



Protein glycosylation profiling in lung adenocarcinoma and precursor lesions: analysis of FFPE tissue sections

Jinggang Cui^{1,3,5} · Yan Chen^{1,5} · Shuang Yue^{2,5} · Jianbo Deng^{2,4,5} · Wenqi Zhang^{2,5} · Shengye Wen^{2,4,5} · Lingbo Zhao^{2,5} · Rumeng Zhang^{2,5} · Shuang Yang^{1,2,4,5} · Junhong Jiang¹

Received: 26 February 2026 / Revised: 15 July 2026 / Accepted: 17 July 2026
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2026

Abstract

Protein glycosylation is a major post-translational modification that regulates tumor initiation and progression; however, its dynamic modeling during multistep evolution of lung adenocarcinoma (LUAD) remains poorly understood, particularly in clinically archived tissues. Here, we established an integrated multi-omics workflow combining global proteomes, N-glycans, and site-specific intact N-glycopeptides to comprehensively characterize glycosylation in formalin-fixed paraffin-embedded (FFPE) specimens spanning four pathological stages of LUAD progression: inflammatory nodules (IN), atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), and invasive adenocarcinoma (IAC). Using optimized protein extraction, hydrophilic interaction liquid chromatography (HILIC)-based glycopeptide enrichment, and high-resolution LC-MS/MS, we achieved large-scale identification of proteins, N-glycans, and intact glycopeptides from archival clinical samples. Integrated analyses revealed progressive remodeling of site-specific N-glycosylation during malignant transformation, characterized by increased glycan branching, fucosylation, and sialylation during the transition from premalignant lesions to invasive cancer. Sialylated glycans reached their highest abundance in the premalignant AAH stage, whereas highly branched and fucosylated complex N-glycans predominated in invasive adenocarcinoma, indicating stage-dependent glycan remodeling throughout disease progression. Functional enrichment analyses linked these glycosylation alterations to extracellular matrix organization, neutrophil degranulation, and immune-associated pathways, while representative glycoproteins, including **CEACAM6** and **FGB**, exhibited coordinated changes in protein abundance and site-specific glycoform micro-heterogeneity across pathological stages. Collectively, this study demonstrates the feasibility of deep glycoproteomic profiling using archived FFPE tissues and provides a comprehensive molecular atlas of glycosylation remodeling during LUAD progression. These findings establish a valuable resource for elucidating disease mechanisms and identifying stage-specific glycosylation biomarkers and potential glycan-targeted therapeutic candidates for early lung adenocarcinoma.

Keywords FFPE · Pathological stage · Glycosylation · Lung cancer · Biomarker

✉ Shuang Yang
shuangyang@stu.edu.cn

✉ Junhong Jiang
jiangjunhong1969@suda.edu.cn

¹ Department of Respiratory and Critical Care Medicine, The Fourth Affiliated Hospital of Soochow University, Suzhou 215000, Jiangsu, China

² Laboratory of Clinical and Molecular Glycobiology, Institute of Glycome Study, Shantou University Medical College, Shantou 515041, Guangdong, China

³ Department of Respiratory and Critical Care Medicine, The Affiliated Suzhou Hospital of Nanjing Medical University, Suzhou Municipal Hospital, Gusu School, Nanjing Medical University, Nanjing 215000, Jiangsu, China

⁴ Clinical Medicine Research Center, The First Affiliated Hospital of Shantou University Medical College, Shantou 515041, Guangdong, China

⁵ Center for Clinical Mass Spectrometry, School of Pharmaceutical Sciences, Soochow University, Suzhou 215123, Jiangsu, China

Introduction

Lung adenocarcinoma (LUAD) is the most common histological subtype of lung cancer and remains a leading cause of cancer-related mortality worldwide [1, 2]. A combination of imaging modalities, such as low-dose computed tomography (LDCT) for high-risk screening [3] and positron emission tomography (PET) scans [4], followed by tissue confirmation, has been routinely practiced in LUAD diagnosis. Although advances in imaging, molecular diagnostics, and targeted therapies have substantially improved clinical management of LUAD [5], patient prognosis remains highly dependent on disease stage at diagnosis, with early-stage lesions exhibiting significantly better survival outcomes following surgical intervention [6]. LUAD develops through a well-defined multistep pathological continuum, progressing from inflammatory nodules (IN) and atypical adenomatous hyperplasia (AAH) to adenocarcinoma in situ (AIS) and ultimately invasive adenocarcinoma (IAC) [7]. However, accurately distinguishing these stages, particularly during the early disease phases, remains a major clinical challenge despite its critical importance for timely diagnosis, risk stratification, and optimal therapeutic decision-making [8]. While the genomic and transcriptomic alterations associated with LUAD progression have been extensively investigated, the post-translational regulatory mechanisms that drive disease initiation and progression remain incompletely understood [9, 10].

Protein glycosylation is a pervasive and dynamic post-translational modification (PTM) essential for protein folding, stability, trafficking, and intercellular communication [11, 12]. In cancer development, aberrant glycosylation, including altered branching, hyperfucosylation, and hypersialylation, plays a pivotal role in promoting tumor cell invasion, immune evasion, angiogenesis, and metastatic dissemination [13–15]. In LUAD, dysregulated N-glycosylation has been implicated in several key oncogenic processes, including cell adhesion [16], extracellular matrix (ECM) remodeling [17], and the organization of the tumor-immune microenvironment [18]. Despite these advances, most studies have focused on advanced-stage disease or have been conducted using cell lines and fresh-frozen tissues, leaving the dynamic evolution of glycosylation across the full pathological spectrum of LUAD largely unexplored. This limitation is particularly significant given that formalin-fixed paraffin-embedded (FFPE) specimens, which represent the standard tissue format in routine clinical practice [19], remain under-investigated in glycoproteomic studies.

Characterizing alterations in protein glycosylation within lung tissues is essential for deciphering the molecular evolution of glycosylation across disease stages [20,

21]; however, the scarcity of fresh tissue—particularly from early-stage patients—presents significant ethical and logistical barriers [22, 23]. FFPE tissues help overcome this challenge as the most abundant and well-annotated resource in clinical archives, enabling large-scale retrospective molecular studies correlated with long-term clinical outcomes [24, 25]. While the extensive protein cross-linking and chemical modifications inherent to formalin fixation historically hindered glycoproteomic depth [26], recent innovations in sample preparation and high-resolution mass spectrometry (MS) have overcome these technical constraints [27, 28]. Specifically, the integration of hydrophilic interaction liquid chromatography (HILIC)-based enrichment strategies, such as Amide-80, has enabled large-scale, site-specific N-glycosylation profiling [29, 30]. This approach allows simultaneous characterization of protein identity, glycosylation site occupancy, and glycan structural heterogeneity, thereby providing a comprehensive view of protein glycosylation directly from archived clinical specimens.

While high-resolution mass spectrometry and HILIC-based enrichment are well-established techniques [31, 32], their application to long-term archived FFPE tissues remains challenging due to extensive formalin-induced cross-linking, progressive sample degradation, and abundant chemical modification artifacts. Consequently, conventional glycoproteomic workflows developed for fresh-frozen tissues often exhibit reduced sensitivity, limited glycopeptide coverage, and elevated false-discovery rates when applied to these compromised clinical specimens [33]. To address this methodological limitation, we developed an integrated workflow specifically tailored for archival FFPE samples, incorporating systematic optimization of de-cross-linking conditions, refinement of the HILIC enrichment microenvironment to reduce FFPE-associated background interference, and an artifact-aware bioinformatics strategy. By minimizing search-space inflation caused by formalin-induced modifications and other chemical artifacts, this workflow substantially improves the depth, accuracy, and confidence of site-specific glycoform identification, thereby enabling robust glycoproteomic analysis of retrospective clinical biobank specimens.

In this study, we employed an integrated multi-omics workflow encompassing global proteomics, N-glycomics, and intact N-glycoproteomics to systematically characterize glycosylation dynamics across four pathological stages of LUAD using archival FFPE tissue specimens (Fig. 1 and Electronic Supplementary Material Figure S1–S2). The experimental workflow included tissue sectioning, slide mounting, deparaffinization, protein extraction, and trypsin digestion, followed by peptide purification, glycopeptide enrichment, and LC–MS/MS analysis. By combining

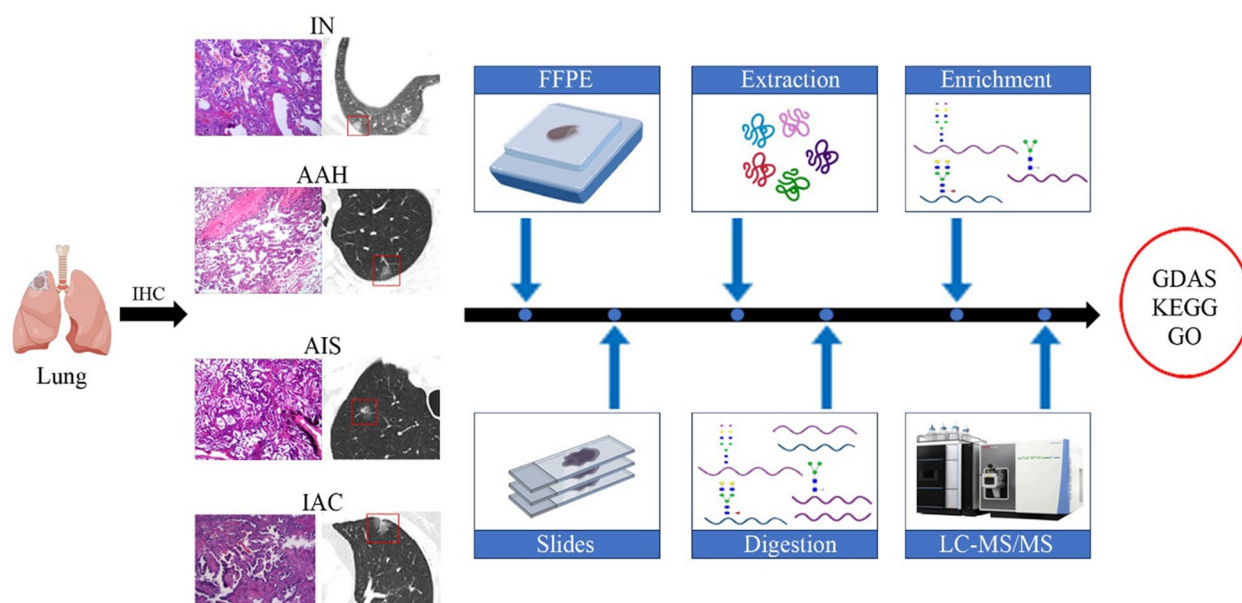


Fig. 1 Schematic workflow for the multi-omics analysis of proteins, glycans, and glycopeptides from FFPE lung tissues. Clinical specimens representing four diagnostic groups— inflammatory nodule (IN), atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), and invasive adenocarcinoma (IAC)—were collected by immunohistochemistry (IHC)-guided tissue dissection. Proteins were extracted from formalin-fixed paraffin-embedded (FFPE) tissue sections through deparaffinization, rehydration, and heat-induced reversal of formalin cross-links. The extracted proteins were subjected to glycoprotein immobilization for glycan extraction (GIG) for glycan

profiling, while parallel aliquots were processed by tryptic digestion for global proteomic and glycoproteomic analyses. Peptides were subsequently enriched using C18 solid-phase extraction (SPE) for proteomic analysis and Amide-80 hydrophilic interaction liquid chromatography (HILIC) for glycopeptide enrichment. Following liquid chromatography–tandem mass spectrometry (LC–MS/MS) analysis, the datasets were processed using GDAS and further annotated through KEGG and GO pathway analyses to characterize molecular alterations and elucidate their potential biological functions during lung adenocarcinoma progression

HILIC-based enrichment with high-resolution LC–MS/MS, we characterized stage-specific alterations in N-glycan structures and site-specific glycoforms, revealing a progressive increase in glycan branching, fucosylation, and sialylation during malignant progression. Functional network analyses based on KEGG (Kyoto Encyclopedia of Genes and Genomes) [34] and GO (Gene Ontology) [35] linked these glycosylation signatures to extracellular matrix (ECM) organization, neutrophil degranulation, and other tumor-associated biological processes. Collectively, our findings highlight the utility of FFPE-based glycoproteomic profiling for elucidating the molecular mechanisms underlying LUAD progression and establish a robust framework for the discovery of disease-specific glycosylation biomarkers for early detection and clinical risk stratification.

Experimental methods

Reagents and materials

All chemicals and materials were ordered from Beyotime (Shanghai, China) unless otherwise noted. Iodoacetamide (IAA), dithiothreitol (DTT), and urea were purchased from

Aladdin Chemical (Shanghai, China). Ammonium bicarbonate (NH_4HCO_3) and formic acid (FA) were obtained from Sinopharm Chemical Reagent Co., Ltd (Shanghai, China). Acetonitrile (ACN) was purchased from Tedia (Fairfield, OH, USA), and trifluoroacetic acid (TFA) was ordered from Macklin (Shanghai, China). High-performance liquid chromatography (HPLC) grade water was purchased from J&K Chemical (Zhejiang, China). Sequencing-grade trypsin was purchased from Promega (Madison, WI, USA), Amide-80 gel slurry was ordered from Tosoh Bioscience (Tokyo, Japan), and Silica C18 resin was purchased from Silicycle (Quebec, QC, Canada). Environmentally friendly transparent dewaxing liquid was purchased from Solarbio (Beijing, China).

Pathological tissue collection

Tissue specimens were obtained from patients undergoing surgical resection and were routinely fixed in formalin and embedded in paraffin for pathological diagnosis. According to the histopathological evaluation, patients were classified into four groups: inflammatory nodule (IN), atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), and invasive adenocarcinoma (IAC) (Table 1). No statistically significant

Table 1 Pathological group classification. These four groups classify the progressive stages of lung adenocarcinoma development, showing the spectrum of lesions from benign inflammation through premalignant changes to fully malignant and invasive cancer

Group	Full name	Description	Significance in progression
IN	Inflammatory node	A benign lesion primarily composed of inflammatory cells, representing the body's reaction to infection, injury, or foreign material	Benign. Not pre-cancerous, but can sometimes mimic early cancer on imaging
AAH	Atypical adenomatous hyperplasia	A premalignant lesion, characterized by a small, localized proliferation of atypical epithelial cells that line the airspaces; cancer cell feature but confined to the alveolar walls	Pre-cancerous, the earliest identifiable precursor to adenocarcinoma
AIS	Adenocarcinoma in situ	A non-invasive malignant lesion, small and composed of malignant cells that grow exclusively along the pre-existing alveolar structures; no evidence of stromal invasion	Noninvasive, surgical removal typically curative
IAC	Invasive adenocarcinoma	A fully malignant lesion where the tumor cells have breached the basement membrane and invaded the surrounding stroma; metastasis	Invasive cancer, worse prognosis than AIS

differences were observed among the four groups with respect to age, sex, smoking history, or other demographic and clinical characteristics. All samples were collected in accordance with protocols approved by the Institutional Review Board (IRB) of the Fourth Affiliated Hospital of Soochow University and Suzhou Municipal Hospital, and written informed consent was obtained from all participants prior to sample collection.

Protein extraction from paraffin-fixed tissue

For protein extraction, several 10- μ m-thick tissue sections were cut from FFPE blocks containing either tumor or inflammatory tissue. These sections were deparaffinized using an environmentally friendly deparaffinization solution (twice, 20 min each) and subsequently rehydrated through a graded ethanol series (100%, 85%, 75%, and 50% EtOH in water) for 10 min at room temperature (RT) at each step, followed by a final wash with DI water [28]. The rehydrated tissues were then scraped from the glass slides and lysed in 1 mL of lysis buffer (RIPA containing 2% SDS and 0.1 M DTT) by incubation at 95 °C for 1 h, followed by an additional incubation at 65 °C for 1 h. The lysates were centrifuged at 13,000 $\times$ g for 10 min at 4 °C to remove insoluble debris, and the resulting supernatant was collected. Proteins were precipitated by adding six volumes of anhydrous ethanol and incubating the samples at -20 °C overnight. Following centrifugation, the supernatant was discarded, and the protein pellets were washed three times with 2 mL of anhydrous EtOH, discarding the solvent after each wash. The final protein pellets were reconstituted in 1 $\times$ PBS, and protein concentrations were measured using a BCA assay.

Intact N-linked glycopeptide enrichment

To enrich intact N-glycopeptides, samples were dissolved in 100 μ L of an 8 M urea solution (prepared in 100 mM NH_4HCO_3), as described in our previous studies [36, 37]. The proteins were reduced by adding 15 μ L of 120 mM TCEP and incubating at 37 °C for 1 h, followed by alkylation with 16.5 μ L of 160 mM IAA at RT for 1 h in the dark. Prior to digestion, samples were diluted fivefold with HPLC-grade water to reduce the urea concentration and then digested overnight at 37 °C with 20 μ g of sequencing-grade trypsin. Digestion was stopped by adjusting the sample pH to less than 3 with 10% formic acid (FA), and the resulting peptides were purified using solid-phase extraction (SPE) cartridges packed with C18 silica resin (200 μ L of slurry prepared in 50% methanol). The C18 column was first conditioned by washing three times with 1 mL of 80% acetonitrile (ACN) in 0.1% trifluoroacetic acid (TFA) and subsequently three times with 1 mL of 0.1% TFA. Peptides were loaded onto the C18 column, and the flow-through was collected and reloaded once to maximize peptide binding. The column was washed five times with 1 mL of 0.1% TFA. Bound peptides were then eluted three times with 300 μ L of 80% ACN in 0.1% TFA.

These purified peptides were subsequently processed for intact glycopeptide enrichment via hydrophilic interaction liquid chromatography (HILIC) using a SPE column packed with Amide-80 resin [29, 38]. The HILIC column was conditioned with three washes of 800 μ L of 0.1% TFA, followed by three washes of 800 μ L of 80% ACN in 0.1% TFA. After the peptide sample was loaded twice, the column was washed three times with 80% ACN in 0.1% TFA. Bound glycopeptides were then sequentially eluted with 800 μ L each of 60% ACN in 0.1% TFA, 40% ACN in 0.1% TFA, and finally 0.1% TFA. The combined eluate was dried in a vacuum concentrator and dissolved in 0.2% FA for subsequent analysis.

Mass spectrometry analysis of intact glycopeptides

Intact glycopeptides were separated using an Easy-nLC 1200 system coupled directly to a Thermo Orbitrap Fusion Lumos mass spectrometer. Approximately 1 µg of intact peptide sample was loaded onto an EASY-Spray™ column (75 µm × 50 cm) packed with 2 µm C18 resin (100 Å) and operated at a flow rate of 250 nL/min. Chromatographic separation was performed over a 120-min gradient using mobile phase A (0.1% FA in 5% ACN) and mobile phase B (0.1% FA in 80% ACN). The gradient began with a 0.1-min hold at 6% B, followed by a linear increase to 22% B over 54 min, then to 50% B over the next 51 min, and 60% B over 5 min. The run was completed with a rapid increase to 90% B within 1 min, followed by a 9-min hold at 90% B to ensure column washing and elution of strongly retained species.

Peptides were ionized by electrospray ionization at a spray voltage of 2.2 kV and analyzed on an Orbitrap mass spectrometer operated in data-dependent acquisition mode. Full MS scans were acquired over an *m/z* range of 300–1600 at a resolution of 60,000 (at *m/z* 200), with a maximum injection time of 20 ms and a dynamic exclusion duration of 30 ms. The 20 most intense precursor ions with signal intensities greater than 6×10^4 counts were selected for high-energy collisional dissociation (HCD) fragmentation using an isolation window of 1.2 *m/z*, while singly charged or unassigned precursor ions were excluded. HCD-MS/MS (MS2) spectra were acquired at a resolution of 15,000 (at *m/z* 200) with a maximum injection time of 30 s. The automatic gain control (AGC) targets were set to 3×10^6 for MS1 and 1×10^5 for MS2, with a normalized collision energy of 27% and an underfilled ratio of 1%. For intact glycopeptide analysis, stepped HCD-MS/MS spectra were acquired using normalized collision energies of 15%, 25%, and 35% for precursor ions with charge states ranging from 2 to 8 and intensities exceeding 5×10^4 . MS/MS spectra were collected at a resolution of 30,000 (at *m/z* 200), and dynamic exclusion was set to 20 s. At least one characteristic glycan oxonium ion (*m/z* 138.0545, 204.0867, or 366.1396) was required to be present among the 20 most abundant fragment ions in the MS/MS spectrum, within a mass tolerance of 15 ppm. The mass spectra were analyzed using GDAS software, as described in detail in the Supporting Information [39]. Glycopeptide identification was performed using a custom N-glycan database comprising 132 human N-glycan compositions, which was established based on our previous glycomics studies of lung tissue [40], serum [41], bronchoalveolar lavage fluid [15], and saliva samples [42] from patients with lung cancer and healthy controls.

Results and discussion

Differential expression of tissue proteome across pathological states

A total of 3835 proteins were identified from four pathological stages of lung tissues (Electronic Supplementary Material Table S1). The total ion intensity for each stage was calculated from nine different patient tissue specimens, in which the intensity was estimated according to the aggregated area under the curve (AUC) for all detected peptides. The heatmap reveals a distinct proteomic progression from IN and AAH stages to the more advanced AIS and IAC. In the early stages, protein expression is largely dominated by pathways related to immune cell chemotaxis and angiogenesis. However, as the pathology transitions to IAC, there is a sharp upregulation of proteins associated with cell proliferation, epithelial-mesenchymal transition (EMT), and chromosomal instability. While the expression profiles of AAH and AIS are relatively similar, reflecting their pre-invasive nature, IAC demonstrates a unique cluster of significantly altered proteins that facilitate basement membrane penetration and stromal invasion (Fig. 2A).

Interestingly, hierarchical clustering of the global proteomic and glycoproteomic datasets (Fig. 2A) revealed distinct molecular subclusters, in which samples from the IN, AAH, AIS, and IAC groups exhibited an intermixed distribution rather than a strictly stage-dependent separation. To determine whether this pattern could be attributed to technical artifacts, experimental variability was carefully controlled through randomized sample processing, standardized analytical procedures, and continuous LC-MS/MS quality control monitoring. Subsequent quality assessments indicated no evidence of significant batch effects that could account for the observed clustering. Instead, the intermixed distribution likely reflects the substantial biological heterogeneity and molecular continuity characteristic of early lung adenocarcinoma development. Because all samples were stringently matched for key clinical parameters, including stage I disease status, presentation as pulmonary nodules detected by chest CT, and tumor size ≤ 3 cm, the overlapping molecular features among AAH, AIS, and IAC lesions suggest that alterations in protein expression and site-specific glycosylation occur along a dynamic pathological progression continuum rather than within discrete clinical stages.

The Venn diagram analysis of IN, AAH, AIS, and IAC in Fig. 2B demonstrates a progressive overlap in differentially expressed proteins (DEPs), reflecting a stepwise molecular evolution. While IN and AAH show relatively distinct profiles, the number of shared genes significantly increases during the transition from AIS to IAC, indicating a common set of drivers for malignancy. KEGG and GO analysis further reveals that while early stages (AAH/AIS) are enriched for

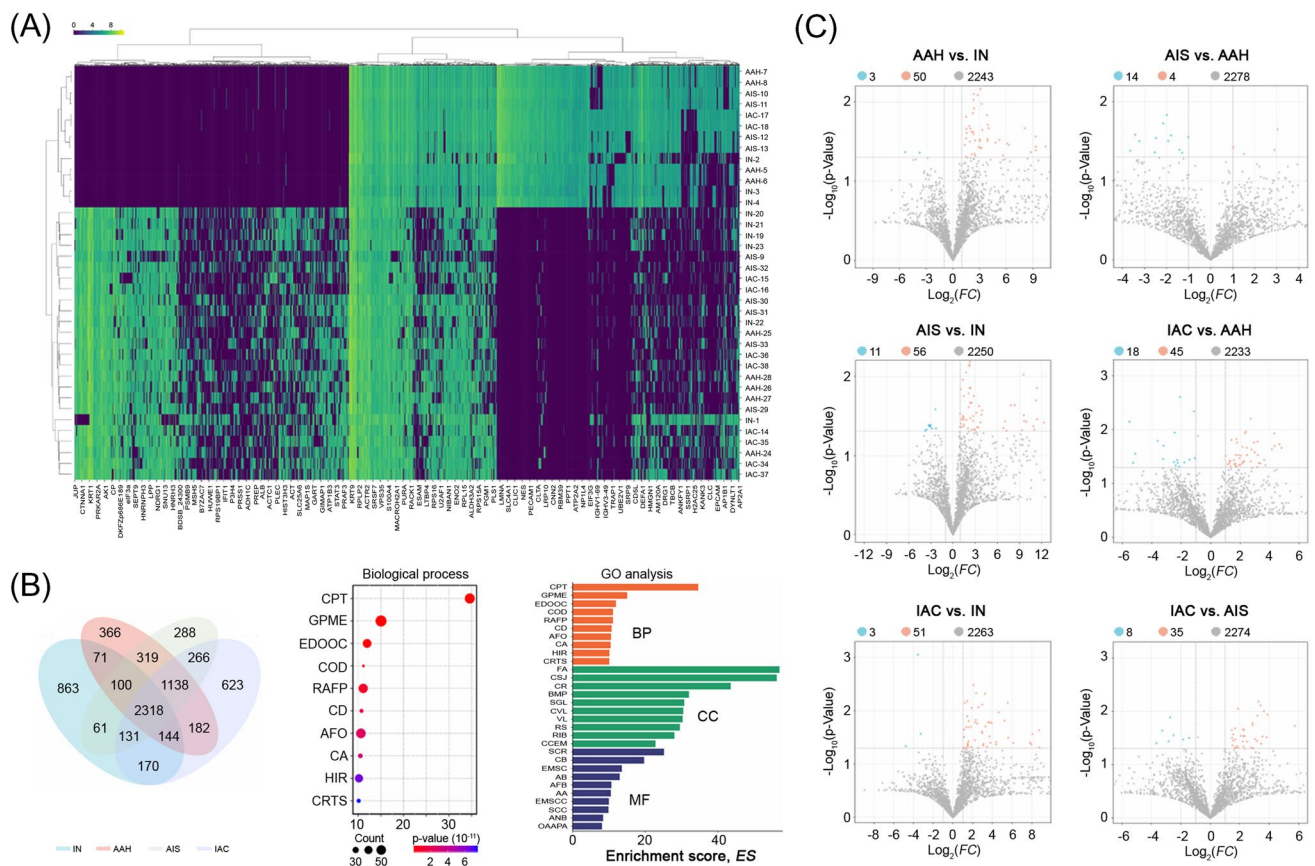


Fig. 2 Global proteomic profiling and differential expression analysis across four lung tissue diagnostic groups. **A** Hierarchical clustering heatmap showing protein expression patterns among inflammatory nodules (IN), atypical adenomatous hyperplasia (AAH), adenocarcinoma in situ (AIS), and invasive adenocarcinoma (IAC) samples. **B** Venn diagram illustrating the overlap and stage-specific distribution of identified proteins, including 2,318 proteins shared across all four groups. Corresponding KEGG and Gene Ontology (GO) enrichment

analyses revealed significantly enriched biological processes (BP), cellular components (CC), and molecular functions (MF) associated with the transition from inflammatory lesions to invasive adenocarcinoma. **C** Volcano plots displaying differentially expressed proteins in pairwise comparisons among diagnostic groups, including AAH, AIS, and IAC versus IN, as well as sequential comparisons of AIS vs. AAH, IAC vs. AAH, and IAC vs. AIS

pathways related to cell–cell adhesion and metabolic shifts, the IAC stage is uniquely characterized by the activation of pathways associated with EMT, focal adhesion, and cell-cycle deregulation. The volcano plots in Fig. 2C illustrate the differential protein expression patterns across the four pathological stages, with blue dots representing upregulated proteins, pink representing downregulated proteins, and gray indicating no significant change. The comparisons—such as AAH vs. IN, AIS vs. AAH, and IAC vs. AIS—reveal that the number of significantly altered proteins varies across these transitions, with a notable shift in expression profiles as the tissue moves from pre-invasive to invasive states. In the transition to IAC, specifically in the IAC vs. AIS and IAC vs. IN plots, there is a distinct emergence of upregulated (blue) and downregulated (pink) proteins, suggesting that the acquisition of invasive potential is driven by specific proteomics changes. Overall, the relatively sparse distribution of

significant dots in comparisons like AIS vs. AAH compared to those involving IAC highlights that the most profound molecular divergence occurs during the final step of malignant progression into invasive disease (Table 2).

N-glycosylation alteration from early stage to the invasive adenocarcinoma

In Fig. 3, the analysis of glycoproteins across the four pathological stages (IN, AAH, AIS, and IAC) reveals significant changes in glycoprotein expression. Figure 3A uses a Venn diagram to illustrate that while a core set of glycoproteins is conserved across all stages, the transition to IAC is marked by a unique set of identified glycoproteins (Electronic Supplementary Material Table S2). KEGG and GO analysis further demonstrates that these glycoproteins are significantly enriched in

Table 2 Summary of clinicopathological characteristics for patients with formalin-fixed paraffin-embedded (FFPE) tissue samples. All specimens were obtained in accordance with institutional protocols following informed patient consent

	IN	AAH	AIS	IAC
Numbers	9	9	10	10
Age $\pm$ SD	55.22 $\pm$ 13.06	56.33 $\pm$ 11.96	58.1 $\pm$ 7.88	61.9 $\pm$ 10.59
Gender				
Male	4	4	5	5
Female	5	5	5	5
CD				
Hypertension	2	0	1	3
Diabetes	0	1	0	1
ND, mm	23.11 $\pm$ 9.85	7.56 $\pm$ 1.24	7.9 $\pm$ 1.66	17.7 $\pm$ 7.99
Nodule property				
GGO	1	8	8	5
Mixed-GGO	1	1	2	2
Solid	7	0	0	3

*IN, inflammatory nodule; AAH, atypical adenomatous hyperplasia; AIS, adenocarcinoma in situ; IAC, infiltrating adenocarcinoma; GGO, ground glass nodule; SD, standard deviation; CD, complicating disease; ND, nodule diameter

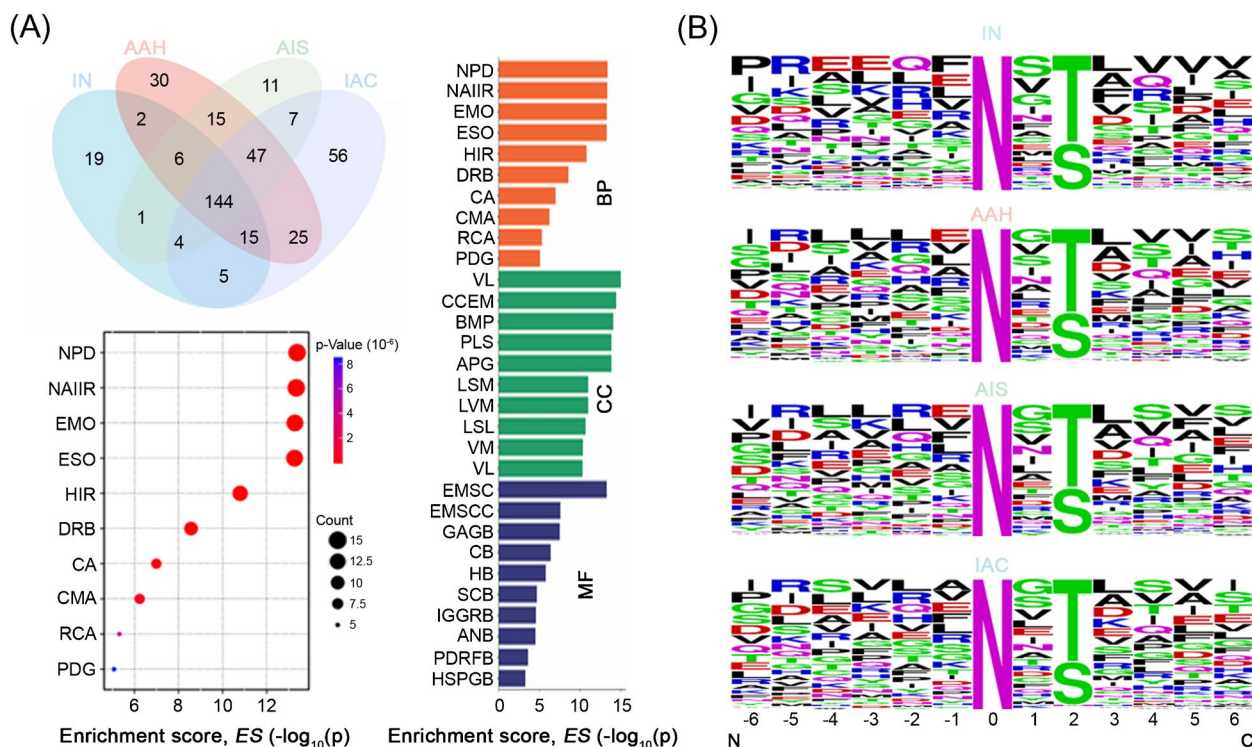


Fig. 3 Glycoprotein profiling and site-specific N-glycosylation motif analysis across lung tissue progression. **A** Venn diagram illustrating the distribution of identified glycoproteins among the four diagnostic groups, including 144 core glycoproteins shared across all stages and stage-specific glycoproteins identified in IN (19), AAH (30), AIS (11), and IAC (56). Corresponding KEGG and Gene Ontology (GO) enrichment analyses reveal significantly enriched biological processes

(BP), such as neutrophil degranulation (NPD) and neutrophil activation involved in immune response (NAIIR), along with key cellular components (CC) and molecular functions (MF). **B** Sequence motif analysis of site-specific N-glycosylation illustrating the frequency and conservation of amino acids surrounding the canonical N-X-S/T sequon, where N represents asparagine, X is any amino acid except proline, and S/T denotes serine or threonine

biological processes (BP) such as cell adhesion, extracellular matrix organization, and metabolic reprogramming, with cellular components (CC) enrichment in the plasma membrane and extracellular region, and molecular functions (MF) enriched in receptor binding and catalytic activities. Figure 3B highlights that the N-glycosylation sequon remains consistent with the canonical N-X-S/T (asparagine-X-serine/threonine, where X is not proline) sequence across all stages; however, the specific amino acid frequencies surrounding these sites (represented by the letters in Fig. 3B) vary, potentially reflecting the aberrant glycosylation patterns—such as increased branching or sialylation—that drive tumor progression and invasive potential in lung adenocarcinoma.

Comparison of the altered N-glycan profile from early-stage to invasive adenocarcinoma

The N-glycan profiling across the four pathological stages reveals a dynamic restructuring of the glyco-landscape as lung tissue progresses from inflammatory nodules (IN) toward invasive adenocarcinoma (IAC) (Fig. 4). Quantitative analysis demonstrates a general trend of upregulation for high-mannose, hybrid, and complex-type glycans—including H5N2F1, S1H5N4, and S3H6N5—in the AAH, AIS, and IAC groups relative to IN. Interestingly, while specific sialylated structures like S1H5N4, S2H6N5, and S2H6N5F1 reach peak abundance early in the AAH stage, complex branched and fucosylated structures such as H5N2F1, H6N3, S1H6N3, and S3H6N5,

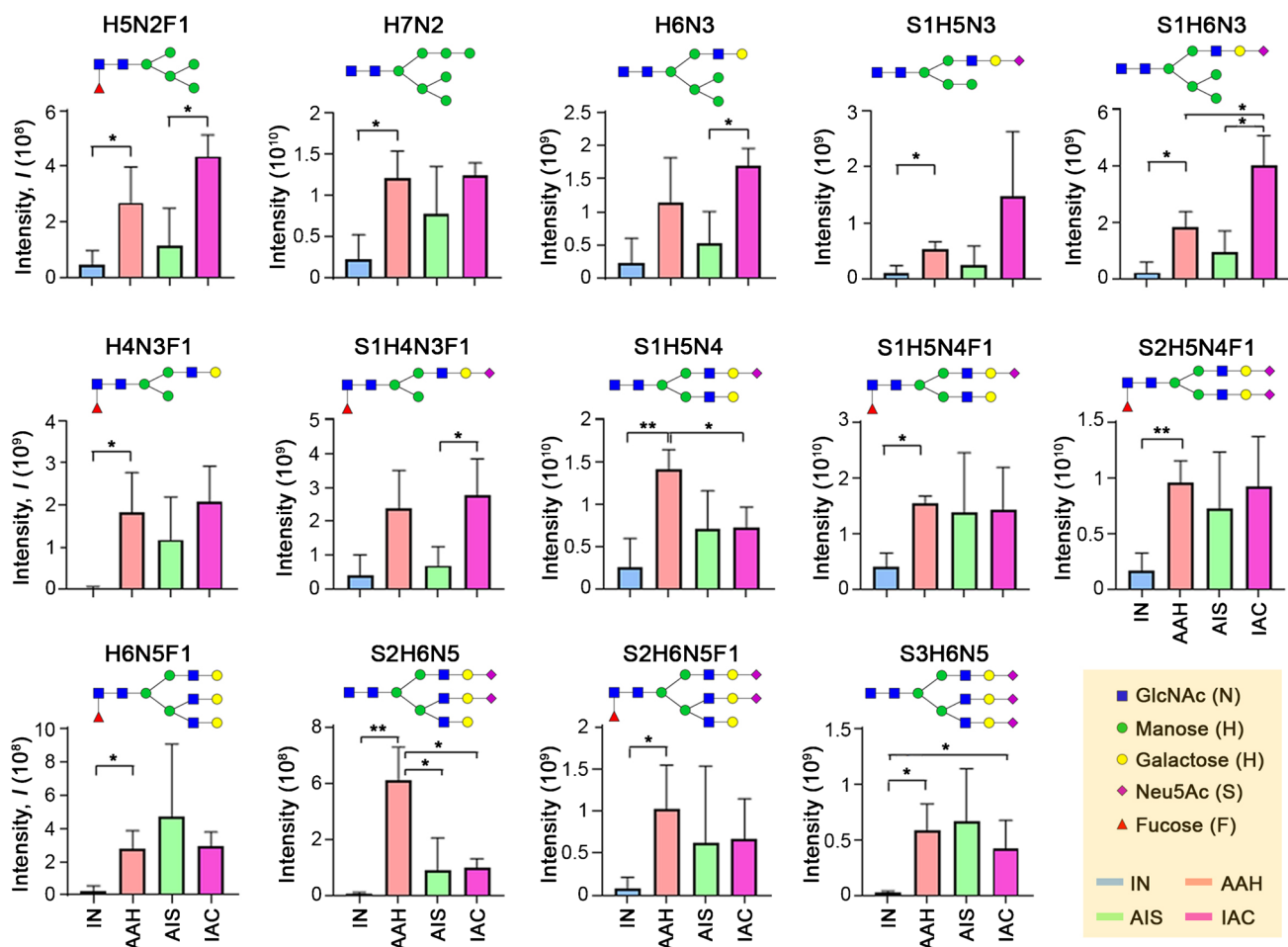


Fig. 4 Profiling of 14 significantly altered N-glycan species across four diagnostic groups. The identified N-glycans, including high-mannose, hybrid, and complex structures (e.g., H5N2F1, S1H5N4, and S3H6N5), exhibited an overall trend of increased abundance in AAH, AIS, and IAC relative to IN. Notably, several sialylated glycans, including S1H5N4, S2H6N5, and S2H6N5F1, reached their highest abundance in the AAH group. In contrast, complex, branched,

and fucosylated glycans, such as H5N2F1, H6N3, S1H6N3, and S1H4N3F1, were most abundant in the IAC group, suggesting a progressive increase in glycan complexity during LUAD development. Statistical significance is indicated by asterisks (* $p < 0.05$, ** $p < 0.01$), and error bars represent the standard deviation of biological replicates

S1H6N3, and S1H4N3F1 exhibit their most significant increases ($p < 0.05$ or $p < 0.01$) in the IAC group (H =hexose, N =HexNAc, S =Neu5Ac, F =fucose). This shift from early metabolic alterations in precursor lesions to the accumulation of highly branched, sialylated, and fucosylated N-glycans in invasive disease suggests a strong correlation between glycan complexity and malignant progression [43, 44]. These aberrant glycosylation patterns likely facilitate the transition to malignancy by modulating cell–cell signaling and enhancing the invasive potential of adenocarcinoma cells [45].

The interplay of biological processes

The network diagram illustrates the complex interplay between several major biological processes through

their shared and unique protein expressions (Fig. 5A). The network identifies key functional clusters centered around humoral immune response (HES), complement activation (CSIV/CCA), and neutrophil degranulation (NDG), alongside structural and regulatory processes like extracellular matrix (ECM) organization and neutrophil activation involved in immune response (NCM). Platelet-related activities, specifically platelet degranulation (PRCM/PRLM) and platelet-derived growth factor binding (PRCA), act as critical bridges within the network, sharing numerous proteins with both immune and structural components. While each process maintains a specific set of proteins (represented in the outer circles), the central overlap highlights highly integrated nodes—such as those involving the ECM and NDG—suggesting a coordinated biological response where structural

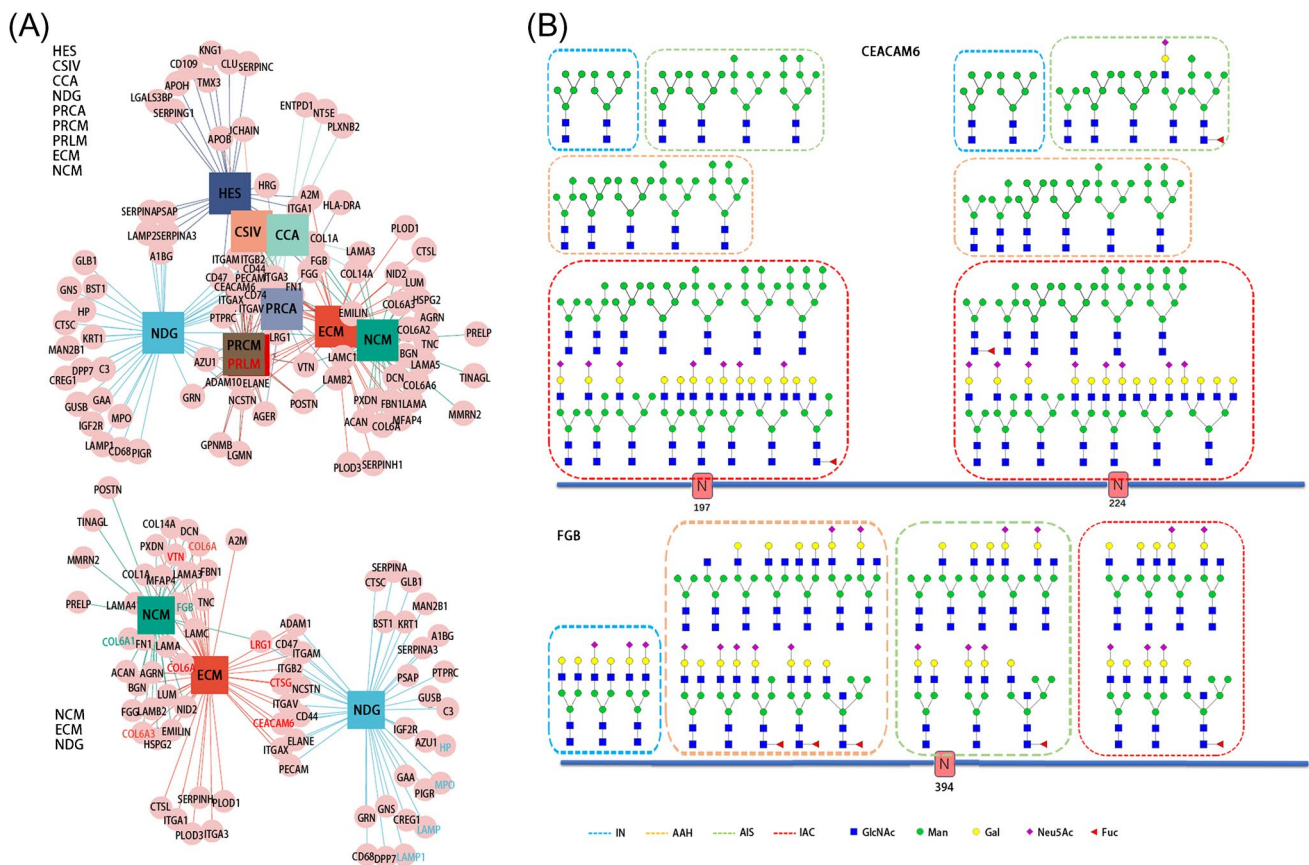


Fig. 5 Network analysis of differentially regulated glycoproteins and representative site-specific glycoforms. **A** Network-based Venn diagram showing differentially regulated glycoproteins associated with key biological processes, including hemostasis (HES), cell surface interactions at the vascular wall (CSIV), cell–cell adhesion (CCA), neutrophil degranulation (NDG), positive regulation of cell adhesion (PRCA), extracellular matrix (ECM) organization, NABA core matrisome (NCM), positive regulation of locomotion (PRLM), and positive regulation of cell migration (PRCM). Glycoproteins shared among the NCM, ECM, and NDG clusters are highlighted to show

functional connectivity and overlap between extracellular matrix remodeling and immune-related pathways. **B** Representative site-specific glycoforms of CEACAM6 and FGB are presented as examples of cancer-specific glycoproteins. CEACAM6 is implicated in ECM and NDG, whereas FGB is associated with NCM and ECM processes. Glycosylation profiles are shown across the four diagnostic groups—IN, AAH, AIS, and IAC—revealing progressive alterations in site-specific glycosylation, characterized by increased glycan branching and structural complexity at glycosylation sites N197, N224, and N394 during disease progression

remodeling and immune cell activation are functionally linked.

As shown in Fig. 5B, site-specific protein glycosylation involved in extracellular matrix organization (ECM) and neutrophil activation (NCM) in immune response is characterized by diverse glycan structures at specific asparagine residues in **CEACAM6** and **FGB**. CEACAM6, which is associated with NCM, displays significant heterogeneity across two specific N-glycosylation sites, N197 and N224; these sites feature high-mannose (Man) and complex-type glycans with various terminal modifications, including galactose (Gal), N-acetylneuraminic acid (Neu5Ac), and fucose (Fuc). These glycoforms are categorized across the four distinct structural patterns—IN, AAH, AIS, and IAC—with the IAC group exhibiting the most complex and branched glycan profiles at both sites. Our integrated multi-omics analysis revealed a distinct two-level dysregulation pattern for CEACAM6. Specifically, proteomic profiling demonstrated significant changes in CEACAM6 protein abundance during the progression from IN to IAC. Concurrently, intact glycoproteomic analysis revealed alterations in its site-specific glycosylation profile, characterized by increased abundances of high-mannose and sialylated glycoforms at the Asn-97 and Asn224 glycosylation sites. These findings indicate that disease progression is associated not only with changes in CEACAM6 protein expression but also with remodeling of its site-specific glycosylation landscape, highlighting the complementary contributions of protein abundance and glycoform micro-heterogeneity to the evolving pathological microenvironment. Interestingly, the survival data indicate that patients with low *CEACAM6* expression exhibited significantly shorter survival times compared to the high CEACAM6 expression group. In contrast, *CEACAM5* expression showed no significant correlation with the overall survival (Electronic Supplementary Material Figure S3). Similarly, the FGB protein, associated with both ECM and NCM, exhibits a diverse glycan structures at site N394, predominantly featuring complex bi-antennary and tri-antennary glycans with sialylation and fucosylation across the same four categorical groups. The site-specific glycoforms for other key glycoproteins—including **MPO**, **PIGR**, **LAMP1**, **COL6A1**, **COL6A2**, **VTN**, **HP**, **A1BG**, and **GRN**—are illustrated in Electronic Supplementary Material Figure S4. Because some of these glycoproteins exhibit distinct glycan profiles in other conditions, such as Parkinson's disease [30] or diabetes [46], they may represent potential disease-specific markers. Consequently,

targeting these unique glycopatterns may improve their utility as diagnostic or prognostic indicators.

Conclusion

In conclusion, this study established an integrated multi-omics workflow for comprehensive profiling of the global proteome and site-specific N-glycosylation in lung adenocarcinoma (LUAD) and its precursor lesions using clinical FFPE tissues. By systematically characterizing molecular alterations across four pathological stages, from inflammatory nodules (IN) to invasive adenocarcinoma (IAC), we revealed dynamic, stage-dependent remodeling of protein N-glycosylation during LUAD progression. Notably, glycoproteomic analyses highlighted coordinated changes in key biological processes, including extracellular matrix organization and neutrophil activation involved in immune response, which were associated with distinct site-specific glycosylation patterns on proteins such as CEACAM6 and FGB.

The integration of proteomic data with large-scale transcriptomic and clinical data from TCGA and GTEx further demonstrated the clinical relevance of several dysregulated glycoproteins. Although both CEACAM5 and CEACAM6 were found to be significantly upregulated in LUAD tissues, only elevated CEACAM6 was significantly associated with improved overall survival. Collectively, these findings provide new insights into the molecular and glycosylation alterations underlying LUAD initiation and progression and identify candidate glycoprotein biomarkers that may facilitate improved diagnosis, prognostic stratification, and the development of glycan-targeted therapeutic strategies.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00216-026-06702-z>.

Acknowledgements We gratefully acknowledge financial support from the Scientific Research Initiation Grant of Shantou University Medical College (SUMC) and the Start-up Fund of the First Affiliated Hospital of SUMC.

Author contribution J.G.C. and S.Y. optimized the protein extraction protocols for FFPE samples. J.G.C., L.B.Z., R.M.Z., and Y.C. performed the preparation of protein and glycoprotein samples for LC–MS analysis. The experiments were conducted by J.G.C., who also developed the standard operating procedures. Data analysis was performed by S.Y., W.Q.Z., and S.Y.W., with J.B.D., W.Q.Z., and S.Y.W. specifically conducting the GO and KEGG enrichment analyses. Funding support was provided by S.Y. and J.H.J. The study was designed by S.Y., who also drafted the initial manuscript. All authors contributed to the editing and final review of the manuscript.

Funding This work was supported by grants from the Jiangsu Science and Technology Plan (BX2022023), the Jiangsu Shuangchuang Boshi Program (JSSCBS20210697), and the Suzhou Medical Innovation Program (SKJY2021141). Additional support was provided by the Frontier Discipline of Pediatric Respiratory Medicine (ML13101123).

Data availability The LC–MS/MS proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD071387 (Access token: LeQrp0USownD).

Declarations

Ethics approval The collection and use of human tissue samples were approved by the Ethics Committee of the Fourth Affiliated Hospital of Soochow University.

Conflict of interest The authors declare no competing interests.

Source of biological material All human specimens were obtained from the Fourth Affiliated Hospital of Soochow University in accordance with approved institutional ethics protocols.

Statement on animal welfare No animals were used in this study; therefore, animal ethics approval was not required.

References

- Myers DJ, Wallen JM. Lung adenocarcinoma. In: StatPearls [Internet]. StatPearls Publishing; 2023. <https://www.ncbi.nlm.nih.gov/books/NBK519578/>. Accessed 26 Feb 2026.
- Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. *CA Cancer J Clin*. 2025;75(1):10–45. <https://doi.org/10.3322/caac.21871>.
- Henschke CI, McCauley DI, Yankelevitz DF, Naidich DP, McGuinness G, Miettinen OS, et al. Early lung cancer action project: overall design and findings from baseline screening. *Lancet*. 1999;354(9173):99–105. [https://doi.org/10.1016/S0140-6736\(99\)06093-6](https://doi.org/10.1016/S0140-6736(99)06093-6).
- Pieterman RM, van Putten JW, Meuzelaar JJ, Mooyaart EL, Vaalburg W, Koëter GH, et al. Preoperative staging of non-small-cell lung cancer with positron-emission tomography. *N Engl J Med*. 2000;343(4):254–61. <https://doi.org/10.1056/NEJM200007273430404>.
- Penault-Llorca F, Socinski MA. Emerging molecular testing paradigms in non-small cell lung cancer management—current perspectives and recommendations. *Oncologist*. 2025. <https://doi.org/10.1093/oncolo/oyae357>.
- Woodard GA, Jones KD, Jablons DM. Lung cancer staging and prognosis. In: Reckamp KL, editor. Lung cancer: treatment and research. Cham: Springer International Publishing; 2016. p. 47–75. https://doi.org/10.1007/978-3-319-40389-2_3.
- Succony L, Rassl DM, Barker AP, McCaughan FM, Rintoul RC. Adenocarcinoma spectrum lesions of the lung: detection, pathology and treatment strategies. *Cancer Treat Rev*. 2021;99:1–10. <https://doi.org/10.1016/j.ctrv.2021.102237>.
- Wang C, Chen B, Liang S, Shao J, Li J, Yang L, et al. China protocol for early screening, precise diagnosis, and individualized treatment of lung cancer. *Sig Transduct Target Ther*. 2025;10(1):1–7. <https://doi.org/10.1038/s41392-025-02256-1>.
- Chakravarthi BVSK, Nepal S, Varambally S. Genomic and epigenomic alterations in cancer. *Am J Pathol*. 2016;186(7):1724–35. <https://doi.org/10.1016/j.ajpath.2016.02.023>.
- Geffen Y, Anand S, Akiyama Y, Yaron TM, Song Y, Johnson JL, et al. Pan-cancer analysis of post-translational modifications reveals shared patterns of protein regulation. *Cell*. 2023;186(18):3945–3967.e3926. <https://doi.org/10.1016/j.cell.2023.07.013>.
- Colley KJ, Varki A, Haltiwanger RS, Kinoshita T. Cellular Organization of Glycosylation. In: Varki A, Cummings RD, Esko JD, Stanley P, Hart GW, Aebi M, Mohnen D, Kinoshita T, Packer NH, Prestegard JH, Schnaar RL, Seeberger PH, editors. *Essentials of Glycobiology*. 4th ed. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 2022. Chapter 4. <https://doi.org/10.1101/glycobiology.4e.4>.
- Wang X, Bo S, Li Q, Li W, Xu L, Hu D, et al. Quantitative proteomic and glycoproteomic analysis identifies CLCA1, FBN1, and FGB as potential biomarkers for ulcerative colitis. *RSC Adv*. 2025;15(45):38044–55. <https://doi.org/10.1039/D5RA05241F>.
- Meany DL, Chan DW. Aberrant glycosylation associated with enzymes as cancer biomarkers. *Clin Proteomics*. 2011;8(1):1–14. <https://doi.org/10.1186/1559-0275-8-7>.
- Gao Z, Xu M, Yue S, Shan H, Xia J, Jiang J, et al. Abnormal sialylation and fucosylation of saliva glycoproteins: characteristics of lung cancer-specific biomarkers. *Curr Res Pharmacol Drug Discov*. 2022;3:1–13. <https://doi.org/10.1016/j.crphar.2021.100079>.
- Ge W, Wen S, Zhou X, Chen Y, Zeng D, Jiang J, et al. FOXA2 loss in TGF- β 1-induced EMT suppresses bisecting-GlcNAc N-glycan synthesis in lung adenocarcinoma. *Proteomics*. 2025;25(8):1–11. <https://doi.org/10.1002/pmic.202400216>.
- Zhao Y-Y, Takahashi M, Gu J-G, Miyoshi E, Matsumoto A, Kitazume S, et al. Functional roles of N-glycans in cell signaling and cell adhesion in cancer. *Cancer Sci*. 2008;99(7):1304–10. <https://doi.org/10.1111/j.1349-7006.2008.00839.x>.
- Naba A. Mechanisms of assembly and remodelling of the extracellular matrix. *Nat Rev Mol Cell Biol*. 2024;25(11):865–85. <https://doi.org/10.1038/s41580-024-00767-3>.
- Fernandes Â, Azevedo CM, Silva MC, Faria G, Dantas CS, Vicente MM, et al. Glycans as shapers of tumour microenvironment: a sweet driver of T-cell-mediated anti-tumour immune response. *Immunology*. 2023;168(2):217–32. <https://doi.org/10.1111/imm.13494>.
- Coscia F, Doll S, Bech JM, Schweizer L, Mund A, Lengyel E, et al. A streamlined mass spectrometry-based proteomics workflow for large-scale FFPE tissue analysis. *J Pathol*. 2020;251(1):100–12. <https://doi.org/10.1002/path.5420>.
- Yang S, Höti N, Yang W, Liu Y, Chen L, Li S, et al. Simultaneous analyses of N-linked and O-linked glycans of ovarian cancer cells using solid-phase chemoenzymatic method. *Clin Proteom*. 2017;14(1):1–11. <https://doi.org/10.1186/s12014-017-9137-1>.
- Lattová E, Skříčková J, Hausnerová J, Frola L, Křen L, Ihnatová I, et al. N-Glycan profiling of lung adenocarcinoma in patients at different stages of disease. *Mod Pathol*. 2020;33(6):1146–56. <https://doi.org/10.1038/s41379-019-0441-3>.
- Oliveira DVNP, Hentze J, O'Rourke CJ, Andersen JB, Høgdall C, Høgdall EV. DNA methylation in ovarian tumors—a comparison between fresh tissue and FFPE samples. *Reprod Sci*. 2021;28(11):3212–8. <https://doi.org/10.1007/s43032-021-00589-0>.
- Chen L, Li Y, Fu Y, Peng J, Mo M-H, Stamatakis M, et al. Role of deregulated microRNAs in breast cancer progression using FFPE tissue. *PLoS ONE*. 2013;8(1):1–11. <https://doi.org/10.1371/journal.pone.0054213>.
- Magdeldin S, Yamamoto T. Toward deciphering proteomes of formalin-fixed paraffin-embedded (FFPE) tissues. *Proteomics*. 2012;12(7):1045–58. <https://doi.org/10.1002/pmic.201100550>.
- Azimzadeh O, Barjaktarovic Z, Aubele M, Calzada-Wack J, Sarioglu H, Atkinson MJ, et al. Formalin-fixed paraffin-embedded (FFPE) proteome analysis using gel-free and gel-based proteomics. *J*

- Proteome Res. 2010;9(9):4710–20. <https://doi.org/10.1021/pr1004168>.
26. Tian Y, Gurley K, Meany DL, Kemp CJ, Zhang H. N-linked glycoproteomic analysis of formalin-fixed and paraffin-embedded tissues. *J Proteome Res*. 2009;8(4):1657–62. <https://doi.org/10.1021/pr800952h>.
 27. Soni RK. Protocol for deep proteomic profiling of formalin-fixed paraffin-embedded specimens using a spectral library-free approach. *STAR Protoc*. 2023;4(3):1–16. <https://doi.org/10.1016/j.xpro.2023.102381>.
 28. Shah P, Zhang B, Choi C, Yang S, Zhou J, Harlan R, et al. Tissue proteomics using chemical immobilization and mass spectrometry. *Anal Biochem*. 2015;469:27–33. <https://doi.org/10.1016/j.ab.2014.09.017>.
 29. Yang S, Wu WW, Shen R-F, Bern M, Cipollo J. Identification of sialic acid linkages on intact glycopeptides via differential chemical modification using IntactGIG-HILIC. *J Am Soc Mass Spectrom*. 2018;29(6):1273–83. <https://doi.org/10.1007/s13361-018-1931-0>.
 30. Zhao L, Hou C, Gao Y, Jin H, Liu C-F, Li S, et al. Mass spectrometry-based comparative analysis of N-glycosylation alterations in three human body fluids in Parkinson's disease. *ACS Chem Neurosci*. 2025;16(22):4364–76. <https://doi.org/10.1021/acscchemneuro.5c00654>.
 31. Ruhaak LR, Xu G, Li Q, Goonatilleke E, Lebrilla CB. Mass spectrometry approaches to glycomic and glycoproteomic analyses. *Chem Rev*. 2018;118(17):7886–930. <https://doi.org/10.1021/acs.chemrev.7b00732>.
 32. Zacharias LG, Hartmann AK, Song E, Zhao J, Zhu R, Mirzaei P, et al. HILIC and ERLIC enrichment of glycopeptides derived from breast and brain cancer cells. *J Proteome Res*. 2016;15(10):3624–34. <https://doi.org/10.1021/acs.jproteome.6b00429>.
 33. Suttapitugsakul S, Sun F, Wu R. Recent advances in glycoproteomic analysis by mass spectrometry. *Anal Chem*. 2020;92(1):267–91. <https://doi.org/10.1021/acs.analchem.9b04651>.
 34. Kanehisa M, Furumichi M, Sato Y, Matsuura Y, Ishiguro-Watanabe M. KEGG: biological systems database as a model of the real world. *Nucleic Acids Res*. 2025;53(D1):D672–7. <https://doi.org/10.1093/nar/gkae909>.
 35. Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D, et al. The gene ontology knowledgebase in 2023. *Genetics*. 2023;224(1):1–14. <https://doi.org/10.1093/genetics/iyad031>.
 36. Zhang W, Wang X, Wen S, Ge W, Deng J, Zhou Z, et al. Glycoproteomics analysis of colorectal cancer reveals differential alterations of glycoproteins and glycans. *Anal Bioanal Chem*. 2025;417(27):6179–90. <https://doi.org/10.1007/s00216-025-06108-3>.
 37. Liu Z, Wu Z, Zhou Y, Xia J, Zhang W, Gao S, et al. Hydrophilic peptide and glycopeptide as immobilized sorbents for glycosylation analysis. *Anal Chem*. 2024;96(4):1498–505. <https://doi.org/10.1021/acs.analchem.3c03944>.
 38. Wang Y, Wu Z, Hu W, Hao P, Yang S. Impact of expressing cells on glycosylation and glycan of the SARS-CoV-2 spike glycoprotein. *ACS Omega*. 2021;6(24):15988–99. <https://doi.org/10.1021/acsomega.1c01785>.
 39. Wen S, Gao Y, Miao X, Deng J, Zhou Y, Ge W, et al. GDAS: a machine learning-driven approach for the high-throughput identification of disease-associated glycoforms. *Anal Chem*. 2026;98(21):15720–9. <https://doi.org/10.1021/acs.analchem.6c01321>.
 40. Yang S, Chen L, Chan DW, Li QK, Zhang H. Protein signatures of molecular pathways in non-small cell lung carcinoma (NSCLC): comparison of glycoproteomics and global proteomics. *Clin Proteomics*. 2017;14(1):1–15. <https://doi.org/10.1186/s12014-017-9166-9>.
 41. Hu W, Ge W, Xia P, Chen Y, Du J, Hu G, et al. Diagnostic potential of serum glycome analysis in lung cancer: a glycopattern study. *J Proteome Res*. 2024;23(1):500–9. <https://doi.org/10.1021/acs.jproteome.3c00645>.
 42. Gao Z, Chen S, Du J, Wu Z, Ge W, Gao S, et al. Quantitative analysis of fucosylated glycoproteins by immobilized lectin-affinity fluorescent labeling. *RSC Adv*. 2023;13(10):6676–87. <https://doi.org/10.1039/d3ra00072a>.
 43. Peixoto A, Ferreira D, Miranda A, Relvas-Santos M, Freitas R, Veth TS, et al. Multilevel plasticity and altered glycosylation drive aggressiveness in hypoxic and glucose-deprived bladder cancer cells. *iScience*. 2025. <https://doi.org/10.1016/j.isci.2025.111758>.
 44. Hakomori S. Glycosylation defining cancer malignancy: new wine in an old bottle. *Proc Natl Acad Sci U S A*. 2002;99(16):10231–3. <https://doi.org/10.1073/pnas.172380699>.
 45. Reily C, Stewart TJ, Renfrow MB, Novak J. Glycosylation in health and disease. *Nat Rev Nephrol*. 2019;15(6):346–66. <https://doi.org/10.1038/s41581-019-0129-4>.
 46. Zhang R, Zhou Y, Wen S, Chen Y, Du J, Ma J, et al. Deciphering disease-specific glycosylation: unraveling diabetes subtypes through serum glycopattern. *Anal Bioanal Chem*. 2025;417(26):5805–20. <https://doi.org/10.1007/s00216-025-06089-3>.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.