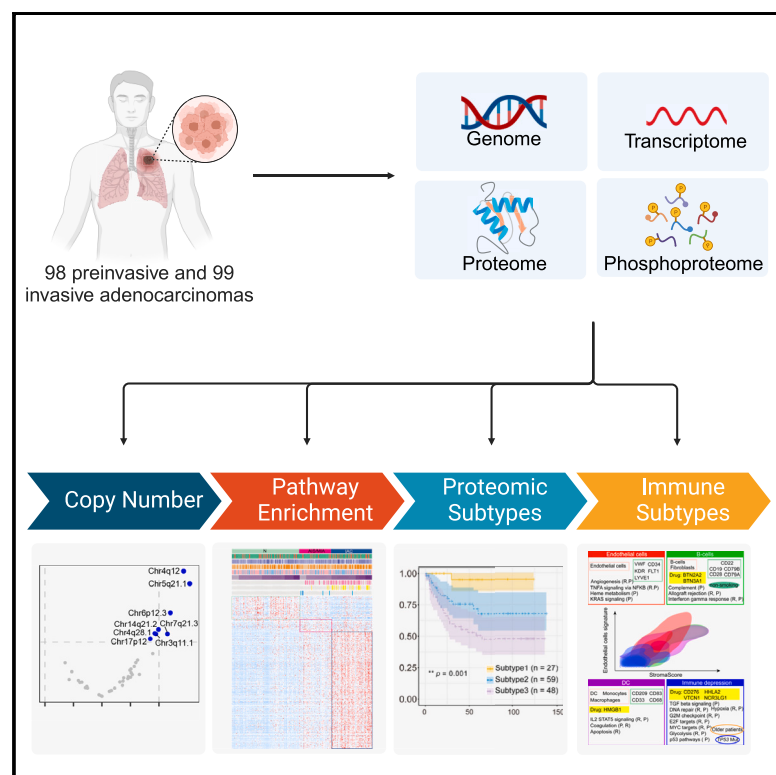


Evolutionary proteogenomic landscape from pre-invasive to invasive lung adenocarcinoma

Graphical abstract



Authors

Yang Zhang, Fangqiu Fu, Qiao Zhang, ..., Yuan Li, Chen Ding, Haiquan Chen

Correspondence

fdzhangyang1987@hotmail.com (Y.Z.),
 fqfu12@fudan.edu.cn (F.F.),
 lumoxuan2009@163.com (Y.L.),
 chend@fudan.edu.cn (C.D.),
 hqchen1@yahoo.com (H.C.)

In brief

Zhang et al. conduct multi-omic profiling on both pre-invasive and invasive stages of lung adenocarcinoma. Their investigations elucidate the molecular architecture linked to the progression of lung adenocarcinoma and reveal the proteomic subtypes. This work establishes a valuable community resource for subsequent research on tumorigenesis and progression.

Highlights

- 4q12 deletion contributes to lung cancer progression via downregulation of SPATA18
- Proteomics reveals diverse enriched pathways in different stages of lung adenocarcinoma
- Three proteomic subtypes linked to lung adenocarcinoma progression are identified
- Four immune clusters are illustrated, characterized by discernible immune signatures



Article

Evolutionary proteogenomic landscape from pre-invasive to invasive lung adenocarcinoma

Yang Zhang,^{1,2,3,7,*} Fangqiu Fu,^{1,2,3,7,*} Qiao Zhang,^{4,7} Lingling Li,^{4,7} Hui Liu,^{4,5,7} Chaoqiang Deng,^{1,2,3,7} Qianqian Xue,^{3,6} Yue Zhao,^{1,2,3} Wenrui Sun,^{1,2,3} Han Han,^{1,2,3} Zhendong Gao,^{1,2,3} Chunmei Guo,⁴ Qiang Zheng,^{3,6} Hong Hu,^{1,2,3} Yihua Sun,^{1,2,3} Yuan Li,^{3,6,*} Chen Ding,^{4,*} and Haiquan Chen^{1,2,3,8,*}

¹Department of Thoracic Surgery and State Key Laboratory of Genetic Engineering, Fudan University Shanghai Cancer Center, Shanghai 200032, China

²Institute of Thoracic Oncology, Fudan University, Shanghai 200032, China

³Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China

⁴State Key Laboratory of Genetic Engineering and Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Institutes of Biomedical Sciences, Human Phenome Institute, Fudan University, Shanghai 200433, China

⁵State Key Laboratory Cell Differentiation and Regulation, Overseas Expertise Introduction Center for Discipline Innovation of Pulmonary Fibrosis (111 Project), College of Life Science, Henan Normal University, Xinxiang, Henan 453007, China

⁶Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai 200032, China

⁷These authors contributed equally

⁸Lead contact

*Correspondence: fdzhangyang1987@hotmail.com (Y.Z.), fqfu12@fudan.edu.cn (F.F.), lumoxuan2009@163.com (Y.L.), chend@fudan.edu.cn (C.D.), hqchen1@yahoo.com (H.C.)

<https://doi.org/10.1016/j.xcrm.2023.101358>

SUMMARY

Lung adenocarcinoma follows a stepwise progression from pre-invasive to invasive. However, there remains a knowledge gap regarding molecular events from pre-invasive to invasive. Here, we conduct a comprehensive proteogenomic analysis comprising whole-exon sequencing, RNA sequencing, and proteomic and phosphoproteomic profiling on 98 pre-invasive and 99 invasive lung adenocarcinomas. The deletion of chr4q12 contributes to the progression from pre-invasive to invasive adenocarcinoma by downregulating SPATA18, thus suppressing mitophagy and promoting cell invasion. Proteomics reveals diverse enriched pathways in normal lung tissues and pre-invasive and invasive adenocarcinoma. Proteomic analyses identify three proteomic subtypes, which represent different stages of tumor progression. We also illustrate the molecular characterization of four immune clusters, including endothelial cells, B cells, DCs, and immune depression subtype. In conclusion, this comprehensive proteogenomic study characterizes the molecular architecture and hallmarks from pre-invasive to invasive lung adenocarcinoma, guiding the way to a deeper understanding of the tumorigenesis and progression of this disease.

INTRODUCTION

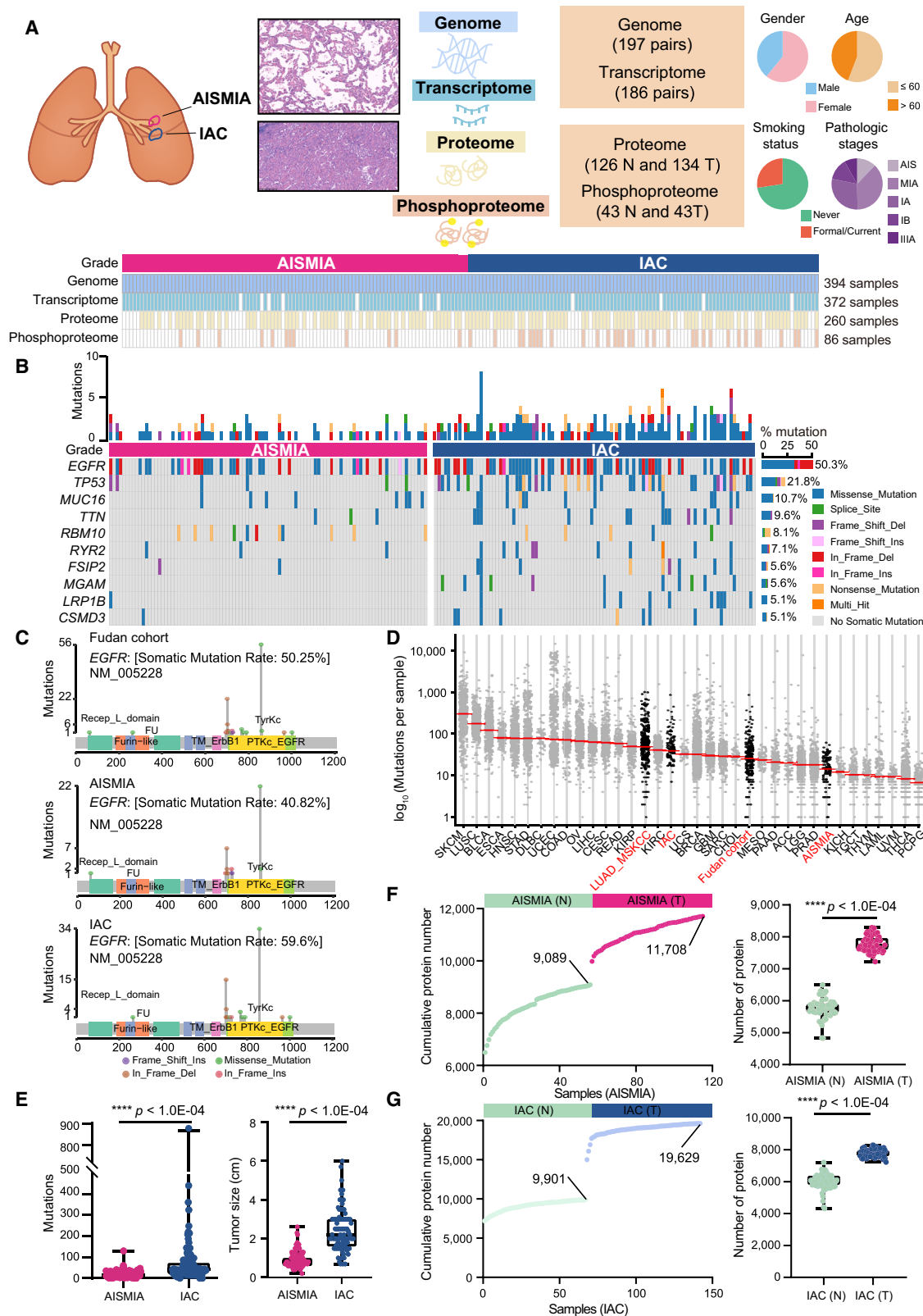
Lung cancer remains the leading cause of cancer deaths worldwide.^{1,2} Along with the wide application of low-dose computed tomography (LDCT) screening, there is an increasing detection rate of early-stage lung adenocarcinoma,^{3–5} which is predominantly attributed to rising cases of adenocarcinoma *in situ* (AIS) and minimally invasive adenocarcinoma (MIA). AIS is defined as a localized small (≤ 3 cm) adenocarcinoma with growth restricted to neoplastic cells along pre-existing alveolar structures without invasion, while MIA is a small adenocarcinoma (≤ 3 cm) with a predominantly lepidic pattern and ≤ 5 mm invasion in the greatest dimension in any one focus.^{6,7} Pathologically, lung adenocarcinoma is generally considered to follow a stepwise progression from AIS/MIA to invasive lesions.⁸ Both AIS and MIA are classified as the pre-invasive stage of lung adenocarcinoma, whereas invasive adenocarcinoma (IAC) represents the invasive stage. Correspondingly, the 5-year recurrence-

free survival (RFS) rate for patients with AIS/MIA after surgical resection could be as high as 100%,⁹ while it decreases dramatically for those with IAC.¹⁰ Therefore, it is critical to identify the molecular architecture and hallmarks contributing to the progression from AIS/MIA to IAC.

Several studies have elucidated molecular changes during lung cancer progression at the dimension of the genome.^{11–15} However, due to the small diameter and limited tissues of pre-invasive lesions, the multi-omics analysis is yet lacking. Our previous study demonstrated that *TP53* mutation, arm-level copy number alterations (CNAs), and *HLA* loss of heterozygosity were increased in frequency during the development of lung adenocarcinoma.¹⁶ Nevertheless, genomic changes need to be translated into alterations at the RNA and protein levels to affect phenotypes, and there remains a gap in the deep investigation of lung adenocarcinoma progression at the multi-omics level.

To decipher the proteogenomic trajectories from normal lung tissues to AIS, MIA, and ultimately IAC, we performed an integrated





(legend on next page)

proteogenomic characterization of pre-invasive and invasive lung adenocarcinoma as well as paired adjacent normal tissues. This proteogenomic study elucidates key molecular events driving the tumorigenesis and progression of lung adenocarcinoma, providing a resource for the scientific community to further delineate the underlying process.

RESULTS

Proteogenomic landscape of AIS/MIA and IAC

We prospectively collected 197 lung tumors and paired normal adjacent tissues (NATs) from east-Asian patients, including 98 AIS/MIAs and 99 IACs (Figures 1A and S1A). No discernible differences were observed between AIS and MIA subtypes (Figures S1B and S1C), and thus AIS/MIA subtypes were classified as one group in this study. The prospective cohort consisted of 72.6% non-smoking and 27.4% smoking patients (Figure S1D), and diverse clinical characteristics are listed in Table S1A. Pathological diagnosis of pre-invasive or invasive lung adenocarcinoma was evaluated by at least two experienced pulmonary pathologists. There was no significant difference in tumor purity evaluated by the ESTIMATE algorithm¹⁷ (Wilcoxon signed-rank test, $p = 0.15$; Figure S1E) or pathologists (Wilcoxon signed-rank test, $p = 0.26$; Figure S1F) between AIS/MIA and IAC. Integrative profiling consisting of whole-exome sequencing (WES), RNA sequencing (RNA-seq), and proteomic and phosphoproteomic analyses was conducted (Figure 1A).

The landscapes of somatic mutations, deletions, and insertions are presented in Figure 1B (Table S2A). Comparative analysis showed that lower tumor mutation burden (TMB) was detected in the AIS/MIA group (Figures 1C, 1D, and 1E) and among the Fudan cohort and other studies.^{12,18} In addition, the top-ranked mutation of *EGFR* was detected as early as in the AIS/MIA subtype and lasted into the IAC subtype (Figures 1C and S1G), indicating that *EGFR* mutation was a key event in the early stage during lung cancer progression. Mutually exclusive analysis is generally used to identify driver genes and pathways and investigate their functional relationships.¹⁹ The top-ranked mutation of *EGFR* was mutually exclusive of other driver gene mutations (e.g., *BRAF*, *ERBB2*) (Figure S1H), in accordance with results from the Cancer Genome Atlas,²⁰ TRACERx,²¹ and the MSKCC cohort.²² *TP53* was the most frequently mutated tumor suppressor in our cohort, accounting for 6.12% of AIS/MIA patients and 37.37% of IAC patients (Figure S1I). Taken together, *EGFR* mutation might already exist in AIS/MIA, whereas *TP53* mutation was prominent in IAC.

To investigate the effect of mutation signatures in AIS/MIA and IAC, mutation spectra were decomposed and compared according to known signatures derived from the Catalogue of So-

matic Mutations in Cancer (COSMIC).²³ Interestingly, cosine similarity of single base substitution (SBS) 1 and SBS2 was higher in IAC compared with AIS/MIA (Figure S1J). Notably, SBS2, attributed to APOBEC activity,²⁴ was associated with tumor progression from AIS/MIA to IAC. This is in line with the previous study revealing the association of APOBEC activity and invasiveness in distinct subtypes of invasive lung adenocarcinoma.²⁵ These findings indicated the exact difference between AIS/MIA and IAC at the DNA level.

Smoking is widely recognized as a significant promoter of lung cancer. In our study, the smoking group exhibited a higher TMB (Figure S1K). Furthermore, we observed a lower prevalence of *EGFR* mutation and a higher prevalence of *TP53* mutation in the former/current smoking group (Wilcoxon signed-rank test, $p = 0.0087$) (Figure S1L).

Proteomic analysis was performed using a label-free quantification strategy.^{26,27} The average (Spearman's) coefficients were 0.77 in the AIS/MIA cohort and 0.76 in the IAC cohort (Figure S1M), allowing the comparison between samples during lung cancer progression. High consistency of mass spectrometry (MS) data for the standards (HEK293T cell lysates) was observed among different batches of HEK293T lysates (Figure S1N), indicating the consistent stability of our MS platform. A total of 11,708 proteins were identified in the AIS/MIA cohort, and 19,629 proteins were detected in the IAC cohort with 1% false discovery rate (FDR) (Figures 1F, 1G, and S1O; Tables S3A and S3B). At the phosphoprotein level, 11,852 phosphoprotein sites were discovered in our cohort (Figure S1P; Table S3C). Notably, more proteins and phosphoprotein sites were detected in the tumor tissues compared with paired normal tissues (Figures 1F, 1G, and S1P).

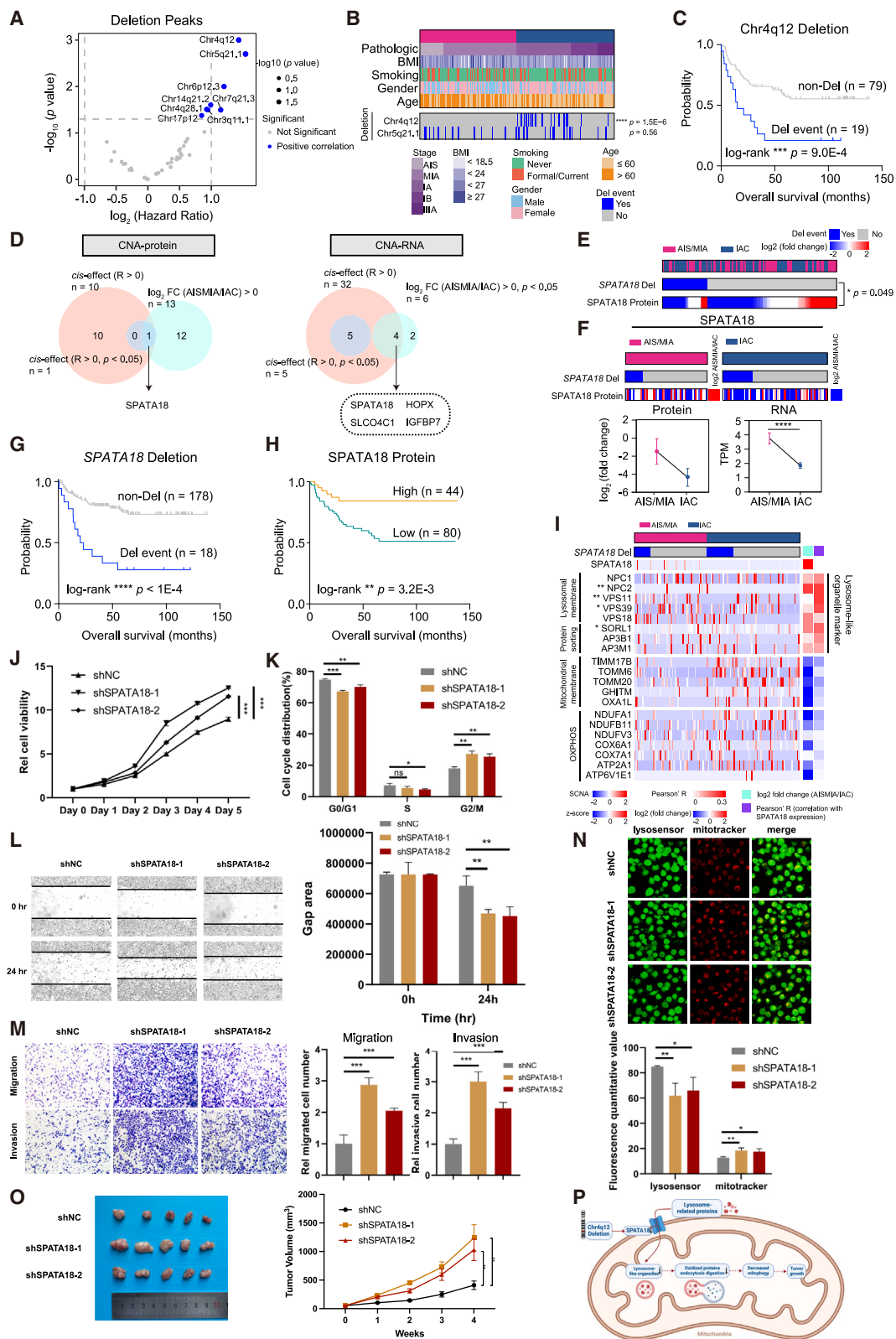
The deletion of chr4q12 was the key event in the transition process from AIS/MIA to IAC

As CNAs play critical parts in activating oncogenes and deactivating tumor suppressor genes during the progression of tumors, deletion peaks connecting to survival outcomes (log rank test, $p < 0.05$) were further investigated. As a result, deletions in chr4q12 and chr5q21.1 were two main alterations associated with clinical outcomes (Figures 2A, 2B, 2C, and S2A). The chr4q12 deletion was detected only in IAC, indicating that the crucial molecular events happened during chr4q12 deletion (Fisher's exact test $p = 1.5E-6$; Figure 2B; Tables S2B and S2C).

To investigate the biological effects of chr4q12 and chr5q21.1 deletions, we undertook investigations into genes displaying *cis*-regulatory effects of somatic CNAs (SCNAs) and downregulation from AIS/MIA to IAC at both the protein and the RNA levels. To this end, only *Spermatogenesis Associated 18* (*SPATA18*, also known as *Mieap*) was identified at both levels (Figure 2D), and

Figure 1. Proteogenomic overview of AIS/MIA and IAC cohorts

- Schematic summary and omics data of AIS/MIA and IAC cohorts.
- Overview of top 10 altered genes stratified by invasive extent of tumor.
- Lollipop plots showing details of *EGFR* mutations in the Fudan cohort (all tumors), as well as the AIS/MIA and IAC cohorts.
- Tumor mutation burden in the Fudan cohort and other public cohorts.
- Numbers of gene mutations (left) and tumor size (right) in AIS/MIA and IAC.
- Cumulative protein number (left) and absolute protein number (right) identified by proteomic analysis in normal tissues and tumors of the AIS/MIA cohort.
- Cumulative protein number (left) and absolute protein number (right) identified by proteomic analysis in normal tissues and tumors of the IAC cohort. The asterisk was defined as follows: **** $p < 1.0E-4$, *** $p < 1.0E-3$, ** $p < 0.01$, * $p < 0.05$.



(legend on next page)

showed *cis* effects on its counterpart protein (Figure 2E). Compared with AIS/MIA, the expression of SPATA18 was lower in IAC at the RNA and protein levels (Figure 2F). In addition, SPATA18 deletion event and low expression of SPATA18 protein were correlated with worse overall survival (OS) (log rank test, $p < 1.0E-4$ for deletion event, $p = 3.2E-3$ for protein level) (Figures 2G and 2H). As well, we performed immunohistochemistry (IHC) analysis of SPATA18 expression in 15 tumors with AIS/MIA and 16 with IAC (Figure S2B), and significantly lower protein level of SPATA18 in IAC was observed (t test, $p = 3.0E-4$; Figure S2C). To further validate the prognostic effect of SPATA18 protein, we collected a new cohort (validation cohort) consisting of 270 patients with IAC, and performed IHC analysis. As a result, low expression of SPATA18 significantly indicated worse OS (log rank test, $p = 2.5E-3$) and RFS (log rank test, $p < 1.0E-4$) in patients with lung adenocarcinoma (Figures S2D and S2E).

SPATA18 is a novel p53-induced protein that has been reported to affect the biological functions of mitochondria.²⁸ SPATA18 is able to induce lysosome-like organelles to eliminate oxidized mitochondrial proteins, which are essential for mitochondrial quality control.^{29,30} Our analysis showed that SPATA18 deletion was associated with downregulation of lysosome-like organelle markers as well as upregulation of mitochondrial membrane molecules and oxidative phosphorylation (OXPHOS) (Figure 2I). Lysosome-like organelle markers (e.g., VPS11 and SORL1) positively correlated with SPATA18 at the protein level (Figures S2F and S2G). Furthermore, negative correlation was noted between SPATA18 and TOP2A as well as DNA replication (Figures S2H and S2I), suggesting the pro-tumor role of SPATA18 downregulation. Two typical markers for mitochondrial membrane, TIMM17B and GHITM, were confirmed to be higher in IAC and affect survival negatively (Figure S2J). In addition, the expression of proliferative markers, such as MCM4 and TOP2A, was higher in IAC and showed negative correlation with the OS of IAC patients (Figure S2K).

SPATA18 has been proven to be an important factor for mitophagy.³¹ We also performed *in vitro* and *in vivo* experiments

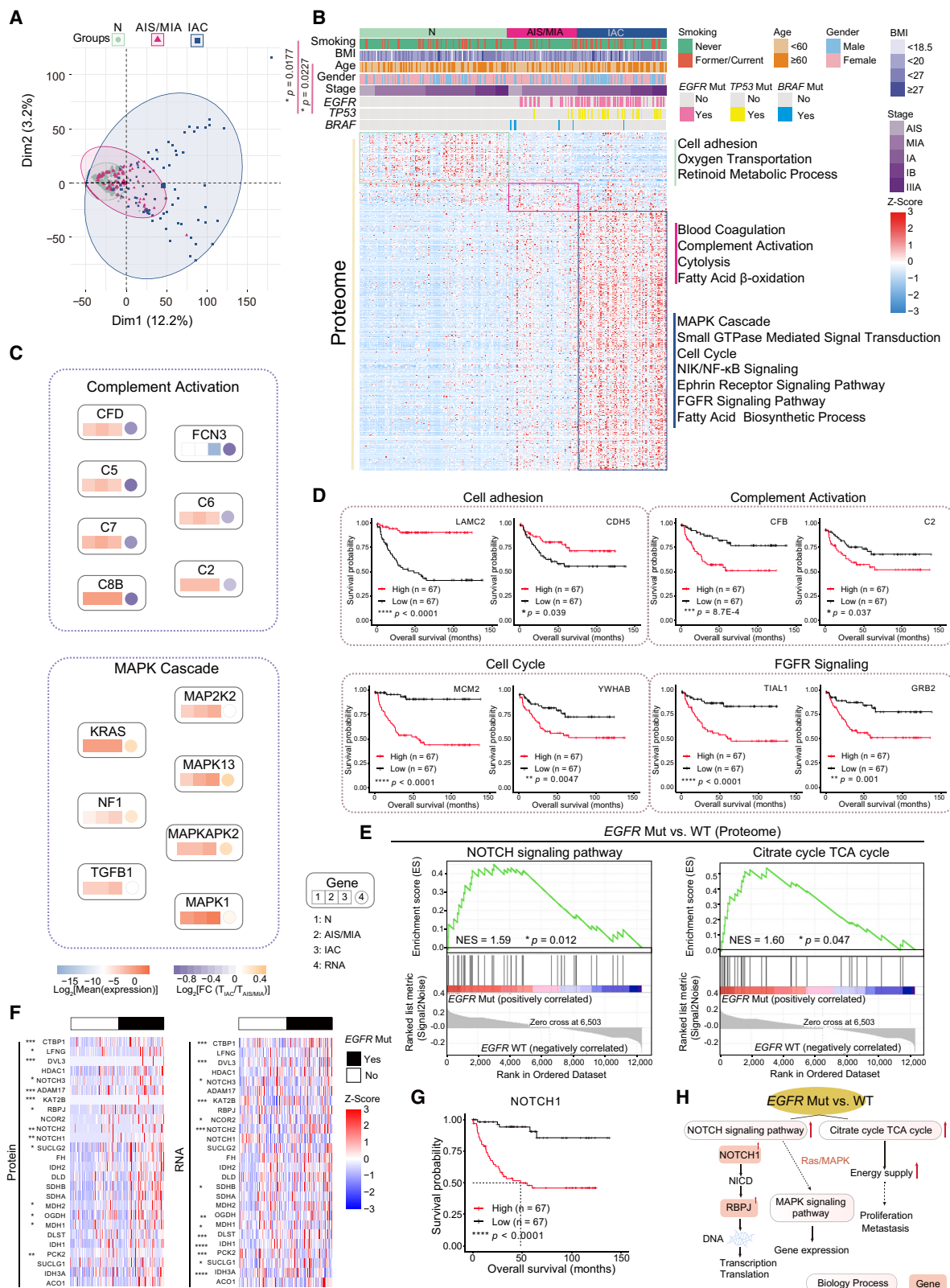
to confirm the biological effect of SPATA18. Downregulation of SPATA18 could significantly promote cell growth and cell cycle (Figures 2J and 2K). Migration and invasion are recognized as critical steps in which tumor cells detach from the primary tumor site and infiltrate lymphatic and blood vessels, ultimately contributing to tumor progression.³² Given the potential role of SPATA18 in the progression from AIS/MIA to IAC, we performed wound healing, migration, and invasion assays. The results demonstrated that the downregulation of SPATA18 enhanced the ability of migration and invasion in lung adenocarcinoma (Figures 2L and 2M). Next, we investigated the effect of SPATA18 on mitophagy. Following transfection with two short hairpin RNAs (shRNAs) targeting shSPATA18, a notable decrease in fluorescence density was observed for LysoSensor, and an increase was detected for MitoTracker (Figure 2N). These findings collectively suggested that the inhibition of mitophagy occurred as a consequence of SPATA18 downregulation. Animal experiments also confirmed the pro-tumor effect of SPATA18 downregulation (Figure 2O). Taken together, the chr4q12 deletion might be the key event in the transition process from AIS/MIA to IAC, in which the SPATA18 deletion advanced tumor growth by suppressing mitophagy (Figure 2P). This study provided evidence for using SPATA18 as a therapy target for lung adenocarcinoma.

Proteomic characterization of normal lung tissues, AIS/MIA, and IAC

Principal-component analysis (PCA) showed the differentiated profiles of normal lung tissues, AIS/MIA, and IAC, at the RNA and protein levels (Figures 3A and S3A–S3C; Tables S3A and S4), which indicated intrinsic differences in distinct stages of lung cancer progression. To understand the association between transcriptomic and proteomic processes, the correlation of mRNA and protein was calculated for NATs and tumor tissues. NATs displayed a fair correlation of 0.20, while tumors had a higher value of 0.39 (Figure S3D). Among tumor tissues, we observed correlation coefficients of 0.36 for AIS/MIA and 0.40

Figure 2. Copy number alterations (CNAs) associated with the progression from AIS/MIA to IAC

- Deletion peaks related to survival in patients with lung adenocarcinoma.
- Distribution of two most significant deletion peaks (chr5q21.1 and chr4q12 deletion) in AIS/MIA and IAC (Fisher's exact test).
- Kaplan-Meier plot comparing overall survival (OS) stratified by chr4q12 deletion in IAC patients receiving whole-exon sequencing. Log rank test was used to calculate the p value.
- cis* CNA-protein and CNA-RNA correlations of genes located in chr4q12 and chr5q21.1.
- Association between SPATA18 deletion and protein-level expression (Pearson's correlation test).
- Comparison of SPATA18 protein and RNA expression in AIS/MIA and IAC.
- Kaplan-Meier plot comparing OS stratified by SPATA18 deletion or not in patients receiving whole-exome sequencing.
- Kaplan-Meier plot comparing OS stratified by the protein level of SPATA18 in patients receiving proteomic analysis.
- Heatmap visualizing the correlation between SPATA18 deletion and genes correlated to lysosome-like organelle and mitochondria markers.
- Cell proliferation analysis was conducted with CCK-8 assays in A549 transfected with shSPATA18 or shNC. The experiment was repeated three times.
- Cell-cycle analysis was performed in A549 transfected with shSPATA18 or shNC. The experiment was repeated three times.
- The effect of downregulated SPATA18 on wound healing. The left shows typical images of wound healing in three groups, and the right shows the quantified results of gap area. The experiment was repeated three times.
- The effect of downregulated SPATA18 on migration and invasion. The left shows typical images of migration and invasion in three groups, and the right shows quantified results of cell number. The experiment was repeated three times.
- Immunofluorescence images of LysoSensor and MitoTracker in A549 transfected with overexpressed SPATA18 or shSPATA18. The density of LysoSensor indicates the pH of the lysosome, and density of MitoTracker represents mitochondrial membrane potential. The experiment was repeated three times.
- The effect of SPATA18 on A549 cell xenografts in nude mice.
- Diagram showing associations among chr4q12 deletion, SPATA18 downregulation, decreased mitophagy, and tumor growth. The asterisk was defined as follows: **** $p < 1.0E-4$, *** $p < 1.0E-3$, ** $p < 0.01$, * $p < 0.05$.



(legend on next page)

for IAC, indicating similar correlations of mRNA and protein between these two groups (Figure S3E).

To explore the molecular characterization of the tumorigenesis and progression of lung adenocarcinoma, gene set enrichment analysis (GSEA) was performed (Figures 3B, 3C, and S3F). As a result, we found that normal lung tissues were predominantly associated with cell adhesion, oxygen transportation, and retinoid metabolic process (Figure 3B), which could be partially explained by physiological functions of lungs, such as pulmonary gas exchange.³³ Specifically, the highly expressed proteins in the normal lung tissues, such as LAMC2 and CDH5 (from cell adhesion), showed positive association with the prognosis of lung cancer patients (Figure 3D). AIS/MIA had upregulated signatures of blood coagulation, complement activation, cytolysis, and fatty acid β -oxidation (Figures 3B and 3C). In IAC, the predominant pathways consisted of MAPK cascade, small-GTPase-mediated signal transduction, cell cycle, NIK/NF- κ B signaling, ephrin receptor signaling pathway, FGFR signaling pathway, and fatty acid biosynthetic process (Figures 3B and 3C). We also validated our findings in a dataset³⁴ that investigated the transcriptomic differences in distinct growth patterns within single lung tumors. Similar enriched pathways were also detected (Figures S3G and S3H). Meanwhile, both AIS/MIA- and IAC-enriched proteins were correlated with worse prognosis in patients with lung adenocarcinoma, including CFB, C2 (from complement activation), MCM2, YWHAB (from cell cycle), TIAL1, and GRB2 (from FGFR signaling) (Figure 3D). To provide potential targets for future studies, we also presented upregulated markers in AIS/MIA, which were associated with survival outcomes (Table S1D). Taken together, these findings revealed the characterization during carcinogenesis and development of lung adenocarcinoma at the proteomic level.

Proteomic characterization of lung adenocarcinoma with EGFR mutation

Target therapy toward EGFR mutation has revolutionized treatments for lung cancer.^{35,36} In the Fudan cohort, EGFR mutation stood out as the top-ranked frequent mutation, co-occurrent with TP53 and RBM10 mutations (Figure S1H). To access the impact of EGFR mutation in tumorigenesis and progression, we then investigated significantly dysregulated hallmarks associated with EGFR mutation. GSEA showed that the citrate cycle TCA cycle ($p = 0.047$) and NOTCH signaling pathway ($p = 0.012$) were significantly enriched in patients with EGFR mutation at the

protein level (Figures 3E and 3F). Similar trends were also observed at the RNA level (Figures 3F, S3I, and S3J). Higher enrichment scores of NOTCH signaling were also detected in the EGFR mutation group from three other datasets (Figure S3K).^{34,37,38} A key regulatory protein from NOTCH signaling, NOTCH1, had negative effects on the survival of patients with lung adenocarcinoma (log rank $p < 1.0E-4$) (Figure 3G). The prognostic effect existed in both the EGFR mutation group (log rank $p = 5.6E-4$) and the EGFR wild-type (WT) group (log rank $p < 1.0E-4$) (Figure S3L). Moreover, the association between EGFR mutation and downregulation of complement and coagulation cascades reported by the Clinical Proteomic Tumor Analysis Consortium (CPTAC)³⁸ was also confirmed in the Fudan cohort (Figure S3M). Specifically, high expression of C1QB and SERPINA1 was associated with poor prognosis in all the patients with lung adenocarcinoma and in the subgroup of patients with EGFR mutation (Figure S3N). Collectively, our results delineated the landscape of downstream proteomic characterization of EGFR mutation (Figure 3H).

Proteomic evolution from AIS/MIA to IAC under the background of EGFR mutation

EGFR mutation is an essential molecular event in the change from normal lung tissue to early-stage lung cancer according to our study, but the development trajectory of EGFR-mutant lung adenocarcinoma remains largely unknown. Therefore, the intrinsic proteomic difference between EGFR-mutant AIS/MIA and EGFR-mutant IAC was further investigated (Figure S4A). A higher proportion of patients over 60 years of age was observed in the EGFR-mutant IAC (chi-squared test, $p = 0.0132$) (Figure S4B). Extracellular matrix (ECM)-receptor interaction, complement and coagulation cascades, and endocytosis were enriched in EGFR-mutant AIS/MIA (Figures 4A, 4B, and S4C). When the disease stage switched to IAC, typical downstream pathways of EGFR signaling were activated, such as PI3K-Akt, MAPK, and mTOR signaling pathways (Figures 4A, 4B, and S4C). Similar trends of enriched pathways were also observed in other lung cancer cohorts, including the Tavernari cohort (Figure S4D).³⁴ Subsequently, we compared kinase abundance between the two groups. Despite harboring EGFR mutations, AIS/MIA had higher expression of MAPK7, while IAC was associated with overexpressed CSNK2A1 and MTOR (Figure 4C). The downstream phosphorylation analyses of enriched kinases revealed that cell cycle was nominated in AIS/MIA with EGFR

Figure 3. Proteogenomic characterization of normal tissues, AIS/MIA, and IAC

(A) Principal-component analysis (PCA) of proteomic data in normal tissues ($n = 126$), AIS/MIA ($n = 59$), and IAC ($n = 75$).

(B) Heatmap of enriched pathways using proteomic data in normal tissues, AIS/MIA, and IAC. (Top) Green box, pathways related to normal tissues; red box, pathways related to AIS/MIA; blue box, pathways related to IAC.

(C) Protein levels and fold changes in RNA in genes belonging to complement activation and MAPK cascade. Squares, $\log_2$ (standard protein expression in normal tissues, AIS/MIA, and IAC); circles, $\log_2$ (fold changes in RNA levels in IAC compared with those in AIS/MIA).

(D) Kaplan-Meier plots comparing overall survival (OS) stratified by protein levels of genes belonging to cell adhesion, complement activation, cell cycle, and FGFR signaling in patients receiving proteomic analysis. Log rank test was used to calculate the p value.

(E) Gene set enrichment analysis (GSEA) based on proteomic data stratified by EGFR mutation status.

(F) RNA and protein expression levels of genes related to NOTCH signaling pathway and citrate cycle TCA cycle stratified by EGFR mutation status.

(G) Kaplan-Meier plot comparing OS stratified by the protein level of NOTCH1, a key molecule of the NOTCH signaling pathway, in patients receiving proteomic analysis.

(H) A network model showing interactions between EGFR mutation and altered pathways. The asterisk was defined as follows: **** $p < 1.0E-4$, *** $p < 1.0E-3$, ** $p < 0.01$, * $p < 0.05$.

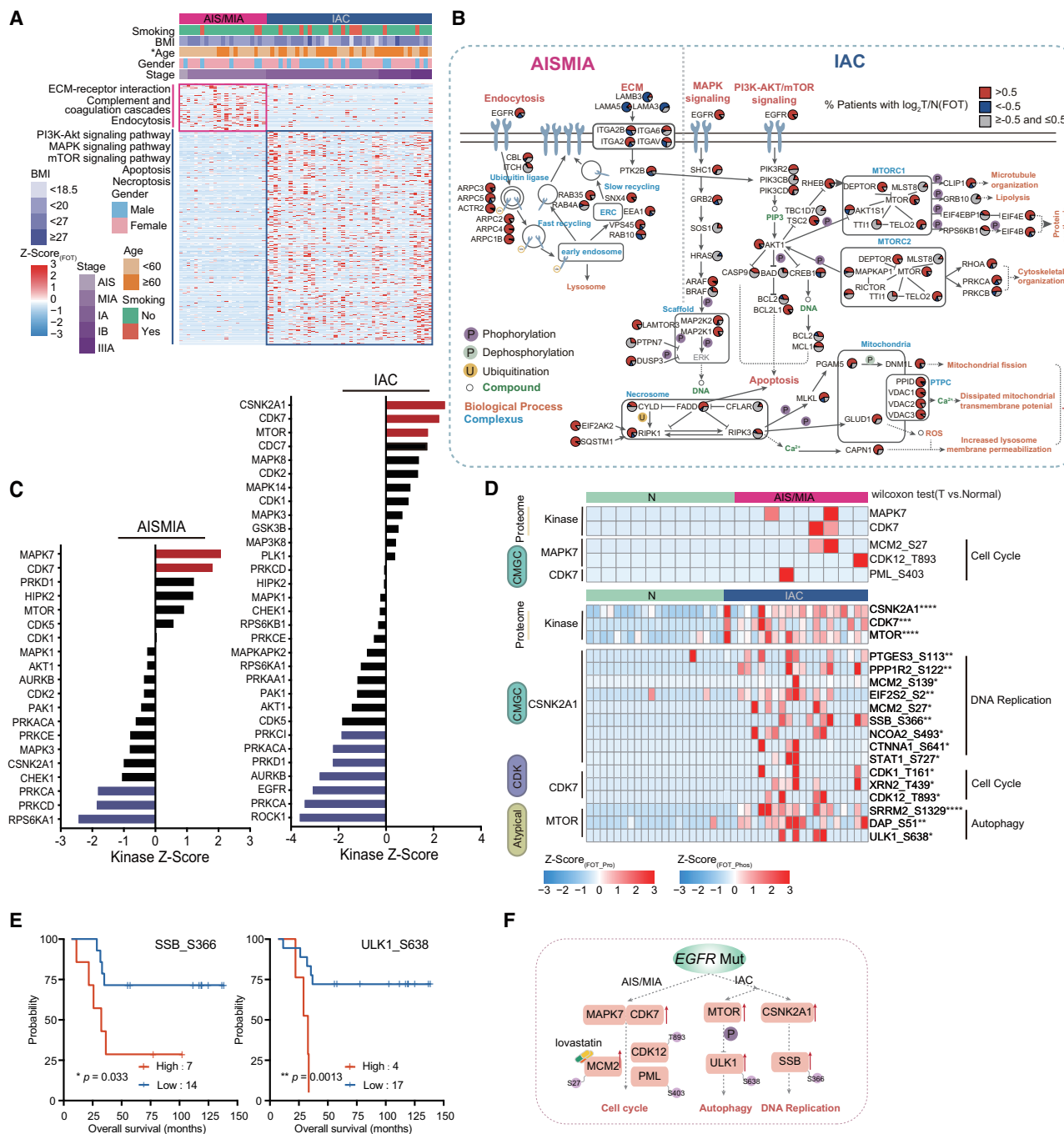


Figure 4. Proteomic characterization from *EGFR*-mutant AIS/MIA to *EGFR*-mutant IAC

(A) Heatmap of enriched pathways using proteomic data in AIS/MIA and IAC under the background of *EGFR* mutation. (Top) Red box, pathways related to AIS/MIA; blue box, pathways related to IAC. Chi-squared test was used to calculate p value for the association between baselines and pathological results. The Benjamini-Hochberg method was used to adjust p values of differentially expressed proteins between *EGFR*^{mut} AIS/MIA and IAC samples.

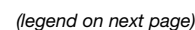
(B) Overview of dysregulated pathways related to *EGFR* mutation in AIS/MIA and IAC.

(C) Protein kinases with aberrant expression in *EGFR*-mutant AIS/MIA and IAC.

(D) Protein level of upregulated kinases and their downstream phosphorylation sites.

(E) Kaplan-Meier plot comparing overall survival stratified by the phosphorylation levels of SSB_S366 and ULK1_S638.

(F) A model depicting the distinct roles of *EGFR* mutation in AIS/MIA and IAC. The asterisk was defined as follows: **** $p < 1.0 \times 10^{-4}$, *** $p < 1.0 \times 10^{-3}$, ** $p < 0.01$, * $p < 0.05$.



mutation. As the disease stage progressed to the IAC stage, DNA replication and autophagy were shown to be activated (Figures 4D and S4E), and the related phosphorylated proteins (e.g., SSB_S366 and ULK1_S638) showed negative association with the prognosis of lung cancer patients (log rank $p = 0.033$ for SSB_S366, log rank $p = 0.0013$ for ULK1_S638) (Figure 4E). As a highly expressed kinase in AIS/MIA and IAC, CDK7 was negatively correlated with survival (log rank $p = 0.048$) (Figure S4F). Collectively, *EGFR* mutation played distinct roles in AIS/MIA and IAC, providing possible explanations for the different biological behaviors of *EGFR*-mutant AIS/MIA and IAC (Figure 4F).

Proteomic subtypes of lung adenocarcinoma

Consensus clustering was performed to specify proteomic subtypes, and three subtypes were identified (S-I, $n = 27$; S-II, $n = 59$; and S-III, $n = 48$) (Figures 5A and S5A). Patients classified as S-III exhibited worse RFS and OS, whereas those classified as S-I demonstrated improved RFS and OS (Figures 5B and S5B). S-I mainly consisted of patients with AIS/MIA (81.5%), while S-III was predominantly composed of patients with IAC (87.5%) (Figures S5C and S5D). S-II was the intermediate group between S-I and S-III. The mutation signatures of the three subtypes were presented (Figure S5E). A previous study also proposed several clusters for lung cancer,^{37–40} and we further compared our proteomic subtypes with those from published studies. There were some overlaps between ours and theirs (Figures S5F and S5G). Given the much higher proportion of AIS/MIA in our study compared with previous studies, we identified S-I as the subtype representing the pre-invasive stage (Figures S5F and S5G). Patients with S-I in our study could not be identified by other clustering systems, highlighting the originality of our proteomic subtypes.

Subtype-specific pathway enrichment was further conducted to disclose different molecular features among the three subgroups. Extracellular matrix organization, cell adhesion, proteolysis, leukocyte migration, and neutrophil degranulation were enriched in S-I (Figures 5C and 5D). S-II was predominantly associated with extracellular matrix organization, cell adhesion, leukocyte migration, neutrophil degranulation, Rho protein signal transduction, LDL particle-mediated signaling, and negative regulation of complement activation (Figures 5C and 5D). S-III was characterized by high levels of Wnt signaling pathway, proteasome-mediated ubiquitin-dependent protein catabolic process, DNA repair, MAPK cascade, autophagy, integrin-mediated signaling pathway, cell cycle, fatty acid metabolic process, TOR signaling, Toll-like receptor signaling pathway, and ERBB2

signaling pathway (Figures 5C and 5D). We also investigated the mutation status in the three subtypes and found that *TP53* was more frequently mutated in S-III than in S-I and S-II (chi-squared test, $p = 0.041$) (Figure S5H). Taken together, the results show that the three proteomic subtypes had distinct molecular backgrounds and represented different stages of tumor progression.

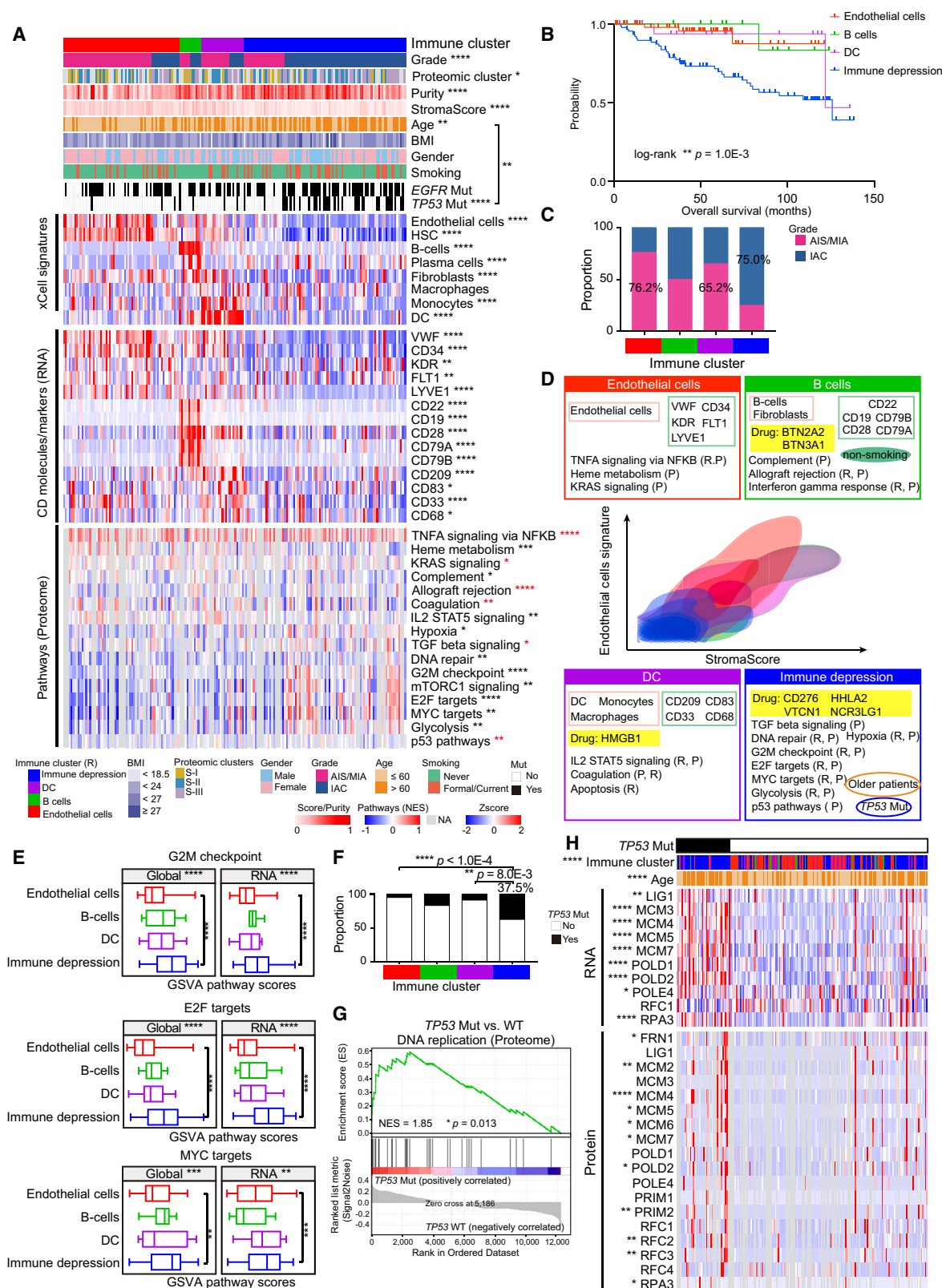
Subsequently, difference in copy number variation was further investigated among the three subtypes. As a result, we found the chr1q21.1, chr5q35.1, chr7p11.2, and chr14q13.3 amplifications differentially occurred in the three subtypes (Figures 5A and S5I), and the appearance of these amplifications indicated worse survival (Figures 5E and S5J). The chr7p11.2 amplification also had a prognostic role in each proteomic subtype (log rank $p = 1.7E-4$) (Figure S5K). SUMF2, EGFR, and GBAS, located at chr7p11.2, were the top three proteins that had significant correlation with the copy number (Figure 5F). SCNA, RNA, and protein levels of these three genes increased from S-I to S-III, and EGFR stood out for its potential role in CRISPR and RNAi databases^{41,42} (Figure 5G). The protein level of EGFR was negatively associated with clinical outcome (Figure 5H). To verify the effect of the chr7p11.2 amplification, the downstream pathway alterations related to the high expression of EGFR were further studied. Wnt signaling, ubiquitin-mediated proteolysis, and cell cycle were upregulated by EGFR (Figures 5I and S5L). Consequently, the chr7p11.2 amplification could lead to the upregulation of EGFR, ultimately promoting tumor progression of lung adenocarcinoma (Figure 5J).

Immune clusters and tumor microenvironment in AIS/MIA and IAC

To gain a deeper understanding of immune characterization in AIS/MIA and IAC, we accessed xCell (<https://xcell.ucsf.edu>)⁴³ with the transcriptomic and proteomic data, which were then analyzed by the consensus clustering (Table S5). Four distinct subtypes were identified among 186 paired tumor samples (Figures 6A and S6A). Based on a Sankey diagram, we found a positive association between proteomic clusters and immune clusters (Fisher's exact test, $p = 0.03$) (Figure S6B). Kaplan-Meier analysis implicated worse survival in patients with the immune depression subtype compared with the other three subtypes (log rank $p = 1E-3$ for OS, $p < 1E-4$ for RFS) (Figures 6B and S6C). The majority of patients with the endothelial cell subtype had AIS/MIA, whereas the majority of patients with the immune depression subtype had IAC (Figure 6C). We further investigated the prevalence of the four immune clusters stratified by

Figure 5. Proteomic subtypes for lung adenocarcinoma

- Heatmap showing proteomic subtypes annotated with clinical features.
- Kaplan-Meier plot comparing recurrence-free survival in different proteomic subtypes. Log rank test was used to calculate the p value.
- Enriched pathways in three proteomic subtypes. The enrichment score was calculated by gene set variation analysis (GSVA).
- Heatmap of differentially expressed proteins among the proteomic subtypes.
- Kaplan-Meier plot comparing overall survival (OS) in patients with 7p11.2 amplification or not.
- Copy number variation (CNV)-protein correlation for genes located in 7p11.2 (Pearson's correlation test).
- SCNA, RNA, and protein levels of SUMF2, EGFR, and GBAS in three proteomic subtypes.
- Kaplan-Meier plot comparing OS in patients stratified by the protein level of EGFR.
- Enriched pathways related to the protein level of EGFR (Pearson's correlation test).
- A working model of 7p11.2 amplification promoting tumor progression via *EGFR* upregulation. The asterisk was defined as follows: **** $p < 1.0E-4$, *** $p < 1.0E-3$, ** $p < 0.01$, * $p < 0.05$.



(legend on next page)

EGFR mutation status. There was no distributional difference in *EGFR* mutation among the four subtypes (Figures 6A and S6D). The endothelial cell subtype had the highest proportion of AIS/MIA in both *EGFR* mutation and WT groups, and the immune depression subtype had the highest proportion of IAC in both groups (Figure S6E). In the endothelial cell subtype, immune signatures including endothelial cells and neutrophils were prominent, as evidenced by the high expression of CD34/VWF/KDR/FLT1/LYVE1 and CD16B (Figure 6D). The B cell subtype displayed more B cells, as evidenced by the overrepresented CD19, CD79B, and CD22 (Figure 6D). In the dendritic cell (DC) subtype, macrophages, monocytes, and DCs were predominant, and CD209, CD83, CD33, and CD68 were highly expressed (Figure 6D). The immune depression subtype was distinct from the others, with low expression of all immune cells and CD molecules (Figure 6D).

Next, we investigated the prognostic significance of these immune cells within the entire cohort. The results revealed that elevated levels of endothelial cells and DCs were significantly linked to improved survival (Figure S6F). Neutrophils were found to exhibit a pro-tumoral role in most studies primarily focused on advanced tumors.⁴⁴ However, contrasting evidence from other investigations has underscored the anti-tumoral potential of neutrophils, particularly in the context of early-stage cancer, such as at the N0 stage.^{45–47} Consistently, we also found that the high expression of neutrophils in patients with N0 diseases indicated better survival (Figure S6G). To investigate biological landscapes and functional processes in each subtype, we conducted gene set variation analysis (GSVA) using transcriptomic and proteomic data. TNFA signaling via NF- κ B was mostly enriched in the endothelial cell subtype, allograft rejection and interferon gamma response in the B cell subtype, IL-2 STAT5 signaling and coagulation in the DC subtype, and pathways related to DNA repair, hypoxia, G2M checkpoint, E2F targets, MYC targets, and glycolysis in the immune depression subtype (Figures 6D and 6E). Given the fact that the G2M checkpoint, E2F target, and MYC target pathways were closely correlated with cell proliferation and cell-cycle activity,^{48,49} drugs targeting the cell cycle, including CDK1 inhibitors (e.g., RO-3306) and CDK4/6 inhibitors (e.g., palbociclib, abemaciclib),^{50,51} might be effective alternatives for the immune depression subtype. Upregulation of DNA replication was also observed in the immune depression subtype (Figure S6H), indicating the potential use of PARP inhibitors in this special subtype.^{52,53}

We also conducted analyses to confirm the predictive value of these immune signatures in patients receiving immunotherapy.⁵⁴ Compared with patients receiving atezolizumab (PD-L1 inhibitor) with short OS (<6 months), the signatures of B cells, plasma cells, and DCs were highly expressed in patients with long OS (>12 months) with atezolizumab (Figure S6I). Moreover, the corresponding markers of these immune signatures were also overrepresented in patients with long OS (>12 months) on atezolizumab (Figure S6J). Importantly, those immune signatures exhibited a positive association with prognosis in patients receiving immunotherapy (Figure S6K). Furthermore, DCs were correlated with improved survival in patients receiving docetaxel (Figure S6L), and the upregulation of CD83 was observed in patients with longer OS (Figure S6J). Taken together, B cells and plasma cells could predict the benefits of immunotherapy in patients with lung adenocarcinoma.

Interestingly, *TP53* mutation and patients over 60 years of age were prominent in the immune depression subtype (Figures 6A, 6F, and S6L). We further investigate the impacts of *TP53* mutation. As a result, GSEA showed that *TP53* mutation had positive effects on DNA replication and base excision repair at the RNA and protein levels (Figures 6G, 6H, and S6M). Furthermore, MCM3, MCM6, and RFC3, the overrepresented proteins from the DNA replication pathway, had a significantly negative effect on the survival of patients with lung adenocarcinoma (log rank $p < 1.0E-4$ for MCM3, $p < 1.0E-4$ for MCM6, and $p < 1.0E-4$ for RFC3) (Figure S6N).

DISCUSSION

Lung adenocarcinoma accounts for 40%–50% of non-small cell lung cancer.^{55,56} Investigations on carcinogenesis and progression of lung adenocarcinoma enable the better understanding of this disease. Although some advancements have been achieved in genomic profiling in lung pre-neoplasia and IAC,^{12,57} gaps between genomic alterations and oncogenic proteins still exist. To fulfill the above limitations, we conducted proteogenomic analysis on 197 tumors and matched NATs, comprising 98 lung pre-invasive (AIS/MIA) adenocarcinomas and 99 IACs. This is a comprehensive proteogenomic study to delineate the molecular architecture and hallmarks of carcinogenesis and tumor progression from AIS/MIA to IAC. Our analysis provides a comprehensive multi-omic landscape for pre-invasive and invasive lung adenocarcinoma, encompassing somatic gene mutations and

Figure 6. Immune clusters of lung adenocarcinoma

(A) Heatmap showing proteogenomic profiles comprising xCell signatures, CD molecules (RNA), potentially targetable immune checkpoints (RNA), and enriched pathways (proteome) in four distinct immune clusters.

(B) Kaplan-Meier plot comparing overall survival for the four immune clusters. Log rank test was used to calculate the p value.

(C) Proportions of AIS/MIA and IAC in the four immune clusters.

(D) Contour plot of density based on stroma score and endothelial cell signature. For each cluster, key upregulated molecules and pathways, as well as potentially targetable immune checkpoints, were reported based on transcriptome (R) or proteome (P). Red, endothelial cells; green, B cells; purple, DCs; blue, immune depression.

(E) Distribution of enrichment scores of G2M checkpoint and E2F targets signaling calculated by gene set variation analysis (GSVA) based on global proteomic and transcriptomic data in the four immune clusters (Kruskal-Wallis test).

(F) Proportions of *TP53* mutation in the four RNA-based immune clusters (chi-squared test).

(G) Gene set enrichment analysis (GSEA) based on proteomic data stratified by *TP53* mutation status.

(H) Molecular profiling of RNA and protein in genes from DNA replication and base excision repair stratified by *TP53* mutation. The asterisk was defined as follows: **** $p < 1.0E-4$, *** $p < 1.0E-3$, ** $p < 0.01$, * $p < 0.05$.

signatures, arm-level events, proteome-based pathways, and the immune microenvironment. Our integrated proteogenomic characterization of AIS/MIA, IAC, and their paired NATs serves as a public resource for the scientific community, paving the way for future studies regarding carcinogenesis and progression of lung adenocarcinoma.

As the utilization of LDCT continues to gain popularity as a screening tool for lung cancer, the proportion of patients with ground-glass opacity (GGO)-featured lung cancer is increasing.⁵ If we can uncover the intrinsic mechanisms that drive tumorigenesis and development from pre-invasive to invasive lung adenocarcinoma, interventions targeting these mechanisms may aid in reducing the prevalence of lung cancer and preventing pre-invasive adenocarcinoma from progressing into IAC. Our study presents the proteogenomic landscape from pre-invasive to invasive lung adenocarcinoma, assisting in the identification of treatment targets for early-stage lung cancer. However, there are also some extrinsic factors that contribute to this process. Recently, William Hill et al. discovered that air pollutants can promote lung cancer by acting on cells that already harbor pre-existing oncogenic mutations in healthy lung tissue.⁵⁸ Oncogenic *EGFR* driver mutations were detected in 18% of healthy tissue samples using ultradeep mutational sequencing.⁵⁸ These findings supported that *EGFR* mutations could be an early molecular event in lung cancer tumorigenesis. Further investigation is warranted to explore other extrinsic factors and the interplay between external and inherent factors.

Moreover, our analysis revealed an interaction between *EGFR* mutation and NOTCH pathways, which has been nominated in neural stem cells.⁵⁹ Traditionally, the NOTCH signaling cascade has a crucial role in shaping development and homeostasis in diverse organs and tissues.⁶⁰ Recent studies report that NOTCH signaling could act as an oncogene or a tumor suppressor in distinct cancers.⁶¹ In non-small cell lung cancer, NOTCH signaling exhibits an oncogenic effect and contributes to drug resistance to tyrosine kinase inhibitors of *EGFR*.^{61,62} Our study indicated that the NOTCH pathway was downstream of the *EGFR* receptor, which requires experimental validation. In addition, a previous study has reported the association between 4q deletion and micrometastasis of lung cancer.⁶³ Our analysis ascertained that chr4q12 deletion is a crucial arm-level event in the progression from AIS/MIA to IAC, resulting in a low protein expression level of p53-induced SPATA18. Downregulation of SPATA18 diminished lysosome-like organelles, causing decreased mitophagy and tumor growth of the lung. Our study put forward a possible explanation of how chr4q12 deletion leads to the progression from AIS/MIA to IAC. Since AIS and MIA represent the indolent type of lung cancer, chr4q12 deletion might be part of the intrinsic differences between indolent and aggressive tumor behavior.

Consistent with previous studies,^{37,64} the correlations between transcriptome and proteome of our study were 0.39 for all the tumor tissues and 0.20 for NATs. Specifically, the study by Xu et al. reported correlations of 0.28 for lung tumor tissues and 0.07 for NATs.³⁷ In the study by Chen et al., they documented a correlation of 0.31.⁶⁴ The proteomic data identified distinct pathways in NATs, AIS/MIA, and IAC. Blood coagulation and fatty acid β -oxidation, linked to tumor growth and dissemina-

tion in previous reports,^{65,66} are upregulated in the pre-invasive stage of lung adenocarcinoma, implying that targeting the above two pathways might be alternative treatments for AIS/MIA. In addition, complement activation was also enriched in AIS/MIA. The complement pathway is an essential part of the innate immune system and interacts with the adaptive immune system to eliminate abnormal cells. Nevertheless, recent studies reveal that complement activation acts as a “double-edged sword” and enables cancer progression.⁶⁷ Future studies are warranted to investigate the actual effect of complement activation in AIS/MIA. In IAC, some classic oncogenic pathways (e.g., cell-cycle and *FGFR* signaling pathways) were enhanced. Taken together, our findings provide insights into the proteomic characterization of pre-neoplasia of lung cancer and pave the way for future studies of lung cancer carcinogenesis.

The unsupervised clustering analysis identified three proteomic subtypes, which differed from the clusters reported in previous studies.^{37–40} Due to the substantial proportion of pre-invasive lesions in our cohort, our proteomic subtypes exhibited a significant association with the invasiveness of lung adenocarcinoma. Specifically, S-I predominantly comprised AIS/MIA, while the majority of S-III was formed by IAC cases. Understanding these proteomic subtypes and their association with tumor invasiveness can offer valuable insights for personalized treatment decisions and prognosis evaluation in patients with lung adenocarcinoma.

Immune escape is recognized as a significant cause of cancer formation and progression,⁶⁸ making immunotherapy a promising treatment approach for most types of cancer.⁶⁹ Malignant cells need to overcome multiple obstacles to achieve the successful formation of a tumor, and it is a significant step to hijack microenvironmental processes to support the initiation and development of tumors. The lesion is less compromised by immune evasion during the initiation, and upregulation of genes involved in immunosuppression was observed as the lesions progressed.^{70,71} Compared with ovarian cancer and other “cold” tumors,⁷² lung adenocarcinoma is not a cold tumor, in general, and has better response to immunotherapy.⁷³ Several studies using IHC and single-cell RNA-seq have confirmed the low infiltration of CD4⁺ and CD8⁺ in early-stage lung cancer.^{74,75} However, some researchers think that PD-L1-mediated immune evasion is already present in the pre-invasive stage,⁷⁴ indicating that PD-1 antibodies might be effective in that stage. Several studies have tried to answer this question, and some of them revealed limited response to immunotherapy for GGO-featured lung cancer.⁷⁶ However, some other studies draw different conclusions.^{77–79} A clinical trial investigated the use of sintilimab (PD-1 inhibitor) in patients with GGOs and found that 15% of patients had a major pathologic response and 46.2% of patients had a partial pathologic response.⁷⁹ Therefore, the actual efficacy of immunotherapy in pre-invasive lung cancer still warrants further investigation. Our study also unveiled that B cells and plasma cells were able to predict the efficacy of immunotherapy in patients with non-small cell lung cancer. This might be useful for the selection of patients receiving immunotherapy.

Recently, several studies have confirmed the pro-tumoral role of neutrophils.^{44,80} However, the anti-tumoral effect of neutrophils has also been observed, particularly at the early stage of

cancer.^{45–47} For instance, a study focusing on head-and-neck cancer demonstrated that neutrophils exhibited an antigen-presenting phenotype and activated T cells at the N0 stage, while at the N1-3 stage, neutrophils displayed immunosuppressive effects and repressed T cell response.⁴⁷ In our study, neutrophils were found to be upregulated at the pre-invasive stage, suggesting a potential anti-tumoral role. In patients with N0 lung adenocarcinoma, high levels of neutrophils indicated improved survival. However, if lymph node metastases occur, neutrophils might switch to a pro-tumoral function. The dual role of neutrophils highlights their complexity and the importance of understanding their distinct functions at different stages of cancer progression.

In summary, our study depicts the comprehensive genomic, proteomic, and phosphoproteomic map from lung neoplasia to IAC, enabling a deep understanding of the molecular architecture and hallmarks of early-stage lung adenocarcinoma. Specifically, we presented a preliminary exploration of multi-omics in the multi-stage carcinogenesis of lung cancer, which can be used to model the key events in lung tumorigenesis. The top-ranked *EGFR* mutation happened as early as in AIS/MIA. The chr4q12 deletion was the key event in the transition process from AIS/MIA to IAC, and the protein-level expression of SPATA18 was decreased, leading to tumor cell proliferation and poor prognosis in patients with lung adenocarcinoma. The data from this project serve as a valuable resource for the scientific community to explore the tumorigenesis and progression of lung adenocarcinoma, guiding the way to a deeper understanding of this lethal disease.

Limitations of the study

Although this is a large-scale proteogenomic study on the initiation and development of lung adenocarcinoma, there are several inherent limitations in this study. First, our cohort included east-Asian patients with lung adenocarcinoma, characterized by the high prevalence of females and never-smokers. Consequently, a higher frequency of *EGFR* mutation was observed compared with Caucasian cohorts. It warrants further investigation as to whether the findings could expand to other races. Second, despite rich resources for the investigation of tumorigenesis and development in lung cancer provided by this study, the hypotheses and conclusions need experimental validation from cell lines, animal models, organoids, or even clinical trials. Third, spatial information was missing due to the nature of the bulk sequencing method. Future analysis incorporating spatial transcriptome into other 'omics might provide a deeper understanding. Fourth, our study utilized NATs as the normal controls for DNA analysis, which may result in slight differences from blood samples. Future researchers should be aware of this issue when incorporating data from our cohort into their analyses.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY

- Lead contact
- Materials availability
- Data and code availability

● EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

- Human samples

● METHOD DETAILS

- Pathological review
- Whole exome sequencing (WES)
- RNA sequencing
- WES alignment and somatic mutation calling
- RNA-seq quantification
- GISTIC and MutSig analysis
- Protein Extraction and tryptic digestion
- First dimensional reversed-phase separation
- Phosphopeptide enrichment
- Proteome analysis
- Phosphoproteome analysis
- Peptide and protein identification
- Label-free-based MS quantification of proteins
- Differential protein analysis
- Pathway enrichment analysis
- Gene Set Enrichment Analysis (GSEA)
- Principal component analysis
- Correlation between subtype and clinical features
- mRNA-protein correlation in tumors and normal samples
- Immune deconvolution
- Immune consensus clustering analysis
- Database searching of MS phosphoproteomic data
- Kinase activity prediction and phosphopeptide analysis
- Immunohistochemistry
- *In vitro* and *in vivo* experiments
- Construction of plasmids and transfection
- Cell viability assay
- Wound-healing assay

● QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.xcrm.2023.101358>.

ACKNOWLEDGMENTS

This work was supported by the Shanghai Rising-Star Program (21QC1400600), the National Natural Science Foundation of the People's Republic of China (82373309, 82102914, 81930073, 81772466, and 31770886), the Shanghai Municipal Science and Technology Major Project (2017SHZDZX01), the National Key R&D Program of China (2017YFA0505102, 2022YFA1303200, 2022YFA1303201, 2022YFA1103900), the Shanghai Technology Innovation Action Project (22ZR1413200, 20JC1417200), the Program of Shanghai Academic/Technology Research Leader (22XD1420100), and the Pilot Project of Fudan University (IDF159045). The project also received support from GENECHM, Shanghai, China.

AUTHOR CONTRIBUTIONS

Y. Zhang, F.F., Y.L., C. Ding, and H.C. conceived and supervised the project. Y. Zhang, F.F., Y. Zhao, H. Han, Z.G., Q.X., Q. Zheng, H. Hu, Y.S., Y.L., and

H.C. coordinated the acquisition and quality control of AIS/MIA and IAC tissues and paired adjacent tissues. Q. Zhang, L.L., H.L., C.G., and C. Ding directed and performed analyses and quality control of MS data. Y. Zhang, F.F., L.L., Q. Zhang, C. Deng, and Y. Zhao performed proteomics and genomics data analyses. W.S. and F.F. performed IHC. Y. Zhang, F.F., L.L., Q. Zhang, and C. Ding. wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no relevant competing interests.

Received: January 24, 2023

Revised: August 29, 2023

Accepted: December 11, 2023

Published: January 5, 2024; corrected online January 12, 2026

REFERENCES

- Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* 68, 394–424.
- Picard, M. (2022). Why Do We Care More About Disease than Health? *Phenomics* 2, 145–155.
- National Lung Screening Trial Research Team; Aberle, D.R., Adams, A.M., Berg, C.D., Black, W.C., Clapp, J.D., Fagerstrom, R.M., Gareen, I.F., Gatsonis, C., Marcus, P.M., and Sicks, J.D. (2011). Reduced lung-cancer mortality with low-dose computed tomographic screening. *N. Engl. J. Med.* 365, 395–409.
- National Lung Screening Trial Research Team; Church, T.R., Black, W.C., Aberle, D.R., Berg, C.D., Clingan, K.L., Duan, F., Fagerstrom, R.M., Gareen, I.F., Gierada, D.S., et al. (2013). Results of initial low-dose computed tomographic screening for lung cancer. *N. Engl. J. Med.* 368, 1980–1991.
- Zhang, Y., Jheon, S., Li, H., Zhang, H., Xie, Y., Qian, B., Lin, K., Wang, S., Fu, C., Hu, H., et al. (2020). Results of low-dose computed tomography as a regular health examination among Chinese hospital employees. *J. Thorac. Cardiovasc. Surg.* 160, 824–831.e4.
- Travis, W.D., Brambilla, E., Noguchi, M., Nicholson, A.G., Geisinger, K.R., Yatabe, Y., Beer, D.G., Powell, C.A., Riehl, G.J., Van Schil, P.E., et al. (2011). 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.
- Travis, W.D., Brambilla, E., Nicholson, A.G., Yatabe, Y., Austin, J.H.M., Beasley, M.B., Chirieac, L.R., Dacic, S., Duhig, E., Flieder, D.B., et al. (2015). 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.
- Noguchi, M. (2010). Stepwise progression of pulmonary adenocarcinoma—clinical and molecular implications. *Cancer Metastasis Rev.* 29, 15–21.
- Yotsukura, M., Asamura, H., Motoi, N., Kashima, J., Yoshida, Y., Nakagawa, K., Shiraishi, K., Kohno, T., Yatabe, Y., and Watanabe, S.I. (2021). Long-Term Prognosis of Patients With Resected Adenocarcinoma In Situ and Minimally Invasive Adenocarcinoma of the Lung. *J. Thorac. Oncol.* 16, 1312–1320.
- Goldstraw, P., Chansky, K., Crowley, J., Rami-Porta, R., Asamura, H., Eberhardt, W.E.E., Nicholson, A.G., Groome, P., Mitchell, A., Bolejack, V., et al. (2016). The IASLC Lung Cancer Staging Project: Proposals for Revision of the TNM Stage Groupings in the Forthcoming (Eighth) Edition of the TNM Classification for Lung Cancer. *J. Thorac. Oncol.* 11, 39–51.
- Izumchenko, E., Chang, X., Brait, M., Fertig, E., Kagohara, L.T., Bedi, A., Marchionni, L., Agrawal, N., Ravi, R., Jones, S., et al. (2015). Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. *Nat. Commun.* 6, 8258.
- Qian, J., Zhao, S., Zou, Y., Rahman, S.M.J., Senosain, M.F., Stricker, T., Chen, H., Powell, C.A., Borczuk, A.C., and Massion, P.P. (2020). Genomic Underpinnings of Tumor Behavior in In Situ and Early Lung Adenocarcinoma. *Am. J. Respir. Crit. Care Med.* 201, 697–706.
- Murphy, S.J., Wigle, D.A., Lima, J.F., Harris, F.R., Johnson, S.H., Halling, G., Asiedu, M.K., Seto, C.T., Terra, S., Kosari, F., et al. (2014). Genomic rearrangements define lineage relationships between adjacent lepidic and invasive components in lung adenocarcinoma. *Cancer Res.* 74, 3157–3167.
- Vinayanuwattikun, C., Le Calvez-Kelm, F., Abedi-Ardekani, B., Zaridze, D., Mukeria, A., Voegelé, C., Vallée, M., Purnomosari, D., Forey, N., Durand, G., et al. (2016). Elucidating Genomic Characteristics of Lung Cancer Progression from In Situ to Invasive Adenocarcinoma. *Sci. Rep.* 6, 31628.
- Fu, F., Tao, X., Jiang, Z., Gao, Z., Zhao, Y., Li, Y., Hu, H., Shen, L., Sun, Y., and Zhang, Y. (2023). Identification of Germline Mutations in East-Asian Young Never-Smokers with Lung Adenocarcinoma by Whole-Exome Sequencing. *Phenomics* 3, 182–189.
- Chen, H., Carrot-Zhang, J., Zhao, Y., Hu, H., Freeman, S.S., Yu, S., Ha, G., Taylor, A.M., Berger, A.C., Westlake, L., et al. (2019). Genomic and immune profiling of pre-invasive lung adenocarcinoma. *Nat. Commun.* 10, 5472.
- Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-Garcia, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* 4, 2612.
- Li, Y., Li, X., Li, H., Zhao, Y., Liu, Z., Sun, K., Zhu, X., Qi, Q., An, B., Shen, D., et al. (2020). Genomic characterisation of pulmonary subsolid nodules: mutational landscape and radiological features. *Eur. Respir. J.* 55, 1901409.
- Hua, X., Hyland, P.L., Huang, J., Song, L., Zhu, B., Caporaso, N.E., Landi, M.T., Chatterjee, N., and Shi, J. (2016). MEGSA: A Powerful and Flexible Framework for Analyzing Mutual Exclusivity of Tumor Mutations. *Am. J. Hum. Genet.* 98, 442–455.
- Campbell, J.D., Alexandrov, A., Kim, J., Wala, J., Berger, A.H., Pedamallu, C.S., Shukla, S.A., Guo, G., Brooks, A.N., Murray, B.A., et al. (2016). Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* 48, 607–616.
- Jamal-Hanjani, M., Wilson, G.A., McGranahan, N., Birkbak, N.J., Watkins, T.B.K., Veeriah, S., Shafi, S., Johnson, D.H., Mitter, R., Rosenthal, R., et al. (2017). Tracking the Evolution of Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* 376, 2109–2121.
- Rizvi, H., Sanchez-Vega, F., La, K., Chatila, W., Jonsson, P., Halpenny, D., Plodkowski, A., Long, N., Sauter, J.L., Rekhtman, N., et al. (2018). Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. *J. Clin. Oncol.* 36, 633–641.
- Tate, J.G., Bamford, S., Jubb, H.C., Sondka, Z., Beare, D.M., Bindal, N., Boutselakis, H., Cole, C.G., Creatore, C., Dawson, E., et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res.* 47, D941–D947.
- Alexandrov, L.B., Kim, J., Haradhvala, N.J., Huang, M.N., Tian Ng, A.W., Wu, Y., Boot, A., Covington, K.R., Gordenin, D.A., Bergstrom, E.N., et al. (2020). The repertoire of mutational signatures in human cancer. *Nature* 578, 94–101.
- Caso, R., Sanchez-Vega, F., Tan, K.S., Mastrogiacomo, B., Zhou, J., Jones, G.D., Nguyen, B., Schultz, N., Connolly, J.G., Brandt, W.S., et al. (2020). The Underlying Tumor Genomics of Predominant Histologic Subtypes in Lung Adenocarcinoma. *J. Thorac. Oncol.* 15, 1844–1856.

26. Jiang, Y., Sun, A., Zhao, Y., Ying, W., Sun, H., Yang, X., Xing, B., Sun, W., Ren, L., Hu, B., et al. (2019). Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. *Nature* 567, 257–261.
27. Ge, S., Xia, X., Ding, C., Zhen, B., Zhou, Q., Feng, J., Yuan, J., Chen, R., Li, Y., Ge, Z., et al. (2018). A proteomic landscape of diffuse-type gastric cancer. *Nat. Commun.* 9, 1012.
28. Sano, H., Futamura, M., Gaowa, S., Kamino, H., Nakamura, Y., Yamaguchi, K., Tanaka, Y., Yasufuku, I., Nakakami, A., Arakawa, H., and Yoshida, K. (2020). p53/Mieap-regulated mitochondrial quality control plays an important role as a tumor suppressor in gastric and esophageal cancers. *Biochem. Biophys. Res. Commun.* 529, 582–589.
29. Kitamura, N., Nakamura, Y., Miyamoto, Y., Miyamoto, T., Kabu, K., Yoshida, M., Futamura, M., Ichinose, S., and Arakawa, H. (2011). Mieap, a p53-inducible protein, controls mitochondrial quality by repairing or eliminating unhealthy mitochondria. *PLoS One* 6, e16060.
30. Nakamura, Y., Kitamura, N., Shinogi, D., Yoshida, M., Goda, O., Murai, R., Kamino, H., and Arakawa, H. (2012). BNIP3 and NIX mediate Mieap-induced accumulation of lysosomal proteins within mitochondria. *PLoS One* 7, e30767.
31. Dan, X., Babbar, M., Moore, A., Wechter, N., Tian, J., Mohanty, J.G., Croteau, D.L., and Bohr, V.A. (2020). DNA damage invokes mitophagy through a pathway involving Spata18. *Nucleic Acids Res.* 48, 6611–6623.
32. Friedl, P., and Wolf, K. (2003). Tumour-cell invasion and migration: diversity and escape mechanisms. *Nat. Rev. Cancer* 3, 362–374.
33. Petersson, J., and Glenny, R.W. (2014). Gas exchange and ventilation-perfusion relationships in the lung. *Eur. Respir. J.* 44, 1023–1041.
34. Tavernari, D., Battistello, E., Dheilly, E., Petruzzella, A.S., Mina, M., Sordet-Dessimoz, J., Peters, S., Krueger, T., Gfeller, D., Riggi, N., et al. (2021). Nongenetic Evolution Drives Lung Adenocarcinoma Spatial Heterogeneity and Progression. *Cancer Discov.* 11, 1490–1507.
35. Ramalingam, S.S., Vansteenkiste, J., Planchard, D., Cho, B.C., Gray, J.E., Ohe, Y., Zhou, C., Reungwetwattana, T., Cheng, Y., Chewaskulyong, B., et al. (2020). Overall Survival with Osimertinib in Untreated, EGFR-Mutated Advanced NSCLC. *N. Engl. J. Med.* 382, 41–50.
36. Wu, Y.L., Tsuboi, M., He, J., John, T., Grohe, C., Majem, M., Goldman, J.W., Laktionov, K., Kim, S.W., Kato, T., et al. (2020). Osimertinib in Resected EGFR-Mutated Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* 383, 1711–1723.
37. Xu, J.Y., Zhang, C., Wang, X., Zhai, L., Ma, Y., Mao, Y., Qian, K., Sun, C., Liu, Z., Jiang, S., et al. (2020). Integrative Proteomic Characterization of Human Lung Adenocarcinoma. *Cell* 182, 245–261.e17.
38. Gillette, M.A., Satpathy, S., Cao, S., Dhanasekaran, S.M., Vasaikar, S.V., Krug, K., Petralia, F., Li, Y., Liang, W.W., Reva, B., et al. (2020). Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. *Cell* 182, 200–225.e35.
39. Soltis, A.R., Bateman, N.W., Liu, J., Nguyen, T., Franks, T.J., Zhang, X., Dalgard, C.L., Viollet, C., Somiari, S., Yan, C., et al. (2022). Proteogenomic analysis of lung adenocarcinoma reveals tumor heterogeneity, survival determinants, and therapeutically relevant pathways. *Cell Rep. Med.* 3, 100819.
40. Roh, W., Geffen, Y., Cha, H., Miller, M., Anand, S., Kim, J., Heiman, D.I., Gainor, J.F., Laird, P.W., Cherniack, A.D., et al. (2022). High-Resolution Profiling of Lung Adenocarcinoma Identifies Expression Subtypes with Specific Biomarkers and Clinically Relevant Vulnerabilities. *Cancer Res.* 82, 3917–3931.
41. Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
42. McDonald, E.R., 3rd, de Weck, A., Schlabach, M.R., Billy, E., Mavrikakis, K.J., Hoffman, G.R., Belur, D., Castelletti, D., Frias, E., Gampa, K., et al. (2017). Project DRIVE: A Compendium of Cancer Dependencies and Synthetic Lethal Relationships Uncovered by Large-Scale, Deep RNAi Screening. *Cell* 170, 577–592.e10.
43. Aran, D., Hu, Z., and Butte, A.J. (2017). xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 18, 220.
44. Hedrick, C.C., and Malanchi, I. (2022). Neutrophils in cancer: heterogeneous and multifaceted. *Nat. Rev. Immunol.* 22, 173–187.
45. Finisguerra, V., Di Conza, G., Di Matteo, M., Serneels, J., Costa, S., Thompson, A.A.R., Wauters, E., Walmsley, S., Prenen, H., Granot, Z., et al. (2015). MET is required for the recruitment of anti-tumoural neutrophils. *Nature* 522, 349–353.
46. Koga, Y., Matsuzaki, A., Suminoe, A., Hattori, H., and Hara, T. (2004). Neutrophil-derived TNF-related apoptosis-inducing ligand (TRAIL): a novel mechanism of antitumor effect by neutrophils. *Cancer Res.* 64, 1037–1043.
47. Pylaeva, E., Korschunow, G., Spyra, I., Bordbari, S., Siakaeva, E., Ozel, I., Domnich, M., Squire, A., Hasenberg, A., Thangavelu, K., et al. (2022). During early stages of cancer, neutrophils initiate anti-tumor immune responses in tumor-draining lymph nodes. *Cell Rep.* 40, 111171.
48. Oshi, M., Newman, S., Tokumaru, Y., Yan, L., Matsuyama, R., Endo, I., Katz, M.H.G., and Takabe, K. (2020). High G2M Pathway Score Pancreatic Cancer is Associated with Worse Survival, Particularly after Margin-Positive (R1 or R2) Resection. *Cancers* 12, 2871.
49. Kent, L.N., and Leone, G. (2019). The broken cycle: E2F dysfunction in cancer. *Nat. Rev. Cancer* 19, 326–338.
50. Sunada, S., Saito, H., Zhang, D., Xu, Z., and Miki, Y. (2021). CDK1 inhibitor controls G2/M phase transition and reverses DNA damage sensitivity. *Biochem. Biophys. Res. Commun.* 550, 56–61.
51. Zhang, J., Xu, D., Zhou, Y., Zhu, Z., and Yang, X. (2021). Mechanisms and Implications of CDK4/6 Inhibitors for the Treatment of NSCLC. *Front. Oncol.* 11, 676041.
52. Pilié, P.G., Tang, C., Mills, G.B., and Yap, T.A. (2019). State-of-the-art strategies for targeting the DNA damage response in cancer. *Nat. Rev. Clin. Oncol.* 16, 81–104.
53. Lord, C.J., and Ashworth, A. (2017). PARP inhibitors: Synthetic lethality in the clinic. *Science* 355, 1152–1158.
54. Patil, N.S., Nabet, B.Y., Muller, S., Koeppen, H., Zou, W., Giltne, J., Au-Yeung, A., Srivats, S., Cheng, J.H., Takahashi, C., et al. (2022). Intratumoral plasma cells predict outcomes to PD-L1 blockade in non-small cell lung cancer. *Cancer Cell* 40, 289–300.e284.
55. Duma, N., Santana-Davila, R., and Molina, J.R. (2019). Non-Small Cell Lung Cancer: Epidemiology, Screening, Diagnosis, and Treatment. *Mayo Clin. Proc.* 94, 1623–1640.
56. Lortet-Tieulent, J., Soerjomataram, I., Ferlay, J., Rutherford, M., Weiderpass, E., and Bray, F. (2014). International trends in lung cancer incidence by histological subtype: adenocarcinoma stabilizing in men but still increasing in women. *Lung Cancer* 84, 13–22.
57. Dejima, H., Hu, X., Chen, R., Zhang, J., Fujimoto, J., Parra, E.R., Haymaker, C., Hubert, S.M., Duose, D., Solis, L.M., et al. (2021). Immune evolution from preneoplasia to invasive lung adenocarcinomas and underlying molecular features. *Nat. Commun.* 12, 2722.
58. Hill, W., Lim, E.L., Weeden, C.E., Lee, C., Augustine, M., Chen, K., Kuan, F.C., Marongiu, F., Evans, E.J., Jr., Moore, D.A., et al. (2023). Lung adenocarcinoma promotion by air pollutants. *Nature* 616, 159–167.
59. Aguirre, A., Rubio, M.E., and Gallo, V. (2010). Notch and EGFR pathway interaction regulates neural stem cell number and self-renewal. *Nature* 467, 323–327.
60. Kopan, R., and Ilgan, M.X.G. (2009). The canonical Notch signaling pathway: unfolding the activation mechanism. *Cell* 137, 216–233.
61. Majumder, S., Crabtree, J.S., Golde, T.E., Minter, L.M., Osborne, B.A., and Miele, L. (2021). Targeting Notch in oncology: the path forward. *Nat. Rev. Drug Discov.* 20, 125–144.

62. Bousquet Mur, E., Bernardo, S., Papon, L., Mancini, M., Fabbriozzi, E., Goussard, M., Ferrer, I., Giry, A., Quantin, X., Pujol, J.L., et al. (2020). Notch inhibition overcomes resistance to tyrosine kinase inhibitors in EGFR-driven lung adenocarcinoma. *J. Clin. Invest.* **130**, 612–624.
63. Wragge, M., Ruosaari, S., Eijk, P.P., Kaifi, J.T., Hollmén, J., Yekebas, E.F., Izbicki, J.R., Brakenhoff, R.H., Streichert, T., Riethdorf, S., et al. (2009). Genomic profiles associated with early micrometastasis in lung cancer: relevance of 4q deletion. *Clin. Cancer Res.* **15**, 1566–1574.
64. Chen, Y.J., Roumeliotis, T.I., Chang, Y.H., Chen, C.T., Han, C.L., Lin, M.H., Chen, H.W., Chang, G.C., Chang, Y.L., Wu, C.T., et al. (2020). Proteogenomics of Non-smoking Lung Cancer in East Asia Delineates Molecular Signatures of Pathogenesis and Progression. *Cell* **182**, 226–244.e17.
65. Lima, L.G., and Monteiro, R.Q. (2013). Activation of blood coagulation in cancer: implications for tumour progression. *Biosci. Rep.* **33**, e00064.
66. Yamamoto, K., Abe, S., Honda, A., Hashimoto, J., Aizawa, Y., Ishibashi, S., Takemura, T., Hanagata, N., Yamamoto, M., Miura, O., et al. (2020). Fatty acid beta oxidation enzyme HADHA is a novel potential therapeutic target in malignant lymphoma. *Lab. Invest.* **100**, 353–362.
67. Kwak, J.W., Laskowski, J., Li, H.Y., McSharry, M.V., Sippel, T.R., Bullock, B.L., Johnson, A.M., Poczoobutt, J.M., Neuwelt, A.J., Malkoski, S.P., et al. (2018). Complement Activation via a C3a Receptor Pathway Alters CD4(+) T Lymphocytes and Mediates Lung Cancer Progression. *Cancer Res.* **78**, 143–156.
68. Galon, J., and Bruni, D. (2020). Tumor Immunology and Tumor Evolution: Intertwined Histories. *Immunity* **52**, 55–81.
69. Billan, S., Kaidar-Person, O., and Gil, Z. (2020). Treatment after progression in the era of immunotherapy. *Lancet Oncol.* **21**, e463–e476.
70. de Visser, K.E., and Joyce, J.A. (2023). The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. *Cancer Cell* **41**, 374–403.
71. Mascaux, C., Angelova, M., Vasaturo, A., Beane, J., Hijazi, K., Anthoine, G., Buttard, B., Rothe, F., Willard-Gallo, K., Haller, A., et al. (2019). Immune evasion before tumour invasion in early lung squamous carcinogenesis. *Nature* **571**, 570–575.
72. Yang, Y., Zhao, T., Chen, Q., Li, Y., Xiao, Z., Xiang, Y., Wang, B., Qiu, Y., Tu, S., Jiang, Y., et al. (2022). Nanomedicine Strategies for Heating "Cold" Ovarian Cancer (OC): Next Evolution in Immunotherapy of OC. *Adv. Sci.* **9**, e2202797.
73. De Guillebon, E., Dardenne, A., Saldmann, A., Séguier, S., Tran, T., Paolini, L., Lebbe, C., and Tartour, E. (2020). Beyond the concept of cold and hot tumors for the development of novel predictive biomarkers and the rational design of immunotherapy combination. *Int. J. Cancer* **147**, 1509–1518.
74. Zhang, C., Zhang, J., Xu, F.P., Wang, Y.G., Xie, Z., Su, J., Dong, S., Nie, Q., Shao, Y., Zhou, Q., et al. (2019). Genomic Landscape and Immune Microenvironment Features of Preinvasive and Early Invasive Lung Adenocarcinoma. *J. Thorac. Oncol.* **14**, 1912–1923.
75. Chen, K., Bai, J., Reuben, A., Zhao, H., Kang, G., Zhang, C., Qi, Q., Xu, Y., Hubert, S., Chang, L., et al. (2021). Multiomics Analysis Reveals Distinct Immunogenomic Features of Lung Cancer with Ground-Glass Opacity. *Am. J. Respir. Crit. Care Med.* **204**, 1180–1192.
76. Wu, F., Li, W., Zhao, W., Zhou, F., Xie, H., Shi, J., Yu, G., Fan, J., Jiang, T., and Zhou, C. (2020). Synchronous ground-glass nodules showed limited response to anti-PD-1/PD-L1 therapy in patients with advanced lung adenocarcinoma. *Clin. Transl. Med.* **10**.
77. Wu, F., Hu, C., Liu, X., Li, Y., Zeng, Y., He, L., and Liu, J. (2022). Ground-glass opacity showed response to immunotherapy in cancer patients. *J. Clin. Oncol.* **40**, e14605.
78. Cheng, B., Cheng, B., Li, C., Zhao, Y., and He, J. (2021). The efficacy of PD-1 antibody sintilimab on ground glass opacity lesions in patients with early-stage multiple primary lung cancer (CCTC-1901, NCT04026841). *J. Clin. Oncol.* **39**, 8545.
79. Xu, L., Shi, M., Wang, S., Li, M., Yin, W., Zhang, J., Zhu, J., Jiang, F., Xia, W., Qiu, N., et al. (2022). Immunotherapy for bilateral multiple ground glass opacities: An exploratory study for synchronous multiple primary lung cancer. *Front. Immunol.* **13**, 1009621.
80. Salcher, S., Sturm, G., Horvath, L., Untergasser, G., Kuempers, C., Fotakis, G., Panizzolo, E., Martowicz, A., Trebo, M., Pall, G., et al. (2022). High-resolution single-cell atlas reveals diversity and plasticity of tissue-resident neutrophils in non-small cell lung cancer. *Cancer Cell* **40**, 1503–1520.e8.
81. Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at arXiv.
82. DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498.
83. Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* **31**, 213–219.
84. Saunders, C.T., Wong, W.S.W., Swamy, S., Becq, J., Murray, L.J., and Cheetham, R.K. (2012). Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* **28**, 1811–1817.
85. Ramos, A.H., Lichtenstein, L., Gupta, M., Lawrence, M.S., Pugh, T.J., Saksena, G., Meyerson, M., and Getz, G. (2015). Oncotator: cancer variant annotation tool. *Hum. Mutat.* **36**, E2423–E2429.
86. Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501.
87. Feng, J., Ding, C., Qiu, N., Ni, X., Zhan, D., Liu, W., Xia, X., Li, P., Lu, B., Zhao, Q., et al. (2017). Firmiana: towards a one-stop proteomic cloud platform for data processing and analysis. *Nat. Biotechnol.* **35**, 409–412.
88. Wang, J. (2018). multiOmicsViz: plot the effect of one omics data on other omics data along the chromosome. R package. Software.
89. Mermel, C.H., Schumacher, S.E., Hill, B., Meyerson, M.L., Beroukhi, R., and Getz, G. (2011). GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41.
90. Mayakonda, A., Lin, D.C., Assenov, Y., Plass, C., and Koeffler, H.P. (2018). Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* **28**, 1747–1756.
91. Kanehisa, M., and Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* **28**, 27–30.
92. Croft, D., Mundo, A.F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M.R., et al. (2014). The Reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477.
93. Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **26**, 1572–1573.
94. Yoshihara, K., Shahmoradgoli, M., Martínez, E., Vegesna, R., Kim, H., Torres-García, W., Treviño, V., Shen, H., Laird, P.W., Levine, D.A., et al. (2013). Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat. Commun.* **4**, 2612.
95. Hornbeck, P.V., Zhang, B., Murray, B., Kornhauser, J.M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res.* **43**, D512–D520.
96. Horn, H., Schoof, E.M., Kim, J., Robin, X., Miller, M.L., Diella, F., Palma, A., Cesareni, G., Jensen, L.J., and Linding, R. (2014). KinomeXplorer: an integrated platform for kinome biology studies. *Nat. Methods* **11**, 603–604.
97. Casado, P., Alcolea, M.P., Iorio, F., Rodríguez-Prados, J.C., Vanhaesebroeck, B., Saez-Rodriguez, J., Joel, S., and Cutillas, P.R. (2013).

- Phosphoproteomics data classify hematological cancer cell lines according to tumor type and sensitivity to kinase inhibitors. *Genome Biol.* **14**, R37.
98. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* **26**, 1367–1372.
 99. Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21.
 100. Li, B., and Dewey, C.N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinf.* **12**, 323.
 101. Li, L., Jiang, D., Zhang, Q., Liu, H., Xu, F., Guo, C., Qin, Z., Wang, H., Feng, J., Liu, Y., et al. (2023). Integrative proteogenomic characterization of early esophageal cancer. *Nat. Commun.* **14**, 1666.
 102. Li, L., Jiang, D., Liu, H., Guo, C., Zhao, R., Zhang, Q., Xu, C., Qin, Z., Feng, J., Liu, Y., et al. (2023). Comprehensive proteogenomic characterization of early duodenal cancer reveals the carcinogenesis tracks of different subtypes. *Nat. Commun.* **14**, 1751.
 103. Tyanova, S., Temu, T., and Cox, J. (2016). The MaxQuant computational platform for mass spectrometry-based shotgun proteomics. *Nat. Protoc.* **11**, 2301–2319.
 104. Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.
 105. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425.
 106. Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. In *J. Proceedings of the National Academy of Sciences*, pp. 15545–15550.
 107. Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. *Nature* **534**, 55–62.
 108. Clark, D.J., Dhanasekaran, S.M., Petralia, F., Pan, J., Song, X., Hu, Y., da Veiga Leprevost, F., Reva, B., Lih, T.S.M., Chang, H.Y., et al. (2019). Integrated Proteogenomic Characterization of Clear Cell Renal Cell Carcinoma. *Cell* **179**, 964–983.e31.
 109. Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* **462**, 108–112.
 110. Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSEA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf.* **14**, 7.
 111. Hornbeck, P.V., Chabra, I., Kornhauser, J.M., Skrzypek, E., and Zhang, B. (2004). Phosphosite: A bioinformatics resource dedicated to physiological protein phosphorylation. *Proteomics* **4**, 1551–1561.
 112. Dinkel, H., Chica, C., Via, A., Gould, C.M., Jensen, L.J., Gibson, T.J., and Diella, F. (2011). Phospho.ELM: a database of phosphorylation sites-update 2011. *Nucleic Acids Res.* **39**, D261–D267.
 113. Yang, C.Y., Chang, C.H., Yu, Y.L., Lin, T.C.E., Lee, S.A., Yen, C.C., Yang, J.M., Lai, J.M., Hong, Y.R., Tseng, T.L., et al. (2008). PhosphoPOINT: a comprehensive human kinase interactome and phospho-protein database. *Bioinformatics* **24**, I14–I20.
 114. Schwartz, D., and Gygi, S.P. (2005). An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. *Nat. Biotechnol.* **23**, 1391–1398.
 115. Casado, P., Rodríguez-Prados, J.C., Cosulich, S.C., Guichard, S., Vanhaesebroeck, B., Joel, S., and Cutillas, P.R. (2013). Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. *Sci. Signal.* **6**, rs6.
 116. Yoshida, A., Tsuta, K., Wakai, S., Arai, Y., Asamura, H., Shibata, T., Furuta, K., Kohno, T., and Kushima, R. (2014). Immunohistochemical detection of ROS1 is useful for identifying ROS1 rearrangements in lung cancers. *Mod. Pathol.* **27**, 711–720.
 117. Hirsch, F.R., Varella-Garcia, M., Bunn, P.A., Jr., Di Maria, M.V., Veve, R., Bremmes, R.M., Barón, A.E., Zeng, C., and Franklin, W.A. (2003). Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J. Clin. Oncol.* **21**, 3798–3807.
 118. Seckinger, A., Meissner, T., Moreaux, J., Depeweg, D., Hillengass, J., Hose, K., Rème, T., Rösen-Wolff, A., Jauch, A., Schnettler, R., et al. (2012). Clinical and prognostic role of annexin A2 in multiple myeloma. *Blood* **120**, 1087–1094.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Rabbit monoclonal anti-SPATA18	Abcam	Catalog: ab180154; RRID: AB_3076533
Biological samples		
Lung adenocarcinoma tissues (bulk DNA, RNA, protein)	The Department of Thoracic Surgery, Fudan University Shanghai Cancer Center, Shanghai, China	This paper
Chemical, peptides, and recombinant proteins		
Phenylmethylsulfonyl fluoride (PMSF)	AMRESCO	Catalog: M145-5G
(2-carboxyl)-phosphine hydrochloride (TCEP)	ALDRICH	Catalog: 4706-10G
Deoxycholic acid sodium salt (DOC)	Solarbio	Catalog: D8330
2-chloroacetamide (CAA)	ALDRICH	Catalog: 22790-250G-F
1M tris-phosphine hydrochloride (Tris)	AMRESCO	Catalog: 0497
Trypsin	Promega	Catalog: V528A
HPLC-grade water	J.T.Baker	Catalog: 4218-03
Acetonitrile (ACN)	J.T. Baker	Catalog: 9829-03
Formic acid (FA)	Sigma	Catalog: F0507
Methanol	J.T. Baker	Catalog: 9830-03
Phosphatase inhibitor cocktail 3	Sigma	Catalog: P0044
NH ₃ .H ₂ O	Aladdin	Catalog:1336-21-6
C18 resin	3M	Catalog: 2215-C18
SepPark C18 cartridges	This paper	N/A (homemade)
Xbridge C18 column	This paper	N/A (homemade)
Phosphopeptide Enrichment Kit	Thermo	Catalog: A32992
Camptothecin	Sigma-Aldrich	Catalog: C9911
Critical commercial assays		
BCA protein assay kit	Beyotime	Catalog: P0011
DNA Assay kit	Life Technologies	Catalog: Q10212
TRIzol reagent	Invitrogen	Catalog: 15596026
Phosphopeptide Enrichment Kit	Thermo	Catalog: A32992
ClonExpress MultiS One Step Cloning Kit	Vazyme Biotech Co., Ltd	Catalog: C113-02
Deposited data		
Proteomics data	This paper	iProx: IPX0003921002
Phosphoproteomics data	This paper	iProx: IPX0003921003
Genomics data (WES)	This paper	EGA: EGAS00001004006
Transcriptomic data (RNA)	This paper	EGA: EGAS00001004006
Images of hematoxylin-eosin staining	This paper	Zenodo: https://doi.org/10.5281/zenodo.7993327
Software and algorithms		
R (version 3.5.1)	The R Project for Statistical Computing	https://www.r-project.org/
BWA-MEM	Li et al. ⁸¹	http://bio-bwa.sourceforge.net/
Picard	Broad Institute	http://broadinstitute.github.io/picard/
Genome Analysis Toolkit	DePristo et al. ⁸²	https://gatk.broadinstitute.org/hc/en-us
MuTect	Broad Institute	https://github.com/broadinstitute/mutect

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MuTect2	Broad Institute ⁸³	https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2
Strelka (version 2.0.13)	Saunders et al. ⁸⁴	https://github.com/Illumina/strelka
Oncotator (version 1.9.1)	Ramos et al. ⁸⁵	https://software.broadinstitute.org/cancer/cga/ncotator
MutSig2CV	Lawrence et al. ⁸⁶	https://github.com/getzlab/MutSig2CV
Firmiana platform	Feng et al. ⁸⁷	https://phenomics.fudan.edu.cn/firmiana/gardener/
multiomicsViz (version 1.6.0)	Wang ⁸⁸	https://git.bioconductor.org/packages/multiOmicsViz
GISTIC2.0	Mermel et al. ⁸⁹	https://www.genepattern.org/modules/docs/GISTIC_2.0
Maftools	Mayakonda et al. ⁹⁰	https://bioconductor.org/packages/release/bioc/vignettes/maftools/inst/doc/maftools.html
KEGG database	Kanehisa et al. ⁹¹	https://www.phosphosite.org/homeAction
Reactome database	Croft et al. ⁹²	https://reactome.org
ConsensusclusterPlus (version 1.46.0)	Wilkerson et al. ⁹³	N/A
ESTIMATE (version 1.0.11)	Yoshihara et al. ⁹⁴	https://bioinformatics.mdanderson.org/estimate/
xCell	Aran et al. ⁴³	https://xcell.ucsf.edu
PhosphoSitePlus	Hornbeck et al. ⁹⁵	https://www.phosphosite.org/homeAction
NetworkKIN 3.0	Horn et al. ⁹⁶	http://networkkin.info
Kinase-Substrate Enrichment Analysis (KSEA) algorithm	Casado et al. ⁹⁷	N/A
ImageQuant TL	GE Healthcare	https://imagequanttl.updatestar.com/
MaxQuant	Cox et al. ⁹⁸	http://www.coxdocs.org/doku.php?id=maxquant:start

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to the lead contact, Dr. Haiquan Chen (hqchen1@yahoo.com)

Materials availability

This study did not generate new unique reagents.

Data and code availability

Raw data from WES and RNA-seq of AIS/MIA and IAC have been deposited at European Genome-phenome Archive (EGA) under the accession code EGAS00001004006. The mass spectrometry proteomics and phosphoproteomics data have been deposited to the iProX Consortium (<https://www.iprox.org/>) with the subproject ID (IPX0003921002 and IPX0003921003, respectively). The images of HE slides have been uploaded to Zenodo (<https://doi.org/10.5281/zenodo.7993326>).

This paper did not report original codes and algorithms. Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Human samples

The specimens of primary tumor and normal adjacent tissue (NAT) were prospectively collected from patients, who underwent surgical resection without neoadjuvant treatments in the Department of Thoracic Surgery, Fudan University Shanghai Cancer Center (FUSCC). This study included patients, diagnosed with adenocarcinoma *in situ* (AIS), minimally invasive adenocarcinoma (MIA), or lung invasive adenocarcinoma (IAC). Tumor tissues and paired NATs, cut from specimens immediately after surgical resection,

were moved into liquid nitrogen within thirty minutes. NATs were defined as lung tissues at least 3cm away from tumor margin, consistent with previous studies³⁷. A total of 197 patients were enrolled in this study, including 98 with AIS/MIA and 99 with IAC. Patient characteristics, such as age, gender, smoking history, and pathological diagnosis, were also acquired prospectively during sample collection. The study was approved by the Institutional Review Board of FUSCC (IRB#090977-1), and informed consents for sample donation to the tissue bank of FUSCC were provided by patients or their direct relatives.

METHOD DETAILS

Pathological review

All the pathological diagnoses were made by at least two experienced pulmonary pathologists independently according to lung adenocarcinoma classification released by WHO⁷. PCA of the proteome in AIS and MIA revealed no discernible differences between the two subtypes (Figure S1B). No significant difference was observed in tumor mutation burden (TMB) between AIS and MIA (*Wilcoxon signed-rank test*, $p = 0.51$) (Figure S1C). Given the above findings and the excellent survival of AIS and MIA, we decided to classify the AIS and MIA as one group. Tumors were staged using the eighth edition of TNM Classification for Lung Cancer¹⁰. Tumor purity was evaluated by hematoxylin-eosin (HE) staining, and only tumors with purity $\geq 50\%$ were further included. HE staining was performed to confirm the normal status for NATs. Each sample was homogenized and subsequently divided into three parts for different analyses if possible: the first part was used for whole-exome sequencing (WES), the second part was applied to RNA sequencing, and the remaining part was prepared for proteomic profiling and phosphoproteomic profiling. The tumor tissue was prioritized for DNA and RNA sequencing before being utilized for protein analysis.

Whole exome sequencing (WES)

Genomic DNA was isolated from frozen tumors and paired NATs using QIAamp DNA Mini Kit (QIAGEN, German) according to the manufacturer's protocols, as previously reported¹⁶. The SureSelect XT Target Enrichment System (Agilent Technologies, USA) was used to construct exon libraries. A total of 1–3 μg genomic DNA was sheared into fragmentations with the average size of approximately 200 bp. SureSelect XT reagents were applied to capture DNA and generate target-enriched library. Constructed libraries were run on the Illumina HiSeq X Ten platform using paired-end methods. The average mappable depth was 200X for tumor tissues and 100X for paired NATs.

RNA sequencing

Total RNA from tumors and paired NATs was extracted using NucleoZOL (Macherey-Nagel, German) and NucleoSpin RNA Set for NucleoZOL (Macherey-Nagel, German) according to the manufacturer's protocols. RNA contamination and degradation were assessed on 1% agarose gels. NanoVue Spectrophotometer (General Electric, USA) was used to determine the concentration of total RNA. Epicenter Ribo-Zero Gold Kits (Epicenter, USA) were applied to remove ribosomal RNA. Quality of RNA was evaluated using the NanoDrop 8000 and the Agilent Bioanalyzer, and RNA integrity number (RIN) score for each specimen was generated. Sequencing libraries were constructed using the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (NEB, USA) following the manufacturer's instructions. Subsequently, libraries were sequenced on the Illumina HiSeq X Ten platform using paired-end method.

WES alignment and somatic mutation calling

BWA-MEM was applied to align sequencing reads to the reference human genome (hg19)⁸¹. PCR duplicates were marked using the Picard tools (<https://broadinstitute.github.io/picard/>). Base quality was recalibrated and local indel was realigned using the Genome Analysis Toolkit⁸². Single nucleotide variants were called by MuTect and MuTect2⁸³, and indels were called by MuTect2 and Strelka v2.0.13⁸⁴. To get reliable results, variants were excluded if found in only one in silico method. Somatic mutations were annotated using Oncotator v1.9.1⁸⁵. MutSig2CV was used to identify significantly mutated genes⁸⁶.

RNA-seq quantification

STAR v2.5.3 was used to align transcriptomic reads to the reference human genome (hg19)⁹⁹. RNA-level expression was normalized to transcripts per million (TPM) using RSEM v1.3.0¹⁰⁰.

GISTIC and MutSig analysis

The Genomic Identification of Significant Targets in Cancer (GISTIC2.0)⁸⁹ method was used to identify significantly amplified or deleted focal-level and arm-level events, with Q values smaller than 0.25 considered significant. Cox proportional hazard models were also used to estimate (HRs) with 95% confidence intervals (CIs) for univariate and multivariate analyses. We screened the most significant amplification and deletion peaks that were associated with hazard ratios to identified. Finally, we filtered the genes by their CN-protein correlation to keep the genes with significant CN cis-effect (p value < 0.05 , Spearman's correlation).

Protein Extraction and tryptic digestion

According to the assessment of the mass of prepared sample pouch and the corresponding (phospho)peptide mass for the mass spectrometry (MS) analysis, 50 μg (~ getting 2 μg peptide mass for MS analysis) and 100 μg (~ getting 500 ng phosphopeptide mass for MS analysis) mass were used for proteomic profiling and phosphoproteomic profiling, respectively. We previously have

achieved deep proteome and phosphoproteome coverage of the tiny samples in early-stage gastrointestinal cancer^{101,102}. For clinical sample preparation, samples were macro-dissected, deparaffinized with xylene, and washed with ethanol. 50 μ L of TCEP buffer (2% deoxycholic acid sodium salt (Solarbio, Catalog: D8330), 40 mM 2-chloroacetamide (ALDRICH, Catalog: 22790-250G-F), 100 mM tris-phosphine hydrochloride (AMRESCO, Catalog: 0497), 10 mM (2-carboxyl)-phosphine hydrochloride (ALDRICH, Catalog: 4706-10G), 1 mM phenylmethylsulfonyl fluoride (AMRESCO, Catalog: M145-5G) mixed with MS water (J.T. Baker, Catalog: 4218-03), pH 8.8) was added into 1.5 mL EP tubes with prepared samples, and then heated in a 99°C metal bath for 30 min. Cool to room temperature, 5.5 μ g trypsin (Promega, Catalog: V528A) was added into each tube and digested for 18 h in a 37°C incubator. Then, 13 μ L of 10% formic acid (FA) (Sigma, Catalog: F0507) was added into each tube and vortexed for 3 min, followed by sedimentation for 5 min (12,000 g). After that, a new 1.5 mL tube with 350 μ L buffer (0.1% FA in 50% acetonitrile [ACN] (J.T. Baker, Catalog: 9830-03)) was needed to collect the supernatant for extraction (vortex for 3 min, and then 12,000 g sedimentation for 5 min). Then, the supernatant was transferred into a new tube for drying in a 60°C vacuum drier.

First dimensional reversed-phase separation

The dried peptides were redissolved with 0.1%FA and loaded into a homemade Durashell Reverse Phase column [2 mg packing (3 μ m, 150 Å, Agela) in a 200 μ L tip], then eluted sequentially with nine gradient elution buffer that contains mobile phases A (2 % acetonitrile (ACN), adjusted pH to 10.0 using NH₃.H₂O (Aladdin, Catalog:1336-21-6) and 6 %, 9 %, 12 %, 15 %, 18 %, 21 %, 25 %, 30 %, 35 % mobile phase B (98 % ACN, adjusted pH to 10.0 using NH₃.H₂O). The nine fractions then were combined into three groups (6 % + 15 % + 25 %, 9 % + 18 % + 30 %, 12 % + 21 % + 35 %) and dried under vacuum for sub-sequential MS analysis.

Phosphopeptide enrichment

For the phosphoproteomic analysis, peptides were extracted from the specimens after trypsin digestion using the methods described as Protein Extraction and tryptic digestion. The tryptic peptides were then enriched with the High-Select™ Fe-NTA Phosphopeptide Enrichment Kit (Thermo Scientific cat. A32992), following the manufacturer's recommendations. Briefly, peptides were suspended in binding/wash buffer (contained in the enrichment kit), mixed with the equilibrated resins, and incubated at 21–25°C for 30 min. After incubation, the resins were washed thrice with binding/wash buffer and twice with water. The enriched peptides were eluted with elution buffer (contained in the enrichment kit) and dried in a 30°C vacuum drier.

Proteome analysis

For the proteome profiling samples, peptides were analyzed on a Q Exactive HF-X Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher Scientific, Rockford, IL, USA) coupled with a high-performance liquid chromatography system (EASY nLC 1200, Thermo Fisher). Dried peptide samples re-dissolved in Solvent A (0.1 % formic acid in water) were loaded to a 2-cm self-packed trap column (100- μ m inner diameter, 3 μ m ReproSil-Pur C18-AQ beads, Dr Maisch GmbH) using Solvent A and separated on a 150- μ m-inner-diameter column with a length of 15 cm (1.9 μ m ReproSil-Pur C18-AQ beads, Dr Maisch GmbH) over a 75 min gradient (Solvent A: 0.1 % Formic acid in water; Solvent B: 0.1 % Formic acid in 80 % ACN) at a constant flow rate of 600 nL/min (0–75 min, 0 min, 4 % B; 0–10 min, 4–15 % B; 10–60 min, 15–30 % B; 60–69 min, 30–50 % B; 69–70 min, 50–100 % B; 70–75 min, 100 % B). The eluted peptides were ionized under 2 kiloVolts and introduced into mass spectrometry. Mass spectrometry was operated under a data-dependent acquisition mode. For the MS1 Spectra full scan, ions with m/z ranging from 300 to 1,400 were acquired by Orbitrap mass analyzer at a high resolution of 120,000. The automatic gain control (AGC)

Target value was set as 3E+06. The maximal ion injection time was 80 ms. MS2 Spectra acquisition was performed in the ion trap in a rapid speed mode. Precursor ions were selected and fragmented with higher energy collision dissociation (HCD) with normalized collision energy of 27 %. Fragment ions were analyzed by ion trap mass analyzer with AGC target at 5E+04. The maximal ion injection time of MS2 was 20 ms. Peptides that triggered MS/MS scans were dynamically excluded from further MS/MS scans for 12 s.

In our study, sample numbers were randomly assigned and subsequently subjected to mass spectrometry analysis. For the purpose of quality control, the HEK293T cell (National Infrastructure Cell Line Resource) lysates were measured as the standards every 15 samples injection to assess the stability of the mass spectrometry performance. PCA indicated that there was no discernible boundary or clustering pattern between samples injected at different profiling time (Figure S3B) or between samples collected at different time points (Figure S3C).

Phosphoproteome analysis

For the phosphoproteomic samples, peptides were analyzed on an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (Thermo Fisher Scientific, Rockford, IL, USA) equipped with an Easy nLC-1000 (Thermo Fisher Scientific, Rockford, IL, USA) and a Nanoflex source (Thermo Fisher Scientific, Rockford, IL, USA). Dried peptide samples re-dissolved in Solvent A (0.1 % formic acid in water) were loaded to a 2-cm self-packed trap column (100- μ m inner diameter, 3 μ m ReproSil-Pur C18-AQ beads, Dr Maisch GmbH) using Solvent A and separated on a 150- μ m-inner-diameter column with a length of 30 cm (1.9 μ m ReproSil-Pur C18-AQ beads, Dr Maisch GmbH) over a 150 min gradient (buffer A: 0.1 % Formic acid in water; buffer B: 0.1 % Formic acid in 80 % ACN) at a constant flow rate of 600 nL/min (0–150 min, 0min, 4 % B; 0–10 min, 4–15 % B; 10–125 min, 15–30 % B; 125–140 min, 30–50 % B; 140–141 min, 50–100 % B; 141–150 min, 100 % B). The eluted phosphopeptides were ionized and detected by an Orbitrap Fusion mass spectrometry. Mass spectra were acquired over the scan range of m/z 350–1500 at a resolution of 120,000 (AUG target value of 5E+05 and max injection

time 50 ms). For the MS2 scan, the higher-energy collision dissociation fragmentation was performed at a normalized collision energy of 30 %. The MS2 AGC target was set to 1E+04 with a maximum injection time of 10 ms, peptide mode was selected for monoisotopic precursor scan and charge state screening was enabled to reject unassigned 1+, 7+, 8+, and > 8+ ions with a dynamic exclusion time of 45 s to discriminate against previously analyzed ions between ± 10 ppm.

Peptide and protein identification

MS raw files were processed with the “Firmiana” (a one-stop proteomic cloud platform)⁸⁷ against the human RefSeq protein database (updated on 04-07-2013, 32,015 entries) maintained by the National Center for Biotechnology Information (NCBI). The maximum number of missed cleavages was set to 2. Mass tolerances of 20 ppm for precursor and 0.5 Da for production were allowed. The fixed modification was cysteine carbamidomethylation and the variable modifications were N-acetylation and oxidation of methionine. For the quality control of proteins identification, the target-decoy-based strategy was applied to confirm that the FDR (False Discovery Rate) of both peptide and protein were lower than 1 %. The program percolator was used to obtain the probability value (q-value), and showed that the FDR (measured by the decoy hits) of every peptide-spectrum match (PSM) was lower than 1 %. Then all peptides shorter than seven amino acids were removed. The cutoff ion score for peptide identification was 20. All of the PSMs in all fractions were combined for protein quality control, which was a stringent quality control strategy. The q-values of both target and decoy peptide sequences were dynamically increased until the corresponding protein FDR was less than 1 % employing the parsimony principle. Finally, to reduce the false positive rate, the proteins with at least one unique peptide were selected for further investigation.

Label-free-based MS quantification of proteins

The one-stop proteomic cloud platform “Firmiana” was further employed for protein quantification. Identification results and the raw data from mzXML file was loaded. Then for each identified peptide, the XIC (extracted-ion chromatogram) was extracted by searching against the MS1 based on its identification information, and the abundance was estimated by calculating the area under the extracted XIC curve. For protein abundance calculation, the nonredundant peptide list was used to assemble proteins following the parsimony principle. Then, the protein abundance was estimated with a traditional label-free, intensity-based absolute quantification (iBAQ) algorithm, which divided the protein abundance (derived from identified peptides’ intensities) by the number of theoretically observable peptides. Then the fraction of total (FOT), a relative quantification value which was defined as a protein’s iBAQ divided by the total iBAQ of all identified proteins in one experiment, was calculated as the normalized abundance of a particular protein among experiments. Finally, the FOT was further multiplied by 1E6 for the ease of presentation and FOTs less than 1E6 were replaced with 1E6 to adjust extremely small values. Match-between-runs (MBR) is applied where peptides identified by tandem mass spectra in one run are transferred to another by inference based on m/z, charge state, retention time, and ion mobility when applicable, was applied to mitigate the missing value problem¹⁰³.

Differential protein analysis

The proteomic data filtered with no missing values ($n = 12293$ genes) in all of samples ($n = 260$) were used as input data for differential expression analysis. A non-paired Mann-Whitney U test was performed on overlapping samples to determine differential abundance of proteins between AIS/MIA, IAC tumor and their matched adjacent tissues, respectively. Upregulated or downregulated proteins in tumors were defined as proteins differentially expressed in tumors compared with their matched adjacent tissues ($T/N > 2$ or $< 1/2$, Mann-Whitney U test, Benjamini-Hochberg adjusted $p < 0.05$). *Wilcoxon signed-rank test* was used to examine whether proteins were differentially expressed between *EGFR*^{Mut} AIS/MIA patients and *EGFR*^{Mut} IAC patients ($T/N > 1.5$ or < 0.67 , Benjamini-Hochberg adjusted $p < 0.05$).

Pathway enrichment analysis

Differentially expressed proteins identified in AIS/MIA, IAC tumor and their matched adjacent were subjected to Gene Ontology and KEGG pathway enrichment analysis in DAVID¹⁰⁴ with a p value/FDR < 0.05 . We used gene sets of molecular pathways from the KEGG⁹¹ /Hallmark¹⁰⁵ /Reactome⁹² databases to compute pathways for both tumor and non-tumor.

Gene Set Enrichment Analysis (GSEA)

GSEA was performed by the GSEA software (<http://software.broadinstitute.org/gsea/index.jsp>)¹⁰⁶. Gene sets including KEGG, Reactome and HALLMARK downloaded from the Molecular Signatures Database (MSigDB v7.1, <http://software.broadinstitute.org/gsea/msigdb/index.jsp>) were set as background.

Principal component analysis

We performed PCA on 59 AIS/MIA tumor samples, 75 IAC samples and 126 normal adjacent samples to illustrate the global proteomic difference between tumor and NAT samples. The PCA function under the FactoMineR R package (version 2.4) was implemented for unsupervised clustering analysis with the parameter ‘n_components = 3’ on the expression matrix of global proteomic data containing 12,293 proteins. The 95 % confidence coverage was represented by a colored ellipse for each group, which was calculated based on the mean and covariance of points in each specific group.

The transcriptomic difference between AIS/MIA and IAC samples were also illustrated by PCA on 68 AIS/MIA tumor samples, 82 IAC samples. The PCA function under the FactoMineR R package (version 2.4) was implemented for unsupervised clustering analysis

with the parameter 'n_components = 2' on the expression matrix of global proteomic data containing 20,208 proteins. The 95 % confidence coverage was represented by a colored ellipse for each group, which was calculated based on the mean and covariance of points in each specific group.

Correlation between subtype and clinical features

In the measurements of correlations between clinical subtypes, immune clusters and clinical features, Fisher's exact test was performed on variables: BMI, age, gender, progression stage, status of mutation, peak events, and immune clusters.

mRNA-protein correlation in tumors and normal samples

Genes, which were detected in more than 30% of tumors or adjacent in both RNA-Seq and MS data, were enrolled to evaluate mRNA-protein correlation. *Pearson* correlation coefficient between paired mRNA expression and protein abundance was measured. In addition, *p* value corresponding to the correlation coefficient was calculated and adjusted by the FDR correction. Functional pathways were enriched by GSEA enrichment analysis using the correlation-ranked list of proteins.

Immune deconvolution

Tumor purity, immune, stroma scores were inferred by R package ESTIMATE v1.0.11¹⁷ using proteome/mRNA data. Though ESTIMATE algorithm was a tool designed to analyze transcriptome data, some studies have used it for proteome analysis^{107,108}. To evaluate the immune characterization of the lung cancer samples, the abundances of 64 different cell types for 196 paired Lung cancer (Tumor /NATs) samples were computed via xCell (<https://xcell.ucsf.edu/>)⁴³, in which 126 paired and 186 paired lung cancer samples were conducted for proteomic and transcriptomic data, respectively. Based on these 64 cell signatures, consensus clustering was performed to identify subtypes of the samples with the same immune/stromal characteristics, and then all the samples were profiled into 4 immune clusters at the RNA and protein levels. Each cell signature and immune/stromal/environmental score were normalized to Z-score and then modeled as function *a*. In addition, we performed the differential analysis of the identified CD/MHC molecules based on global transcriptomic and proteomic data. In addition, pathway scores were computed using ssGSEA¹⁰⁹ via Gene Set Variation Analysis (GSVA) package¹¹⁰.

Immune consensus clustering analysis

The protein expression matrix of the 186 and 126 paired lung cancer samples were used to identify the transcriptomic and proteomic subtypes, respectively, using the consensus cluster method. Consensus clustering was performed using the ConsensusClusterPlus R package⁹³ with the 64 cell signatures from xCell (<https://xcell.ucsf.edu/>)⁴³. Consensus Cluster Plus parameters were reps=1,000, pitem = 0.8, pFeature = 1, clusterAlg = "hc", distance = "euclidean", plot="PDF". Euclidean distance and 1,000 resampling repetitions in the range of 2 to 10 clusters. As a result, the consensus matrix of *k* = 4 showed clear separation among clusters. The empirical cumulative distribution function (CDF) plot initially showed optimal separation. Clustering by *k* = 4 had the lowest proportion of ambiguous clustering (PAC) and showed significant association with the patients' survival. A consensus matrix with *k* = 4 appeared to have the clearest cut between clusters and showed significant association with the patients' survival. Taken together, proteomic and transcriptomic clusters were defined using *k*-means consensus clustering with *k* = 4. As summarized in Figure S6A, the clustering analysis of the tumors (vertical column) by protein/TPM abundance (horizontal rows) divided all tumors into 4 proteomic/transcriptomic subgroups defined by silhouette analyses. A consensus matrix with *k* = 4 appeared to have the clearest cut between clusters and showed significant association with the patients' survival.

Database searching of MS phosphoproteomic data

MS phosphoproteomic raw data were processed with firmiana platform against the human RefSeq protein database (updated on 04-07-2013, 32,015 entries) in the National Center for Biotechnology Information (NCBI).

Kinase activity prediction and phosphopeptide analysis

The phospho-proteome data were searched against the same database with MaxQuant. The phosphorylation of S or T or Y was set as variable modification, in which three mis-cleavages were allowed, with a minimum Andromeda score of 40 for spectra matches. The ratios of identified phosphorylation sites of all samples were used to estimate the kinase activities by Kinase-Substrate Enrichment Analysis (KSEA) algorithm⁹⁷. The information of kinase-substrate relationships was obtained from publicly available databases, including PhosphoSite¹¹¹, Phospho.ELM¹¹² and PhosphoPOINT¹¹³. The information of substrate motifs was obtained either from the literature¹¹⁴ or from an analysis of the KSEA dataset with Motif (SP)¹¹⁵. The kinase-substrate-motif network analysis was referenced from PhosphoSitePlus (PSP, <https://www.phosphosite.org/homeAction>)⁹⁵ and NetworkKIN 3.0⁹⁶. Statistical analysis was performed in R (version 3.5.1) with Kruskal-Wallis test.

Immunohistochemistry

Distinct protein expressions of SPATA18 in AIS/MIA and IAC were validated through immunohistochemistry (IHC) of SPATA18. FPPE tumor tissues were sliced into 4-μm slices, which were stained with SPATA18 antibody (dilution 1:100, Abcam, Catalog: ab180154)

using Anti-mouse/rabbit Immunohistochemistry Detection Kit (Proteintech, USA, Cat No. PK10006) according the manufacture's protocol. To explore survival effect of SPATA18 expression, IHC on tissue microarrays were also conducted.

For the assessment of immunostaining intensity, a semi-quantitative method, HistoScore (H-score, 0-300) was calculated by multiplying the intensity of positive staining (no staining = 0, weak staining = 1, moderate staining = 2, and strong staining = 3) by the proportion of positive cells (0-100%)^{64,116,117}. Two independent pathologists, who were blinded to the pathological results and survival outcomes, scored H-score. For cases with conflicting results, consensus were reached through discussion.

In vitro and in vivo experiments

The A549 human LUAD cell line was purchased from the the Cell Bank of the Shanghai Institutes of Biological Sciences, Chinese Academy of Sciences. The A549 cells were cultured in 1640 medium (Thermo Fisher Scientific, USA) containing 10% fetal bovine serum (Gibco, USA) and 100 units penicillin–streptomycin in a humidified atmosphere with 5% CO₂ at 37°C.

Construction of plasmids and transfection

PLL3.7 plasmids encoding SPATA18-specific shRNAs [GCAACAGTTAGACTCTAAT (shSPATA18-1) and GGAACCTCACTCAAGCAGAA (shSPATA18-2)] or a control scrambled shRNA (shNC) were developed. Cells were transfected with the plasmids using Lipofectamine 3000 (Thermo Fisher Scientific, Tokyo, Japan), in accordance with the manufacturer's instructions. Cells stably expressing the SPATA18 shRNA or control shRNA were selected using puromycin.

Cell viability assay

The cell viability assays were performed with a Cell Counting Kit-8 Assay (DOJINDO Laboratories, Japan). The LUAD cells were seeded into 96-well plates (1 × 10³ cells per well) in 100 μL medium incubated at 37 °C, 5%CO₂ in humidified incubator. After the indicated number of days, the supernatants were removed. 90 μL medium and 10 μL CCK-8 were added to each well, and the plates were then incubated for 1 h. The absorbance at 450 nm was detected with a microplate reader (BioTek, USA).

Wound-healing assay

Cells were cultured until confluence and then wounded using a 200μL pipette tip. Cells were treated with Oncostatin M (R&D systems) at 100-200ng/mL every 48h during 4 days. Three wounds were made for each sample, and migration distance was photographed and measured at time 0h and 24h.

The 24-well transwell plates (BD, USA) with 8 μm pore filters were used for measuring cell migration and invasion. According to the manufactures' instructions, 1 × 10⁶ LUAD cells per well were seeded in the serum-free medium into upper insert, with (invasion assay) or without (migration assay) Matrigel Matrix (BD, USA), and the subjacent compartment was incubated with medium, with 10% FBS at 37 °C in an incubator. After incubation for 24 h, LUAD cells migrating to the lower surface were stained with 1.0% crystal violet and photographed.

The LUAD cells were collected and fixed using 75% pre-cooled ethanol at 4 °C overnight to perform cell cycle analysis. After being washed three times with phosphate buffered saline (PBS) and resuspended with 500 μL PI staining solution (Vazyme Biotech, China) at room temperature for 30 min. The cell cycle analysis was detected via a flow cytometry (FACS LSRII, BD Bioscience, USA).

To confocal Confocal laser microscopy of mitochondrial markers, cells (5 × 10⁵ /sample) were seeded in 35-mm dishes and then stained with 1 μM LysoSensor Green DND-189 (L7535; Invitrogen) to detect the pH of lysosomes; 50 nM MitoTracker Red FM (M22425; Invitrogen) to stain active mitochondria and measure mitochondria membrane potential. Images were acquired using an LSM-880 confocal laser microscope (Carl Zeiss).

Animal studies were performed in accordance with the guidelines approved by the Fudan University Shanghai Cancer Center (FUSCC-IACUC-S2023-0021). All mice were maintained in a standard cage on a 12-h light-dark cycle, in a temperature-controlled room and were allowed access to a standard chow diet and water. A549 cells with SPATA18 ablation were trypsinized and resuspended in sterile 1 × phosphate-buffered saline (PBS). Cells were counted and diluted with 1 × PBS to a concentration of 1 × 10⁷ cells/mL, and then 100 μL containing 1 × 10⁶ total cells were subcutaneously injected to BALB/c nude mice. Tumors were measured for the longest and shortest diameters.

QUANTIFICATION AND STATISTICAL ANALYSIS

Wilcoxon signed-rank test was used to as a nonparametric test to test the difference of variables with nonnormal distribution. Benjamini-Hochberg method was used to adjust p values for multiple testing. Kaplan-Meier survival curves (log-rank test) were used for overall survival or recurrence-free survival (RFS) of the molecular abundance and patients. Coefficient value, which equals to ln (HR), was calculated from Cox proportional hazards regression analysis. *P*-values less than 0.05 were considered as significantly different and were chosen to enter Cox regression multivariate analysis. Prior to the log-rank test of a given protein, phosphoprotein, or phosphosite, survminer R package (version:0.4.9) with maxstat (maximally selected rank statistics) (<http://r-addict.com/2016/11/21/Optimal-Cutpoint-maxstat.html>) was used to determine the optimal cut-off point for the selected samples according to the previous study¹¹⁸. RFS or OS curves were then calculated (Kaplan-Meier analysis, log-rank test) based on the optimal cut-off point. Proteins associated with survival outcomes were also presented (Tables S1B and S1C).