Up-regulated tumor intrinsic growth potential and decreased immune function orchestrate the evolution of lung adenocarcinoma

Yue Zhao^{1,2,3*}, Jun Shang^{4*}, Jian Gao^{1,2,5*}, Han Han^{1,2,3}, Zhendong Gao^{1,2,3}, Yueren Yan^{1,2,3}, Qiang Zheng^{3,6}, Ting Ye^{1,2,3}, Fangqiu Fu^{1,2,3}, Chaoqiang Deng^{1,2,3}, Zelin Ma^{1,2,3}, Yang Zhang^{1,2,3}, Difan Zheng^{1,2,3}, Shanbo Zheng^{1,2,3}, Yuan Li^{3,6}, Zhiwei Cao^{7#}, Leming Shi^{2,4#} and Haiquan Chen^{1,2,3#}

1. Departments of Thoracic Surgery and State Key Laboratory of Genetic Engineering, Fudan University Shanghai Cancer Center, Shanghai, China.
2. Institute of Thoracic Oncology, Fudan University, Shanghai, China.
3. Department of Oncology, Shanghai Medical College, Fudan University, Shanghai, China.
4. State Key Laboratory of Genetic Engineering, Human Phenome Institute, School of Life Sciences and Shanghai Cancer Center, Fudan University, Shanghai, China.
5. International Human Phenome Institutes (Shanghai), Shanghai, China.
6. Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai, China.
7. School of Life Sciences, Fudan University, Shanghai 200092, China.

* Yue Zhao, Jun Shang and Jian Gao contributed equally to this work and are regarded as co-first authors.

Haiquan Chen, Leming Shi and Zhiwei Cao contributed equally to this work and are regarded as co-corresponding authors.

Correspondence to: Haiquan Chen (hqchen1@yahoo.com), Leming Shi (lemingshi@fudan.edu.cn) and Zhiwei Cao (zwcao@fudan.edu.cn).

This work was supported by the National Natural Science Foundation of China (81930073), Shanghai Science and Technology Innovation Action Project (20JC1417200), Shanghai Municipal Science and Technology Major Project (2017SHZDZX01, VBH1323001/026), Shanghai Municipal Key Clinical Specialty Project (SHSLCZDZK02104) and Pilot Project of Fudan University (IDF159045).

Abstract

Introduction

Lung adenocarcinoma is the most common pathological subtype of lung cancer. Precursors of lung adenocarcinoma, namely adenocarcinoma *in situ* and minimally invasive adenocarcinoma, have a superb 5-year survival rate after surgical resection. A deeper understanding of the key genetic changes driving the progression of lung adenocarcinoma is needed.

Methods

In this study, we performed whole-exome sequencing and RNA-sequencing on surgically resected 24 AIS, 74 MIA, 99 LUAD specimens and their adjacent paired normal tissues. Radiological, clinical, and pathological characteristics were recorded. Gene expression patterns were identified to find key pathways driving the progression of lung adenocarcinoma. Furthermore, genomic alterations and differential expression analyses were performed to compare tumors with different radiological manifestations. Finally, a progressive index was developed to quantitatively measure the level of imbalance between tumor intrinsic growth potential and immune microenvironment.

Results

12 patterns of gene expression were identified. Pathways associated with tumor growth and metastasis were found to be up-regulated as tumors progressed, while pathways associated with immune function were found to be down-regulated. Deconvolution of RNA-seq data also showed a decrease of CD8+ T cells and an increase of Tregs as the tumors progressed. Furthermore, tumors with more solid components on CT scan had a higher mutation frequency of tumor suppressor genes, higher tumor mutation burden and higher frequency of somatic copy number alterations. Finally, tumor progressive index demonstrated an increasing trend with the progression of lung adenocarcinoma.

Discussion

Our results reveal the imbalance of tumor intrinsic factors and immune function orchestrate the progression of lung adenocarcinoma.

Introduction

Lung cancer is one of the deadliest disease worldwide, with a 5-year survival of only ~19%¹. Lung adenocarcinoma (LUAD) is the most common pathological subtype. Pre-invasive stages of lung adenocarcinoma, namely adenocarcinoma *in situ* (AIS) and minimally invasive adenocarcinoma (MIA), have a nearly 100% 5-year survival rate after complete surgery^{2,3}. Although there have been studies on genomic and immune profiling of patients with AIS, MIA and LUAD, there lack a systematic study focusing on key molecular events that drive the evolution of lung adenocarcinoma⁴⁻⁸.

With the development of thoracic computed tomography (CT) scanning and the application of low-dose CT screening, an increasing number of small pulmonary nodules, especially subsolid nodules have been detected⁹⁻¹². The prognostic impact of solid components for lung adenocarcinomas presented as ground-glass opacities (GGO) on CT scanning has been under extensive investigation, where tumors manifesting as GGOs were found to have indolent clinical course^{13,14}. Our two previous studies further provided evidence on the prognostic value of lung adenocarcinomas manifesting as GGOs on CT scan^{15,16}. However, there still lack comprehensive genomic and transcriptomic studies comparing the differences between lung adenocarcinomas having GGO components and their counterparts not having GGO components on CT scan.

In this study, we performed whole-exome sequencing and RNA-sequencing on 197 surgically resected lung adenocarcinomas and divided them by pathological characteristics and radiological manifestations, aiming to find key genetic factors that drive the evolution of lung adenocarcinoma. 12 expression patterns were identified based on expression profiles, and pathway analysis was performed to reveal the biological functions for each pattern. Tumor intrinsic growth potential and immune microenvironment were assessed, and immune cell infiltration was calculated using transcriptomic data. Finally, a tumor progressive index was developed to provide a quantitative measurement of the level of unbalance between tumor intrinsic growth potential and immune function.

Methods

Study cohort

197 lung adenocarcinoma patients who underwent surgery between September 2011 and May 2016 at the Department of Thoracic Surgery, Fudan University Shanghai Cancer Center were retrospectively included in this study (figure 1A). None of the patients received neoadjuvant therapy. This study was approved by the Committee for Ethical Review of Research (Fudan University Shanghai Cancer Center Institutional Review Board No. 090977-1). Informed consents of all patients for donating their samples to the tissue bank of Fudan University Shanghai Cancer Center were obtained from patients themselves or their

relatives. Radiological, histological evaluation and Follow-up protocol are described in detail in supplementary methods.

Radiological and histological evaluation

Whole lung CT scans were performed on each patient included before surgery as previously described¹⁵. For each nodule, the maximum diameter of both the entire nodule and solid component on the single largest axial dimension was recorded on lung window. Pulmonary nodules were further divided into 3 groups: pure GGOs, where there was no solid component in one pulmonary nodule; subsolid nodules, where both solid and GGO components existed in one pulmonary nodule; solid nodules, where the pulmonary nodule contained only solid component (figure 1A). CT scans were reviewed by two radiologists independently, and interobserver and intraobserver agreements were measured to quantify the reproducibility and accuracy between the two radiologists as previously described¹⁵.

Intraoperative frozen section diagnosis was made after the tumor was resected, and postoperative diagnosis was made after surgery by two independent pathologists. According to the IASLC/ATS/ERS classification, tumors were classified as adenocarcinoma *in situ* (AIS), minimally invasive adenocarcinoma (MIA) or invasive lung adenocarcinoma (LUAD) based on their histological presentations. Invasive lung adenocarcinoma subtypes were further analyzed in a semi-quantitative manner, where the components of different subtypes (lepidic, acinar, papillary, micropapillary, solid and invasive mucinous) were recorded in 5% increments. The predominant subtype was the one with the largest percentage (not necessarily 50% or higher)¹⁷. Pathological stage of disease was determined according to the 8th TNM staging system.

Follow-up protocol

Patients were followed up regularly after surgery as we previously described¹⁵. Briefly, patients were followed up every 3 months after surgery for the first 2 years, where physical examination, chest CT scans and abdominal ultrasonography were performed every 3 to 6 months. The follow-up interval was changed to every 6 months for the third year and once a year from then on. Brain CT or magnetic resonance imaging (MRI) and bone scintigraphy were performed every 6 months for patients with invasive adenocarcinoma in the first 3 years. In addition, positron emission tomography-CT (PET-CT) scans was performed if necessary. Recurrence-free survival (RFS) was defined as the time between surgery and first recurrence or last follow-up. Patients with no recurrence but died from other causes were censored on that date. Overall survival (OS) was defined as the time between surgery and death or last follow-up.

Whole-exome sequencing

Genomic DNA from tumors and paired adjacent normal tissues were extracted and prepared using the QIAamp DNA Mini Kit (Qiagen, Germany) following the manufacturer's instructions. Exon libraries were constructed using the SureSelect XT Target Enrichment System. A total amount of 1-3 µg genomic DNA for each samples was fragmented into an average size of ~200 bp. DNA was hybridized, captured and amplified using SureSelect XT reagents and protocols to generate indexed, target-enriched library amplicons. Constructed libraries were then sequenced on the Illumina HiSeq X Ten platform and 150 bp paired-end reads were generated.

Alignment, mutation calling and somatic copy number alteration calling

Sequence reads from the exome capture libraries were aligned to the reference human genome (hg19) using BWA-MEM¹⁸. Picard tools was then used for marking PCR duplicated and the Genome Analysis Toolkit (GATK) was used to perform base quality recalibration and local indel re-alignments¹⁹. Single nucleotide variants (SNVs) were called using MuTect and MuTect2²⁰. Indels were called using MuTect2 and Strelka v2.0.13²¹. Variants were filtered if called by only one tool. After the variants were called, Oncotator v1.9.1 was used for annotating somatic mutations²², and significantly mutated genes were identified using MutSig2CV²³. Tumor mutation burden (TMB) was calculated as the total number of nonsynonymous SNVs and indels per sample divided by 30, given the total coverage of ~30 Mb. CNVkit v0.9.7 with default parameters was used to perform somatic copy number alteration (SCNA) analysis for alignment reads²⁴. Amplification and deletion peaks were identified using GISTIC2.0 from segment files²⁵. Amplification and deletion thresholds were set 0.1 and -0.1, respectively. Frequency distribution of amplification and deletion was shown using R package copynumber v1.26.0²⁶.

RNA-sequencing and calculation of expression

Total RNA from tumors and paired adjacent normal tissues was extracted and prepared using NucleoZOL (Macherey-Nagel, Germany) and NucleoSpin RNA Set for NucleoZOL (Macherey-Nagel, Germany) following the manufacturer's instructions. A total amount of 3 µg RNA per sample was used as initial material for RNA sample preparations. Ribosomal RNA was removed using Epicenter Ribo-Zero Gold Kits (Epicenter, USA). Sequencing libraries were generated using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, Ipswich, USA) according to the manufacturer's instructions. Libraries were then sequenced on the Illumina HiSeq X Ten platform and 150 bp paired-end reads were generated. As a quality-control step, only those samples with RNA integrity number (RIN) >= 5.0 were included in this study.

After the data were obtained, RNA-seq reads were aligned to the reference human genome (hg19) using STAR v2.5.3²⁷. Expression values were normalized to the transcripts per million (TPM) estimates using RSEM v1.3.0²⁸.

Identification of gene expression patterns

150 samples with RIN ≥ 5.0 were divided into 4 groups: normal (n = 150), AIS (n = 16), MIA (n = 52) and LUAD (n = 82). Analysis of variance (ANOVA) was used to determine differentially expressed genes. First, paired comparison was performed between two adjacent groups (normal vs AIS, AIS vs MIA and MIA vs LUAD), then genes with $p < 0.0001$ and $|\text{fold change}| \geq 2$ in at least 1 of 3 comparisons were considered significant and put into downstream analyses. 12 patterns of gene expression were then identified based on their trends of expression between each two adjacent groups in the study cohort (up trend, no significant change or down trend, figure 1B and figure 2). Finally, KOBAS was used to perform KEGG pathway enrichment analysis for the patterns identified in the study cohort²⁹.

Differential expression analysis and gene set enrichment analysis

Differential expression analysis was performed using DESeq2³⁰. False discovery rate (FDR) was controlled using the Benjamini-Hochberg method. Genes with $\log_2\text{-fold change} \geq 1$ or ≤ -1 , and adjusted p value < 0.05 were considered differentially expressed between two groups. Significantly differentially expressed genes were then included in the downstream gene set enrichment analysis (GSEA). GSEA was performed using the GSEA software (version 3.0)^{31,32}. Briefly, normalized enrichment scores (NES) and p values were calculated to sort the pathways enriched in each group. Cancer hallmark pathways were used for this analysis, and gene set permutations were performed 1000 times for each analysis. Enrichment maps were generated to visualize the results.

Development of tumor progressive index

Tumor progressive index was developed for quantitatively measuring the level of imbalance between tumor intrinsic growth potential and immune response. To minimize the possible confounding effect introduced by genes with low expression, we first filtered out genes with mean TPM < 1.0 across all samples. Next, genes for calculating the tumor index were selected from Pattern 1 where there was a significant increase between each two adjacent groups. Genes for calculating the immune index were selected from Pattern 8 where there was a significant decrease from normal tissue to AIS and from MIA to LUAD. Those genes that were both selected from Pattern 8 and were in an *a priori* immune-related gene list containing 730 genes was used to calculate the immune index⁷. For both tumor index and immune index, expression for those genes across all samples were first transformed to Z-scores, then for each sample, tumor index was calculated as the mean value of Z-scores for the tumor index genes, while

immune index was calculated as the mean value of Z-scores for immune index genes. Finally, tumor progressive index was calculated as the subtraction of tumor index by immune index:

$$\text{Tumor progressive index} = \text{tumor index} - \text{immune index}$$

Statistical analysis

Clinical and pathological characteristics were recorded and compared among three groups. Pearson χ^2 test and Fisher's exact test were used to compare categorical variables wherever applicable. Non-parametric Wilcoxon signed-rank test was used to compare medians of groups of continuous variables. Kaplan-Meier survival curves and log-rank p values were calculated for patients' RFS and OS. All statistical analyses and graphing work were performed using R (version 3.6.0, R Foundation for Statistical Computing, Vienna, Austria). Two-tailed $p < 0.05$ was considered significant for all statistical analyses unless stated otherwise.

Results

Association between pathological stage and radiological manifestations

Tumor samples were divided into 3 groups based on their pathological findings: 24 AIS, 74 MIA and 99 LUAD, and divided into 3 groups based on their radiological manifestations: 69 pure GGOs, 63 subsolid nodules and 65 solid nodules (figure 1A). Clinical and pathological characteristics, including sex, smoking status, tumor location, pathology and adenocarcinoma subtypes were compared (table 1). Of note, we found that there was an association between radiological and pathological presentations. 66 of 69 (95.7%) of pulmonary nodules manifesting as pure GGOs on CT scan were either adenocarcinoma *in situ* (AIS) or minimally invasive adenocarcinoma (MIA), while for pulmonary nodules manifesting as pure solid nodules, only 1 out of 65 (1.5%) was either AIS or MIA (MIA in this case). For pulmonary nodules manifesting as subsolid nodules on CT scan, 31 out of 63 (49.2%) were either AIS or MIA, and 32 (50.8%) were invasive adenocarcinoma. Compared to male patients, female patients had significantly higher frequency of pure GGO lesions ($p = 0.025$). For predominant adenocarcinoma subtypes, lepidic subtype was significantly enriched in the pure GGO group ($p < 0.001$), while papillary subtype was significantly enriched in the pure solid group ($p = 0.004$). Solid subtype was only found predominant in the pure solid group. Moreover, for presenting subtypes, lepidic subtype was significantly enriched in the pure GGO group ($p = 0.001$), and solid subtype was significantly enriched in the pure solid group ($p = 0.014$). Micropapillary subtype was also more likely to be found in the pure solid group, though without a significant difference (pure GGO vs subsolid vs solid: 0.0% vs 3.1% vs 9.4%, $p = 0.472$). Taken together, these results demonstrated there was a link between

radiological and pathological findings, therefore, radiological manifestations can, at least partly, help predict the invasiveness of lung adenocarcinoma.

12 expression patterns were identified as tumors progress

One aim of this study is to find key genetic factors that drive the evolution of lung adenocarcinoma. 150 samples with RNA integrity number (RIN) ≥ 5.0 were included in this part. We developed an ANOVA model with pathological stage (normal tissue, AIS, MIA and LUAD) as a factor and patient as a random effect. 2023 genes with $p < 0.0001$ and $|\text{fold change}| > 2$ in pairwise comparisons of two adjacent groups of patients (normal tissue vs AIS, AIS vs MIA, and MIA vs LUAD, respectively) were divided into 12 following patterns: Pattern 1 (9 genes): increase from normal to AIS, from AIS to MIA and from MIA to LUAD; pattern 2 (97 genes): increase from normal to AIS and from MIA to LUAD, no change from AIS to MIA; pattern 3 (446 genes): increase from normal to AIS alone; pattern 4 (54 genes): no change from normal to AIS, increase from AIS to MIA and from MIA to LUAD; pattern 5 (20 genes): increase from AIS to MIA alone; pattern 6 (309 genes): increase from MIA to LUAD alone; pattern 7 (3 genes): decrease from normal to AIS, from AIS to MIA and from MIA to LUAD; pattern 8 (175 genes): decrease from normal to AIS and from MIA to LUAD, no change from AIS to MIA; pattern 9 (771 genes): decrease from normal to AIS alone; pattern 10 (6 genes): no change from normal to AIS, decrease from AIS to MIA and from MIA to LUAD; pattern 11 (3 genes): decrease from AIS to MIA alone; pattern 12 (130 genes): decrease from MIA to LUAD alone (figure 2 and supplementary table 1). Pathway enrichment analysis showed that pathways associated with tumor invasiveness and cell growth were enriched in up-trend patterns (patterns 1 to 6). Of note, PI3K-AKT signaling pathway and NF- κ B signaling pathway were up-regulated in AIS compared with normal tissue, suggesting malignant behaviors could exist in as early as AIS. Epithelial-mesenchymal transition, which was highly associated with tumor metastasis, was also found to be increased from AIS to MIA, suggesting an acquisition of metastatic potential as invasive subtypes of lung adenocarcinoma emerge³³. Moreover, cell cycle was found to be increased from MIA to LUAD, consistent with the fact that AIS/MIA behaved more indolent than LUAD. On the other hand, for down-trend patterns, natural killer mediated cytotoxicity and TGF- β signaling pathway were found to be decreased from normal lung tissue to AIS, suggesting an inhibited immune response against tumor. Chemokine signaling, IL2-STAT5 signaling, TNF- α signaling via NF- κ B and leukocyte transendothelial migration were also found to be enriched in down-trend patterns, further suggesting an impaired immune function in tumor microenvironment. Interestingly, Wnt signaling pathway, which was often up-regulated in cancer, was found to be decreased from AIS to MIA and from MIA to LUAD. A deeper look into pattern 10 showed that this finding was contributed by WIF1 (Wnt inhibiting factor 1), which was an inhibitor of Wnt signaling pathway. Taken together, our results suggest that both an increased

tumor intrinsic growth potential and an inhibited immune microenvironment contributed to the development and progression of lung adenocarcinoma.

Decreased number of cytotoxic CD8+ T cells and increased number of Tregs in LUAD

Next, we used CIBERSORT, a deconvolution method to estimate the infiltration of immune cells in different pathological stages³⁴. We found that the number of CD8+ T cells decreased as tumor stage increased (figure 3A), which was consistent with Dejima et al. on the immune evolution from neoplasia to invasive lung adenocarcinoma⁷. Furthermore, we found that the number of Tregs increased as tumor stage increased, suggesting an enhanced immunosuppression as disease progressed (figure 3B), which was also validated by the Dejima dataset (figure 3C). Interestingly, we found an increased number of activated NK cells and a decreased number of resting NK cells as disease progressed, and validated this result using the Dejima dataset, suggesting NK cells might be important in the progression of lung adenocarcinoma (figure 3D and 3E). A more detailed staging of LUAD also showed similar results, with tumors at earlier stages having more CD8+ T cells and less Tregs, while tumors at later stages having less CD8+ T cells and more Tregs (supplementary figure 1). Taken together, our findings showed that a suppressed immune microenvironment plays an important role in the progression of lung adenocarcinoma.

Genomic alterations of tumors with different radiological manifestations

We next assessed the genomic alterations of tumors with different radiological manifestations. The landscape of somatic mutations for all patients included in this study was shown in Figure 4a. The most frequently mutated gene in this cohort was *EGFR* (50%), followed by *TP53* (22%), *RBM10* (8%), *ERBB2* (5%), *BRAF* (5%), *RB1* (5%), *KRAS* (4%) and *NF1* (3%), etc (figure 4A). A further comparison of mutation frequency in driver genes and tumor suppressor genes found that *BRAF* and *ERBB2* mutations were significantly enriched in pure GGOs and subsolid nodules ($p = 0.046$ and 0.016 , respectively), while *ALK* fusions were detected only in solid nodules (3 out of 65, $p = 0.045$). There was no significant difference in total frequency of driver genes, however, a significant difference was observed in the frequency of tumor suppressor genes. The number of mutations in common tumor suppressor genes was 8 (11.6%) for the pure GGOs, 19 (30.2%) for subsolid nodules and sharply rose to 50 (76.9%) for solid nodules ($p < 0.001$, supplementary table 2). For individual tumor suppressor genes, frequency of *TP53* and *RB1* mutations was found to be significantly higher in solid nodules ($p < 0.001$ and $p = 0.038$, respectively, figure 4B, 4C and supplementary table 1). Tumor mutation burden was significantly higher in solid nodules compared to GGOs and subsolid nodules on CT scan (figure 4D).

Frequency of somatic copy number alterations were higher in solid nodules

To explore the somatic copy number alterations (SCNAs) of pure GGOs, subsolid and solid nodules, we identified SCNAs from raw sequencing reads of samples from the three groups. Notably, the SCNA frequency including amplification and deletion increased gradually from pure GGOs to solid nodules (figure 4E). We used GISTIC2.0 to identify significant arm-level SCNAs in samples of pure GGOs, subsolid and solid nodules. The union of significant arm-level event of the three groups was used for further analysis. As a result, the ratio of amplification and deletion in arm-level SCNAs also increased from pure GGOs to solid nodules. These results indicated that the frequency of somatic copy number alterations was positively associated with solid component in pulmonary nodules.

Expression profile, differential expression analysis and gene set enrichment analysis (GSEA)

RNA-seq was performed to better understand the expression profiles of those radiologically different pulmonary nodules. Principle component analysis (PCA) showed that the transcriptomic profiles of the 3 groups were different (supplementary figure 2). Differential expression analysis demonstrated that 1359 genes were differentially expressed between the solid group and the GGO/subsolid group ($|\log_2\text{-fold change}| \geq 1$, adjusted p value < 0.05 , figure 5A and 4B). Of the 1359 differentially expressed genes, 870 were up-regulated and 489 were down-regulated (supplementary table 3). Furthermore, subgroup analyses showed that compared with the pure GGOs, subsolid nodules had 269 significantly up-regulated genes and 76 down-regulated genes, while compared with the subsolid group, the solid group had 668 significantly up-regulated genes and 320 down-regulated genes (supplementary figures 3-4 and supplementary table 4-5). There were 119 up-regulated genes and 20 down-regulated genes that were in common in the subgroup comparisons.

To investigate the functional basis of the changes between GGOs/subsolid nodules and solid nodules, we performed gene set enrichment analysis (GSEA) on the RNA-seq data. Among all the predefined cancer hallmark pathways, genes associated with cellular proliferation and tumor growth were significantly enriched in the solid group, including glycolysis (NES=2.05, $p < 0.001$), G2M checkpoint (NES = 1.86, $p = 0.006$), E2F targets (NES = 1.84, $p = 0.004$), the MTORC1 signaling pathway (NES = 1.79, $p = 0.004$), MYC targets v1 (NES = 1.71, $p = 0.012$) and mitotic spindle (NES = 1.50, $p = 0.048$). Moreover, genes associated with angiogenesis (NES = 1.57, $p = 0.041$) and hypoxia (NES = 1.56, $p = 0.031$) were also found to be significantly enriched in the solid group (figure 5C). Taken together, these results suggest that compared with pure GGOs and

subsolid nodules, solid nodules have a higher proliferation rate and tumor cells behave more malignantly.

Tumor progressive index showed an imbalance between tumor intrinsic growth potential and immune response

Finally, we developed a tumor progressive index which quantitatively measures the level of imbalance between tumor intrinsic growth potential and immune response against tumors. Based on our previous findings, we used the genes in Pattern 1 and immune-related genes in Patterns 7 & 8 with mean TPM ≥ 1 across all samples (methods). 4 genes that were associated with apoptosis, tumor metastasis and progression (*BCL2L15*, *COMP*, *CST1* and *FAM83A*) reflected tumor intrinsic growth potential, while 8 genes (*ITLN2*, *MARCO*, *C8B*, *MASP1*, *CD36*, *TAL1*, *PPBP* and *CDH5*) that were associated with immune microenvironment reflected immune response against tumors. A negative tumor progressive index indicates that the immune system is competent enough to suppress the progression of tumors, while a positive tumor progressive index indicates that the immune system can no longer suppress the growth of tumor cells. In our study cohort, tumor progressive index was negative in normal tissues but positive in AIS and stages onwards, indicating immune escape already existed in AIS, the precursor stage of lung adenocarcinoma, and became more severe with the progression of disease (figure 6A). Same increasing trend was observed in another dataset which showed a significant increase of tumor progressive index from normal tissue to atypical adenomatous hyperplasia (AAH) then to LUAD (figure 6B)³⁵. Interestingly, there was a negative tumor progressive index in the stage of AAH, where tumors could not overcome the immune system to metastasize further. Survival analyses showed that patients with a high tumor progressive index had poorer RFS and OS (figure 6C and 6D). Taken together, our results suggested that increased tumor intrinsic growth potential and impaired immune response against tumor work together to drive the progression of lung adenocarcinoma (figure 6E).

Discussions

In this study, we provided a comprehensive analysis integrating clinical, radiological, pathological, genomic and transcriptomic analysis of 197 pulmonary lesions with different radiological and pathological manifestations.

We first assessed the transcriptomic profiles of our cohort. Based on the differentially expressed genes between each two adjacent groups, 12 expression patterns were identified (figure 2). Genes associated with cell cycle and G2M checkpoint were found to be significantly up-regulated from MIA to LUAD, but not different from normal tissue to MIA. This indicates that pre-invasive stages of lung adenocarcinoma had more indolent behaviors, which might explain why AIS and MIA had a nearly 100% 5-year survival rate after surgical resection.

Histologically, AIS is defined as a ≤ 3 cm adenocarcinoma lacking invasive patterns, while MIA is a ≤ 3 cm adenocarcinoma with invasive patterns of no more than 5 mm in size³³. PI3K-AKT signaling pathway, a classical pathway that were up-regulated in various cancer types, were found to be increased from normal tissue to AIS³⁶. Though AIS is the precursor of LUAD, this finding shows that cells in this stage already have an increased potential for growth. On the other hand, we found that epithelial-mesenchymal transition, a biological process known to increase metastatic potential of cancer cells³⁷, were found to be significantly increased from AIS to MIA, consistent with histological differences between AIS and MIA. Though AIS and MIA both have a nearly perfect prognosis, our results suggest that MIA should be surgically intervened before it progresses to the next stage as invasive patterns already exist. To our surprise, several immune related pathways were found in down-trend expression patterns, indicating a decreased or impaired immune response as tumors progress. Consistent with this finding, using CIBERSORT to deconvolute our bulk-sequencing data, we found a decrease in the number of CD8+ cytotoxic T cells and an increase in the number of Treg cells, which was also validated by another dataset that compared immune cells among normal tissue, atypical adenomatous hyperplasia (AAH), AIS, MIA and LUAD⁷. Taken together, our results demonstrated that increased tumor intrinsic growth signals and decreased immune response orchestrate the evolution of lung adenocarcinoma.

We next demonstrated that there was an association between radiological and pathological presentations. Most nodules that were GGOs on CT scan were AIS or MIA, while most solid nodules were invasive lung adenocarcinoma (table 1). This might provide explanation of why pulmonary nodules manifesting as GGOs on CT scan usually have indolent clinical courses. This result was also consistent with previous studies and provided a link between histological and radiological manifestations^{38,39}. For major driver mutations, we found that frequency of neither *EGFR* nor *KRAS* mutations was significantly different among the 3 groups (figure 4B and supplementary table 2). *BRAF* mutations were more enriched in the pure GGO group, while *ALK* fusion were only seen in the solid group. On the other hand, the frequency of mutations in tumor suppressor genes were significantly different among the 3 groups, with a sharp increase from 2 (2.9%) in the pure GGO group, to 8 (12.7%) in the subsolid group and 33 (50.8%) in the solid group (figure 4C and supplementary table 2), suggesting a pivotal role that tumor suppressor genes play in the progression of lung adenocarcinoma. For individual tumor suppressor genes, *TP53* and *RB1* were significantly different among the 3 groups ($p < 0.001$ and $p = 0.038$, respectively, figure 3C and supplementary table 2). Interestingly, we incidentally found that there was a significant difference in *EGFR* mutation types among the 3 groups. *EGFR* exon 19 deletion was more enriched in the GGO group, while *EGFR* L858R was more often seen in the solid group (supplementary figure 5). Although in clinical

settings, treatment for lung adenocarcinoma patients harboring *EGFR* L858R and *EGFR* exon 19 deletion is the same, our results indicates that those two major types of *EGFR* mutations might render tumors different malignant potentials.

Previous studies have discussed the genomic alterations in pulmonary nodules manifesting as GGO. In 2015, Kobayashi et al. evaluated *EGFR*, *KRAS*, *ALK* and *HER2* mutations in 104 pulmonary nodules manifesting as GGO. They found that *EGFR* was mutated in 64% of all the 104 samples, while *KRAS*, *ALK* and *HER2* were mutated in 4%, 3% and 4% of the samples, respectively⁴⁰. In 2018, Lu et al. reported that *EGFR* was mutated in 75 out of 156 (48.1%) lung adenocarcinoma patients, which was similar to 50% in our cohort⁴¹. However, we did not observe a significant difference in *EGFR* mutation frequency among those 3 groups (figure 4B and supplementary table 2). In 2020, Li et al. performed whole-exome sequencing on 154 pulmonary subsolid nodules from 120 patients and found that *EGFR* were the most frequently mutated gene, followed by *RBM10*, *TP53*, *STK11* and *KRAS*. They also found that frequency of *EGFR*, *RBM10* and *TP53* was significantly different between pure GGO and subsolid nodules⁴². While not significantly different for *EGFR* and *RBM10* in our cohort, there was indeed a significant difference of the frequency of *TP53* mutations (figure 4C and supplementary table 2).

For somatic copy number alterations (SCNAs), we found that the frequency of SCNAs increased from pure GGO to subsolid and solid nodules. As the solid components increased, the genome became more unstable. Frequency of arm-level alterations also followed the same trend, where pure GGO had fewer significant arm-level events and solid nodules had more significant arm-level events. Genomic instability was widely reported to be associated with increased rate of tumor proliferation and progression among different cancer types⁴³⁻⁴⁵. Due to a higher level of genomic instability, solid nodules behave more malignant than GGO and subsolid nodules. Another unique advantage of our study is the application of RNA-seq in our cohort. By adding another aspect of analysis, we found that there were 1359 differentially expressed genes between the GGO/subsolid group and the solid group. GSEA analysis revealed that genes associated with cellular proliferation and tumor growth were significantly enriched in the solid group, including glycolysis, G2M checkpoint, E2F targets, the MTORC1 signaling pathway, MYC targets v1 and mitotic spindle, along with angiogenesis and hypoxia, which had a well-established role as a microenvironmental factor promoting tumor metastasis⁴⁶.

Tumor intrinsic growth potential and immune microenvironment are both associated with tumor evolution and progression^{47,48}. Based on our findings, a tumor progressive index was designed to quantitatively measure the level of imbalance between tumor intrinsic growth potential and immune

microenvironment. Interestingly, AAH, a stage of pre-cancerous hyperplasia, had a negative tumor progressive index, indicating some key alteration would be needed to activate the malignant transformation of cells. AIS, MIA and LUAD all had a progressive index of more than 0, and the index increased as the tumor developed to the next stage. Furthermore, patients with a high tumor progressive index tended to have a poorer recurrence-free survival and overall survival. Taken together, tumor progressive index can be used as a predictor for disease progression and prognosis.

In summary, this study integrates the genomic alterations, transcriptomic profiles, and histological and radiological progression of lung adenocarcinoma, providing deeper understandings of the evolution of this disease.

References

- 1 Siegel, R. L., Miller, K. D. & Jemal, A. Cancer statistics, 2020. *CA Cancer J Clin* **70**, 7-30, doi:10.3322/caac.21590 (2020).
- 2 Maeshima, A. M. *et al.* Histological scoring for small lung adenocarcinomas 2 cm or less in diameter: a reliable prognostic indicator. *J Thorac Oncol* **5**, 333-339, doi:10.1097/JTO.0b013e3181c8cb95 (2010).
- 3 Borczuk, A. C. *et al.* Invasive size is an independent predictor of survival in pulmonary adenocarcinoma. *Am J Surg Pathol* **33**, 462-469, doi:10.1097/PAS.0b013e318190157c (2009).
- 4 Izumchenko, E. *et al.* Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. *Nat Commun* **6**, 8258, doi:10.1038/ncomms9258 (2015).
- 5 Murphy, S. J. *et al.* Genomic rearrangements define lineage relationships between adjacent lepidic and invasive components in lung adenocarcinoma. *Cancer Res* **74**, 3157-3167, doi:10.1158/0008-5472.CAN-13-1727 (2014).
- 6 Hu, X. *et al.* Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma. *Nat Commun* **10**, 2978, doi:10.1038/s41467-019-10877-8 (2019).
- 7 Dejima, H. *et al.* Immune evolution from preneoplasia to invasive lung adenocarcinomas and underlying molecular features. *Nat Commun* **12**, 2722, doi:10.1038/s41467-021-22890-x (2021).
- 8 Chen, H. *et al.* Genomic and immune profiling of pre-invasive lung adenocarcinoma. *Nat Commun* **10**, 5472, doi:10.1038/s41467-019-13460-3 (2019).
- 9 Walter, J. E. *et al.* Characteristics of new solid nodules detected in incidence screening rounds of low-dose CT lung cancer screening: the NELSON study. *Thorax* **73**, 741-747, doi:10.1136/thoraxjnl-2017-211376 (2018).
- 10 Horeweg, N. *et al.* Lung cancer probability in patients with CT-detected pulmonary nodules: a prespecified analysis of data from the NELSON trial of low-dose CT screening. *Lancet Oncol* **15**, 1332-1341, doi:10.1016/S1470-2045(14)70389-4 (2014).
- 11 McWilliams, A. *et al.* Probability of cancer in pulmonary nodules detected on first screening CT. *N Engl J Med* **369**, 910-919, doi:10.1056/NEJMoa1214726 (2013).
- 12 Fu, F., Zhou, Y., Zhang, Y. & Chen, H. Lung cancer screening strategy for non-high-risk individuals: a narrative review. *Transl Lung Cancer Res* **10**, 452-461, doi:10.21037/tlcr-20-943 (2021).
- 13 Hattori, A., Matsunaga, T., Takamochi, K., Oh, S. & Suzuki, K. Prognostic impact of a ground glass opacity component in the clinical T classification of non-small cell lung cancer. *J Thorac Cardiovasc Surg* **154**, 2102-2110 e2101, doi:10.1016/j.jtcvs.2017.08.037 (2017).
- 14 Hattori, A. *et al.* Prognostic impact of a ground-glass opacity component in clinical stage IA non-small cell lung cancer. *J Thorac Cardiovasc Surg* **161**, 1469-1480, doi:10.1016/j.jtcvs.2020.01.107 (2021).
- 15 Ye, T. *et al.* Lung Adenocarcinomas Manifesting as Radiological Part-Solid Nodules Define a Special Clinical Subtype. *J Thorac Oncol* **14**, 617-627, doi:10.1016/j.jtho.2018.12.030 (2019).
- 16 Fu, F. *et al.* Distinct Prognostic Factors in Patients with Stage I Non-Small Cell Lung Cancer with Radiologic Part-Solid or Solid Lesions. *J Thorac Oncol* **14**, 2133-2142, doi:10.1016/j.jtho.2019.08.002 (2019).

- 17 Travis, W. D. *et al.* International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J Thorac Oncol* **6**, 244-285, doi:10.1097/JTO.0b013e318206a221 (2011).
- 18 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997 (2013). <<https://ui.adsabs.harvard.edu/abs/2013arXiv1303.3997L>>.
- 19 DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-498, doi:10.1038/ng.806 (2011).
- 20 Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* **31**, 213-219, doi:10.1038/nbt.2514 (2013).
- 21 Saunders, C. T. *et al.* Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811-1817, doi:10.1093/bioinformatics/bts271 (2012).
- 22 Ramos, A. H. *et al.* Oncotator: cancer variant annotation tool. *Hum Mutat* **36**, E2423-2429, doi:10.1002/humu.22771 (2015).
- 23 Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495-501, doi:10.1038/nature12912 (2014).
- 24 Talevich, E., Shain, A. H., Botton, T. & Bastian, B. C. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol* **12**, e1004873, doi:10.1371/journal.pcbi.1004873 (2016).
- 25 Mermel, C. H. *et al.* GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol* **12**, R41, doi:10.1186/gb-2011-12-4-r41 (2011).
- 26 Nilsen, G. *et al.* Copynumber: Efficient algorithms for single- and multi-track copy number segmentation. *BMC Genomics* **13**, 591, doi:10.1186/1471-2164-13-591 (2012).
- 27 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 28 Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).
- 29 Bu, D. *et al.* KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. *Nucleic Acids Research* **49**, W317-W325, doi:10.1093/nar/gkab447 (2021).
- 30 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 31 Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550, doi:10.1073/pnas.0506580102 (2005).
- 32 Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **34**, 267-273, doi:10.1038/ng1180 (2003).
- 33 Travis, W. D. *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* **10**, 1243-1260, doi:10.1097/JTO.0000000000000630 (2015).
- 34 Newman, A. M. *et al.* Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods* **12**, 453-457, doi:10.1038/nmeth.3337 (2015).

- 35 Sivakumar, S. *et al.* Genomic Landscape of Atypical Adenomatous Hyperplasia Reveals Divergent Modes to Lung Adenocarcinoma. *Cancer Res* **77**, 6119-6130, doi:10.1158/0008-5472.CAN-17-1605 (2017).
- 36 Jiang, N. *et al.* Role of PI3K/AKT pathway in cancer: the framework of malignant behavior. *Mol Biol Rep* **47**, 4587-4629, doi:10.1007/s11033-020-05435-1 (2020).
- 37 Dongre, A. & Weinberg, R. A. New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat Rev Mol Cell Biol* **20**, 69-84, doi:10.1038/s41580-018-0080-4 (2019).
- 38 Zhang, Y., Fu, F. & Chen, H. Management of Ground-Glass Opacities in the Lung Cancer Spectrum. *Ann Thorac Surg* **110**, 1796-1804, doi:10.1016/j.athoracsur.2020.04.094 (2020).
- 39 Sun, F. *et al.* Ground glass opacities: Imaging, pathology, and gene mutations. *J Thorac Cardiovasc Surg* **156**, 808-813, doi:10.1016/j.jtcvs.2018.02.110 (2018).
- 40 Kobayashi, Y., Mitsudomi, T., Sakao, Y. & Yatabe, Y. Genetic features of pulmonary adenocarcinoma presenting with ground-glass nodules: the differences between nodules with and without growth. *Ann Oncol* **26**, 156-161, doi:10.1093/annonc/mdu505 (2015).
- 41 Lu, Q. *et al.* Epidermal growth factor receptor mutation accelerates radiographic progression in lung adenocarcinoma presented as a solitary ground-glass opacity. *J Thorac Dis* **10**, 6030-6039, doi:10.21037/jtd.2018.10.19 (2018).
- 42 Li, Y. *et al.* Genomic characterisation of pulmonary subsolid nodules: mutational landscape and radiological features. *Eur Respir J* **55**, doi:10.1183/13993003.01409-2019 (2020).
- 43 Coschi, C. H. & Dick, F. A. Chromosome instability and deregulated proliferation: an unavoidable duo. *Cell Mol Life Sci* **69**, 2009-2024, doi:10.1007/s00018-011-0910-4 (2012).
- 44 Endesfelder, D. *et al.* Chromosomal instability selects gene copy-number variants encoding core regulators of proliferation in ER+ breast cancer. *Cancer Res* **74**, 4853-4863, doi:10.1158/0008-5472.CAN-13-2664 (2014).
- 45 Watkins, T. B. K. *et al.* Pervasive chromosomal instability and karyotype order in tumour evolution. *Nature* **587**, 126-132, doi:10.1038/s41586-020-2698-6 (2020).
- 46 Black, J. C. *et al.* Hypoxia drives transient site-specific copy gain and drug-resistant gene expression. *Genes Dev* **29**, 1018-1031, doi:10.1101/gad.259796.115 (2015).
- 47 Rosenthal, R. *et al.* Neoantigen-directed immune escape in lung cancer evolution. *Nature* **567**, 479-485, doi:10.1038/s41586-019-1032-7 (2019).
- 48 Pitt, J. M. *et al.* Targeting the tumor microenvironment: removing obstruction to anticancer immune responses and immunotherapy. *Ann Oncol* **27**, 1482-1492, doi:10.1093/annonc/mdw168 (2016).

Acknowledgements

This work was supported by the National Natural Science Foundation of China (81930073), Shanghai Science and Technology Innovation Action Project (20JC1417200), Shanghai Municipal Science and Technology Major Project (2017SHZDZX01, VBH1323001/026), Shanghai Municipal Key Clinical Specialty Project (SHSLCZDZK02104) and Pilot Project of Fudan University (IDF159045).

Authors' contributions:

Haiquan Chen, Leming Shi, Zhiwei Cao and Yue Zhao designed the study. Yue Zhao, Jun Shang and Jian Gao performed data analysis. Yue Zhao, Han Han and Zhendong Gao performed experiments. Qiang Zheng and Yuan Li reviewed pathological slides. Yueren Yan, Ting Ye, Fangqiu Fu, Chaoqiang Deng, Zelin Ma, Yang Zhang, Difan Zheng and Shanbo Zheng collected samples. Jun Shang and Leming Shi helped with quality control of sequencing data. Yue Zhao, Jun Shang and Jian Gao drafted the manuscript. All authors proofread the manuscript.

Figure legends:

Figure 1. Study design. A) a total number of 197 tumor samples, including 24 adenocarcinoma *in situ* (AIS), 74 minimally invasive adenocarcinoma (MIA), 99 lung adenocarcinoma (LUAD) and their paired adjacent normal lung tissue underwent whole-exome sequencing and RNA-sequencing. Genomic and transcriptomic data were analyzed and compared among different groups. Samples were further divided into 3 groups according to their radiological manifestations: 69 pure ground-glass opacities (GGOs), 63 subsolid nodules and 65 solid nodules. B) Identification of 12 expressional patterns and development of tumor progressive index based on genomic and transcriptomic data.

Figure 2. Identification of 12 expression patterns. Pattern 1 (9 genes): increase from normal to AIS, from AIS to MIA and from MIA to LUAD; pattern 2 (97 genes): increase from normal to AIS and from MIA to LUAD, no change from AIS to MIA; pattern 3 (446 genes): increase from normal to AIS alone; pattern 4 (54 genes): no change from normal to AIS, increase from AIS to MIA and from MIA to LUAD; pattern 5 (20 genes): increase from AIS to MIA alone; pattern 6 (309 genes): increase from MIA to LUAD alone; pattern 7 (3 genes): decrease from normal to AIS, from AIS to MIA and from MIA to LUAD; pattern 8 (175 genes): decrease from normal to AIS and from MIA to LUAD, no change from AIS to MIA; pattern 9 (771 genes): decrease from normal to AIS alone; pattern 10 (6 genes): no change from normal to AIS, decrease from AIS to MIA and from MIA to LUAD; pattern 11 (3 genes): decrease from AIS to MIA alone; pattern 12 (130 genes): decrease from MIA to LUAD alone. Gene set enrichment analysis (GSEA) was performed to assess the functional significance for those patterns.

Figure 3. Immune cell infiltration prediction. A) Prediction of number of CD8+ T cells in the Fudan University Shanghai Cancer Center (FUSCC) cohort. B) Prediction of number of Tregs in the FUSCC cohort. C) Prediction of number of Tregs in Dejima's cohort. D) Prediction of number of activated and resting natural killer (NK) cells in the FUSCC cohort. E) Prediction of number of activated and resting natural killer (NK) cells in Dejima's cohort.

Figure 4. Comparison of genomic alterations for different radiological groups. A) Waterfall plot showing the landscape of genomic alterations in each radiological group. B) Comparison of mutation frequency in major driver genes among different radiological groups. C) Comparison of mutation frequency in tumor suppressor genes among different radiological groups. D) Comparison of tumor mutation burden among different radiological groups. E) Comparison of somatic copy number alterations among different radiological groups.

Figure 5. Differential expression analysis between the solid group and the pure GGO + subsolid group. A) Expression of differentially expressed genes between the solid group and the pure GGO + subsolid group. B) Volcano plot showing the

number of significantly differentially expressed genes. C) Gene-set enrichment analysis (GSEA) showing the pathways that were enriched by differentially expressed genes.

Figure 6. Comparison of tumor progressive index for different groups and the prognostic value of tumor progressive index. A) boxplot showing the index increased as tumor progressed in the FUSCC cohort. B) boxplot showing the index increased as tumor progressed in Dejima's cohort. C) Recurrence-free survival for patients with solid vs subsolid and pure GGO/subsolid nodules. D) Overall survival for patients with solid vs subsolid and pure GGO/subsolid nodules.

Supplementary figure 1. Prediction of number of A) CD8+ T cells, B) Tregs and C) natural killer (NK) cells for different tumor stages.

Supplementary figure 2. Principle component analysis (PCA) of expressional profiles different pathological groups.

Supplementary figure 3. Differential analysis between solid and subsolid nodules. A) Expression of differentially expressed genes between solid and subsolid nodules. B) Volcano plot showing the number of significantly differentially expressed genes.

Supplementary figure 4. Differential analysis between GGOs and subsolid nodules. A) Expression of differentially expressed genes between solid and subsolid nodules. B) Volcano plot showing the number of significantly differentially expressed genes.

Supplementary figure 5. Comparison of *EGFR* mutation type for tumors with different radiological manifestations.

A

197 tumor samples and their paired normal tissues with different pathological stages and different radiological manifestations:

Normal lung tissue
(n = 197)

AIS
(n = 24)

MIA
(n = 74)

LUAD
(n = 99)

Pure ground-glass opacity
(n = 69)

Part-solid nodules
(n = 63)

Solid nodules
(n = 65)

Whole-exome sequencing
and RNA sequencing

B

Identify expressional patterns

Pathway analysis

Tumor Progressive Index

A

B

C

	Pure GGO (n = 69)	Subsolid (n = 63)	Solid (n = 65)	P-value
Sex				0.025
Female	48 (69.6%)	41 (65.1%)	31 (47.7%)	
Male	21 (30.4%)	22 (34.9%)	34 (52.3%)	
Smoking status				0.212
Former/current	16 (23.2%)	15 (23.8%)	23 (35.4%)	
Never	53 (76.8%)	48 (76.2%)	42 (64.6%)	
Tumor location				0.661
LUL	18 (26.1%)	13 (20.6%)	11 (16.9%)	
LLL	7 (10.1%)	6 (9.5%)	10 (15.4%)	
RUL	26 (37.7%)	31 (49.2%)	23 (35.4%)	
RML	7 (10.1%)	5 (7.9%)	8 (12.3%)	
RLL	11 (15.9%)	8 (12.7%)	13 (20.0%)	
Pathology				< 0.001
AIS/MIA	66 (95.7%)	31 (49.2%)	1 (1.5%)	
IAD	3 (4.3%)	32 (50.8%)	64 (98.5%)	
Predominant subtype				
Lepidic	2 (66.7%)	9 (28.1%)	3 (4.7%)	< 0.001
Acinar	1 (33.3%)	17 (53.1%)	40 (62.5%)	0.453
Papillary	0 (0.0%)	6 (9.4%)	10 (15.6%)	0.004
Micropapillary	0 (0.0%)	0 (0.0%)	0 (0.0%)	-
Solid	0 (0.0%)	0 (0.0%)	9 (14.1%)	0.067
IMA	0 (0.0%)	0 (0.0%)	2 (3.1%)	0.123
Presenting subtype				
Lepidic	2 (66.7%)	12 (37.5%)	6 (9.4%)	0.001
Acinar	2 (66.7%)	24 (75.0%)	50 (78.1%)	0.863
Papillary	0 (0.0%)	9 (28.1%)	22 (34.4%)	0.407
Micropapillary	0 (0.0%)	1 (3.1%)	6 (9.4%)	0.472
Solid	0 (0.0%)	1 (3.1%)	17 (26.6%)	0.014
IMA	0 (0.0%)	0 (0.0%)	2 (3.1%)	0.572

Table 1. Clinical and pathological characteristics of the study cohort (n = 197).

Abbreviations: AIS, adenocarcinoma *in situ*; GGO, ground-glass opacity; IAD, invasive adenocarcinoma; IMA, invasive mucinous adenocarcinoma; MIA, minimally invasive adenocarcinoma; LLL, left lower lobe; LUL, left upper lobe; RLL, right lower lobe; RML, right middle lobe; RUL, right upper lobe.