

Proteogenomic characterization reveals tumorigenesis and progression of lung cancer manifested as subsolid nodules

Received: 31 May 2024

Accepted: 20 February 2025

Published online: 11 March 2025

 Check for updates

Hang Su^{1,9}, Li Chen^{2,3,9}, Jun Wu^{4,9}, Zhongyi Cheng^{1,5,9}, Jing Li⁶, Yijiu Ren¹, Junfang Xu^{2,3}, Yifang Dang^{2,3}, Mengge Zheng^{2,3}, Yajuan Cao^{2,3}, Jiani Gao¹, Chenyang Dai¹, Xuefei Hu¹, Huikang Xie⁷, Jianxia Chen³, Tao Luo⁵, Jun Zhu⁵, Chunyan Wu⁷✉, Wei Sha⁸✉, Chang Chen¹✉ & Haipeng Liu^{2,3}✉

Lung adenocarcinoma (LUAD) radiologically displayed as subsolid nodules (SSNs) is prevalent. Nevertheless, the precise clinical management of SSNs necessitates a profound understanding of their tumorigenesis and progression. Here, we analyze 66 LUAD displayed as SSNs covering 3 histological stages including adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA) and invasive adenocarcinoma (IAC) by incorporating genomics, proteomics, phosphoproteomics and glycoproteomics. Intriguingly, cholesterol metabolism is aberrantly regulated in the preneoplastic AIS stage. Importantly, target ablation of proprotein convertase subtilisin/kexin type 9 (PCSK9) promotes the initiation of LUAD. Furthermore, **sustained endoplasmic reticulum stress is demonstrated to be a hallmark and a reliable biomarker of AIS progression to IAC.** Consistently, target promotion of ER stress profoundly retards LUAD progression. Our study provides comprehensive proteogenomic landscape of SSNs, sheds lights on the tumorigenesis and progression of SSNs and suggests preventive and therapeutic strategies for LUAD.

The utilization of low-dose computed tomography (LDCT) screening has significantly boosted the identification of early-stage lung cancer presenting as radiological subsolid nodules (SSNs)^{1,2}. It is widely acknowledged that lung cancer appearing as SSNs represents a distinct subtype characterized by an indolent nature and a more favorable prognosis compared to radiologically solid lung cancer^{3–8}. The management of pulmonary SSNs is often a dilemma in clinical decision making, due to the SSNs exhibit significant heterogeneous growth

trends. SSNs that demonstrate a tendency for rapid progression necessitate surgical excision, while those exhibiting stability or slow growth warrant regular CT surveillance. Therefore, it is of utmost importance to differentiate between stable SSNs with those high potential progression to facilitate precise clinical management.

Those SSNs are adenocarcinoma in situ (AIS), minimally invasive adenocarcinoma (MIA), and sometimes invasive adenocarcinoma (IAC)^{2,9,10}. AIS is defined as a ≤ 3 cm adenocarcinoma lacking invasion,

¹Department of Thoracic Surgery, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China. ²Central Laboratory, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China. ³Shanghai Key Laboratory of Tuberculosis, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China. ⁴Center for Bioinformatics and Computational Biology, and the Institute of Biomedical Sciences, School of Life Sciences, East China Normal University, Shanghai 200241, China. ⁵Jingjie PTM BioLab (Hangzhou). Co. Inc, Hangzhou 310000, China. ⁶Department of Urology, Changhai Hospital, Second Military Medical University, Shanghai 200433, China. ⁷Department of Pathology, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China. ⁸Department of tuberculosis, Shanghai Pulmonary Hospital, Tongji University School of Medicine, Shanghai 200433, China. ⁹These authors contributed equally: Hang Su, Li Chen, Jun Wu, Zhongyi Cheng.

✉ e-mail: wuchunyan581@sina.com; shfksw@tongji.edu.cn; changchenc@tongji.edu.cn; haipengliu@tongji.edu.cn

while MIA is a ≤ 3 cm adenocarcinoma with ≤ 5 mm invasion¹¹. It has been proposed that AIS, the preneoplasia of invasive lung cancer, may progress to MIA and eventually IAC. Some studies have depicted the genomic, immune and metabolic landscape of these SSNs and revealed the molecular events underlying the initiation and progression trajectories of lung cancer manifests as SSNs^{12–16}; however, thorough understanding of proteomics and post-translational modification (PTMs) characteristics from preneoplasia to invasive lung cancer is still undetermined. In tumorigenesis and cancer progression, alterations in protein abundance or function play a pivotal role, impacting key biological processes such as cell proliferation, metabolism, immune response, and mechanical stresses¹⁷. Notably, recent research highlights the critical role of glycosylation in cancer advancement, influencing cell-cell adhesion, growth, ligand-receptor interactions, and acting as essential biomarkers, offering specific targets for therapeutic interventions^{18–23}.

In this work, we perform an integrated multi-omics analyses using proteomics, phosphoproteomics, glycoproteomics, and genomics data collect from 66 resected lung cancer and their paired normal adjacent tissues (NATs), delineate the pivotal molecular events driving lung cancer initiation and progression and provide valuable biological and clinical insights. The findings present a detailed proteogenomic landscape of lung cancer with a focus on SSNs, offering testable hypotheses pertaining to lung cancer initiation and identifying potential therapeutic vulnerabilities for further exploration.

Results

Proteogenomic landscape of LUAD manifested as SSNs

We investigated the proteogenomic landscape of 66 lung adenocarcinomas (LUAD) radiologically displaying as SSNs and 66 paired NATs. These SSNs encompassed consecutive histological stages consisting of 17 AIS, 24 MIA, and 25 IAC, following the histopathologic classification of LUAD, as reported by the International Association for the Study of Lung Cancer (IASLC)¹¹. Representative images of radiology and pathology for LUAD manifested as SSNs were shown in Fig. 1A and Supplementary Fig. 1A, B. Clinicopathological characteristics of patients and tumors were presented in Supplementary Table 1. The consolidation to tumor ratio (CTR)-the ratio of the maximum solid component size to the whole tumor size on CT scans-was employed in clinical assessments (Supplementary Fig. 1A). A lower CTR corresponds with less aggressive phenotype²⁴. Progressing from AIS to MIA and IAC, there was a stepwise increase in whole tumor size, solid component size and CTR (Supplementary Fig. 1C). After validation of LUAD histopathology by two senior pathologists, portions of cryopulverized tissue underwent proteomic, phosphoproteomic, and glycoproteomic characterization (Fig. 1B and Supplementary Fig. 1D). Whole exome sequencing (WES) was conducted on formalin-fixed paraffin-embedded (FFPE) samples (Supplementary Fig. 1D and Supplementary Data 1). To minimize artifacts and false positives, additional processing was employed for mutation calling. We also included other two variant-calling pipelines (Sanger and DKFZ), which were used as the core algorithms in the PCAWG projects²⁵.

In our cohort, proteomics, phosphoproteomics, and glycoproteomics analyses identified and quantified a total of 10,255 proteins, 27,283 phosphosites and 12,480 glycopeptides (Supplementary Data 2). On average, the LUAD proteome had ~6700 proteins per sample, ranging from a minimum of ~6000 in NATs to a maximum of ~7300 in tumors (Fig. 1C). Furthermore, a total of ~7500 phosphosites were identified per sample with a confident site localization score (probability, >0.75) (Fig. 1D), and ~3,800 glycopeptides were identified per sample with a confident site localization score (probability, >0.75) (Fig. 1E). Excellent reproducibility and data quality were maintained across the entire dataset (Supplementary Fig. 1E). Moreover, a significant heterogeneity in the glycans found on the examined glycoproteins was noted, with approximately 28% of glycoproteins having

multiple glycosites and 47% of glycosites being observed to undergo modification by multiple glycans (Fig. 1F–H).

The landscapes of somatic mutations, deletions, and insertions were depicted in Fig. 1I. There were some similarities and inconsistency between our genomic profiling of SSNs and recent large-scale genomic analyses of LUAD. The frequencies of epidermal growth factor receptor (EGFR), tumor protein 53 (TP53), RNA binding motif protein 10 (RMB10), and mucin 16 (MUC16) mutation in our study were 51%, 19%, 17% and 17%, respectively, in accordance with results from recent large-scale genomic analyses of LUAD (Yang Zhang's study²⁶, Chao Zhang's study²⁷ and Xin Hu's study¹³). It is worth noting that the frequency of EGFR mutation among Asian LUAD populations was about 50–60%^{28,29} contrasting with 15–20% in white populations^{30,31}. Moreover, the frequency of titin (TTN) mutation in our study (29%) was significantly higher than the 9% in Zhang Yang's study²⁶, which is likely due to the relatively high proportion of pre-invasive lesions in our study population (62.1%). The total mutational burden (TMB) was observable across all three histologic stages, exhibiting a tendency for greater enrichment in later stages (Fig. 1J), albeit not reaching statistical significance, likely due to the low TMB in the LUAD displaying as SSNs. This finding aligns with a prior study by Xin Hu, et al.¹³.

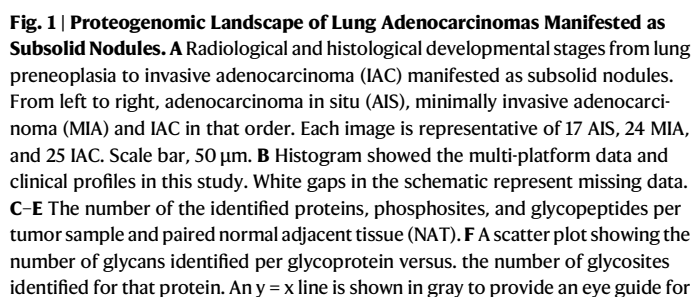
Connecting driver mutations to proteome, phosphoproteome, glycoproteome landscape

We examined selected mutated genes that affects expression of either the cognate gene product (cis effects) or other gene products (trans effects), specifically of a defined set of cancer-related genes (CAGs)³². Intriguingly, we identified 10 genes with significant (FDR < 0.05) cis or trans effects in protein, phosphoprotein or glycoprotein data. EGFR mutations reduced cognate protein, whereas elevated cognate glycoprotein abundance (Figs. 2A, C). EGFR mutant tumors showed remarkably tyrosine phosphorylation of PTPN11/SHP2 at Y62, but no effect was observed at the protein or glycoprotein levels. Intriguingly, we observed a higher level of Y62 phosphorylation of SHP2 in the EGFR mutant cell line (PC-9) compared to the EGFR WT cell lines (A549, NCI-H226, and NCI-H1730) (Fig. 2D), indicating a link between EGFR mutation and Y62 phosphorylation of SHP2. The phosphorylation of SHP2 at Y542 is thought to relieve basal inhibition and stimulate SHP-2 tyrosine phosphatase activity³³. We also observed a similar phosphorylation pattern for SHP2 at Y542 as that observed for SHP2 at Y62 (Fig. 2D), which aligns with a previous study demonstrating that phosphorylation of SHP2 at Y62 stabilizes SHP2 in an open, active conformation³⁴.

Expression of sushi domain containing 2 (SUSD2) has been reported in a variety of human cancers, such as breast³⁵, ovarian³⁶, and lung cancer³⁷. In this study, EGFR regulates the glycosylation of SUSD2 at N494, and EGFR mutant tumors showed significantly reduced glycosylation at N494 (Fig. 2C), although the exact function and mechanism remain unclear. **Additional analysis indicated that EGFR mutation correlates with increased PTPN11 phosphorylation and decreased SUSD2 glycosylation transitioning from AIS to invasive LUAD (Fig. 2E).** Our findings offer insights and directions for future investigation into glycosylation and phosphorylation patterns as potential targets to impede the progression from AIS to invasive LUAD.

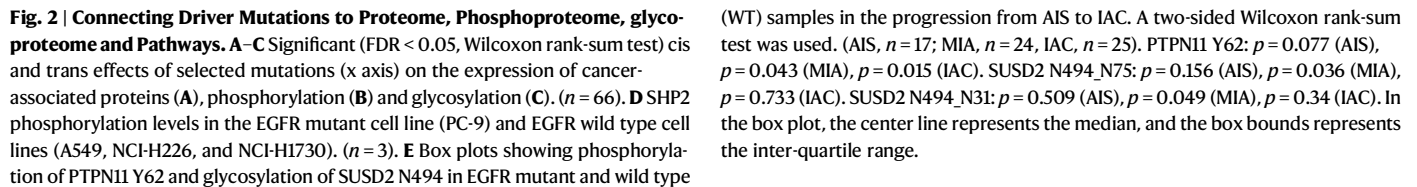
Multi-omics analysis of LUAD manifested as SSNs for therapeutic intervention and early detection

To investigate the intrinsic structure of the proteogenomics data, non-negative matrix factorization (NMF)-based unsupervised clustering was performed on protein, phosphosite and glycopeptide, collectively as multi-omics clustering. All the patients were stratifying into three clusters at multi-omics. 27 patients were identified as cluster 1, 18 patients with cluster 2, 21 patients with cluster 3, the samples of the clusters were significantly associated with distinctive clinical and molecular features (Fig. 3A). Cluster 1, aligned with ECM receptor

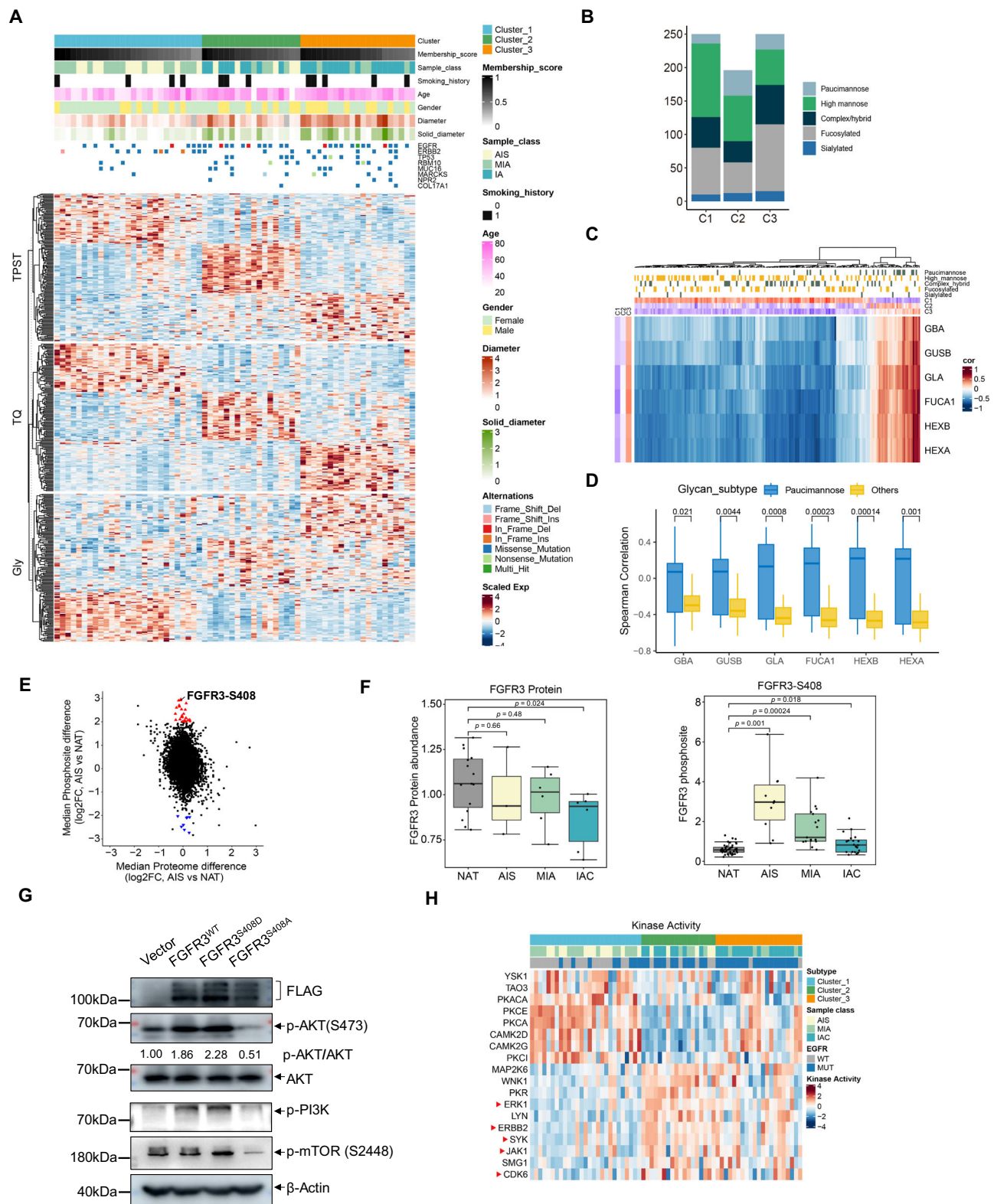


interaction, leukocyte transendothelial migration, renin-angiotensin system and cell adhesion molecules, was enriched for female patients and smaller tumor diameter. Cluster 2, aligned with ribosome, spliceosome, lysosome, and protein digestion and absorption. Cluster 3, aligned with apoptosis, autophagy, ferroptosis, glycosis/gluconeogenesis, amino acid metabolism and pyruvate metabolism, which displaying a stress-adaptive self-eating process, was enriched for male patients and larger tumor diameter. Our findings were well supported

by previous studies which have delineated various multi-omics clusters in LUAD (Supplementary Figs. 2A and B). Cluster 1 represents the pre-invasive stage, aligning with the S-I subtype (AIS/MIA) identified in Yang Zhang's study²⁶. Additionally, our cluster 1 corresponded with the S-I subtype emphasizing high environmental and metabolic factors in Jun-Yu Xu's study³⁸. Moreover, Cluster 1 in our investigation similarly aligned with the C4 subtype (terminal respiratory unit) identified in Michael A. Gillette's study, characterized by EGFR mutations and a



Through glycoproteomic analysis, 1627 intact glycopeptides were identified as differentially expressed in tumors compared to NATs (Supplementary Fig. 2G). NAPSA is one of the most widely used immunomarkers for the differential diagnosis of LUAD and other



pathological types of lung cancer⁴⁴. Notably, the glycopeptides of NAPSA featuring paucimannose and high mannose glycans exhibited a more pronounced distinction between tumors and NATs than the protein expression of NAPSA (Supplementary Fig. 2H). This suggests that monitoring NAPSA glycosylation levels rather than protein abundance could potentially augment the sensitivity of LUAD diagnosis.

Comparison of differentially expressed proteins and phosphorylation sites in AIS and NATs revealed distinct expression patterns of certain phosphopeptides compared to their corresponding protein levels (Fig. 3E). These altered phosphopeptides represent potential therapeutic targets (Fig. 3E, Supplementary Data 4). For instance, FGFR3 exhibited a noteworthy increase in phosphorylation at S408 with no detectable impact on protein level (Fig. 3E and F), between the

Fig. 3 | Multi-omics Analysis of Lung Adenocarcinomas for Therapeutic Intervention and Early Detection. **A** Integrative classification of tumors into three non-negative matrix factorization-derived clusters (multi-omics cluster 1 [C1] to cluster 3 [C3]). Within each cluster, tumors are sorted by cluster membership scores, decreasing from left to right. The heatmap shows the top 50 differential proteins, phosphoproteins, and glycosylated proteins for each multi-omics cluster. **B** Selection of the top 250 upregulated intact glycopeptides in each cluster ($p < 0.05$) for the analysis of glycan distribution within tumor clusters. If the number falls below 250, all qualifying features were included (Fisher's exact test). ($n = 66$). **C** The hierarchical-clustered correlation matrix of intact glycopeptides and glycosylation enzymes. The glycan types and relative abundance of intact glycopeptides among three multi-omics tumor clusters were highlighted in the top rows. The relative abundance of protein expression of selected glycohydrolases among the tumor clusters were represented in the left. **D** Correlation between six selected glycohydrolases and intact glycopeptides with/without paucimannose glycans (paucimannose and others). A two-sided Wilcoxon rank-sum test was used. Sixty-six tumor samples and paired normal adjacent tissues (NATs) were used in the analysis.

$p = 0.021$ (GBA), $p = 0.0044$ (GUSB), $p = 0.0008$ (GLA), $p = 0.00023$ (FUCA1), $p = 0.00014$ (HEXB), $p = 0.001$ (HEXA). **E** Median phosphosite fold change compared to the protein fold change in adenocarcinoma in situ (AIS) tumor compared to normal adjacent tissues (NATs). **F** The abundance changes of global protein expression (left panel) and phosphosite of fibroblast growth factor receptor 3 (FGFR 3) in NATs, AIS, minimally invasive adenocarcinoma (MIA) and invasive adenocarcinoma (IAC). A two-sided Wilcoxon rank-sum test was used. (AIS, $n = 3$; MIA, $n = 5$; IAC, $n = 6$; NATs, $n = 14$). FGFR 3: $p = 0.66$ (AIS), $p = 0.48$ (MIA), $p = 0.024$ (IAC). FGFR 3: $p = 0.66$ (AIS), $p = 0.48$ (MIA), $p = 0.024$ (IAC). FGFR3-S408D: $p = 0.001$ (AIS), $p = 0.00024$ (MIA), $p = 0.018$ (IAC). **G** The effects of FGFR3-WT, FGFR3-S408D, and FGFR3-S408A on AKT, PI3K, and mTOR phosphorylation levels in PC-9 cells. The numbers below the gel lanes represent the pAKT/AKT protein level. ($n = 3$). **H** Heatmaps showing the kinase activity score of selected kinases. Food and Drug Administration (FDA)-approved drug targets were indicated with a red arrow. In the box plot **D** and **F**, the center line represents the median, and the box bounds represents the inter-quartile range.

AIS and NATs. Additionally, levels of FGFR3_S408 were elevated in AIS patients relative to MIA and IAC (Fig. 3F). FGFR3 encodes a transmembrane receptor tyrosine kinase with six autophosphorylation sites on the tyrosine residues⁴⁵. Intriguingly, our phosphopeptide-enriched dataset revealed a phosphorylation site at S408, distinct from the known autophosphorylation sites of FGFR3. To assess the influence of FGFR3 S408 phosphorylation in activating the PI3K-AKT-mTOR pathway, we overexpressed wild type FGFR3 (FGFR3 WT), a phosphorylation-mimicking FGFR3 mutant (S408D), and a phosphorylation-deficient FGFR3 mutant (S408A) in PC-9 cells. Notably, both FGFR3 WT and FGFR3 S408D markedly enhanced AKT phosphorylation at S473, while the FGFR3 S408A mutation exerted a dominant negative effect on phosphorylation of AKT (Fig. 3G). Consistently, we observed a similar pattern of phosphorylation for PI3K and mTOR as seen with phosphorylation of AKT at S473 (Fig. 3G), indicating that the phosphosite S408 of FGFR3 is functional in activating PI3K-AKT-mTOR pathway. Additionally, overexpression of both FGFR3 WT and FGFR3 S408D notably enhanced proliferation in PC-9 cells, whereas transfection with FGFR3 S408A moderately reduced proliferation (Supplementary Fig. 2I). However, the mutation FGFR3 at S408 did not significantly affect the invasion of PC-9 cells (Supplementary Fig. 2J, K).

In our pursuit of uncovering alternative therapeutic targets for the progression of LUAD, we standardized kinase activities by evaluating the abundance of their phosphorylation substrates. We identified five kinases (ERK1, ERBB2, SYK, JAK1, and CDK6) known to be associated with FDA-approved inhibitors, showcasing heightened activity in cluster 2 and cluster 3 in contrast to cluster 1 (Fig. 3H). Using stringent cutoffs for quantitative difference, significance, and consistency ($\log_2$ FC > 1.5 , FDR < 0.05 , and differential in $\geq 50\%$ of all tumor-NAT pairs), we identified 69, 48 and 46 proteins up-regulated of AIS, MIA and IAC, respectively (Supplementary Data 5). Furthermore, up to 40 differentially regulated proteins were illustrated in Supplementary Fig. 3A–C. Paraoxonase 2 (PON2)⁴⁶, PON3⁴⁷, SLC34A2⁴⁸, Mucin 1 (MUC1)⁴⁹, Golgi membrane protein 1 (GOLM1)⁵⁰, crystallin mu (CRYM)⁵¹ were collectively observed as upregulated proteins in AIS, MIA, and IAC, all previously associated with roles in tumorigenesis according to these studies.

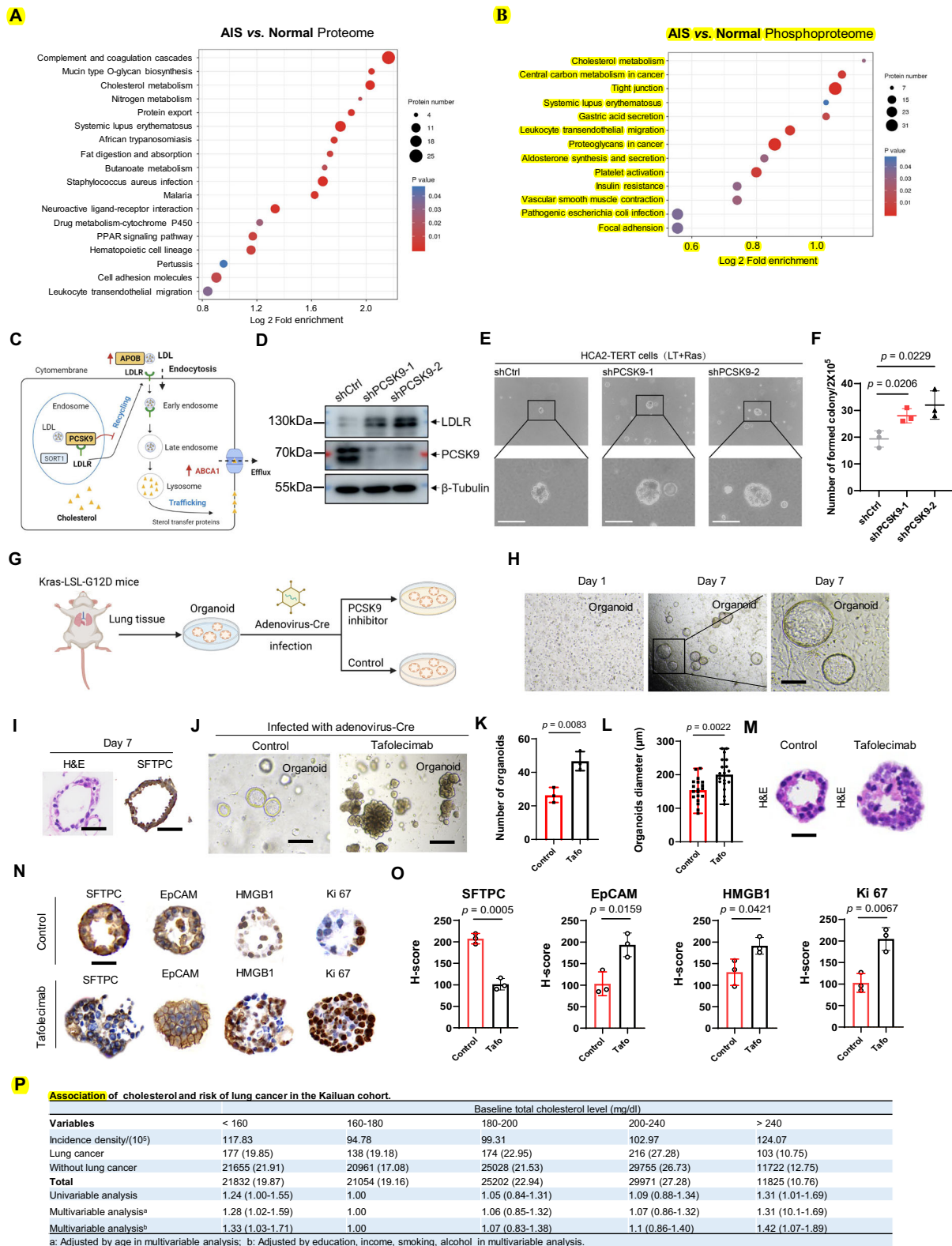
Moreover, we screened up-regulated secretory and soluble proteins from tumors as potential biomarkers for LUAD screening. Totally, five soluble proteins-MUC1, AGR2, SFN, CD38, and CD5-were identified as potential biomarkers at the proteomic level (Supplementary Fig. 3D). CA15-3, a shed and soluble subunit of MUC1, was targeted for measuring MUC1 levels using an analytic assay⁵². Initially, we assessed the concentrations of these candidate biomarkers in the blood of 20 LUAD patients and 10 patients with benign disease using ELISA kits. We observed that CA15-3 showed promise as a diagnostic marker, while

the other four proteins had undetectable levels (AGR2, CD5 and CD38) or limited specificity (SFN) (Supplementary Fig. 3D). Subsequently, MUC1 expression was evaluated in six tumor and paired NATs using immunohistochemistry (IHC) assays. The analysis revealed a significantly elevated level of MUC1 in tumors compared to NATs (Supplementary Fig. 3E–F). Carcinoembryonic antigen (CEA), a commonly utilized tumor marker in clinical settings, served as a reference standard. To confirm the diagnosis efficacy of CA15-3 in LUAD, we tested CA15-3 and CEA in the blood of 38 patients with benign disease and 101 patients with LUAD (test cohort, $n = 139$), and found that CA15-3 and CEA levels were significantly higher in LUAD than in benign disease (Supplementary Fig. 3G). For the diagnostic performance of CA15-3 and CEA for LUAD manifesting as SSNs, area of the receiver operating characteristic curve (ROC) was 0.705 and 0.733, respectively (Supplementary Fig. 3H). In addition, we found that the diagnostic performance is best when the CA15-3 threshold is set at 17 ng/ml by using the Jordon index analysis, and the sensitivity and specificity were 56.7% and 78.8%, respectively in the test cohort. Next, this threshold was used to confirm its diagnostic accuracy for distinguishing adenocarcinomas and benign conditions in an external cohort comprising 41 patients with benign diseases and 124 LUAD patients ($n = 165$) (Supplementary Fig. 3I). For the diagnostic performance of CA15-3 and CEA for LUAD in the external cohort, area of the ROC was 0.709 and 0.756, respectively (Supplementary Fig. 3J). The diagnostic sensitivity and specificity were 56.7% and 78.8% at the threshold of 17 ng/ml in the external cohort. Detailed clinicopathological characteristics of patients in both the test cohort and the external cohort are outlined in Supplementary Table 2.

Proteogenomic characterization reveals tumorigenesis of lung preneoplasia

We identified the cholesterol metabolism was a hallmark in the pre-neoplastic AIS stage at proteomic and phosphoproteomic levels, which may be a pivotal molecular event driving LUAD initiation (Fig. 4A and B). To verify this result, we selected proprotein convertase subtilisin/kexin type 9 (PCSK9), an important target in cholesterol metabolism pathway, and a soft agar assay⁵³ showed that inhibition of PCSK9 promoted the transformation of HCA2-TERT cells, an immortalized human foreskin fibroblast cell line, to tumor cells, demonstrating the critical role of cholesterol metabolism in tumorigenesis (Fig. 4C–F). Consistently, PCSK9 promotes the degradation of LDL receptors (Fig. 4D), thereby diminishing the clearance of LDL from the circulation⁵⁴.

To further elucidate the role of PCSK9 in tumorigenesis, we generated mouse lung organoids using lung tissue from Kras-LSL-G12D mice in the presence of adenovirus-Cre and analyzed tumor formation in organoids with or without a PCSK9 inhibitor (Tafelomicimab)⁵⁵



(Fig. 4G). These Kras-LSL-G12D lung organoids formed epithelial spheres after 7 days of culture (Fig. 4H). Subsequently, immunohistochemistry (IHC) was employed to assess the expression of pulmonary surfactant proteins (SFTPC), known marker of alveolar type II cells (AT2)⁵⁶. High expression of SFTPC in the lung tissue organoids (Fig. 4I) confirmed the predominance of AT2 cells, which were considered as the cells of origin for LUAD^{57,58}. After expansion, organoids were

divided equally, and infected with adenovirus-Cre (Fig. 4J). Interestingly, a higher number of organoids were observed in the PCSK9 inhibitor group compared to controls (Fig. 4K), with the former group displaying larger sizes (Fig. 4L), indicating that the PCSK9 inhibitor promoted the growth of organoids in vitro. Histological characteristics of the organoids were analyzed by H&E staining (Fig. 4M). Intriguingly, inhibition of PCSK9 markedly reduced the expression level of SFTPC

Fig. 4 | Proteogenomic characterization reveals tumorigenesis of lung pre-neoplasia. **A, B** KEGG enrichment analysis of adenocarcinoma in situ (AIS) compared to paired normal adjacent tissues at proteomic (**A**) and phosphoproteomic (**B**) levels (Fisher's exact test) (AIS, $n = 17$). **C** Cholesterol metabolic pathway and the expression of target molecules on the pathway by pathway analysis. Red arrows indicated the up-regulated proteins at proteomic level. **D** The effects of PCSK9 knockdown on LDLR and PCSK9 levels in HCA2-TERT cells. ($n = 3$). **E, F** Soft agar assays of anchorage-independent colony formation of HCA2-TERT cells transfected with vectors encoding SV40 large tumor antigen (LT) and HRAS V12 (Ras) together with shCtrl or shPCSK9 followed by plating in soft agar. **E** 2×10^5 cells were plated in 0.4% Noble agar and colonies were counted 5 weeks after seeding. Each image is representative of 15 fields from 3 independent experiments. **F** Quantitative data of **E**. (two-sided Student's t -test, mean $\pm$ SD). $n = 3$, $p = 0.0206$ (shPCSK9-1), $p = 0.0229$ (shPCSK9-2). Scale bar, 90 μ m. **G** Schematic illustration of the workflow for establishing organoids from Kras-LSL-G12D mice lung tissues. **H** Representative images from bright-field microscopy of organoids after 7 days of culture. Scale bar, 100 μ m. **I** The morphological and histological characteristics (surfactant protein C,

SPPTC) of the organoids at day 7. Scale bar, 100 μ m. **J** The organoids were infected with adenovirus-Cre in the absence or presence of PCSK9 inhibitor (Tafocicimab). Scale bar, 100 μ m. **K** The number of organoids formed in the indicated group after infected with adenovirus-Cre 3 weeks. (two-sided Student's t -test, mean $\pm$ SD). $n = 3$, $p = 0.0083$. **L** Bar graph showing quantifications of the diameter in organoids with indicated groups. (two-sided Student's t -test, mean $\pm$ SD). $n = 19$, $p = 0.0022$. **M** The morphology of the organoids formed in the PCSK9 inhibitor group and control group. Scale bar, 100 μ m. **N** Representative images of SFTPC, EpCAM and HMGB1 and Ki 67 on the organoids in PCSK9 inhibitor group and control group by immunohistochemistry (IHC). Scale bar, 100 μ m. **O** Quantitative data of **N**. (two-sided Student's t -test, mean $\pm$ SD). $n = 3$, $p = 0.0005$ (SFTPC), $p = 0.0159$ (EpCAM), $p = 0.0421$ (HMGB1), $p = 0.0067$ (Ki 67). **P** Association of total cholesterol and risk of lung cancer among 109,884 participants in the Kailuan cohort from 2006-2014. Each image is representative of one organoid from 3 independent experiments in (**I, M, N**). Each image is representative of 9 fields from 3 independent experiments in (**H**) and (**J**). **C** and **G** were created in BioRender⁹³.

and enhanced the expression level of epithelial cell adhesion molecule (EpCAM), indicating the formation of epithelial tumors (Fig. 4N and O). Furthermore, treatment with PCSK9 inhibitor significantly enhanced the expression levels of high mobility group box-1 protein (HMGB1)⁵⁹ and Ki-67 (Fig. 4N and O), underscoring the significant role of PCSK9 in regulating tumorigenesis as verified with this lung epithelial organoid model.

To validate the correlation between cholesterol intensity and lung cancer risk, an analysis was conducted on the blood total cholesterol (TC) intensity and lung cancer incidence within the Kailuan cohort, consisting of both employees and retirees from the Kailuan group in Tangshan, China, totaling 109,884 male participants⁶⁰. The association and nonlinear relationship between baseline TC levels and lung cancer risk within this cohort were evaluated using the Cox proportional hazards regression model (Fig. 4P). Following adjustments for age, education, income, smoking, and alcohol, the hazard ratios (HR) (with 95% confidence intervals) for lung cancer risk linked to lower TC levels (<160 mg/dl) and higher TC levels (>240 mg/dl) were found to be 1.34 (1.04-1.72) and 1.45 (1.09-1.92) respectively, in comparison to individuals with normal TC levels (160-180 mg/dl) (Fig. 4P). Individuals with either lower or higher TC levels exhibited an increased risk of lung cancer, indicating a possible association between TC levels and heightened lung cancer risk, pending further clinical validation.

Endoplasmic reticulum stress is a hallmark of AIS progress to IAC and a drug target for LUAD

KEGG pathway analysis revealed that the activation of focal adhesion, protein processing in endoplasmic reticulum (ER), phagosome, platelet activation, and proteoglycans in cancer pathway were enriched in tumors in comparison to the NATs (Fig. 5A). To further delineate the evolutionary trajectories in the airway progress from preneoplasia to invasive LUAD, we performed mFuzz clustering using the differential proteins and phosphosites, and glycopeptides among three stages (Kruskal-Wallis tests, FDR < 0.1) and identified four distinct clusters at proteomic, phosphoproteomic and glycosproteomic levels, respectively (Supplementary Data 6). **We focused on the proteins in cluster 3, displayed a constant upward trend along with the tumor progression,** and KEGG pathway analysis revealed that the protein processing in ER, ribosome biogenesis in eukaryotes, spliceosome and lysosome pathways were enriched in cluster 3 (Supplementary Fig. 4B). At phosphoproteomic level, phosphorylation modification of proteins in cluster 4 showed a stepwise increase from AIS to MIA and IAC, displaying a constant upward trend along with the tumor progression (Supplementary Fig. 4A and Supplementary Data 6). KEGG pathway analysis revealed that the protein processing in ER, apoptosis, and autophagy pathways were enriched in cluster 4 (Supplementary Fig. 4C). At glycosproteomic level, glycosylation modification of

proteins in cluster 2 showed a stepwise increase from AIS to MIA and IAC, displaying a constant upward trend along with the tumor progression (Supplementary Fig. 4A and Supplementary Data 6). KEGG pathway analysis revealed that the protein processing in ER, apoptosis, and autophagy, proteoglycans in cancer pathways were enriched in cluster 2 (Supplementary Fig. 4D).

Collectively, protein processing in ER pathway enrichment is a hallmark in the progression from AIS to MIA and IAC. Several key proteins involved in ER quality control such as OS9⁶¹ and Bap31⁶² showed elevated expression, which implies the accumulation of misfolded protein in the ER (Supplementary Fig. 4E). During the malignant transformation from preneoplasia to invasive LUAD, the accumulation of misfolded proteins exceeds cellular tolerance limits, thereby activating endoplasmic reticulum (ER)-resident sensors that initiate the unfolded protein response (UPR). What can be echoed is that apoptosis, autophagy, spliceosome and ribosome biogenesis in eukaryotes, the UPR-related biological process, are also stepwise increased in the process of AIS to MIA and IAC (Supplementary Fig. 4B-D). Consequently, we propose that the dysregulation of correct protein folding within the endoplasmic reticulum may influence tumor progression from AIS to IAC, presenting a potential therapeutic target.

Protein disulfide isomerase (PDI) is a disulfide bond-modulating ER chaperone, induced during ER stress and required for proper protein folding in the ER^{63,64}. To investigate the impact of PDI inhibition on anticancer effects, we recruited a small molecule inhibitor called CCF642 to increase ER stress⁶⁵ (Supplementary Fig. 4F). Firstly, we established patient-derived tumor xenograft (PDX) models from LUAD manifesting as SSNs, which were used to test the in vivo efficacy of CCF642. Treatment with CCF642 resulted in reduced tumor growth and significantly prolonged survival in mice (Fig. 5B-F). In addition, to further investigate the role of ER stress hyperactivation in the anti-tumor effects of PDI inhibition in PDX model (PDX-P2), we using IHC assay evaluated reliable markers of the ER stress, including PERK and ATF4. We found that the expression of PERK and ATF4 in the tumors treated with CCF642 was significantly higher than that in the control group (Supplementary Fig. 4G and H). We previously employed a patient-derived tumor-like cell cluster model (PTC model) to assess drug efficacy for lung cancer and this model has demonstrated high accuracy in predicting the effects of chemotherapy⁶⁶. Here, PTCs from 17 patients with LUAD were generated to evaluate the effect of CCF642. We observed that CCF642 effectively restrained tumor growth in PTCs, and this effect was concentration-dependent (Fig. 5G). Furthermore, we found similar suppressive effect of CCF642 on the growth of xenografted LLC cells (Fig. 5H-J).

Subsequently, we treated cultured LLC cells with 1 μ M CCF642 and observed an inhibitory effect on cell proliferation (Fig. 5K). Importantly, when we specifically targeted Pdia3 (Fig. 5L), a member of

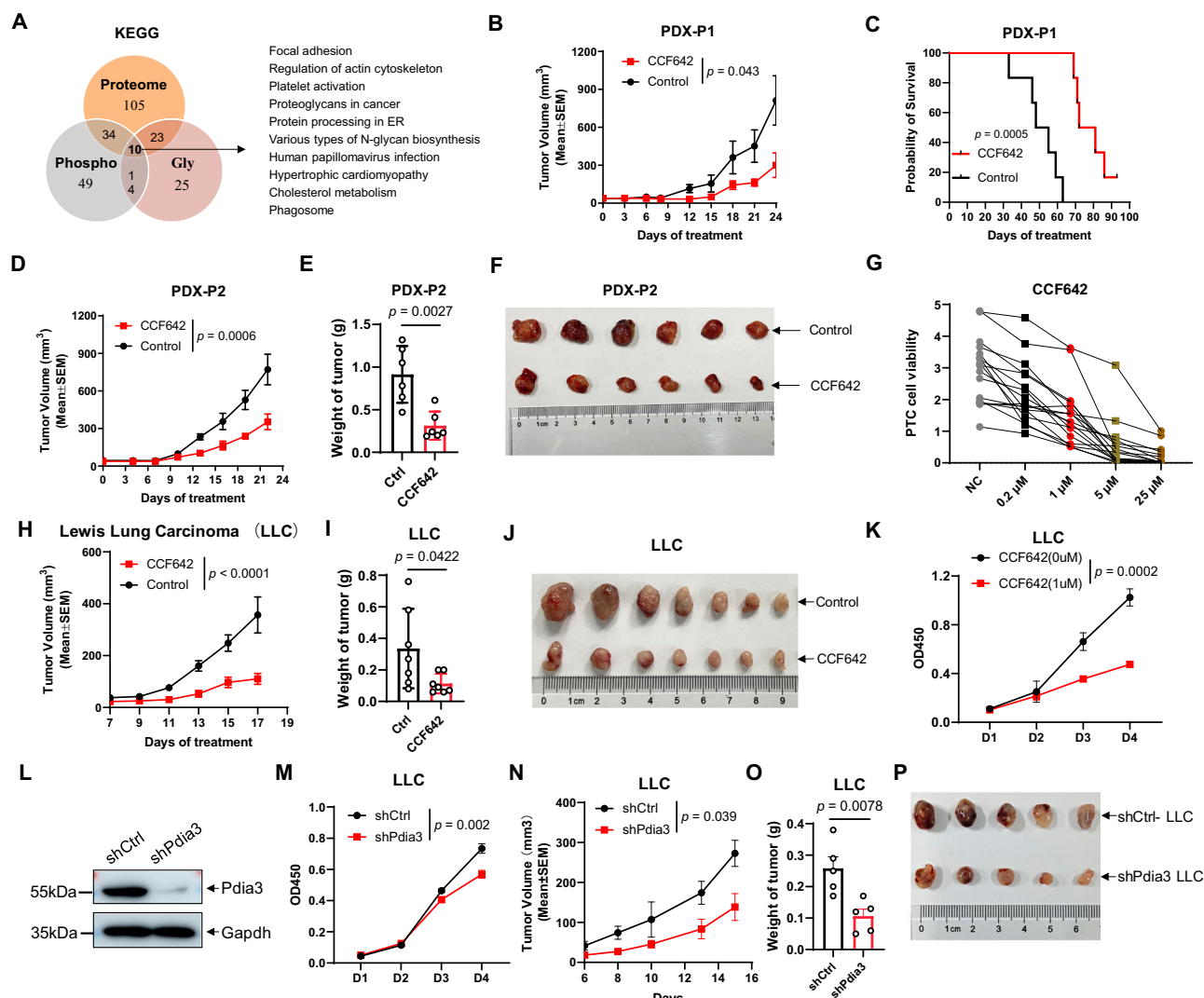


Fig. 5 | Endoplasmic reticulum stress is a hallmark and drug target of AIS progress to IAC. A The differentially regulated proteins, phosphosites and intact glycopeptides between tumors and paired normal adjacent tissues (NATs) and the pathways. **B** Growth curve of tumors formed by patient-derived xenografts (PDX) from lung adenocarcinoma (PDX-P1) left untreated or treated with CCF642 in BALB/c Nude mice. $n = 6$ mice per group. $p = 0.043$. **C** The effects of CCF642 on the survival of PDX-bearing mice in (B). Log-rank test was used. $n = 6$ mice per group. $p = 0.0005$. **D** Growth curve of tumors formed by PDX from another lung adenocarcinoma (PDX-P2) left untreated or treated with CCF642 in BALB/c Nude mice. $n = 6$ mice per group. $p = 0.0006$. **E** The effects of CCF642 on the tumor weight formed by PDX (PDX-P2). $n = 6$. $p = 0.0027$. **F** Tumors formed by PDX-P2 implantation left untreated or treated with CCF642 in BALB/c Nude mice. **G** Drug-response profile of patient-derived tumor-like cell clusters (PTCs) from seventeen lung adenocarcinoma under CCF642. **H** Growth curve of tumors formed by allografted

WT (Lewis Lung Carcinoma) LLC cells left untreated or treated with CCF642 in C57BL/6 mice. $n = 7$ mice per group. $p < 0.0001$. **I** The effects of CCF642 on the tumor weight formed in (H). $n = 7$. $p = 0.0422$. **J** Tumors formed by allografted WT LLC cells left untreated or treated with CCF642 in C57BL/6 mice. **K** The effects of CCF642 on cell proliferation in LLC cells. (two-sided Student's t -test, mean $\pm$ SEM, $n = 3$). $p = 0.0002$. **L** The effects of Pdia3 knockdown on Pdia3 levels in LLC cells. ($n = 3$). **M** The effects of Pdia3 knockdown on cell proliferation in LLC cells. (two-sided Student's t -test, mean $\pm$ SEM, $n = 3$). $p = 0.002$. **N** Growth curve of tumors formed by WT and Pdia3 knockdown LLC cells in C57BL/6 mice. $n = 5$ mice per group. $p = 0.055$. **O** The effects of Pdia3 on the tumor weight formed in (N). $n = 5$. $p = 0.0078$. **P** Tumors formed in (N). Two-way ANOVA followed by Tukey's post hoc test, mean $\pm$ SEM, was used for **B**, **D**, **H**, **N**. Two-sided Student's t -test, mean $\pm$ SD, was used for **E**, **I**, **O**.

the PDI family, we consistently observed a strong inhibitory effect (Fig. 5M), whereas knockdown of Pdia3 had no discernible effect on apoptosis (Supplementary Fig. 4I and J). Moreover, our data revealed that Pdia3 knockdown did not significantly alter the relative LC3-II/LC3-I protein level in the presence or absence of the autophagy inhibitor chloroquine (CQ), indicating that Pdia3 had no discernible effect on autophagy (Supplementary Fig. 4K). Moreover, silencing Pdia3 significantly suppressed the growth rates of LLC xenografts and reduced the weight of LLC tumors compared to the control group (Fig. 5N–P). Overall, these findings indicated that inhibiting PDI could effectively impede the progression of lung cancer.

We presented up-to-date evidence and our own opinions regarding the management of SSNs, classifying AIS into two categories based on the risk of progression. A1 subtype is rapid progression to MIA and IAC, A2 subtype is no progression for a long time and remains in a stable state (Supplementary Fig. 4L). In our previous results, sustained ER stress was demonstrated to be a hallmark of AIS progress to IAC (Fig. 5A and Supplementary Fig. 4A–D). To further validated whether ER stress could be utilized as a reliable biomarker for AIS progression to IAC, we collected A1 and A2 tumors from clinical settings. Growth curve showed that A1 subtype tumors rapid growth by evaluating the tumor size on CT scans (Fig. 6A), and A2 subtype tumors

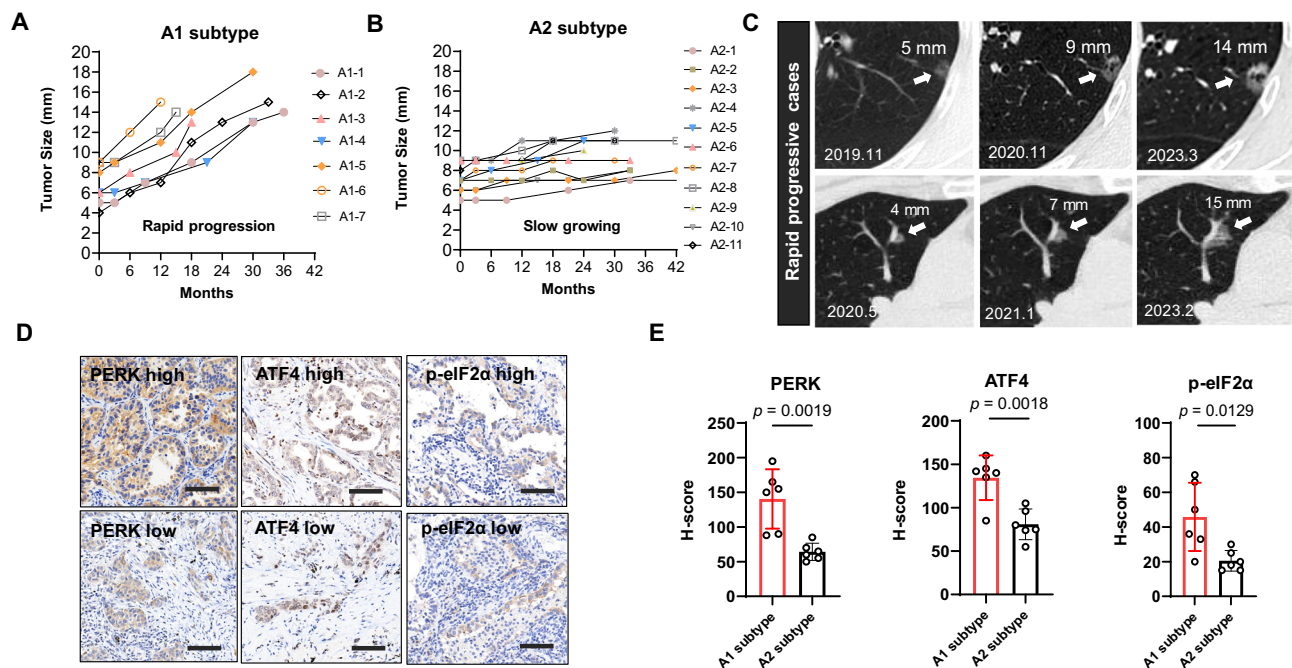


Fig. 6 | Endoplasmic reticulum stress is a hallmark of AIS progress to IAC in the clinic. **A, B** Growth curve of lung adenocarcinoma manifested as SSNs in A1 and A2 subtype by tumor size on CT scans. A1 subtype: *n* = 7, A2 subtype: *n* = 11. **C** Representative CT scans of rapidly progressing cases. Top: All images presented are representative of one tumor among seven tumors classified as A1 subtype. Bottom: All images presented are representative of another tumor among seven

tumors classified as A1 subtype. **D** Representative immunohistochemistry (IHC) stains of PERK, ATF4 and p-eIF2S1 of A1 and A2 subtype tumors. Each image is representative of 30 fields from 6 samples. Scale bar, 100 μm. **E** The quantification data of PERK, ATF4 and p-eIF2S1 of A1 and A2 subtype tumors. Two-sided Student's *t*-test, mean ± SD. *n* = 6 tumors per group. *p* = 0.0019 (PERK), *p* = 0.0018 (ATF4), *p* = 0.0129 (p-eIF2α).

in a nearly stable state (Fig. 6B). Representative CT images of rapidly progressing cases were shown in Fig. 6C. Then, we assessed reliable ER stress markers (PERK, ATF4, and eIF2S1 phosphorylation) simultaneously via IHC assay. We found that the expression of PERK, ATF4 and eIF2S1 phosphorylation in A1 subtype tumors were significantly higher than those in A2, pinpointing a pivotal role of ER stress in the tumor progression (Fig. 6D and E). In addition, we validated the predictive value of PERK in a distinct patient cohort that included 110 patients with LUAD manifesting as SSNs. The clinical characteristics of patients were depicted in Supplementary Table 3. Of the 110 LUAD cases, 31 (28.1%) exhibited high PERK expression (defined as more than 10% PERK positive cells). Patients with elevated PERK expression experienced a notably increased postoperative recurrence rate when contrasted with those displaying lower PERK levels (Supplementary Fig. 4M), indicating a positive correlation between PERK expression and heightened invasive and progressive tendencies.

Discussion

This research delineated the evolutionary trajectory from AIS to MIA and IAC through a comprehensive integrative proteogenomic analysis of tissue specimens. Our findings underscore **the critical involvement of cholesterol metabolism in the onset of AIS and the enduring impact of ER stress in driving the transition from AIS to MIA and IAC, highlighting ER stress as a robust biomarker for predicting tumor progression risk**. The in-depth proteomic profiling of LUAD in this investigation offers a profound understanding of SSNs' tumorigenesis and progression, presenting new avenues for personalized therapeutic strategies (Supplementary Fig. 5).

The progression of LUAD from AIS to invasive forms is hypothesized to occur through incremental stages modulated by a range of genetic, metabolic, and immunological alterations in addition to those initiating the malignant transformation^{13,67,68}. Herein, we identified dysregulated cholesterol metabolism in the preneoplastic AIS stage. Cholesterol serves as an essential regulator in tumor cell proliferation,

primarily through its pivotal role in membrane biogenesis. Neoplastic cells acquire cholesterol via two distinct pathways: LDL uptake and de novo biosynthesis. Beyond its structural function as a membrane component, cholesterol serves as a metabolic precursor for bile acids and steroid hormones, which have been implicated in the initiation and progression of multiple malignancies, particularly colorectal, breast, and prostate cancers^{69–71}. Cholesterol exerts regulatory effects on oncogenic signaling pathways through two distinct molecular mechanisms: firstly, by mediating post-translational modifications of key signaling proteins such as Hedgehog and Smoothened through covalent attachment^{72,73}; and secondly, by orchestrating the assembly and maintenance of specialized membrane microdomains that serve as critical signaling platforms^{74,75}. Utilizing a mouse organoid model derived from alveolar type II cells, a proposed origin cell for LUAD, our study highlighted the indispensable contribution of PCSK9 in tumorigenesis, underscoring cholesterol as a promising target for therapeutic interventions in preventing and managing LUAD. Elevated cholesterol levels may enhance tumor proliferation by providing increased cholesterol for cellular processes, promoting tumor progression. The relationship between low cholesterol levels and the progression of lung adenocarcinoma is not yet fully understood. Prior studies have indicated that reduced cholesterol levels might increase the risk of certain cancers, potentially due to cholesterol's critical function in maintaining cell membrane stability and facilitating cellular signaling⁷⁶. Similar regulatory mechanisms are evident in other biological systems, such as growth hormone (GH) dynamics. Elevated GH levels are associated with systemic health issues, including a higher susceptibility to gastrointestinal tumors, while GH deficiency is frequently accompanied by complications such as pituitary tumors^{77,78}.

Our study revealed a progressive pattern of differential proteins, phosphosites, and glycopeptides from AIS to MIA and IAC. Notably, the enrichment of protein processing in the ER was observed across proteomics, phosphoproteomics, and glycoproteomics, suggesting a crucial role of this pathway in invasive progression. Sustained

activation of ER stress sensors endows malignant cells with greater tumorigenic and metastatic capacity⁷⁹. Robust ER stress responses are evident in various human cancers, spanning breast, pancreatic, lung, skin, prostate, brain, and hematological malignancies⁸⁰. Treatment with CCF642 significantly enhances PERK activation through promoting its dimerization and subsequent phosphorylation, while simultaneously inducing IRE1- α oligomerization in KMS-12-PE cells⁶⁵. A fundamental therapeutic strategy involves the targeted induction of ER stress and subsequent activation of the UPR, thereby triggering apoptotic pathways specifically in malignant cells.

Accurate differentiation between stable SSNs and those with high potential for progression is crucial to enable precise clinical management. Our study identifies ER stress as a dependable biomarker for assessing the risk of tumor progression; tumors exhibiting high ER stress are more prone to rapid growth and advancement. In the context of LUAD development, Nour Ghaddar et al. emphasize the adaptive PERK/p-eIF2 α branch of the integrated stress response (ISR) as a vital component of tumorigenesis and a promising target for therapeutic interventions in mutant KRAS lung cancer treatment⁸¹. SSNs are typically characterized by slow growth clinically. Hasegawa et al. showed that the tumor doubling times of SSNs and solid LUAD were 813 ± 375 and 149 ± 125 days, respectively⁸². Early resection of SSNs before significant growth may prevent the progression to invasive LUAD. Combining clinical and proteogenomic data can improve scoring systems designed to identify pre-invasive nodules at high risk of malignant progression.

Several limitations should be acknowledged. A notable constraint shared by both the current investigation and most prior studies lies in their exclusive reliance on resected tumor specimens. This approach provides only a static, single-timepoint molecular profile, thereby limiting our ability to comprehensively capture the dynamic evolutionary trajectory of individual lesions. The conventional model assumes a linear pathological progression from AIS to MIA and IAC. However, critical knowledge gaps remain regarding whether all AIS lesions inevitably progress to MIA or IAC, and whether every IAC necessarily follows this sequential evolutionary pathway through AIS and MIA stages. Secondly, lung cancer manifested as radiological SSNs is prevalent in East Asian populations, and all of these findings are in Chinese Asian patients and it will need to be confirmed if they are also found in other ethnic, racial groups. Thirdly, the absence of the validation of those preclinical experiments on human tumors or tissues is a major limitation of this study. Future human specimen-related studies are needed to further consolidate our findings.

Methods

The study received approval from the Research Ethics Committee of Shanghai Pulmonary Hospital (Institutional Review Board: L21-021) and adheres to all applicable ethical guidelines. Detailed methodologies are outlined in the supplemental information.

Experimental model and subject details

Sample collection. The study was approved by the Research Ethics Committee of Shanghai Pulmonary Hospital (Institutional Review Board: L21-021), and written informed consent was obtained from each patient. Participants were fully informed during the consent process that no compensation would be provided, and their willingness to participate was entirely voluntary. The tumor and NATs used in this study were prospectively collected from Shanghai Pulmonary Hospital from December 2020 to March 2021. Totally, 71 patients were enrolled in this study, including 18 AIS, 26 MIA and 27 IAC patients. Of them, one AIS, two MIA and two IAC and their paired non-cancerous NATs were used for the construction of the common reference pool. Finally, 17 AIS, 24 MIA and 25 IAC tumor tissues and paired NATs were used for proteomics, phosphoproteomics and glycoproteomics analysis. Patients receiving any anti-cancer treatments before surgery were

excluded. Primary tumor tissues and paired non-cancerous NATs (>3 cm apart from tumor edge, morphologically negative for malignant cells) were surgically resected and promptly transferred to liquid nitrogen during the operation. Subsequently, they were stored in -80°C refrigerators until they are used for protein extraction. Rigorous pathology quality control, including histological subtype, tumor purity and degree of necrosis, was applied by histologic sections obtained from the top and bottom portions of each case and reviewed by two lung cancer pathologists (Chunyan Wu and Huikang Xie). All samples used in this study had tumor purity of more than 50% and less than 20% necrosis. Selected specimens were cryopulverized using the Covaris CryoPREP instrument and the material was aliquoted for subsequent molecular characterization. For proteomic analyses, the material was sent to the proteomic characterization center at Jingjie Biotechnology Co., Ltd, Hangzhou, China.

Processing of FFPE specimens by laser capture microdissection.

Formalin-fixed Paraffin-embedded (FFPE) samples were used for whole exome sequencing. All specimens were laser microdissected from hematoxylin and eosin (H&E) slides after the histologic evaluation by two lung cancer pathologists⁸³. The 6-micron H&E slides were loaded onto the MMI CellCut Laser Microdissection system. Using MMI CellCut software, we performed a comprehensive slide scan with the aid of the CellScan toolbar at $4\times$ magnification to make navigation easier. We used a closed-shape manual drawing tool to select our region of interest via the computer mouse. We then focused our laser at $350\text{ }\mu\text{m}$ and performed an automated cutting using a 60% laser power setting moving at $50\text{ }\mu\text{m}/\text{second}$. All samples were collected and processed for DNA extraction.

Mice. BALB/c Nude mice and C57BL/6 mice were purchased from the Shanghai Laboratory Animal Center. Kras-LSL-G12D mice were kindly provided by the laboratory of Prof Fei Li, School of Basic Medical Sciences, Fudan University. All mice in the experiments were 6–8 weeks old, matched for age and sex, and kept under specific pathogen-free (SPF) conditions at the Laboratory Animal Center of Tongji University. All experiments were approved by Tongji University School of Medicine Animal Care and Use Committee and were conducted in accordance with the National Institutes of Health Guidelines for the Care and Use of Laboratory Animals.

Gene knockdown by lentiviral particles. The short hairpin RNA (shRNA) sequence targeting mouse Pdia3 or human PCSK9 was cloned into pLKO lentiviral vector, with the target sequences provided in Supplementary Table 4. To prepare lentiviral particles, the plasmid was co-transfected into 293 T cells with psPAX2, and pMD2.G by using Lipofectamine 3000 (ThermoFisher). Viral supernatants were collected 48 hours following transfection and used to transduce target cells. Stable gene knockdown cell lines were selected using puromycin

Gene overexpression by lentiviral particles. The cDNAs of FGFR3-WT, FGFR3-S408D and FGFR3-S408A were cloned into the pCDH vector. After packaging lentiviral particles as described above, PC-9 cell were transduced with the viral particles mixed with polybrene ($8\text{ }\mu\text{g}/\text{mL}$). Stable gene overexpression cell lines were selected using puromycin.

Method details

Mass spectrometry methods

Protein extraction. The sample was ground with liquid nitrogen into cell powder and then transferred to a 5-mL centrifuge tube. After that, four volumes of lysis buffer (1% Triton X-100, 1% protease inhibitor cocktail, 1% phosphatase inhibitor) was added to the cell powder, followed by sonication three times on ice using a high intensity ultrasonic processor (Scientz). The remaining debris was removed by

centrifugation at 12,000 g at 4 °C for 10 min. Finally, the supernatant was collected and the protein concentration was determined with BCA kit according to the manufacturer's instructions.

Tryptic digestion. The sample was slowly added to the final concentration of 20% (m/v) TCA to precipitate protein, and an appropriate amount of standard protein was added, then vortexed to mix and incubated for 2 h at 4 °C. The precipitate was collected by centrifugation at 4500 g for 5 min at 4 °C. The precipitated protein was washed with pre-cooled acetone for 3 times and dried for 1 min. The protein sample was then redissolved in 200 mM TEAB and ultrasonically dispersed. Trypsin was added at 1:50 trypsin-to-protein mass ratio for the first digestion overnight. The sample was reduced with 5 mM dithiothreitol for 60 min at 37 °C and alkylated with 11 mM iodoacetamide for 45 min at room temperature in darkness. Finally, the peptides were desalted by C18 SPE column.

TMT/TMT labeling TMT labeling of peptides

Tryptic peptides were firstly dissolved in 0.5 M TEAB. Each channel of peptide was labeled with their respective TMT reagent according to the manufacturer's protocol, (Thermo Scientific), and incubated for 2 hours at room temperature. Five microliters of each sample were pooled, desalted and analyzed by MS to check labeling efficiency. Then, samples were quenched by adding 5% hydroxylamine. The pooled samples were then desalted with Strata X C18 SPE column (Phenomenex) and dried by vacuum centrifugation. The detailed steps for labeling are as follows.

- Dissolve TMT reagent: Immediately before use, equilibrate the TMT Label Reagents to room temperature. For the 5 mg vials, add 256 μ L of anhydrous acetonitrile to each tube. Allow the reagent to dissolve for 5 minutes with occasional vortexing. Briefly centrifuge the tube to gather the solution.
- Label peptides: 300 μ g or 367 μ g sample peptides were labeled with 3.2 mg TMT reagent. Peptides were dissolved in 400 μ L of 0.5 M TEAB solution and labeling reagent was added in 164 μ L of acetonitrile. Incubate the reaction for 2 hours at room temperature.
- Label efficiency: For each labeling group, 5 μ L from each sample were combined and measured by LC-MS/MS. The labeling efficiency of each group was over 99%.
- Desalt: Add 32 μ L of 5% hydroxylamine to the sample and incubate for 15 minutes to quench the reaction. Differentially labeled peptides were then mixed (11*300 μ g or 9*367 μ g), dried down via vacuum centrifuge and desalted with Strata X C18 SPE column (Phenomenex).

Construction of the common reference pool. We used TMT10-plex and TMT11-131C to perform TMT labeling on 142 (17 pairs in the AIS group, 26 pairs in the MIA group, and 28 pairs in the IAC group) samples. The 142 samples were divided into 15 labeling groups (Supplementary Data 7), so that the samples of the AIS, MIA, and IAC were relatively evenly distributed among the different labeling groups. To ensure comparisons between different labeling groups, each labeling group contained the same control sample mix, and as many samples as possible were allowed to participate in the construction of the control sample mix as long as the amount of peptide could meet the experimental requirements. 106 (Supplementary Data 7) samples of the 142 samples participated in the mix construction, and equal amounts of peptide (55.6 μ g) were taken respectively to be mixed as mix samples for labeling.

HPLC Fractionation. The sample was fractionated into fractions by high pH reverse-phase HPLC using Agilent 300Extend C18 column (5 μ m particles, 4.6 mm ID, 250 mm length). The wavelength was set to 214 nm, and the column oven temperature was maintained at

35 °C. The column was equilibrated with 95% buffer A (2% ACN, pH 9.0 adjusted by ammonia) and 5% buffer B (98% ACN, pH 9.0 adjusted by ammonia) for at least 30 min. Sent the stepwise gradient method, after the baseline became flat. Then, added 1 ml buffer A to the peptide sample and vortexed to dissolved it. Centrifuged at 12000 g for 5 min and transferred the supernatant to a new tube. After another centrifugation step, the supernatant was loaded to the HPLC, and sent the method to separate the sample while simultaneously started the automatic collector. Collected the sample at 1 min/tube and from the 11th to 46th tubes total of 36 tubes. At last, the peptides were combined into 4 fractions and dried by vacuum centrifuging.

Bio-material-based PTM enrichment (for Phosphorylation). Peptide mixtures were first incubated with immobilized metal affinity chromatography (IMAC) microspheres suspension with vibration in loading buffer (50% acetonitrile/0.5% acetic acid). To remove the non-specifically adsorbed peptides, the IMAC microspheres were washed sequentially with 50% acetonitrile/0.5% acetic acid and 30% acetonitrile/0.1% trifluoroacetic acid. To elute the enriched phosphopeptides, the elution buffer containing 10% NH₄OH was added and the enriched phosphopeptides were eluted with vibration. The supernatant containing phosphopeptides was collected and lyophilized for LC-MS/MS analysis.

Bio-material-based PTM enrichment (for N-glycopeptides). Tryptic peptides were redissolved in 200 μ L washing buffer (80% ACN, 5% TFA) and then loaded onto the column and then washed with washing buffer for three times. Glycopeptides were eluted with 0.1% TFA, 50 mM ammonium bicarbonate and 50% ACN two times. The eluted glycopeptides were desalted using C18 Zip Tips according to the manufacturer's instructions and then dried for further MS analysis.

LC-MS/MS analyses LC-MS/MS analyses for proteome

The tryptic peptides were dissolved in solvent A and subsequently loaded onto a homemade reversed-phase analytical column (25 cm length, 100 μ m i.d.). The mobile phase consisted of solvent A (0.1% formic acid, 2% acetonitrile/in water) and solvent B (0.1% formic acid, 90% acetonitrile/in water). Peptides were separated with the following gradient: 0–2 min, 8–10% B; 2–40 min, 10–23% B; 40–54 min, 23–33% B; 54–57 min, 33–80% B; 57–60 min, 80% B, and all at a constant flow rate of 500 nL/min on a EASY-nLC 1200 UPLC system (ThermoFisher Scientific). The separated peptides were analyzed in Q Exactive HF-X mass spectrometer equipped with a nano-electrospray ion source. An electrospray voltage of 2100 V was applied. Precursors and fragments were analyzed at the Orbitrap detector. The full MS scan resolution was set to 120,000 for a scan range of 400–1500 m/z. The fragments were detected in the Orbitrap at a resolution of 45,000 and fixed first mass was set as 100.0 m/z. Up to 20 most abundant precursors were then selected for further MS/MS analyses with 30.0 s dynamic exclusion. The HCD fragmentation was performed at a normalized collision energy (NCE) of 28%. Automatic gain control (AGC) target was set at 5e4, with an intensity threshold of 5.8e4 ions/s and a maximum injection time of 86 ms.

Database search. The resulting MS/MS data were processed using the MaxQuant search engine (v.1.6.15.0). Tandem mass spectra were searched against the human SwissProt database (20395 entries) concatenated with a reverse decoy database. Trypsin/P was specified as the cleavage enzyme, allowing up to 2 missing cleavages. The mass tolerance for precursor ions was set as 20 ppm for the First search and 4.5 ppm for the Main search, while the mass tolerance for fragment ions was set as 20 ppm. Carbamidomethyl on Cys was specified as fixed modification, while Acetylation on protein N-terminal, oxidation on Met and deamidation (NQ) were specified as variable modifications.

TMT-IIplex quantification was performed. FDR was adjusted to <1% and the minimum score for peptides was set >40.

LC-MS/MS analyses for phosphoproteome. The tryptic peptides were dissolved in solvent A, directly loaded onto a home-made reversed-phase analytical column (25-cm length, 100 μ m i.d.). The mobile phase consisted of solvent A (0.1% formic acid, 2% acetonitrile/in water) and solvent B (0.1% formic acid, 90% acetonitrile/in water). Peptides were separated with the following gradient: 0–62 min, 3–20% B; 62–82 min, 20–32% B; 82–86 min, 32–80% B; 86–90 min, 80% B, and all at a constant flow rate of 500 nL/min on a EASY-nLC 1200 UPLC system (ThermoFisher Scientific). The separated peptides were analyzed in Q Exactive HF-X mass spectrometer equipped with a nano-electrospray ion source. The electrospray voltage applied was 2100 V. Precursors and fragments were analyzed at the Orbitrap detector. The full MS scan resolution was set to 120,000 for a scan range of 400–1500 m/z. The fragments were detected in the Orbitrap at a resolution of 45,000 and fixed first mass was set as 100.0 m/z. Up to 15 most abundant precursors were then selected for further MS/MS analyses with 30.0 s dynamic exclusion. The HCD fragmentation was performed at a normalized collision energy (NCE) of 28%. Automatic gain control (AGC) target was set at 5e4, with an intensity threshold of 5.6e4 ions/s and a maximum injection time of 90 ms.

Database search. The resulting MS/MS data were processed using the MaxQuant search engine (v.1.6.15.0). Tandem mass spectra were searched against the human SwissProt database (20395 entries) concatenated with the reverse decoy database. Trypsin/P was specified as a cleavage enzyme, allowing up to 2 missing cleavages. The mass tolerance for precursor ions was set as 20 ppm in the First search and 4.5 ppm in Main search, and the mass tolerance for fragment ions was set as 20 ppm. Carbamidomethyl on Cys was specified as a fixed modification. Acetylation on protein N-terminal, oxidation on Met, deamidation (NQ) and phosphorylation on Ser, Thr, Tyr were specified as variable modifications. TMT-IIplex quantification was performed. FDR was adjusted to <1% and minimum score for peptides was set > 40.

LC-MS/MS analyses for glycoproteome. The tryptic peptides were dissolved in solvent A, directly loaded onto a home-made reversed-phase analytical column (25-cm length, 100 μ m i.d.). The mobile phase consisted of solvent A (0.1% formic acid, 2% acetonitrile/in water) and solvent B (0.1% formic acid, 90% acetonitrile/in water). Peptides were separated with the following gradient: 0–3 min; 2–7% B; 3–43 min, 7–22% B; 43–53 min, 22–34% B; 53–56 min, 34–80% B; 56–60 min, 80% B, and all at a constant flow rate of 500 nL/min on a EASY-nLC 1200 UPLC system (ThermoFisher Scientific). The separated peptides were analyzed in Q Exactive HF-X mass spectrometer equipped with a nano-electrospray ion source. The electrospray voltage applied was 2100 V. Precursors and fragments were analyzed at the Orbitrap detector. The full MS scan resolution was set to 120,000 for a scan range of 700–2000 m/z. The fragments were detected in the Orbitrap at a resolution of 45,000 and fixed first mass was set as 100.0 m/z. Up to 20 most abundant precursors were then selected for further MS/MS analyses with 30.0 s dynamic exclusion. The HCD fragmentation was performed at a normalized collision energy (NCE) of 28%. Automatic gain control (AGC) target was set at 2e5, with an intensity threshold of 2.5e4 ions/s and a maximum injection time of 200 ms.

Database search. The resulting MS/MS data were processed using MSFragger (v2.3). Tandem mass spectra were searched against the human SwissProt database (20395 entries) concatenated with reverse decoy database. Trypsin/P was specified as cleavage enzyme allowing up to 2 missing cleavages. The mass tolerance for precursor and fragment ions was set as 20 ppm. Carbamidomethyl on Cys was specified as fixed modification. Acetylation on protein Nterminal,

oxidation on Met, and deamidation (NQ) were specified as variable modifications. The default N-glycosylation mass list was used as mass offsets. FDR was adjusted to <1%.

Whole exome sequencing

DNA extraction. We performed DNA extraction from laser micro-dissected FFPE lung cancer samples. The DNA was isolated from the FFPE using the DNeasy Blood and Tissue Kit (QIAGEN) according to the manufacturer's instructions. Targeted capture pulldown and exon-wide libraries were generated from native DNA using the xGen® Exome Research Panel (Integrated DNA Technologies, Inc., Skokie, Illinois, US) and TruePrep DNA Library Prep Kit V2 for Illumina (Vazyme). The paired-end sequence data were generated using Illumina HiSeq machines with an average sequencing depth of 150 $\times$ for tumor tissues and NATs. The sequence data were aligned to the human reference genome (NCBI build 37) using Burrows-Wheeler Aligner (BWA), and polymerase chain reaction (PCR) duplicates were sorted and removed using sambamba.

Variant calling pipeline. Raw Sequencing reads were preprocessed by FASTP (v.0.23.2) with default parameters for quality control, including adapter removal, quality trimming and read filtering. The clean reads were aligned to the Human Genome Reference Consortium build 38 (GRCh38) using Burrows-Wheeler Aligner (BWA, v.0.7.17). Picard (v.2.27.5) was used to process PCR duplicates for mapped BAM files. After that, we used the GATK4 (v.4.2.4.0) *BaseRecalibrator* and *ApplyBQSR* modules to generate recalibrated bam files for each BAM file. To reduce the artifacts and false positives, we have applied two different variant calling methods, which are Mutect2 module in GATK4 and Strelka2 (v.2.9.10), to detect the somatic mutations including single-nucleotide variants (SNV) and small insert and deletions (indels). These two methods were used as the core algorithms in the PCAWG projects²⁵.

For Mutect2, we first built a panel of normal (PON) file using all 76 normal samples. Somatic mutations were then identified by comparing each tumor sample with its matched normal counterpart and the PON file. Considering our samples were obtained using FFPE technology, we called somatic single nucleotide variants (SNVs) and small insertion deletions (Indels) using MuTect2 with the argument “-f1r2-tar-gz” and then used LearnReadOrientationModel module to annotate the artifacts. The raw mutations first filtered using the FilterMutectCalls module in the GATK4 and only mutations annotated as “PASS” were retained. We selected mutations with variant allele frequency (VAF) $\geq$ 10% in tumors, supporting variant reads $\geq$ 8, and reads at the variant position $\geq$ 20, and excluded mutations occurring within ENCODE blacklist regions. For Strelka2, somatic mutations were called with the flag ‘PASS’. We added an additional quality filter to tighten filtering for low allelic frequency variants: quality score $\times$ allele frequency > 1.3. We filtered any variant that was supported by three or more reads in the reference sample in at least three patients. The intersection of the Mutect2 and Strelka2 results was used as the final set of somatic mutations.

After filtering for artifacts and defining a final set of mutations, the Mutation Annotation Format (MAF) files was analyzed to identify the significantly mutated genes (SMGs). This was accomplished using the MutSigCV and the q-value less than 0.01 was set as a cutoff. We also generated a blacklist according to the MutSigCV paper and filtered the SMGs included in the blacklist. Additionally, genes with an average read per kilobase per million (RPKM) less than 1.5 were also excluded.

Annotar (version 2019Oct24)⁸⁴ with the following databases for hg38: refGene, clinvar_20210501, 1000g2015aug_all, exac03, and avsnp150 was used to annotate VCF (Variant Call Format). To ensure that no potential driving mutations were mistakenly filtered out during post-processing, candidate mutations in previously implicated cancer genes were manually reviewed using IGV (v.2.16.2)⁸⁵.

Tumor Mutation burden. The tumor mutation burden (TMB) of a tumor sample is calculated by the number of non-synonymous somatic mutations (single nucleotide variants and small insertions/deletions) per mega-base in coding regions. We defined the non-synonymous somatic mutations as follows: firstly, select the mutations that appeared within the consensus coding sequence region designed by the CDS project; secondly, remove variant with variant count <3, allele frequency <5% or total depth <25; at last, filter out synonymous and splice region mutations, and retain only nonsynonymous mutations (missense, nonsense, insertions and deletions mutations).

Proteomics data analysis

Quantitative analysis-proteomic. The raw reporter intensities of peptides across samples were extracted from the MaxQuant result files. Protein relative quantitative abundance was then calculated based on intensity by the following steps: Firstly, the intensities of peptides across all samples were centralized, and then transformed into relative abundance in each sample. To adjust the systematic bias of the identified peptide amount among different samples, the relative quantitative abundance of peptide was subjected to median normalization. The final relative protein abundance was calculated as the median value of the relative abundance from each protein's constituent peptides.

Quantitative analysis-phosphoproteomic and glycoproteomic. The raw reporter intensities of peptides across samples were extracted from the MaxQuant/MSFragger result files. The relative quantitative abundance of sites/glycopeptides were then calculated by the following steps: Firstly, the intensities of sites/glycopeptides were centralized and transformed into relative quantitative abundance in each sample. To account for sample-specific biases in the modified sites/glycopeptides analysis, we normalized the relative abundance of the modified sites/glycopeptides by dividing by the relative abundance of proteins in each sample.

Glycan types characterization. Five glycan types referred to Riley et al. were defined and investigated in this study⁸⁶: Paucimannose, High mannose Complex/hybrid, Sialylated, Fucosylated. Specifically, any glycan with a NeuAc moiety was categorized as sialylated, fucosylated glycan type group contains any glycan that contains a fucose moiety and also is not sialylated, complex/hybrid glycan class are neither fucosylated nor sialylated. The high mannose glycan type represents glycans containing two N-acetylhexosamine and hexose without additional N-acetylhexosamine. The remaining glycans are paucimannose.

Systems biology analysis

Dataset filtering. Proteins (global proteome), phosphosites and glycopeptides present in fewer than 30% of samples (i.e., missing in >70% of samples) were removed from the respective datasets. The remaining missing values were imputed via k-nearest neighbor (kNN) imputation implemented in the impute R-package using the 5 nearest neighbors.

Unsupervised multi-omics clustering using NMF. We used non-negative matrix factorization (NMF) implemented in the NMF R-package to perform unsupervised clustering of tumor samples. The filtered proteins, phosphosites and glycopeptides data tables were concatenated and performed a log transformation to facilitate the interpretation of the expression data. Prior to conducting NMF clustering, uninformative features were excluded from the dataset by removing features with the lowest standard deviation (bottom 5th percentile) across all samples. Each row in the data matrix was further scaled and standardized such that all features from different data types were represented as z-scores. To meet the non-negative input requirement of NMF, we transformed the data matrix by setting all

negative numbers to zero. Then the matrix was subjected to NMF v.0.23.0 for unsupervised clustering. To determine the optimal factorization rank k (number of clusters) for the multi-omic data matrix, we tested a range of clusters between $k = 2$ and $k = 10$, and factorized matrix using 200 iterations for each value of k . The rank survey profiles of the cophenetic score and silhouette width—along with the consensus membership heat maps—indicated a three-subtype solution.

Functional characterization. To define the signatures of each subtype, we used the Wilcoxon rank-sum test to extract differentially expressed features (proteins, phosphosites and glycopeptides) in each subtype against those in the other subtypes. The following cut-off criteria were used: (1) p values should be less than 0.05; (2) fold change (expressed as $\log_2$ (ratio between subtypes)) ≤ -0.58 for down regulation or fold change (expressed as $\log_2$ (ratio between subtypes)) ≥ 1.5 for up regulation;

Top 250 differentially expressed features (proteins, phosphosites and glycopeptides) for each subtype were performed pathway enrichment analysis using GO and KEGG.

To gain further insight into biological implications, we performed single sample Gene Set Enrichment Analysis (ssGSEA) to identify the pathway alterations. We calculated normalized enrichment scores (NES) and p value by projecting the matrix of fold change onto MSigDB database using ssGSEA. We used the ssGSEA implementation available on <https://github.com/broadinstitute/ssGSEA2.0>.

Kinase activity prediction. Kinase-substrate relationships were predicted by previously reported iGPS method⁸⁷. We only considered kinases with at least five substrates observed in our phosphoproteomic data. To calculate the kinase activity score for each sample, we performed a Wilcoxon rank sum test comparing the abundance of substrates of a particular kinase with that of the remaining phosphosites observed in our dataset. The normalized test statistic of the Wilcoxon test was used as the activity score for each kinase.

Glycosylation enzymes analysis. The intact glycopeptide expression was hypothesized to be influenced at least by the expression of substrate glycoproteins and glycosylation enzymes. The Spearman's rank correlation coefficient was used to assess the correlation between the abundance ($\log_2$ ratio values) of intact glycopeptides and the abundance of glycosylation enzymes collected from GlyGen database (<https://glygen.org/>). The correlation matrix was constructed and visualized by performing hierarchical clustering on glycopeptides (columns) and glycosylation enzymes (rows).

Mutation-based cis- and trans-effects. After excluding silent mutations, samples were separated into mutated and WT groups. We used the Wilcoxon rank-sum test to report differentially expressed features (proteins, phosphosites and glycopeptides) between the two groups. Differentially enriched features passing an p value < 0.05 cut-off were separated into two categories based on cis- and trans- effects.

CNA-driven cis and trans effects. Correlations between copy number alterations (CNA) proteome, phosphoproteome and glycoproteome were determined using Pearson correlation of common genes present in CNA-proteome, CNA-phosphoproteome and CNA-glycoproteome datasets. In addition, p -values (corrected for multiple testing using Benjamini-Hochberg FDR) for assessing the statistical significance of the correlation values were also calculated. CNA trans-effects for a given gene were determined by identifying genes with statistically significant (FDR < 0.05) positive or negative correlations.

Quantification and statistical analysis. Statistical details of experiments and analyses can be found in the figure legends and main text above. Statistical significance tests, including Fisher's exact test,

Chi-square test, Student's t-test, Anova test, and Pearson or Spearman correlation test were performed using R, as denoted in each analysis. All statistical tests were two-sided, and statistical significance was considered when P value < 0.05 . To account for multiple-testing, the P values were adjusted using the Benjamini-Hochberg FDR correction. We applied the Mfuzz R package to analyze clusters of features based on their histological subtype.

Cell culture. Human lung squamous cell carcinoma (NCI-H226; ATCC, CRL-5826; NCI-H1703; ATCC, CRL-5889) and human lung adenocarcinoma (A549; ATCC, CCL-185; PC-9; ECACC, 90071810) cells were cultured in Rosewell Park Memorial Institute (RPMI) 1640 medium (Gibco). The mouse Lewis lung carcinoma cells (LLC; ATCC, CRL-1642) and human embryonic kidney epithelial cells (HEK293T; ATCC CRL-11268) and were grown in Dulbecco's Modified Eagle's Medium (DMEM; Gibco). Both of the DMEM and RPMI 1640 media were supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS; Gibco) and 1% (v/v) penicillin-streptomycin (Gibco). HCA2-TERT cells⁸⁸ (an immortalized foreskin fibroblast cell line) were cultured in Minimum Essential Medium (MEM; Hyclone) supplemented with 10% (v/v) FBS (Gibco), 1% (v/v) penicillin-streptomycin (Gibco) and 1% MEM Nonessential Amino Acids (NEAA; Gibco). The human cell lines were authenticated by short tandem repeat (STR) profiling within 6 months prior to use and only cell lines with $\geq 80\%$ match to the ATCC/DSMZ database records were utilized. The LLC cells were authenticated through species-specific PCR using primers targeting the *Mus musculus* cytochrome b gene and functional validation was performed by assessing tumor formation capability in C57BL/6 mice. Cells were maintained at 37 °C in 5% CO₂. All cells included in this research were free of mycoplasma confirmed by the Look Out Mycoplasma PCR Detection Kit (MP0035, Sigma-Aldrich).

Anchorage-independent growth and tumor formation. Growth of cells in soft agar was performed as previously described⁸⁸. In brief, one million exponentially growing HCA2-TERT fibroblasts were electroporated with vectors encoding SV40 large tumor antigen (LT) (5 µg) and H-Ras V12 (Ras) (5 µg) together with pLKO.1-shCtrl (5 µg) or pLKO.1-shPCSK9 (5 µg) on a Lonza 4D machine. After transfection, cells were seeded and allowed to recover for 24 h. Then 2×10^5 cells were seeded into 0.4% Noble agar and colonies were photographed after 5 weeks of growth. Cultures were coded and the colony numbers were scored in a blinded fashion by a second observer.

Organoid culture and manipulation. Organoids were generated from 8-10 weeks Kras-LSL-G12D mice. Experimental protocol was based on previous studies^{55,89}. In brief, the lung tissue was dissected from mouse and washed 2 times with phosphate buffered saline (PBS). The tissues were minced by scissors and then digested in collagenase D and DNase I in Hank's Balanced Salt Solution at 37 °C for 30 minutes. After incubation, the digested tissues were passed through a 70 µm cell strainer to obtain single-cell suspensions. Then, cells were resuspended in organoid media. Lung epithelial organoids were maintained for successive passages. After expansion, these organoids were subsequently infected with adenovirus-Cre. Organoids were split into 2 or 3 equal aliquots, depending on the experiment, pelleted by pulse spin and resuspended in 100 mL MTEC/Plus media containing 6×10^7 PFU/ml of Adenovirus-Cre, in 100 mL per 100,000 cells. The cells were incubated for 1 h at 37 °C, 5% CO₂ in 1.5 mL tubes. Cells were then pelleted by pulse spin and resuspended in 1x phosphate-buffered saline (PBS). This step was repeated twice for a total of three washing steps.

Allografted tumor model in mice. C57BL/6 female, 6-8-week-old mice were divided into 2 groups (with 6-7 animals per group) for vehicle treatment and CCF642 treatments. Mice were housed in polycarbonate cages, and provided free access to food and water with a 12-h light: dark cycle. All handling procedures were performed by trained

personnel to minimize stress and discomfort. Two million LLC cells mixed in Matrigel (Corning) were injected into each flank of the 6 mice to generate 12-14 allograft tumors per treatment group. When the tumor volumes reached approximately 60-80 mm³, we then randomly assigned them to the different treatment arms. Then, the drugs (PBS and CCF642) with a total volume of 100 µL per treatment were injected directly into the peritoneal. Treatments were carried out twice per week for 4 weeks. On the dosing day, tumors were first measured with calipers, body weights were recorded, and then the drugs were injected directly into the peritoneal. The mice were sedated with isoflurane prior to measurements and treatments. CCF642 was injected at a dose of 10 mg/kg per animal. Tumor volumes were calculated by the formula (length $\times$ width²/2) as described previously⁸⁸. The allografted tumor models were also set up with Pdia3 KD LLC and MANS peptide in the same methods as described above. For MARCKS inhibitor treatment, the drugs (PBS and MANS peptide at 50 nmoles) with a total volume of 100 µL per treatment were intraperitoneal injected every three days in C57BL/6 mice. The maximal tumor size/burden permitted by the ethics committee was defined as a tumor diameter not exceeding 1.5 cm in any dimension for subcutaneous models. In all experiments, we confirmed that the maximal tumor size/burden was not exceeded. Animals were continuously monitored, and mice were euthanized via asphyxiation when any of the following endpoints were met: study termination, tumor ulceration, or moribund appearance. All experiments were approved by Tongji University School of Medicine Animal Care and Use Committee and were conducted in accordance with the National Institutes of Health Guidelines for the Care and Use of Laboratory Animals.

Culture of patient-derived tumor xenograft (PDX) for lung cancer. Female BALB/c Nude mice were purchased from Shanghai Laboratory Animal Center. Mice were 6-8 weeks of age for all experiments, matched for age and sex, and kept under specific pathogen-free (SPF) conditions at the Laboratory Animal Center of Tongji University. PDX models were generated from primary tumors as described previously⁹⁰. Briefly, mice were implanted subcutaneously with patient-derived LUAD tumor fragments. When the average tumor volume was 60-80 mm³, mice were randomized into two groups and administered by PBS or CCF642.

Culture of patient-derived tumor-like cell clusters (PTCs) for lung cancer. Collected surgically resected samples were conditioned in ice-cold PBS with 10 mM HEPES and 100 U/mL penicillin-streptomycin (Thermo Fisher Scientific). Tissues were washed with PBS at least 5 times. Necrotic areas and adipose tissue were removed as possible. Tissues were minced into small pieces and digested in 5 mL PBS/EDTA 1 mM containing collagenase I, II, and IV (Thermo Fisher Scientific) 200 U/mL each, at 37 °C for 1 hour. The digestion was pipetted every 15 minutes to facilitate cell release. 40 µm filters were used to collect dissociated cells. After 10 minutes' centrifugation (300 $\times$ g, 4 °C), cell pellets were resuspended in PTC growth medium and seeded in low-attachment-surface dish at the concentration of 10⁵ cells/cm². Cells were cultured in an incubator at 37 °C, 5% CO₂. PTC growth medium was refreshed every 2-3 days.

PTC drug assay. The PTC drug screens were conducted in 96-well cell culture plates following the procedure described previously^{91,92}. Briefly, PTCs that were more than 40 µm in diameter were collected using 40-µm filters (BD Falcon), centrifuged at 300 $\times$ g and 4 °C for 10 minutes, washed with PBS, and resuspended with G/I-PTC growth medium. Then, 100 µL of a medium containing 30-50 PTCs was seeded into a Teflon-modified chip (GeneX Health, GX-01). Next, 50 µL of PTC growth medium containing the drug was added to the well. Images of each well were screened with the Nikon Ti-U microscope system. The plates were incubated at 37 °C and 5% CO₂. After drug treatment, the

plates were screened again with the Nikon Ti-U microscope system. Compounds were sourced from commercial vendors and stored as 10 mM aliquots.

Cell proliferation. The proliferative capacity of the cells was measured by CCK8 assay (Dojindo, Japan). A total of 5000 cells in a volume of 100 μ L per well were seeded in a 96-well plate. Then, 90 μ L culture medium was mixed with 10 μ L CCK8 reagent to replace the original culture medium in the wells and incubated in 5% CO₂ at 37 °C for 2 h. The absorbance was measured by enzyme-linked immunosorbent assay with a wavelength of 450 nm. We performed this assay at 24 h, 48 h, 72 h and 96 h.

Western blot. Cells were lysed using RIPA Lysis Buffer (Beyotime Biotechnology, China) supplemented with protease inhibitor cocktail (P8340, Sigma-Aldrich), 1 mM of PMSF and phosphatase inhibitor cocktail (P5726, Sigma-Aldrich). The lysates were centrifuged at 13,000 $\times$ g for 10 min and the cellular debris was discarded. For immunoblotting, the protein sample lysate or precipitates were denatured in 1 $\times$ sodium dodecyl sulfate (SDS) protein sample buffer at 95 °C for 10 min and then were resolved by electrophoresis through a 6% or 10% SDS-polyacrylamide gel. Separated proteins were transferred onto polyvinylidene difluoride membranes and were incubated with the prespecified antibodies at indicated dilutions. An enhanced chemiluminescence reagent (Thermo Fisher Scientific) was applied for immunoblotting. The uncropped and unprocessed scans of the most important blots were supplied in the Source Data file.

Reagent or resource. Antibodies, chemicals, peptides, experimental models, cell lines, critical commercial assays, and software information were provided in Supplementary Table 5.

Statistics & reproducibility. Data from independent experiments were expressed as the mean $\pm$ SD and one-way ANOVA followed by Dunnett's post hoc test or two-way ANOVA followed by Tukey's post hoc test was performed for statistical analysis by using GraphPad Prism 8 (GraphPad, San Diego, CA). For all analyses, statistical significance was defined as $p < 0.05$.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The raw proteomics data generated in this study have been deposited in the PRIDE Archive PRoteomics IDentifications database under accession code [PXD046637](https://www.ebi.ac.uk/PRIDE/archive/PXD046637). The raw WES data generated in this study have been deposited in the Genome Sequence Archive (GSA) database under accession code [HRA005849](https://www.genome.gov.cn/hra/HRA005849). The raw data is available under restricted access due to the policy related to human genetic resources in China. Access to the data is granted to researchers for non-commercial academic use upon approval of a data access proposal. Data applicants must meet one of the following criteria: i) be a principal investigator (PI) leading an independent research group at a scientific research organization or hold a senior academic title; ii) hold a professorial or associate professorial position at a university; or iii) serve as a hospital director or associate chief physician. Requests must be submitted through the repository's managed access system (GSA database), accompanied by a signed data use agreement specifying the research purpose and data protection commitments. Data access requests will be processed within 7 working days. The data will be available to the applicants for 2 years once access has been granted. The remaining data are available within the Article, Supplementary

Information or Source Data file. Source data are provided with this paper.

References

- Walter, J. E. et al. New subsolid pulmonary nodules in lung cancer screening: the nelson trial. *J. Thorac. Oncol.* **13**, 1410–1414 (2018).
- Zhang, Y., Fu, F. & Chen, H. Management of ground-glass opacities in the lung cancer spectrum. *Ann. Thorac. Surg.* **110**, 1796–1804 (2020).
- Choi, S., et al. Importance of Lymph Node Evaluation in $\leq$ 2-cm Pure-Solid Non-Small Cell Lung Cancer. *Ann. Thorac. Surg.* **117**, 586–593 (2023).
- Ye, T. et al. Lung adenocarcinomas manifesting as radiological part-solid nodules define a special clinical subtype. *J. Thorac. Oncol.* **14**, 617–627 (2019).
- Flores, R. M. et al. Computed tomography screening for lung cancer: mediastinal lymph node resection in stage ia nonsmall cell lung cancer manifesting as subsolid and solid nodules. *Ann. Surg.* **265**, 1025–1033 (2017).
- Su, H. et al. Which T descriptor is more predictive of recurrence after sublobar resection: whole tumour size versus solid component size? *Eur. J. Cardiothorac. Surg.* **54**, 1028–1036 (2018).
- Su, H. et al. Risk factors of recurrence in patients with clinical stage ia adenocarcinoma presented as ground-glass nodule. *Clin. lung cancer* **19**, e609–e617 (2018).
- Kamigaichi, A. et al. The prognostic impact of the ground-glass opacity component in nearly pure-solid stage IA non-small-cell lung cancer. *Eur. J. Cardiothorac. Surg.* **62**, e2ac166 (2022).
- Hiramatsu, M. et al. Pulmonary ground-glass opacity (GGO) lesions-large size and a history of lung cancer are risk factors for growth. *J. Thorac. Oncol.* **3**, 1245–1250 (2008).
- Sun, F. et al. Ground glass opacities: Imaging, pathology, and gene mutations. *J. Thorac. Cardiovasc Surg.* **156**, 808–813 (2018).
- Travis, W. D. et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. *J. Thorac. Oncol.: Off. Publ. Int. Assoc. Study Lung Cancer* **6**, 244–285 (2011).
- Dejima, H. et al. Immune evolution from preneoplasia to invasive lung adenocarcinomas and underlying molecular features. *Nat. Commun.* **12**, 2722 (2021).
- Hu, X. et al. Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma. *Nat. Commun.* **10**, 2978 (2019).
- Nie, M. et al. Evolutionary metabolic landscape from preneoplasia to invasive lung adenocarcinoma. *Nat. Commun.* **12**, 6479 (2021).
- Xing, X. et al. Decoding the multicellular ecosystem of lung adenocarcinoma manifested as pulmonary subsolid nodules by single-cell RNA sequencing. *Sci. Adv.* **7**, eabd9738 (2021).
- Lu, T. et al. Single-cell transcriptome atlas of lung adenocarcinoma featured with ground glass nodules. *Cell Discov.* **6**, 69 (2020).
- Li, W., Li, F., Zhang, X., Lin, H. K. & Xu, C. Insights into the post-translational modification and its emerging role in shaping the tumor microenvironment. *Signal Transduct. Target. Ther.* **6**, 422 (2021).
- Scupakova, K. et al. Clinical importance of high-mannose, fucosylated, and complex N-glycans in breast cancer metastasis. *JCI Insight* **6**, e146945 (2021).
- Marsico, G., Russo, L., Quondamatteo, F. & Pandit, A. Glycosylation and Integrin Regulation in Cancer. *Trends Cancer* **4**, 537–552 (2018).
- Sun, X. et al. ALG3 contributes to stemness and radioresistance through regulating glycosylation of TGF-beta receptor II in breast cancer. *J. Exp. Clin. Cancer Res* **40**, 149 (2021).

21. Wang, Y. N., Lee, H. H., Hsu, J. L., Yu, D. & Hung, M. C. The impact of PD-L1 N-linked glycosylation on cancer therapy and clinical diagnosis. *J. Biomed. Sci.* **27**, 77 (2020).
22. Pinho, S. S. & Reis, C. A. Glycosylation in cancer: mechanisms and clinical implications. *Nat. Rev. Cancer* **15**, 540–555 (2015).
23. Hart, G. W. & Copeland, R. J. Glycomics hits the big time. *Cell* **143**, 672–676 (2010).
24. Hattori, A., Matsunaga, T., Takamochi, K., Oh, S. & Suzuki, K. Importance of Ground Glass Opacity Component in Clinical Stage IA Radiologic Invasive Lung Cancer. *Ann. Thorac. Surg.* **104**, 313–320 (2017).
25. ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature* **578**, 82–93 (2020).
26. Zhang, Y. et al. Evolutionary proteogenomic landscape from pre-invasive to invasive lung adenocarcinoma. *Cell Rep. Med* **5**, 101358 (2024).
27. Zhang, C. et al. Genomic landscape and immune microenvironment features of preinvasive and early invasive lung adenocarcinoma. *J. Thorac. Oncol.: Off. Publ. Int. Assoc. Study Lung Cancer* **14**, 1912–1923 (2019).
28. Shi, Y. et al. A prospective, molecular epidemiology study of EGFR mutations in Asian patients with advanced non-small-cell lung cancer of adenocarcinoma histology (PIONEER). *J. Thorac. Oncol.: Off. Publ. Int. Assoc. Study Lung Cancer* **9**, 154–162 (2014).
29. Zhang, X. C. et al. Comprehensive genomic and immunological characterization of Chinese non-small cell lung cancer patients. *Nat. Commun.* **10**, 1772 (2019).
30. D'Angelo, S. P. et al. Incidence of EGFR exon 19 deletions and L858R in tumor specimens from men and cigarette smokers with lung adenocarcinomas. *J. Clin. Oncol.: Off. J. Am. Soc. Clin. Oncol.* **29**, 2066–2070 (2011).
31. Rosell, R. et al. Screening for epidermal growth factor receptor mutations in lung cancer. *N. Engl. J. Med.* **361**, 958–967 (2009).
32. Bailey, M. H. et al. Comprehensive characterization of cancer driver genes and mutations. *Cell* **173**, 371–385.e318 (2018).
33. Lu, W., Gong, D., Bar-Sagi, D. & Cole, P. A. Site-specific incorporation of a phosphotyrosine mimetic reveals a role for tyrosine phosphorylation of SHP-2 in cell signaling. *Mol. cell* **8**, 759–769 (2001).
34. Pfeiffer, A. et al. Phosphorylation of SHP2 at Tyr62 Enables Acquired Resistance to SHP2 Allosteric Inhibitors in FLT3-ITD-Driven AML. *Cancer Res.* **82**, 2141–2155 (2022).
35. Patrick, M. E. & Egland, K. A. SUSD2 Proteolytic Cleavage Requires the GDPH Sequence and Inter-Fragment Disulfide Bonds for Surface Presentation of Galectin-1 on Breast Cancer Cells. *Int. J. Mol. Sci.* **20**, 3814 (2019).
36. Sheets, J. N. et al. SUSD2 expression in high-grade serous ovarian cancer correlates with increased patient survival and defective mesothelial clearance. *Oncogenesis* **5**, e264 (2016).
37. Guo, W. et al. Loss of SUSD2 expression correlates with poor prognosis in patients with surgically resected lung adenocarcinoma. *J. Cancer* **11**, 1648–1656 (2020).
38. Xu, J. Y. et al. Integrative proteomic characterization of human lung adenocarcinoma. *Cell* **182**, 245–261.e217 (2020).
39. Gillette, M. A. et al. Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* **182**, 200–225.e235 (2020).
40. Chen, C. H. et al. Targeting myristoylated alanine-rich C kinase substrate phosphorylation site domain in lung cancer. Mechanisms and therapeutic implications. *Am. J. Respir. Crit. Care Med* **190**, 1127–1138 (2014).
41. Reddy, S. P., Natarajan, V. & Dudek, A. Z. MARCKS is marked in combating lung cancer growth and acquired resistance. *Am. J. Respir. Crit. Care Med* **190**, 1084–1086 (2014).
42. Chen, C. H. et al. A peptide that inhibits function of Myristoylated Alanine-Rich C Kinase Substrate (MARCKS) reduces lung cancer metastasis. *Oncogene* **33**, 3696–3706 (2014).
43. Hu, Y. et al. Integrated proteomic and glycoproteomic characterization of human high-grade serous ovarian carcinoma. *Cell Rep.* **33**, 108276 (2020).
44. Ueno, T., Linder, S. & Elmberger, G. Aspartic proteinase napsin is a useful marker for diagnosis of primary lung adenocarcinoma. *Br. J. Cancer* **88**, 1229–1233 (2003).
45. Hart, K. C., Robertson, S. C. & Donoghue, D. J. Identification of tyrosine residues in constitutively activated fibroblast growth factor receptor 3 involved in mitogenesis, Stat activation, and phosphatidylinositol 3-kinase activation. *Mol. Biol. Cell* **12**, 931–942 (2001).
46. Whitt, A. G. et al. Paraoxonase 2 (PON2) plays a limited role in murine lung tumorigenesis. *Sci. Rep.* **13**, 9929 (2023).
47. Bacchetti, T., Ferretti, G. & Sahebkar, A. The role of paraoxonase in cancer. *Semin Cancer Biol.* **56**, 72–86 (2019).
48. Wang, Y. et al. The effects and mechanisms of SLC34A2 in tumorigenesis and progression of human non-small cell lung cancer. *J. Biomed. Sci.* **22**, 52 (2015).
49. Piyush, T. et al. Interaction of galectin-3 with MUC1 on cell surface promotes EGFR dimerization and activation in human epithelial cancer cells. *Cell death Differ.* **24**, 1937–1947 (2017).
50. Pu, Y. et al. GOLM1 restricts colitis and colon tumorigenesis by ensuring Notch signaling equilibrium in intestinal homeostasis. *Signal Transduct. Target. Ther.* **6**, 148 (2021).
51. Aksoy, O., Hantusch, B. & Kenner, L. Emerging role of T3-binding protein μ -crystallin (CRYM) in health and disease. *Trends Endocrinol. Metab.* **33**, 804–816 (2022).
52. Duffy, M. J., Evoy, D. & McDermott, E. W. CA 15-3: uses and limitation as a biomarker for breast cancer. *Clin. Chim. Acta* **411**, 1869–1874 (2010).
53. Hahn, W. C. et al. Creation of human tumour cells with defined genetic elements. *Nature* **400**, 464–468 (1999).
54. Sabatine, M. S. PCSK9 inhibitors: clinical evidence and implementation. *Nat. Rev. Cardiol.* **16**, 155–165 (2019).
55. Dost, A. F. M. et al. Organoids Model Transcriptional Hallmarks of Oncogenic KRAS Activation in Lung Epithelial Progenitor Cells. *Cell Stem Cell* **27**, 663–678.e668 (2020).
56. Barkauskas, C. E. et al. Type 2 alveolar cells are stem cells in adult lung. *J. Clin. Invest* **123**, 3025–3036 (2013).
57. Lin, C. et al. Alveolar type II cells possess the capability of initiating lung tumor development. *PLoS One* **7**, e53817 (2012).
58. Xu, X. et al. Evidence for type II cells as cells of origin of K-Ras-induced distal lung adenocarcinoma. *Proc. Natl Acad. Sci. USA* **109**, 4910–4915 (2012).
59. Lv, G. et al. Multiple functions of HMGB1 in cancer. *Front Oncol.* **14**, 1384109 (2024).
60. Zheng, M. et al. Arterial Stiffness Preceding Diabetes: A Longitudinal Study. *Circ. Res* **127**, 1491–1498 (2020).
61. Alcock, F. & Swanton, E. Mammalian OS-9 is upregulated in response to endoplasmic reticulum stress and facilitates ubiquitination of misfolded glycoproteins. *J. Mol. Biol.* **385**, 1032–1042 (2009).
62. Namba, T. BAP31 regulates mitochondrial function via interaction with Tom40 within ER-mitochondria contact sites. *Sci. Adv.* **5**, eaaw1386 (2019).
63. Jaronen, M., Goldsteins, G. & Koistinaho, J. ER stress and unfolded protein response in amyotrophic lateral sclerosis—a controversial role of protein disulphide isomerase. *Front Cell Neurosci.* **8**, 402 (2014).
64. Perri, E. R., Thomas, C. J., Parakh, S., Spencer, D. M. & Atkin, J. D. The unfolded protein response and the role of protein disulfide isomerase in neurodegeneration. *Front Cell Dev. Biol.* **3**, 80 (2015).

65. Vatolin, S. et al. Novel protein disulfide isomerase inhibitor with anticancer activity in multiple myeloma. *Cancer Res* **76**, 3340–3350 (2016).
66. Liu, H. et al. Pharmacological boosting of cGAS activation sensitizes chemotherapy by enhancing antitumor immunity. *Cell Rep.* **42**, 112275 (2023).
67. Sakamoto, H. et al. **Disproportionate representation of KRAS gene mutation in atypical adenomatous hyperplasia, but even distribution of EGFR gene mutation from preinvasive to invasive adenocarcinomas.** *J. Pathol.* **212**, 287–294 (2007).
68. Chen, H. et al. Genomic and immune profiling of pre-invasive lung adenocarcinoma. *Nat. Commun.* **10**, 5472 (2019).
69. Attard, G., Cooper, C. S. & de Bono, J. S. Steroid hormone receptors in prostate cancer: a hard habit to break? *Cancer cell* **16**, 458–462 (2009).
70. Degirolamo, C., Modica, S., Palasciano, G. & Moschetta, A. Bile acids and colon cancer: Solving the puzzle with nuclear receptors. *Trends Mol. Med* **17**, 564–572 (2011).
71. Finlay-Schultz, J. & Sartorius, C. A. Steroid hormones, steroid receptors, and breast cancer stem cells. *J. Mammary Gland Biol. Neoplasia* **20**, 39–50 (2015).
72. Porter, J. A., Young, K. E. & Beachy, P. A. Cholesterol modification of hedgehog signaling proteins in animal development. *Sci. (N. Y., N. Y.)* **274**, 255–259 (1996).
73. Xiao, X. et al. Cholesterol modification of smoothened is required for hedgehog signaling. *Mol. cell* **66**, 154–162.e110 (2017).
74. Oneyama, C. et al. Transforming potential of Src family kinases is limited by the cholesterol-enriched membrane microdomain. *Mol. Cell Biol.* **29**, 6462–6472 (2009).
75. Sheng, R. et al. Cholesterol selectively activates canonical Wnt signalling over non-canonical Wnt signalling. *Nat. Commun.* **5**, 4393 (2014).
76. Li, M. et al. The association and joint effect of serum cholesterol, glycemic status with the risk of incident cancer among middle-aged and elderly population in china cardiometabolic disease and cancer cohort (4C)-study. *Am. J. Cancer Res* **10**, 975–986 (2020).
77. Battistone, M. F. et al. Increased risk of preneoplastic colonic lesions and colorectal carcinoma in acromegaly: multicenter case-control study. *Pituitary* **24**, 96–103 (2021).
78. Esposito, D., Ragnarsson, O., Johannsson, G. & Olsson, D. S. Incidence of benign and malignant tumors in patients with acromegaly is increased: a nationwide population-based study. *J. Clin. Endocrinol. Metab.* **106**, 3487–3496 (2021).
79. Cubillos-Ruiz, J. R., Bettigole, S. E. & Glimcher, L. H. Tumorigenic and immunosuppressive effects of endoplasmic reticulum stress in cancer. *Cell* **168**, 692–706 (2017).
80. Wang, M. & Kaufman, R. J. The impact of the endoplasmic reticulum protein-folding environment on cancer development. *Nat. Rev. Cancer* **14**, 581–597 (2014).
81. Ghaddar, N. et al. The integrated stress response is tumorigenic and constitutes a therapeutic liability in KRAS-driven lung cancer. *Nat. Commun.* **12**, 4651 (2021).
82. Hasegawa, M. et al. Growth rate of small lung cancers detected on mass CT screening. *Br. J. Radio.* **73**, 1252–1259 (2000).
83. Shen, S. et al. Parallel, high-quality proteomic and targeted metabolomic quantification using laser capture microdissected tissues. *Anal. Chem.* **93**, 8711–8718 (2021).
84. Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* **38**, e164 (2010).
85. Robinson, J. T., Thorvaldsdóttir, H., Wenger, A. M., Zehir, A. & Mesirov, J. P. Variant review with the integrative genomics viewer. *Cancer Res.* **77**, e31–e34 (2017).
86. Riley, N. M., Hebert, A. S., Westphall, M. S. & Coon, J. J. Capturing site-specific heterogeneity with large-scale N-glycoproteome analysis. *Nat. Commun.* **10**, 1311 (2019).
87. Song, C. et al. Systematic analysis of protein phosphorylation networks from phosphoproteomic data. *Mol. Cell. Proteom.: MCP* **11**, 1070–1083 (2012).
88. Liu, H. et al. Nuclear cGAS suppresses DNA repair and promotes tumorigenesis. *Nature* **563**, 131–136 (2018).
89. Pan, Y. et al. KMT2D deficiency drives lung squamous cell carcinoma and hypersensitivity to RTK-RAS inhibition. *Cancer cell* **41**, 88–105.e108 (2023).
90. Daniel, V. C. et al. A primary xenograft model of small-cell lung cancer reveals irreversible changes in gene expression imposed by culture in vitro. *Cancer Res.* **69**, 3364–3373 (2009).
91. Pauli, C. et al. Personalized in vitro and in vivo cancer models to guide precision medicine. *Cancer Discov.* **7**, 462–477 (2017).
92. Vlachogiannis, G. et al. Patient-derived organoids model treatment response of metastatic gastrointestinal cancers. *Science* **359**, 920–926 (2018).
93. Li, C. <https://BioRender.com/n67b524> (2025).

Acknowledgements

This project was supported from the project of National Key R&D Program of China (2023YFC2307002 to H.L.), Creative Research Groups of the National Natural Science Foundation of China (32421002 to H.L.), Shanghai Municipal Health Commission (2023ZZ02025), National Natural Science Foundation of China (82271882 to H.L., 82303553 to L.C.), and Noncommunicable Chronic Diseases-National Science and Technology Major Project (2023ZD0511700), Shanghai Science and Technology Commission (24ZR1460300 to H.S., 22S11900700 to H.L.), and the Shanghai Sailing Program (22YF1451400 to L.C.). H.L. is sponsored by the National Science Fund for Excellent Young Scholars (81922030). We thank Professor Fei Li's laboratory at the School of Basic Medical Sciences, Fudan University, for providing the Kras-LSL-G12D mice.

Author contributions

H.L. and C.C. conceived the study and designed the experiments. H.L., C.C. and W.S. analyzed the data and wrote the original draft. H.S., and L.C. performed most of the experiments with help from Y.R., J.X., Y.D., M.Z., Y.C., J.G. and J.C. T.L., J.Z. J.W., J.L. and Z.C. performed the bioinformatics analysis. C.W., H.X., C.D. and X.H. provided clinical samples.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-025-57364-x>.

Correspondence and requests for materials should be addressed to Chunyan Wu, Wei Sha, Chang Chen or Haipeng Liu.

Peer review information *Nature Communications* thanks Charles Powell, who co-reviewed with Jeremy Mudd, and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025