores, we conducted spatial transcriptomics using 10× Visium on two of the MIA samples. The pathologist carefully annotated regions, including the adjacent normal tissue, MIA_nonInv, and MIA_Inv (Fig. 3A; Supplementary Fig. S12A). Cell-type deconvolution analysis confirmed the high abundance of epithelial cells in the tumor region (Fig. 3B; Supplementary Fig. S12B). Additionally, it revealed higher abundances of ECs and fibroblasts in the adjacent normal tissue than in the tumor region (Fig. 3C and D; Supplementary Fig. S12C and S12D). Within the tumor, ECs were notably enriched in the invasive regions and fibroblast abundance was slightly increased in the invasive region, although the distinction was somewhat unclear. These findings were consistent with the ssGSEA patterns observed among the normal tissue, MIA_nonInv, and MIA_Inv for ECs and stromal cells from the RNA-seq data (Fig. 2F and G). Notably, plasma cells were primarily located in the tumor region (Fig. 3E; Supplementary Fig. S12E), a topic that will be explored further in the forthcoming analysis of immune infiltration.

Overall, the nonlinear variations of these cell types suggest that there might be two different phases during the early progression of lung adenocarcinoma. In phase I, tumor cells grow, and this likely leads to decreases in the relative abundances of ECs and stromal cells from normal to AAH and to the early stage of AIS (AIS1/2). In phase II, ECs and stromal cells start growing at the late stage of AIS (AIS3), and this likely leads to increases in relative abundances of ECs and stromal cells from AIS2 to AIS3. These changes were not only associated with the increased probability of observing solid components on CT but were also molecular events that immediately preceded tissue invasion (MIA), highlighting their roles in the development of lung adenocarcinoma invasiveness.

Integrative analysis of bulk RNA-seq and scRNA-seq data revealed associations of tip ECs and CAFs with AIS3, implicating angiogenesis during the late stage of AIS

We further investigated the changes in cell types during disease progression by focusing on AAH and AIS3 because these samples had comparable EC ssGSEA scores (Fig. 2F). Based on our hypothesis about the two phases of lung adenocarcinoma progression, even though these samples had similar EC ssGSEA scores, their scores had different biological meanings. The EC ssGSEA score in AAH reflected the relative abundances of remaining healthy vasculature from normal tissues that were altered by tumor growth; the EC ssGSEA score in AIS3 reflected the relative abundances of newly developing vasculature from AIS2 and were related to tumor angiogenesis. To examine this hypothesis, we utilized SCISSOR, a method for identifying phenotype-associated cell subpopulations by integrating bulk RNA-seq and scRNA-seq data (33). SCISSOR uses a binary phenotype (0/1) that is annotated for bulk samples and

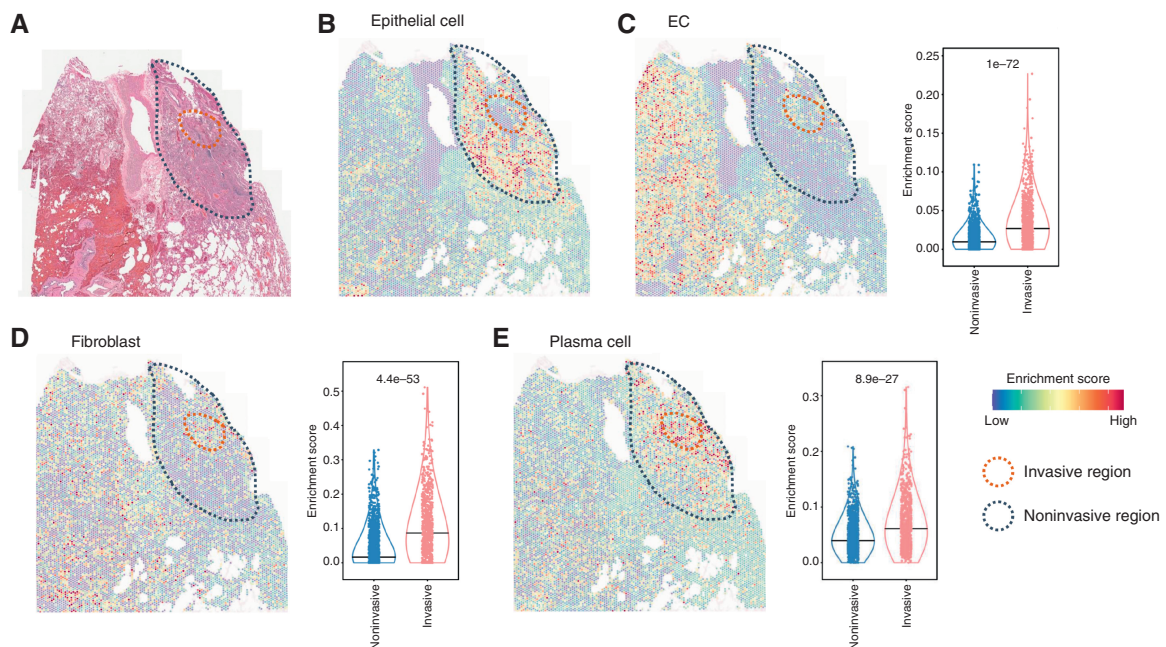


Figure 3.

Distribution of epithelial cells, ECs, fibroblasts, and plasma cells based on spatial transcriptomic analysis of an MIA sample. **A**, The tumor region (black dots) was annotated by the pathologist based on hematoxylin and eosin staining. Within the tumor, the area marked by red dots represents the invasive component (MIA_Inv) and the remaining region is the noninvasive component (MIA_nonInv). **B–E**, Spatial feature plots show enrichment scores for epithelial cells (**B**), ECs (**C**), fibroblasts (**D**), and plasma cells (**E**). All enrichment scores were calculated using xCell, and the distributions of enrichment scores for spots from the invasive and noninvasive components were compared using the Wilcoxon rank-sum test.

then identifies a subset of cells associated with each group (0 or 1) within a cell population in scRNA-seq data based on correlations in expression between bulk samples and individual cells and on the connectivity of cells in the cell population. We then used SCISSOR to compare the ECs in AAH and AIS3 based on retrieval of scRNA-seq data (**Fig. 4A**). The results showed that AIS3-associated ECs were enriched for cells that had high expression of markers for tip ECs, a subpopulation of ECs that is unique to tumors and is associated with tumor angiogenesis in lung adenocarcinoma (34).

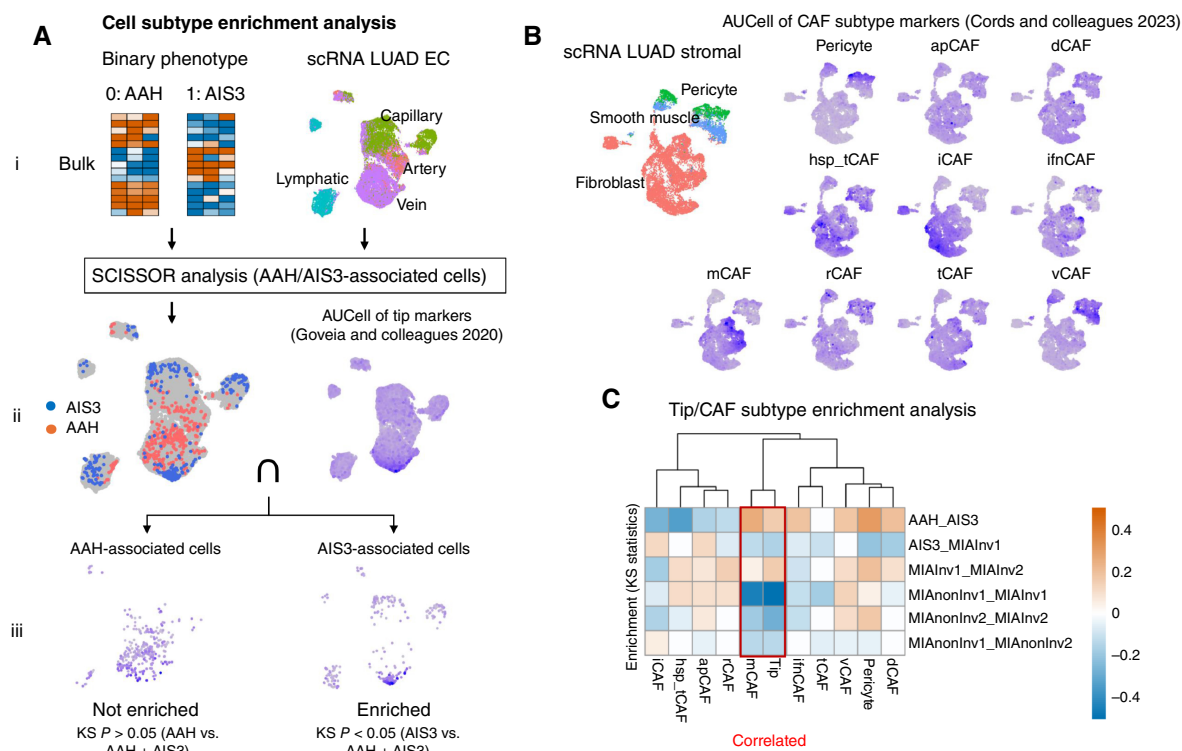
The ssGSEA scores for stromal cells and ECs had similar patterns during disease progression (**Fig. 2F and G**). We therefore examined whether any subtype of stromal cells potentially contributed to angiogenesis in concert with ECs (in particular tip ECs). A recent scRNA-seq study reported 10 CAF subtypes in breast cancer, spanning fibroblasts, pericytes, and smooth muscle cells, the same cell types in our stromal cell population (35). Most of the 10 CAFs were also confirmed in non-small cell lung cancer (39). CAFs contribute to the epithelial–mesenchymal transition (EMT; ref. 40) and angiogenesis (41) during cancer progression, and these processes promote tissue invasion (42, 43). We therefore computed cell scores based on markers of these 10 subtypes of CAFs for our stromal cell population (**Fig. 4B**).

To investigate the connections between tip ECs and CAF subtypes after the potential onset of angiogenesis in AIS3, we performed the same SCISSOR analysis as above (shown in **Fig. 3A**) by comparing pairs of molecular subtypes. In particular, within each pair, we identified the EC and stromal cell subpopulations that were associated with each subtype (Supplementary Figs. S13 and S14). We then computed enrichment scores for tip ECs and the 10 subtypes of CAFs to determine whether tip ECs or any of the CAF

subtypes were enriched in cell subpopulations that were associated with either of these molecular subtypes (**Fig. 4C**). Among the 10 subtypes of CAFs, matrix CAFs (mCAF) co-varied with tip ECs in all analyzed pairs. In addition to SCISSOR analysis, representative genes of mCAF and tip ECs exhibited similar trends across these subtypes (Supplementary Fig. S15), with higher expression in AIS3 than in AAH. The mCAF subtype is characterized as high expression of matrix and collagen proteins and upregulation of the TGF- β , KRAS, and the EMT pathways (35). In summary, our analysis of the AIS3 subtype highlights the potential roles of tip ECs and mCAF during angiogenesis, the EMT, and tissue invasion during the early progression of lung adenocarcinoma.

Infiltration of immune cells is associated with changes in nonimmune cells in the TME during early progression of lung adenocarcinoma

Next, we examined immune infiltration using RNA-seq data. For consistency with the previous analysis of nonimmune cells, we computed ssGSEA scores for each sample using the top 50 markers of each immune cell type from the scRNA-seq data. We also estimated immune cell composition using TIMER (<http://timer.cistrome.org>), a tool that profiles immune infiltration in the TME (32). Dejima and colleagues (22) used TIMER and reported a linear increase in the CD4/CD8 T-cell ratio during progression from normal tissue to AAH, to AIS, and to MIA in their study of immune infiltration of lung preneoplasia. We observed the same trend in ssGSEA scores from normal to AAH, to AIS, and to MIA_nonInv (but not MIA_Inv; **Fig. 5A**). The ratio we estimated by TIMER, however, did not have a linear increase, in that AIS had a slightly lower ratio than AAH (**Fig. 5B**). Zhao and colleagues (14) studied

**Figure 4.**

Enrichment of cell subtypes from scRNA-seq data for two groups of bulk samples using SCISSOR. **A**, Three steps used for enrichment of tip ECs in the AAH and AIS3 types: (i) AAH and AIS3 samples and the EC population from scRNA-seq were used to compute AAH- and AIS3-associated ECs using SCISSOR; (ii) tip EC scores were calculated for AAH- and AIS3-associated ECs; and (iii) enrichment of tip ECs in AAH or AIS3 was determined by a one-sided Kolmogorov-Smirnov (KS) test to assess the distribution of tip EC scores in the union of AAH-associated and AIS3-associated ECs with AAH-associated ECs (left) and/or AIS3-associated ECs (right). **B**, Uniform Manifold Approximation and Projection of stromal cells from scRNA-seq data and their CAF subtype scores. apCAF, antigen-presenting CAF; dCAF, dividing CAF; hsp_tCAF, heat-shock protein-high tCAF; iCAF, inflammatory CAF; ifnCAF, interferon-response CAF; LUAD, lung adenocarcinoma; mCAF, matrix CAF; rCAF, reticular-like CAF; tCAF, tumor-like CAF; vCAF, vascular CAF. **C**, Enrichment of EC and CAF subtypes between paired bulk sample subtypes (from **A**).

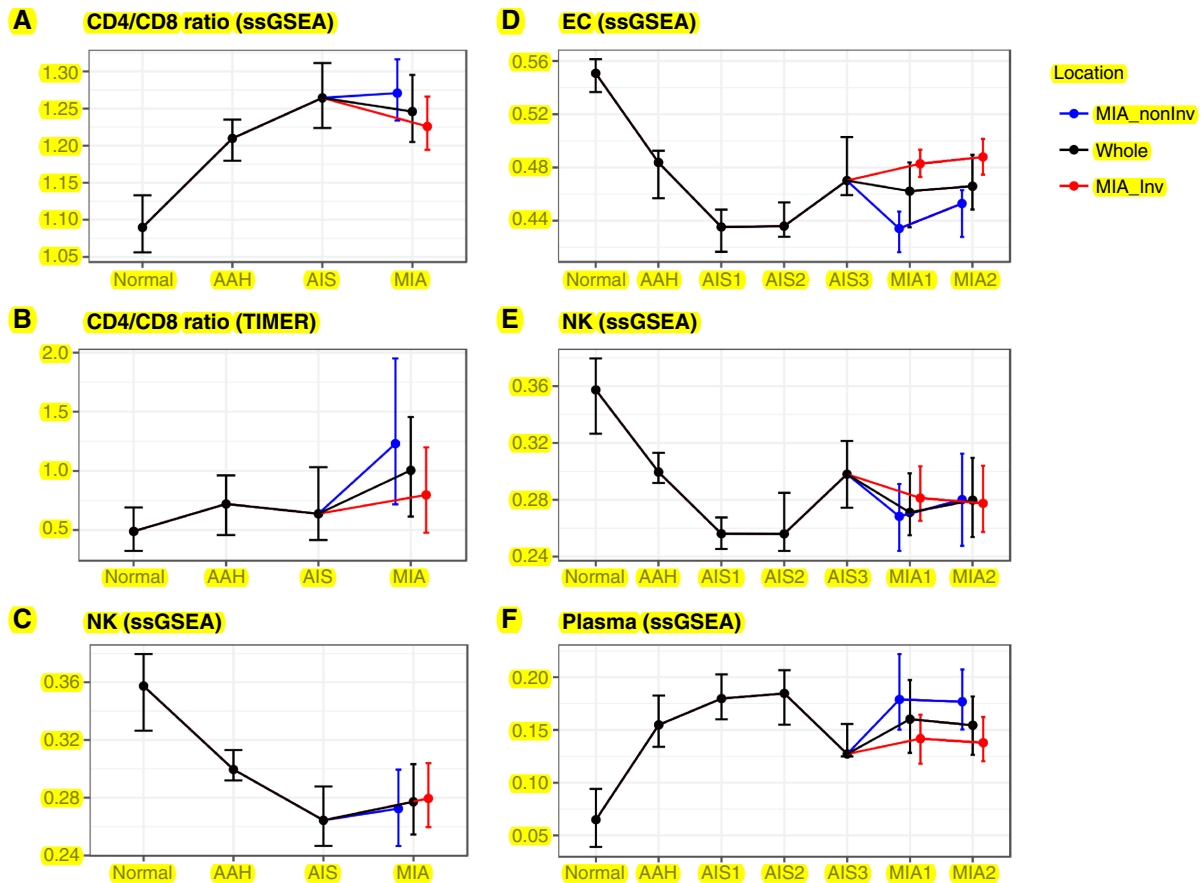
lung preneoplasia with no AAH and found that the proportion of resting NK cells decreased from normal tissue to AIS but increased slightly from AIS to MIA; however, this proportion had a linear decrease from normal tissue to AAH, to AIS, and to MIA in the study of Dejima and colleagues (22). Our results showed that changes in the ssGSEA score for NK cells were consistent with the findings of Zhao and colleagues (Fig. 5C). Taken together, these results suggest that the immune cell profiles of lung preneoplasia vary among patient cohorts and may also be affected by biopsy site (invasive vs. noninvasive tissues for MIA) and the computational approaches used for estimation.

To investigate the possible coordination of changes in immune and nonimmune cells in the TME, we examined the ssGSEA scores of immune cell types for the seven subtypes previously defined by nonimmune cell markers. Surprisingly, some immune cell types varied in concert with nonimmune cell types. For example, changes in the ssGSEA score of NK cells according to subtype were similar to the changes in the ssGSEA score of ECs (Fig. 5D and E), with a minor variation observed from AIS3 to MIA_Inv2 (see more discussion in Supplementary Text). A comparable trend was also noted in CD8 T cells (Supplementary Fig. S16). On the other hand, there were similar changes in the ssGSEA scores of plasma cells and malignant cells (Figs. 2H and 5F). Plasma cells are depleted in

normal lung tissues but are highly enriched in lung adenocarcinoma (44). Representative gene markers of these cell types also corresponded to the ssGSEA scores (Supplementary Fig. S9). Analysis of the spatial transcriptomic data of the MIA sample showed that plasma cells were predominantly in the tumor region (Fig. 3E; Supplementary Fig. S12E). This spatial distribution helps explain the covariation of ssGSEA scores between plasma cells and malignant cells. Our analysis suggests that the infiltration of plasma cells into the TME might be an early event during the progression of lung adenocarcinoma that is potentially triggered by the abnormal growth of tumor cells. We observed the same trends using markers from another scRNA-seq dataset, from deconvolution-based estimates with scRNA-seq data as the reference, and from cell-type compositions estimated by another deconvolution method, iSort [the optimized version of CIBERSORTx (45); Supplementary Figs. S17–S19]. The ssGSEA scores for other major immune cells are provided in Supplementary Fig. S20.

AAH and MIA nodules are associated with a higher TMB and smaller TS

After exclusion of samples that failed sequencing quality control metrics (see Materials and Methods), we included 159 of 167 patients in the WES analysis of highly mutated genes (Fig. 6A). We also compared the mutation rates of these gene mutations with two

**Figure 5.**

Coordinated variations of ECs and immune cells. Pathologic types or subtypes according to the ratio of CD4 T and CD8 T ssGSEA scores (A), ratio of CD4 T and CD8 T fractions estimated by TIMER (B), NK ssGSEA score (C), EC ssGSEA score (D), NK ssGSEA score (E), and plasma ssGSEA score (F), with symbols indicating medians and the 25th and 75th percentiles.

other external datasets: a cohort of patients with lung adenocarcinoma from TCGA and a cohort of patients with lung adenocarcinoma from Chen and colleagues (named Chen2020; ref. 19), a previously described Taiwanese cohort mostly consisting of nonsmoking patients with stage I lung adenocarcinoma. The mutation rates of precancerous lesions were generally lower in our study than in these two previous studies (Fig. 6B). There was a characteristically high *EGFR* mutation rate in East Asian patients with lung cancer from our study and the Chen2020 cohort compared with TCGA lung adenocarcinoma cohort, demonstrating molecular differences between nonsmoking East Asian patients and patients with European ancestry.

Hu and colleagues (13) reported that the TMB increased linearly from AAH to AIS and to MIA. However, in the present study, we observed that the TMB was higher in AAH than in AIS and was comparable for AAH and MIA, although with a slightly higher median TMB in AAH (Fig. 7A). This was unexpected because AAH is presumably the earliest type of lesion. We therefore examined whether our results could be attributed to ascertainment bias (i.e., the AAH nodules resected in the present study were not representative of AAH nodules in general but were biased toward those with a high TMB). In this study, we collected three types of nodules based on surgical criteria: (i) nodules with any solid components on

CT (non-pGGO), (ii) large pGGO nodules (TS > 0.8 cm, pGGO_Large), and (iii) small pGGO nodules (TS ≤ 0.8 cm) that enlarged during follow-up (pGGO_SmallGrow; Fig. 7B). As expected, the AAH nodules were mainly in the third group because they often presented as small pGGO nodules on CT (5). However, within this group, there was a high TMB in the smaller tumors, and these were highly associated with AAH or MIA nodules but not AIS nodules (one-sided $P = 2 \times 10^{-6}$; Fig. 7C). It is possible that surgical intervention for very small but fast-growing pGGO nodules (TS < 0.5 cm in Fig. 7C) was more likely to sample AAH nodules with a high TMB and to miss AAH nodules with a low TMB, which often presented as small and dormant pGGO nodules on CT (pGGO_other group in Fig. 7B). Much investigation is needed to elucidate the connection between the TMB and the growth rate of tumors at the very beginning of lung adenocarcinoma progression.

In addition, we found that AIS nodules were associated with larger TS in the pGGO_SmallGrow group and in the non-pGGO group ($P = 1 \times 10^{-5}$) and were also enriched in the pGGO_Large group ($P = 3.6 \times 10^{-4}$). As expected, AIS nodules were significantly larger in the entire cohort ($P = 1.2 \times 10^{-11}$). When all nodules were sorted into three groups according to TS (< 0.5 cm; 0.5 cm ≤ TS ≤ 0.8 cm; 0.8 cm < TS), the AAH and MIA nodules were highly enriched in the smallest group, and the AIS nodules were depleted

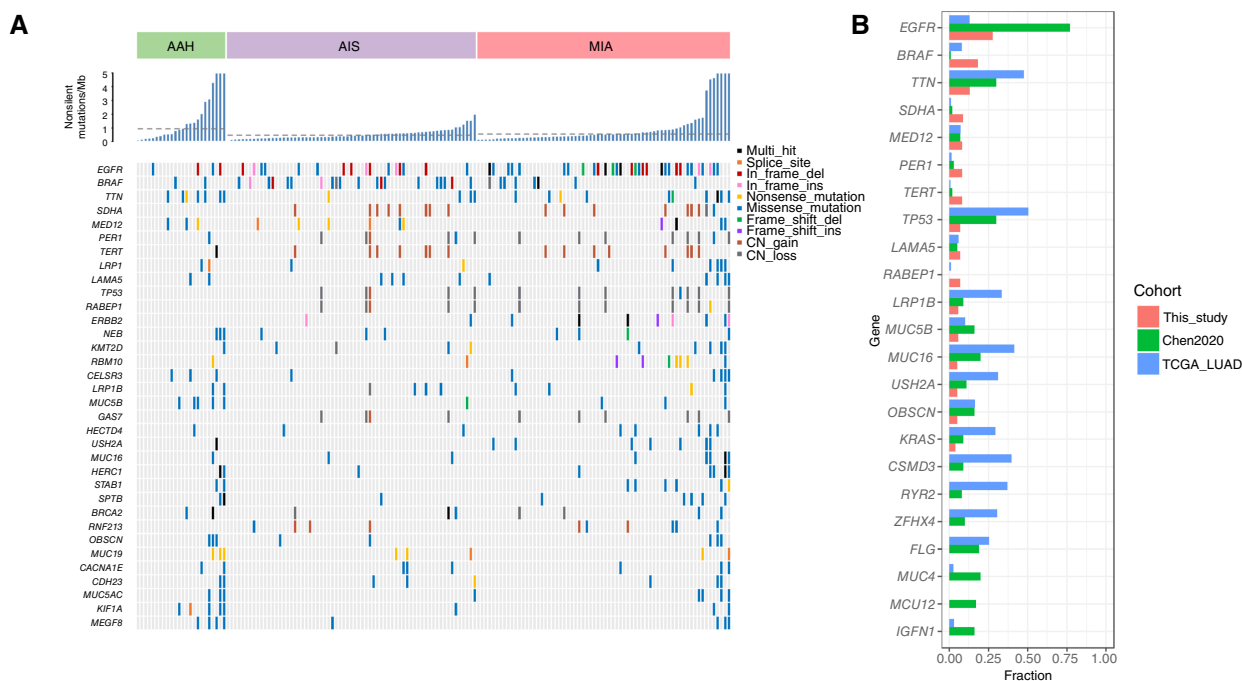


Figure 6.

Mutational tracking during clinicopathologic progression of lung adenocarcinoma. **A**, TMB (top) and somatic mutations (bottom) in different genes (OncoPrint). **B**, Gene mutation rates of the 10 most mutated genes in this study, the cohort from Chen2020 (a Taiwanese nonsmoking cohort; ref. 19), and the cohort from TCGA. LUAD, lung adenocarcinoma.

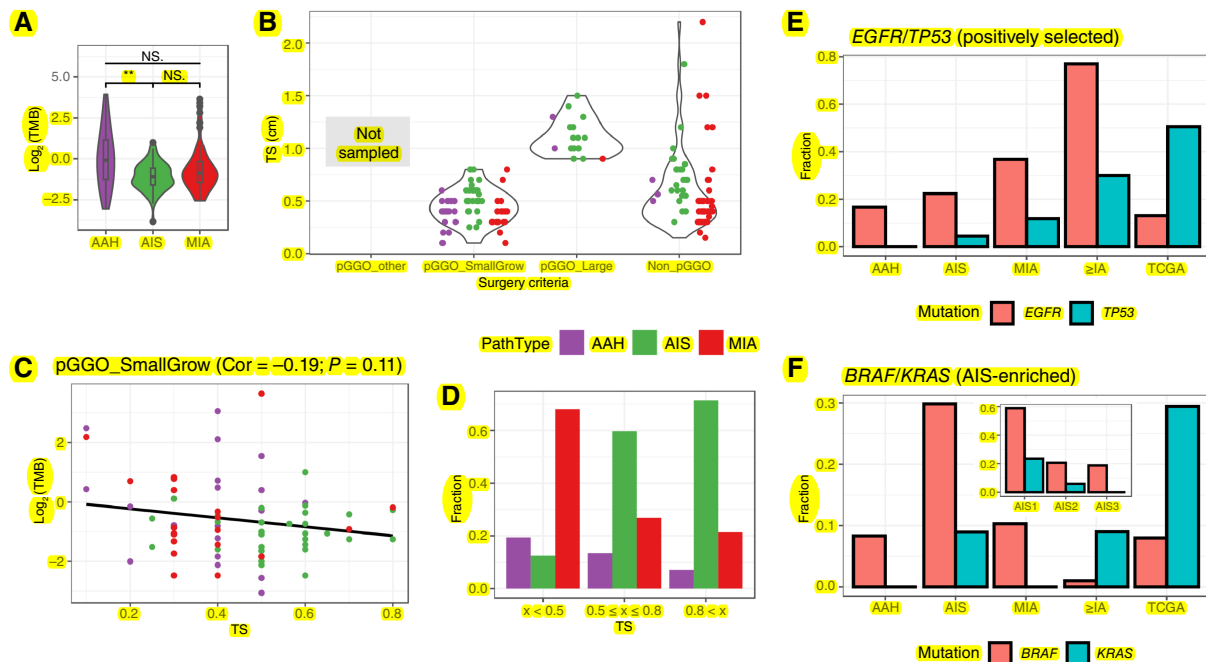
in the smallest group regardless of radiological type and surgical criteria. The fraction of AIS nodules increased dramatically as tumor size increased (Fig. 7D). Because pathology was determined after surgery, the uneven distribution among the groups with different tumor sizes could not be due to ascertainment bias based on pathologic types. Therefore, this suggests that under our surgical criteria, very small nodules (<0.5 cm) were more common at the early stage of progression (small and fast-growing AAH) or the late stage of progression (MIA) but were less common in the middle stage (AIS). In contrast, large nodules had the opposite trend, in that they were more common in AIS and less common in AAH and MIA.

Positive selection for *EGFR* and *TP53* mutations occurred throughout lung adenocarcinoma progression, but Ras mutations were highly enriched in AIS

We examined mutation rates in our cohort (including AAH, AIS, and MIA nodules) and in the Chen2020 cohort (nonsmoking Taiwanese patients with stage I lung adenocarcinoma; ref. 19) to study the variation of gene mutation rates during the lung adenocarcinoma progression in nonsmokers. We focused on the 90 genes that were mutated in six or more patients in our study (Supplementary Table S5). The results showed that there was positive selection for mutations in five genes, in that their mutation rates increased linearly as lung adenocarcinoma progressed from AAH to AIS, to MIA, and to (stage I) lung adenocarcinoma. These genes included two well-known oncogenes, *EGFR* (mutated in 44 patients) and *TP53* (mutated in 11 patients; Fig. 7E); *KMT2D*, a gene recently reported to drive progression of lung squamous cell carcinoma (mutated in 10 patients; ref. 46); and two additional genes

implicated in lung cancer, *FCGBP* (mutated in 6 patients; ref. 47) and *RNF213* (mutated in 8 patients; ref. 48; Supplementary Fig. S21). More observations for other mutations identified in this study were summarized in Supplementary Text. The complete lists of somatic mutations and copy-number alterations are provided in Supplementary Tables S6 and S7.

Interestingly, *BRAF* (mutated in 29 patients) and *KRAS* (mutated in 6 patients) were two outliers (Fig. 7F), in that they were highly mutated in AIS nodules (Fisher exact test: $P = 0.001$ and $P = 0.005$, respectively), the same pathologic type that had the lowest median TMB relative to AAH and MIA types (Fig. 7A). Within AIS, *BRAF* and *KRAS* were highly mutated in the AIS1 subtype (Fig. 7F). Because these two genes are well-known oncogenes, their high mutation rates in AIS could be biologically meaningful. Although our observations could be attributable to bias from the small sample size, similar mutation patterns were identified in other studies of early-stage lung adenocarcinoma. For example, Zhao and colleagues (14) examined AIS, MIA, and lung adenocarcinoma samples and reported that the mutation rates in *EGFR* and *TP53* increased linearly from AIS to MIA and to lung adenocarcinoma (although *TP53* had a slightly lower mutation rate in MIA than in AIS), but *BRAF* had the opposite mutation rate pattern. Sivakumar and colleagues (12) examined AAH and lung adenocarcinoma samples and found that *EGFR* had a higher mutation rate in lung adenocarcinoma than in AAH, whereas *BRAF* was only mutated in AAH and not in lung adenocarcinoma. Xiang and colleagues (49) found enrichment of *BRAF* mutations in AIS using a targeted 68-gene sequencing panel. These results are similar to the mutation patterns of *EGFR*, *TP53*, and *BRAF* identified in the present study, suggesting they are not cohort specific and are applicable to nonsmoking patients. In

**Figure 7.**

Relationships of TMB, TS, surgical criteria, pathologic types, and gene mutation rate. **A**, $\text{Log}_2(\text{TMB})$ distributions for AAH, AIS, and MIA. **B**, TS for nodules resected based on different surgical criteria. Non-pGGO, nodules with any solid or partially solid components; pGGO_Large, large pGGO nodules (>0.8 cm); pGGO_SmallGrow, pGGO nodules that are small (<0.8 cm) but enlarged during follow-up; pGGO_other, other pGGO nodules not collected in this study. **C**, Pearson correlation of $\text{log}_2(\text{TMB})$ and TS within the pGGO_SmallGrow group (from **B**). **D**, Fractions of nodules of different pathologic types in different size groups. **E** and **F**, Fractions of patients with *EGFR* and *TP53* mutations (**E**) and *BRAF* and *KRAS* mutations (**F**) according to pathologic types and in two separate lung adenocarcinoma cohorts: Chen2020 (a Taiwanese nonsmoking cohort; ref. 19) and TCGA. **F**, inset, fraction of patients with *BRAF* and *KRAS* mutations according to AIS subtypes. **, $P < 0.01$.

addition, we also noted that *EGFR* and *BRAF* were the two most highly mutated genes, but their mutations tended to be mutually exclusive, and only one patient who had a mutant *EGFR* also had a mutant *BRAF* (copy-number loss). Differences in tumor size and mutational patterns among AAH, AIS, and MIA nodules in this study might provide insights into the progression paths during early lung adenocarcinoma development (see more discussion in Supplementary Text).

Different pathologic types are associated with different TME ecotypes

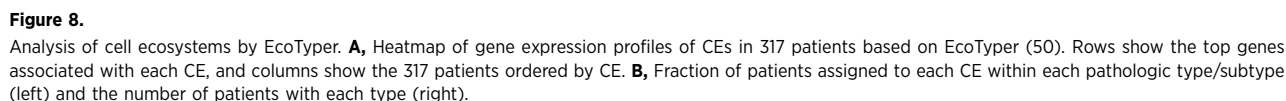
Finally, we applied EcoTyper (50) to our RNA-seq data to classify each sample into one of the 10 carcinoma ecotypes (CE) that were previously described based on studies of more than 6,000 tumor and normal transcriptomes in 16 types of human carcinomas. Each CE is a cell ecosystem composed of a frequently observed set of coordinated cell states in human carcinomas. EcoTyper analysis provides a deeper and more comprehensive view of the TME than analysis of individual cell types. Luca and colleagues (50) assigned almost all normal samples from the Genotype-Tissue Expression study and all adjacent normal samples in pan-cancer studies to CE6 but assigned most human carcinomas to CE1, CE8, or CE10.

We successfully assigned 317 of the 406 RNA-seq samples to 1 of the 10 CEs (**Fig. 8A**). Among these 317 samples, almost all normal samples were assigned to CE6, whereas most MIA samples were assigned to one of the carcinoma-associated CEs (CE1, CE8, or CE10; **Fig. 8B**). The similar findings of our study and the original EcoTyper study demonstrated the robustness of this analysis among cohorts. Interestingly, our

AIS1/2 nodules were highly enriched in CE3, although the study of Luca and colleagues only rarely assigned this ecotype to tumors. The enrichment of CE3 in our AIS1/2 subtypes demonstrated the unique cell ecosystem in the TME, in contrast to MIA and carcinomas. The AIS3 and AAH nodules had similar CE patterns, with enrichment in CE6 (normal), but they also had enrichment of some CEs that were more frequent in the MIA type. These results suggest that AIS3 is a subtype that is molecularly distinct from AIS1/2 and that the AAH samples in the present study might consist of a mixture of TME ecosystems from normal tissues and MIA nodules.

Discussion

In this study, we investigated the molecular basis of the early progression of lung adenocarcinoma by analyzing WES and RNA-seq data for AAH, AIS, and MIA nodules and adjacent normal tissues in a cohort of East Asian patients most of whom were nonsmokers. We used markers of different cell types from scRNA-seq data to classify AIS and MIA nodules into molecular subtypes that were associated with different radiological types. Based on ssGSEA and deconvolution analyses, we found that variation in the composition of different cell types during disease progression had a nonlinear pattern. These results suggest that the progression of lung adenocarcinoma has two phases. First, tumor cells grow from normal to AAH and then to early AIS as the relative abundances of ECs and stromal cells decrease. Second, during the later steps of AIS, angiogenesis occurs with involvement of tip ECs and EMT-driving CAFs, and this coincides with an increased probability of



Our finding of an association of angiogenesis and CAFs with the non-pGGO phenotype is consistent with several previous studies of lung adenocarcinoma. One study of patients with lung cancer reported that the amount of ctDNA in the peripheral blood correlated with the size of non-pGGO components but not with total tumor volume (10). This is likely because the release of ctDNA from tumors into the peripheral blood requires development of vasculature (angiogenesis) in the TME, strengthening the connection between angiogenesis and non-pGGO components. Another recent study reported that circulating proteins in prediagnostic blood samples that were related to tissue invasion and angiogenesis were associated with a diagnosis of lung cancer during the following 3 years (51), also highlighting the role of angiogenesis during lung adenocarcinoma progression from a precancerous lesion to an invasive adenocarcinoma. There is evidence that nodules with more solid components on CT contain more tumor cells and proliferating fibroblasts that disrupted alveolar structures and that the

fibrosis area of the entire tumor area is associated with the size of the invasive area (52). These observations support the presence of a link between the increased activities of CAFs and the increased non-pGGO components on CT during the late stage of AIS (AIS3) and during subsequent tissue invasion in MIA. Finally, there is also evidence that tumor angiogenesis promotes cell motility, invasion, and metastasis, all of which contribute to tumor progression (42, 53).

In this study, we observed that *EGFR* and *TP53* mutations were positively selected from AAH to AIS, to MIA, and to lung adenocarcinoma, whereas *BRAF* mutations were particularly enriched for AIS nodules with sizes significantly larger than AAH and MIA nodules. These observations seem consistent with the established functions of these gene mutations in cancer. The main effect of *BRAF* oncogenic mutations is promotion of cell proliferation (54, 55), and this seems to explain their association with larger lung tumors. *EGFR* mutations are known to promote tumor angiogenesis by overexpressing the *VEGF*, which subsequently stimulates ECs and leads to the formation of new blood vessels (56–58). On the other hand, the loss of *TP53* function can promote the EMT during cancer progression (59). *EGFR* and *TP53* mutations were functionally linked to two events that are implicated in AIS3: growth of tip ECs (which are responsible for angiogenesis) and increased activity of EMT-driving CAFs (mCAFs). These links explain the positive selection of *EGFR* and *TP53* mutations during cancer progression biologically.

This study has several limitations. As mentioned above, most of the AAH nodules we examined were small and fast-growing pGGO nodules, and this may not be typical of AAH nodules; typical AAH nodules may include many more small, pGGO, and dormant (but not fast-growing) nodules. This is a challenging topic to investigate because small, pGGO, and dormant nodules often do not require surgical intervention. Some studies collected AAH nodules from lesions that were near resected lung adenocarcinoma tumors in the same patients (12). However, whether those AAH nodules were typical of AAH nodules remains unclear because they might have been altered by nearby primary tumors. In addition, all patients in our cohort were East Asian, and most were nonsmokers. It is unknown whether our findings can be generalized to other racial or ethnic groups or to patients who smoke. Finally, our data were derived from resected nodules and were not longitudinal data but essentially a snapshot of disease progression in different patients. We therefore delineated the molecular trajectory of progression processes based on variations among nodules. It remains unknown whether all tumors follow the same trajectory.

In conclusion, we classified nodules within different pathologic types into molecular subtypes and provided a more granular view of the early progression of lung adenocarcinoma from a precancerous lesion to an adenocarcinoma. We characterized each subtype based on its molecular characteristics and identified angiogenesis during a late stage of AIS as a key event during the early progression of lung adenocarcinoma. This progression was associated with non-pGGO components on CT and could potentially contribute to tissue invasion. Angiogenesis increases vascularity in the TME, and this can increase the amount of ctDNA and the release of tumor proteins into the peripheral blood, allowing for early detection (60). In addition, there are well-known therapies that target vasculature in the TME, such as antiangiogenic agents and agents that promote vessel normalization (61, 62). These therapies have

potential as an early intervention for lung adenocarcinoma because they may prevent progression and facilitate infiltration by immune cells. Overall, our findings provide new insights into the early diagnosis and management of early-stage lung adenocarcinoma.

Authors' Disclosures

Y.-C. Chen reports current employment with Johnson & Johnson, as well as holding shares in the company. H.-M. Wang reports employment with Johnson & Johnson at the time of analysis. M. Pirooznia reports current employment with Johnson & Johnson, as well as holding shares in the company. R. Yang reports other support from Johnson & Johnson outside the submitted work. J.-Y. Shih reports grants from National Science and Technology Council during the conduct of the study, as well as grants from Roche and personal fees from AbbVie, ACT Genomics, Amgen, AstraZeneca, Bayer, Boehringer Ingelheim, Bristol Myers Squibb, Chugai Pharmaceutical, Daiichi Sankyo, Eli Lilly and Company, Guardant Health, Illumina, Janssen, Lotus, Merck Sharp & Dohme, Merck, Ono Pharmaceutical, Orient EuroPharma, Pfizer, Roche, Takeda, TSH Biopharm, and TTY Biopharm outside the submitted work. No disclosures were reported by the other authors.

Authors' Contributions

Y.-C. Chen: Conceptualization, software, formal analysis, visualization, methodology, writing—original draft, writing—review and editing. **C.-L. Hsu:** Conceptualization, software, formal analysis, visualization, methodology, writing—review and editing. **H.-M. Wang:** Conceptualization, software, formal analysis, visualization, methodology, writing—original draft, writing—review and editing. **S.-G. Wu:** Conceptualization, formal analysis, methodology, writing—original draft, writing—review and editing. **Y.-L. Chang:** Conceptualization, resources, data curation, methodology. **J.-S. Chen:** Conceptualization, data curation, methodology. **Y.-C. Wu:** Data curation, software, formal analysis. **Y.-T. Lin:** Conceptualization, data curation, formal analysis, methodology. **C.-Y. Yang:** Conceptualization, data curation, formal analysis, methodology. **M.-W. Lin:** Conceptualization, data curation, methodology. **J.-M. Lee:** Conceptualization, data curation, methodology. **S.-W. Kuo:** Conceptualization, data curation, methodology. **K.-C. Chen:** Conceptualization, data curation, methodology. **H.-H. Hsu:** Conceptualization, data curation, methodology. **P.-M. Huang:** Conceptualization, data curation, methodology. **Y.-L. Huang:** Conceptualization, data curation. **C.-J. Yu:** Conceptualization, formal analysis, methodology, writing—review and editing. **M. Pirooznia:** Conceptualization, formal analysis, methodology, writing—review and editing. **B.E. Huang:** Conceptualization, software, formal analysis, methodology, writing—review and editing. **R. Yang:** Conceptualization, formal analysis, methodology, writing—review and editing. **J.-Y. Shih:** Conceptualization, formal analysis, methodology, writing—review and editing. **P.-C. Yang:** Conceptualization, resources, formal analysis, methodology, writing—original draft, writing—review and editing.

Acknowledgments

We thank the Sequencing Core, Department of Medical Research, National Taiwan University Hospital, for technical support and the National Center for High-performance Computing of Taiwan for providing computational and storage resources. We also thank Dr. Silvia Chen for comments on the manuscript. This study was supported by grants from the Ministry of Science and Technology, Taiwan (MOST 109-3111-8-002-001 and MOST 110-2622-8-002-016), as well as research funding from Johnson & Johnson.

Note

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

Received March 14, 2024; revised July 20, 2024; accepted November 5, 2024; published first November 21, 2024.

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* 2018;68:394–424.
- Wu B, Chen J, Zhang X, Feng N, Xiang Z, Wei Y, et al. Prognostic factors and survival prediction for patients with metastatic lung adenocarcinoma: a population-based study. *Medicine (Baltimore)* 2022; 101:e32217.

3. Myers DJ, Wallen JM. Lung adenocarcinoma. In: StatPearls. Treasure Island (FL): StatPearls Publishing; 2023.
4. Sakurai H, Maeshima A, Watanabe S-i, Suzuki K, Tsuchiya R, Maeshima AM, et al. Grade of stromal invasion in small adenocarcinoma of the lung: histopathological minimal invasion and prognosis. *Am J Surg Pathol* 2004;28:198–206.
5. Noguchi M. Stepwise progression of pulmonary adenocarcinoma—clinical and molecular implications. *Cancer Metastasis Rev* 2010;29:15–21.
6. Travis WD, Brambilla E, Nicholson AG, Yatabe Y, Austin JHM, Beasley MB, et al. The 2015 World Health organization classification of lung tumors: impact of genetic, clinical and radiologic advances since the 2004 classification. *J Thorac Oncol* 2015;10:1243–60.
7. Divisi D, Barone M, Bertolaccini L, Rocco G, Solli P, Crisci R, et al. Standardized uptake value and radiological density attenuation as predictive and prognostic factors in patients with solitary pulmonary nodules: our experience on 1,592 patients. *J Thorac Dis* 2017;9:2551–9.
8. Kobayashi Y, Mitsudomi T. Management of ground-glass opacities: should all pulmonary lesions with ground-glass opacity be surgically resected? *Transl Lung Cancer Res* 2013;2:354–63.
9. Kobayashi Y, Sakao Y, Deshpande GA, Fukui T, Mizuno T, Kuroda H, et al. The association between baseline clinical-radiological characteristics and growth of pulmonary nodules with ground-glass opacity. *Lung Cancer* 2014;83:61–6.
10. Chen K, Bai J, Reuben A, Zhao H, Kang G, Zhang C, et al. Multiomics analysis reveals distinct immunogenomic features of lung cancer with ground-glass opacity. *Am J Respir Crit Care Med* 2021;204:1180–92.
11. 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.
12. Sivakumar S, Lucas FAS, 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.
13. Hu X, Fujimoto J, Ying L, Fukuoka J, Ashizawa K, Sun W, et al. Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma. *Nat Commun* 2019;10:2978.
14. Zhao Y, Shang J, Gao J, Han H, Gao Z, Yan Y, et al. Increased tumor intrinsic growth potential and decreased immune function orchestrate the progression of lung adenocarcinoma. *Front Immunol* 2022;13:921761.
15. Valle JW, Kelley RK, Nervi B, Oh D-Y, Zhu AX. Biliary tract cancer. *Lancet* 2021;397:428–44.
16. Wang Z, Li Z, Zhou K, Wang C, Jiang L, Zhang L, et al. Deciphering cell lineage specification of human lung adenocarcinoma with single-cell RNA sequencing. *Nat Commun* 2021;12:6500.
17. Zhu J, Fan Y, Xiong Y, Wang W, Chen J, Xia Y, et al. Delineating the dynamic evolution from preneoplasia to invasive lung adenocarcinoma by integrating single-cell RNA sequencing and spatial transcriptomics. *Exp Mol Med* 2022;54:2060–76.
18. Zhou F, Zhou C. Lung cancer in never smokers—the East Asian experience. *Transl Lung Cancer Res* 2018;7:450–63.
19. Chen Y-J, Roumeliotis TI, Chang Y-H, Chen C-T, Han C-L, Lin M-H, et al. Proteogenomics of non-smoking lung cancer in East Asia delineates molecular signatures of pathogenesis and progression. *Cell* 2020;182:226–44.e17.
20. Sui Q, Liang J, Hu Z, Chen Z, Bi G, Huang Y, et al. Genetic and microenvironmental differences in non-smoking lung adenocarcinoma patients compared with smoking patients. *Transl Lung Cancer Res* 2020;9:1407–21.
21. Chang G-C, Chiu C-H, Yu C-J. Initial results of LDCT screening for lung cancer risk among never-smokers with or without a family history of lung cancer. *Lancet Res Med*. Forthcoming 2023.
22. Dejima H, Hu X, Chen R, Zhang J, Fujimoto J, Parra ER, et al. Immune evolution from preneoplasia to invasive lung adenocarcinomas and underlying molecular features. *Nat Commun* 2021;12:2722.
23. MacMahon H, Naidich DP, Goo JM, Lee KS, Leung ANC, Mayo JR, et al. Guidelines for management of incidental pulmonary nodules detected on CT images: from the Fleischner society 2017. *Radiology* 2017;284:228–43.
24. Yoshihara K, Shahmoradgol M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612.
25. Hafemeister C, Satija R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 2019;20:296.
26. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol* 2017;18:220.
27. Leek JT, Storey JD. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* 2007;3:1724–35.
28. Sinjab A, Han G, Treekitkarnmongkol W, Hara K, Brennan PM, Dang M, et al. Resolving the spatial and cellular architecture of lung adenocarcinoma by multiregion single-cell sequencing. *Cancer Discov* 2021;11:2506–23.
29. Salcher S, Sturm G, Horvath L, Untergasser G, Kuempers C, Fotakis G, et al. High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. *Cancer Cell* 2022;40:1503–20.e8.
30. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinformatics* 2013;14:7.
31. Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. *Nat Methods* 2017;14:1083–6.
32. Facchinetti F, Lacroix L, Mezquita L, Scazecz J-Y, Lorient Y, Tselikas L, et al. Molecular mechanisms of resistance to BRAF and MEK inhibitors in BRAF^{V600E} non-small cell lung cancer. *Eur J Cancer* 2020;132:211–23.
33. Sun D, Guan X, Moran AE, Wu L-Y, Qian DZ, Schedin P, et al. Identifying phenotype-associated subpopulations by integrating bulk and single-cell sequencing data. *Nat Biotechnol* 2022;40:527–38.
34. Goveia J, Rohlenova K, Taverna F, Treps L, Conradi L-C, Pircher A, et al. An integrated gene expression landscape profiling approach to identify lung tumor endothelial cell heterogeneity and angiogenic candidates. *Cancer Cell* 2020;37:21–36.e13.
35. Cords L, Tietscher S, Anzeneder T, Langwieder C, Rees M, de Souza N, et al. Cancer-associated fibroblast classification in single-cell and spatial proteomics data. *Nat Commun* 2023;14:4294.
36. Becker WR, Nevins SA, Chen DC, Chiu R, Horning AM, Guha TK, et al. Single-cell analyses define a continuum of cell state and composition changes in the malignant transformation of polyps to colorectal cancer. *Nat Genet* 2022;54:985–95.
37. Hoogstrate Y, Draaisma K, Ghisai SA, van Hijfte L, Barin N, de Heer I, et al. Transcriptome analysis reveals tumor microenvironment changes in glioblastoma. *Cancer Cell* 2023;41:678–92.e7.
38. Gardiner N, Jogai S, Wallis A. The revised lung adenocarcinoma classification—an imaging guide. *J Thorac Dis* 2014;6(Suppl 5):S537–46.
39. Cords L, Engler S, Haberecker M, Rüschoff JH, Moch H, de Souza N, et al. Cancer-associated fibroblast phenotypes are associated with patient outcome in non-small cell lung cancer. *Cancer Cell* 2024;42:396–412.e5.
40. Szabo PM, Vajdi A, Kumar N, Tolstourov MY, Chen BJ, Edwards R, et al. Cancer-associated fibroblasts are the main contributors to epithelial-to-mesenchymal signatures in the tumor microenvironment. *Sci Rep* 2023;13:3051.
41. Sun H, Wang X, Wang X, Xu M, Sheng W. The role of cancer-associated fibroblasts in tumorigenesis of gastric cancer. *Cell Death Dis* 2022;13:874.
42. Lugano R, Ramachandran M, Dimberg A. Tumor angiogenesis: causes, consequences, challenges and opportunities. *Cell Mol Life Sci* 2020;77:1745–70.
43. Son H, Moon A. Epithelial-mesenchymal transition and cell invasion. *Toxicol Res* 2010;26:245–52.
44. Laumont CM, Banville AC, Gilardi M, Hollern DP, Nelson BH. Tumour-infiltrating B cells: immunological mechanisms, clinical impact and therapeutic opportunities. *Nat Rev Cancer* 2022;22:414–30.
45. Steen CB, Liu CL, Alizadeh AA, Newman AM. Profiling cell type abundance and expression in bulk tissues with CIBERSORTx. *Methods Mol Biol* 2020;2117:135–57.
46. Pan Y, Han H, Hu H, Wang H, Song Y, Hao Y, et al. KMT2D deficiency drives lung squamous cell carcinoma and hypersensitivity to RTK-RAS inhibition. *Cancer Cell* 2023;41:88–105.e8.
47. Zhou C, Chen H, Han L, Xue F, Wang A, Liang YJ. Screening of genes related to lung cancer caused by smoking with RNA-seq. *Eur Rev Med Pharmacol Sci* 2014;18:117–25.
48. Jiang N, Zhou J, Zhang W, Li P, Liu Y, Shi H, et al. RNF213 gene mutation in circulating tumor DNA detected by targeted next-generation sequencing in the assisted discrimination of early-stage lung cancer from pulmonary nodules. *Thorac Cancer* 2021;12:181–93.
49. Xiang C, Ji C, Cai Y, Teng H, Wang Y, Zhao R, et al. Distinct mutational features across preinvasive and invasive subtypes identified through comprehensive profiling of surgically resected lung adenocarcinoma. *Mod Pathol* 2022;35:1181–92.

50. Luca BA, Steen CB, Matusiak M, Azizi A, Varma S, Zhu C, et al. Atlas of clinically distinct cell states and ecosystems across human solid tumors. *Cell* 2021;184:5482–96.e28.
51. Lung Cancer Cohort Consortium LC3. The blood proteome of imminent lung cancer diagnosis. *Nat Commun* 2023;14:3042.
52. Saito H, Kameda Y, Masui K, Murakami S, Kondo T, Ito H, et al. Correlations between thin-section CT findings, histopathological and clinical findings of small pulmonary adenocarcinomas. *Lung Cancer* 2011;71:137–43.
53. Jiang X, Wang J, Deng X, Xiong F, Zhang S, Gong Z, et al. The role of microenvironment in tumor angiogenesis. *J Exp Clin Cancer Res* 2020;39:204.
54. Caronia LM, Phay JE, Shah MH. Role of BRAF in thyroid oncogenesis. *Clin Cancer Res* 2011;17:7511–7.
55. Croce L, Coperchini F, Magri F, Chiovato L, Rotondi M. The multifaceted anti-cancer effects of BRAF-inhibitors. *Oncotarget* 2019;10:6623–40.
56. Hung M-S, Chen I-C, Lin P-Y, Lung J-H, Li Y-C, Lin Y-C, et al. Epidermal growth factor receptor mutation enhances expression of vascular endothelial growth factor in lung cancer. *Oncol Lett* 2016;12:4598–604.
57. Minder P, Zajac Ethyc], Quigley JP, Deryugina EI. EGFR regulates the development and microarchitecture of intratumoral angiogenic vasculature capable of sustaining cancer cell intravasation. *Neoplasia* 2015;17:634–49.
58. Palumbo G, Esposito G, Carillio G, Manzo A, Montanino A, Sforza V, et al. Angiogenesis and epidermal growth factor receptor inhibitors in non-small cell lung cancer. *Explor Target Antitumor Ther* 2020;1:117–30.
59. Chang C-J, Chao C-H, Xia W, Yang J-Y, Xiong Y, Li C-W, et al. p53 regulates epithelial-mesenchymal transition and stem cell properties through modulating miRNAs. *Nat Cell Biol* 2011;13:317–23.
60. Elazezy M, Joosse SA. Techniques of using circulating tumor DNA as a liquid biopsy component in cancer management. *Comput Struct Biotechnol J* 2018;16:370–8.
61. Carmeliet P, Jain RK. Principles and mechanisms of vessel normalization for cancer and other angiogenic diseases. *Nat Rev Drug Discov* 2011;10:417–27.
62. Lopes-Coelho F, Martins F, Pereira SA, Serpa J. Anti-angiogenic therapy: current challenges and future perspectives. *Int J Mol Sci* 2021;22:3765.