Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Integrative Multi-omics Analysis Identifies B3GNT3 as a Prognostic Biomarker Linking Microbiota Dysbiosis to Adverse Outcomes in Lung Adenocarcinoma

Weitao Song
TH Hospital of Joint Logistics Support Force
Jinhe Xu
Fuzong Clinical Medical College of Fujian Medical University
Jianqing Lian
TH Hospital of Joint Logistics Support Force
Zhangyi Bao
TH Hospital of Joint Logistics Support Force
Shuaicai Lin
TH Hospital of Joint Logistics Support Force
Jun Lu
TH Hospital of Joint Logistics Support Force
Duo Zhang
duozhang80@xmu.edu.cn

TH Hospital of Joint Logistics Support Force

Research Article

Keywords: Lung adenocarcinoma, Tumor microbiome, B3GNT3, Biomarker

Posted Date: June 26th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-10044173/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

The upstream regulatory mechanisms governing aberrant protein glycosylation in lung adenocarcinoma (LUAD) remain poorly understood. Through integrative multi-omics profiling, we identified the glycosyltransferase B3GNT3 as an adverse prognostic biomarker intimately associated with tumor microbial dysbiosis, characterized by the specific enrichment of Porphyromonas. In an independent The Cancer Genome Atlas (TCGA) cohort, elevated B3GNT3 expression significantly correlated with inferior overall survival ($p = 0.022$) and the hyperactivation of the mTORC1 signaling pathway. Crucially, to substantiate these in silico predictions, we performed experimental validation using an independent clinical LUAD cohort. Quantitative real-time PCR (qRT-PCR) and Western blot analyses explicitly confirmed the significant upregulation of B3GNT3 at the transcriptional level and the constitutive activation of the mTORC1 pathway in tumor tissues. Ultimately, this study delineates a novel mechanistic axis linking microbial dysbiosis to host metabolic reprogramming and aberrant glycosylation, providing robust, multi-level evidence for microbiome-mediated tumor progression and nominating B3GNT3 as a promising therapeutic target in LUAD.

1. Introduction

Lung cancer remains the leading cause of cancer-related mortality worldwide(Bray et al. 2024). Non-small cell lung cancer (NSCLC) constitutes approximately 85% of cases, with lung adenocarcinoma (LUAD) being the most prevalent histological subtype (Herbst et al. 2018). Despite advances in targeted therapies and immune checkpoint blockade, which have extended survival in selected patients, intratumoral heterogeneity, therapeutic resistance, and the lack of reliable early-detection strategies continue to constrain five-year overall survival to approximately 18–20%(Thai et al. 2021). The identification of novel molecular drivers, prognostic biomarkers, and therapeutic vulnerabilities in LUAD therefore represents an unmet clinical need.

Tumor progression is orchestrated by the tumor microenvironment (TME), a dynamic ecosystem encompassing immune infiltrates, cancer-associated fibroblasts, and extracellular matrix that collectively sustain proliferation, invasion, metastasis, and immune evasion (Hanahan and Weinberg 2011). Emerging evidence has positioned the intratumoral microbiome as an integral TME component. Next-generation sequencing studies have established the existence of a distinct, low-biomass bacterial community within the lower respiratory tract, and its dysbiosis has been repeatedly implicated in lung carcinogenesis and disease progression (Shah et al. 2021; Li et al. 2024). Notably, enrichment of specific genera—including Veillonella, Prevotella, and Streptococcus—has been documented in lung tumor tissue and correlated with advanced stage and inferior clinical outcome (Tsay et al. 2021).

Aberrant cellular metabolism represents a hallmark of malignancy, enabling rapid biomass accumulation and adaptive signaling (Hanahan 2022). Protein glycosylation, which is tightly coupled to nutrient flux (particularly hexosamine and nucleotide-sugar pathways), is profoundly altered in cancer, thereby modulating receptor tyrosine kinase activity, cell adhesion, and immune recognition (Pinho and Reis

2015; Thomas et al. 2021). Dysregulated glycosyltransferases have accordingly emerged as candidate therapeutic targets.

Integrative analyses concurrently examining microbial composition, metabolic reprogramming, and glycosylation abnormalities in LUAD have been limited. In the present study, we harnessed publicly available multi-omics datasets to comprehensively profile the LUAD-associated microbiome, pinpoint glycosyltransferases whose expression correlates with microbial dysbiosis and adverse prognosis, and delineate candidate upstream regulatory circuits. Specifically, we identified *B3GNT3* as a crucial prognostic biomarker associated with the enrichment of *Porphyromonas*, and experimentally verified in an independent clinical cohort that its upregulation is tightly coupled with the hyperactivation of the mTORC1 signaling pathway. This work thus elucidates a previously underappreciated microbiome–metabolism–glycosylation axis in LUAD pathogenesis and nominates B3GNT3 as a promising diagnostic and therapeutic candidate.

2. Materials and Methods

2.1 Microbiome Data Acquisition and Processing

The 16S rRNA gene amplicon sequencing data were retrieved from the NCBI Sequence Read Archive (SRA) under BioProject PRJNA303190. The dataset consists of single-end FASTQ files derived from bronchoalveolar lavage fluid (BALF) of lung adenocarcinoma (LUAD) patients. A paired-study design was employed, with BALF samples collected from the tumor-bearing lung lobe (Tumor group) and the contralateral healthy lung lobe (Normal group) of each patient.

Raw FASTQ files were processed using QIIME 2 (version 2024.2)(Bolyen et al. 2019). Data were imported as QIIME 2 artifacts, followed by quality control, denoising, chimera removal, and ASV table construction using the DADA2 plugin (qiime dada2 denoise-single)(Callahan et al. 2016). Sequences were truncated at 150 bp (--p-trunc-len 150) based on interactive quality plots; all other parameters were kept at default settings. Representative ASV sequences were used for all downstream analyses. Taxonomic classification was performed with a naive Bayes classifier pre-trained on the SILVA 138.1 99% OTU full-length sequence database (qiime feature-classifier classify-sklearn)(Quast et al. 2012). The phylogenetic tree was constructed using the qiime phylogeny align-to-tree-mafft-fast tree pipeline, which sequentially invokes MAFFT for multiple sequence alignment(Katoh 2002) and FastTree for tree inference(Swart et al. 2020).

2.2 TCGA Data Acquisition and Preprocessing

LUAD transcriptome data and clinical information were downloaded from cBioPortal using the “Lung Adenocarcinoma (TCGA, PanCancer Atlas)” dataset (study ID: luad_tcga_pan_can_atlas_2018). This dataset represents the final, batch-corrected version of the TCGA PanCancer Atlas project, comprising 515 primary tumors and 59 matched adjacent normal tissues, alongside RSEM-normalized mRNA expression data and comprehensive clinical annotations.

Expression matrices and clinical data were loaded and integrated in R 4.3.2. Differential expression analysis was restricted to samples labeled “Primary Tumor” (n = 515) and “Solid Tissue Normal” (n = 59). Survival and GSEA analyses used only “Primary Tumor” samples. To meet DESeq2 input requirements, RSEM values were rounded to the nearest integer and used as pseudocounts.

2.3 Bioinformatics and Statistical Analyses

Unless otherwise specified, all statistical analyses and visualization were performed in R 4.3.2. Statistical significance was defined as $p < 0.05$ (two-sided tests).

Microbiome analysis: ASV abundance table, taxonomic assignments, phylogenetic tree, and sample metadata were imported into R and combined into a phyloseq object(McMurdie and Holmes 2013). Alpha diversity was calculated using the Shannon index and compared between groups with paired Wilcoxon signed-rank test. Beta diversity was assessed using weighted UniFrac distance, visualized by principal coordinates analysis (PCoA), and tested for significance with PERMANOVA (vegan::adonis2, 999 permutations)(Chepkemoi et al. 2024). Differential abundance analysis at genus level was performed with DESeq2(Love et al. 2014) using the model ~PatientID + Group to account for the paired design.

Differential expression and functional enrichment analysis

Differential expression between TCGA-LUAD tumors and adjacent normal tissues was conducted using DESeq2[8]. Genes with $|log2FoldChange| > 0.5$ and $padj < 0.05$ were considered significantly differentially expressed. GO and KEGG pathway enrichment analyses were performed with clusterProfiler(Wu et al. 2021) after converting gene symbols to Entrez IDs using org.Hs.eg.db.

Gene set enrichment analysis (GSEA)

Patients were stratified into B3GNT3 high- and low-expression groups based on the median expression in TCGA tumor samples. Genes were pre-ranked by the Wald statistic from DESeq2, and GSEA was conducted on the MSigDB Hallmark gene sets (v2023.2.Hs)(Liberzon et al. 2015)using the fgsea(Korotkevich et al. 2016). Pathways with $padj < 0.05$ were considered significantly enriched.

Survival analysis

Patients were divided into B3GNT3 high- and low-expression groups using the median expression value. Kaplan-Meier curves were generated and compared with the log-rank test using the survival and survminer packages.

2.4 Validation of B3GNT3 Expression and mTORC1 Pathway Activation in Clinical LUAD Samples

2.4.1 Clinical Sample Collection and Processing

Fresh tumor tissues and paired adjacent normal lung tissues (> 2 cm from the tumor margin) were collected from 10 pathologically confirmed LUAD patients who had not received prior neoadjuvant therapy. Following surgical resection, the specimens were immediately snap-frozen in liquid nitrogen and cryopreserved at − 80°C in the biobank of the 900th Hospital of Joint Logistics Support Force. Prior to experiments, tissue blocks were rapidly sectioned on ice: ~30–50 mg for RNA extraction and ~ 50–100 mg for protein extraction. All procedures were performed on ice to minimize RNA and protein degradation.

2.4.2 Quantitative Real-Time PCR and Western Blot Analysis

For real-time quantitative PCR (qRT-PCR) and western blotting analyses, 30–50 mg of tissue was homogenized in 1 mL of TRIzol™ Reagent (Cat. No. 15596026, Invitrogen, Carlsbad, CA, USA) on ice using a high-speed tissue disruptor. Total RNA was extracted following the standard TRIzol chloroform-isopropanol protocol, washed with 75% ethanol, air-dried briefly, and dissolved in RNase-free water. RNA concentration and purity were assessed using a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA), and only samples with an A260/A280 ratio of 1.8–2.0 were utilized. For cDNA synthesis, 1 μg of total RNA was reverse-transcribed using the PrimeScript™ RT Reagent Kit with gDNA Eraser (Cat. No. RR047A, Takara Bio, Kusatsu, Japan), which included genomic DNA elimination at 37°C for 5 min, reverse transcription at 42°C for 15 min, and enzyme inactivation at 85°C for 5 s. The synthesized cDNA was diluted to ~ 10 ng/μL with RNase-free water. qRT-PCR was performed using TB Green® Premix Ex Taq™ II No. RR820A, Takara Bio). The primer sequences used were as follows: B3GNT3 forward, 5'-TGCTGGGCTACGACTTCAAC-3', and reverse, 5'-GTCGATGGCGATGCAGTTG-3'; β-actin forward, 5'-CATGTACGTTGCTATCCAGGC-3', and reverse, 5'-CTCCTTAATGTCACGCACGAT-3'. The 20 μL reaction system consisted of 10 μL of TB Green Premix, 0.8 μL of each primer (10 μM), 2 μL of cDNA, and 6.4 μL of ddH_2O. Cycling conditions on an Applied Biosystems™ 7500 Real-Time PCR System (Thermo Fisher Scientific, Foster City, CA, USA) were set at 95°C for 30 s, followed by 40 cycles of 95°C for 5 s and 60°C for 34 s, concluding with a melting curve analysis. Triplicate technical replicates and no-template controls were included for all samples. Relative B3GNT3 mRNA expression was calculated using the 2^{-ΔΔCt} method with β-actin as the reference gene. For protein analysis, 50–100 mg of tissue was lysed on ice in RIPA buffer supplemented with protease and phosphatase inhibitor cocktails (Cat. No. 11697498001 and 4906845001, Roche Diagnostics, Mannheim, Germany). The lysates were centrifuged at 12,000 rpm for 15 min at 4°C, and the protein concentrations of the supernatants were quantified using a BCA Protein Assay Kit (Cat. No. 23225, Thermo Fisher Scientific). Equal amounts of total protein (30 μg) were denatured in 5× SDS loading buffer at 95°C for 10 min, separated by 10% SDS-polyacrylamide gel electrophoresis (SDS-PAGE), and wet-transferred (100 V, 90 min) onto 0.45 μm PVDF membranes (Cat. No. IPVH00010, MilliporeSigma, Burlington, MA, USA). The membranes were blocked with 5% non-fat milk in TBST for 1 h at room temperature and then incubated overnight at 4°C with the following primary antibodies: rabbit anti-phospho-S6 Ribosomal Protein (Ser235/236) (1:1000, Cat. No. 4858, Cell Signaling Technology, Danvers, MA, USA), rabbit anti-S6 Ribosomal Protein (1:1000, Cat. No. 2217, Cell Signaling Technology), and mouse anti-β-actin (1:5000, Cat. No. 66009-1-Ig, Proteintech

Group, Rosemont, IL, USA). After three washes with TBST, the membranes were incubated with HRP-conjugated anti-rabbit IgG (1:5000, Cat. No. 7074, Cell Signaling Technology) or anti-mouse IgG (1:5000, Cat. No. 7076, Cell Signaling Technology) secondary antibodies for 1 h at room temperature. Protein signals were developed using the SuperSignal™ West Pico PLUS Chemiluminescent Substrate (Cat. No. 34580, Thermo Fisher Scientific) and visualized on an Amersham™ Imager 600 (GE Healthcare, Chicago, IL, USA). The mTORC1 activation level was evaluated by calculating the band intensity ratio of p-S6 to total S6 using ImageJ software (version 1.53, NIH, Bethesda, MD, USA), with normalization to β-actin. Paired Student's t-tests were utilized to compare mRNA and protein expression differences between tumor and adjacent normal tissues using GraphPad Prism 9.0 (GraphPad Software, San Diego, CA, USA), and a p-value < 0.05 was considered statistically significant.

3. Results

3.1The Overall Ecological Diversity Remains Stable in the LUAD Tumor Microenvironment

To investigate the ecological features of the microbiome in lung adenocarcinoma (LUAD), we first performed alpha- and beta-diversity analyses on bronchoalveolar lavage fluid (BALF) samples collected from paired tumor-bearing and contralateral healthy lung segments within the same patients ("Tumor" and "Normal" groups, respectively).

Alpha-diversity analysis was conducted to assess the species richness and evenness within individual samples. Calculation of the Shannon index revealed no statistically significant difference in alpha diversity between the Tumor and Normal groups (Wilcoxon signed-rank test, $p > 0.05$) (Fig. 1A). This result suggests that the presence of a tumor does not significantly alter the overall richness of the microbial community. Consistently, when we further evaluated the community structure differences between samples using beta diversity, Principal Coordinates Analysis (PCoA) based on the weighted UniFrac distance demonstrated that the Tumor samples (yellow) and Normal samples (blue) were largely overlapping in the two-dimensional space without forming distinct clusters (Fig. 1B). This was further confirmed by Permutational Multivariate Analysis of Variance (PERMANOVA), which showed no significant difference in the overall microbial community structure between the two groups ($p > 0.05$).

Taken together, these findings reveal that the macroscopic ecological diversity and overall community structure remain relatively stable in the LUAD microenvironment. Nevertheless, this global stability does not preclude potential critical alterations among specific bacterial taxa. This prompted us to perform a higher-resolution differential abundance analysis to identify specific microbes altered in the tumor microenvironment.

1. (A) Comparison of alpha diversity of the microbiome between tumor tissues and adjacent normal tissues. The boxplot displays the distribution of the Shannon Diversity Index. "ns" indicates no statistical significance (Wilcoxon signed-rank test, $p > 0.05$).
2. (B) Beta diversity Principal Coordinates Analysis (PCoA) based on the weighted UniFrac distance. Each point represents an individual sample (blue: adjacent normal tissue; yellow: tumor tissue). Ellipses represent the 95% confidence intervals. The p-value ($p > 0.05$) calculated by PERMANOVA indicates no significant difference in overall community structure. The percentage in the parentheses on each axis indicates the proportion of the total variance explained by that principal coordinate.

3.2 Severe Baseline Dysbiosis and Identification of Key Differential Genera in the LUAD Microbiome

To understand the underlying basis of the macroscopic stability observed in the diversity analyses, we first evaluated the taxonomic composition of the samples at the phylum level. A striking feature emerged: the microbial communities in both adjacent normal and tumor tissues were overwhelmingly dominated by the phylum Proteobacteria (p__Proteobacteria), with its relative abundance approaching 100% in the vast majority of samples (Fig. 2A). This absolute dominance likely accounts for the lack of overall structural shifts detected by beta diversity, while simultaneously revealing a state of severe, baseline dysbiosis prevalent in the lower respiratory tract microenvironment of these lung cancer patients.

Despite this highly uniform macroscopic profile at the phylum level, a more granular differential abundance analysis unveiled a drastic and decisive remodeling of the microbial membership at the genus level. The volcano plot (Fig. 2B) clearly illustrates the altered distribution of specific genera in tumor tissues relative to adjacent normal tissues. We identified multiple genera whose abundances were statistically significantly altered (padj < 0.05). Notably, the genera *Caulobacter* (g__Caulobacter) and *Porphyromonas* (g__Porphyromonas) showed highly significant enrichment in tumor tissues (log2 Fold Change > 10, padj < 0.001). Conversely, several genera, including *Comamonas* (g__Comamonas), *Fusobacterium* (g__Fusobacterium), and *Prevotella* (g__Prevotella), were significantly depleted in the tumor microenvironment.

Collectively, these results indicate that while the LUAD tumor microenvironment remains uniformly dominated by Proteobacteria at the phylum level, it undergoes specific, fine-scale taxonomic remodeling at the genus level, marked by the enrichment of *Caulobacter* and *Porphyromonas*. These significantly enriched opportunistic pathogens are potential key biological drivers of TME alterations, thereby warranting further investigation as the focus of our subsequent mechanistic exploration.

1. (A) Stacked bar plot showing the taxonomic composition at the phylum level for tumor and adjacent normal tissues. Each vertical bar represents an individual sample, and the Y-axis represents the relative abundance.
2. (B) Volcano plot of differential abundance at the genus level. The X-axis represents the Log2 Fold Change of tumor versus normal groups, and the Y-axis represents the -Log10 Adjusted P-value. Yellow points indicate genera significantly enriched in tumors, blue points indicate genera significantly depleted in tumors, and grey points indicate genera with no significant difference.

3.3 The Glycosyltransferase B3GNT3 is a Poor Prognostic Biomarker in LUAD Associated with mTORC1 Pathway Activation

To identify key host factors that may mediate the effects of microbiome dysbiosis, we performed a differential expression analysis on transcriptome data from an independent TCGA-LUAD cohort. A heatmap clearly illustrated a significant difference in gene expression profiles between tumor and adjacent normal tissues, with a systematically activated "upregulated gene module" enriched in multiple genes related to cell proliferation and metabolism (Fig. 3A). Within this module, we focused on the key enzyme responsible for protein glycosylation, β-1,3-N-acetylglucosaminyltransferase 3 (B3GNT3). Considering that aberrant glycosylation is a core hallmark of cancer, we further investigated the clinical significance of B3GNT3. Strikingly, Kaplan-Meier survival analysis revealed that the expression level of B3GNT3 was significantly associated with the overall survival of LUAD patients. Compared to patients with low B3GNT3 expression, the high-expression group exhibited a significantly lower survival probability (Log-rank test, $p = 0.022$), indicating that B3GNT3 is a potential independent poor prognostic biomarker for LUAD (Fig. 3B). To explore the potential molecular mechanisms underlying high B3GNT3 expression, we next performed Gene Set Enrichment Analysis (GSEA) on tumor samples stratified by B3GNT3 expression. The results strongly demonstrated that in B3GNT3-high tumors, the mTORC1 signaling pathway (HALLMARK_MTORC1_SIGNALING), a central pathway regulating cell growth, metabolism, and protein synthesis, was profoundly and significantly enriched ($NES > 0$, $padj < 0.01$) (Fig. 3C). In summary, our bioinformatics analysis not only identifies B3GNT3 as a novel biomarker of poor prognosis in LUAD but, more importantly, provides the first evidence linking its overexpression to the activation of the core metabolic regulatory pathway mTORC1.

1. (A) Heatmap of the top 30 most significantly differentially expressed genes in the TCGA-LUAD cohort. Each column represents a sample, and each row represents a gene. The color scale indicates the relative gene expression level (Z-score), with yellow representing high expression and blue representing low expression.
2. (B) Kaplan-Meier survival curve for overall survival of LUAD patients based on B3GNT3 expression levels. Patients were divided into high-expression (yellow) and low-expression (blue) groups based on the median expression of B3GNT3. The p-value was calculated using the Log-rank test.

3. (C) Gene Set Enrichment Analysis (GSEA) plot showing the significant enrichment of the mTORC1 signaling pathway in the B3GNT3-high expression group.

3.4 Experimental Validation of B3GNT3 Upregulation and mTORC1 Pathway Activation in LUAD

To substantiate our *in silico* findings regarding *B3GNT3* and the mTORC1 pathway, we performed experimental validation using an independent clinical cohort comprising 10 matched pairs of LUAD and adjacent normal tissues. First, we evaluated the transcriptional levels of *B3GNT3* via qRT-PCR. Consistent with the TCGA dataset analysis, *B3GNT3* mRNA expression was significantly upregulated in LUAD tumor tissues compared to their paired normal counterparts (paired t-test, $p < 0.05$) (Fig. 4A).

Next, to confirm the hyperactivation of the mTORC1 signaling pathway predicted by our GSEA analysis, we assessed the phosphorylation status of the S6 ribosomal protein (p-S6)—a canonical and robust downstream readout of mTORC1 activity—across the clinical samples via Western blotting. Representative immunoblots from four patient pairs visually illustrated a marked accumulation of p-S6 in the tumor tissues (T) relative to the adjacent normal tissues (N), whereas the levels of total S6 and the internal control β-actin remained essentially unchanged (Fig. 4B). To ensure statistical robustness, subsequent semi-quantitative analysis was rigorously conducted across the entire cohort of 10 matched pairs. This comprehensive analysis confirmed that the ratio of p-S6 to total S6 was significantly elevated in the tumor group ($p < 0.05$) (Fig. 4C), providing definitive protein-level evidence for constitutive mTORC1 activation in LUAD.

Collectively, these independent experimental data robustly corroborate our bioinformatics-derived hypothesis, confirming that *B3GNT3* upregulation is intimately accompanied by constitutive mTORC1 pathway activation in the LUAD microenvironment.

1. (A) The relative mRNA expression of B3GNT3 in paired LUAD tumor tissues (Tumor) and adjacent normal tissues (Normal) was measured by qRT-PCR (n = 10 pairs). The boxplot overlays jittered points representing individual samples. Statistical significance was determined using a paired Student's t-test. *p < 0.05.
2. (B) Representative Western blot analysis of phosphorylated S6 (p-S6), total S6, and β-actin in four paired clinical samples. β-actin served as the loading control. N, adjacent normal tissue; T, tumor tissue.
3. (C) Semi-quantitative analysis of the relative p-S6 to total S6 protein ratio across all 10 pairs of matched samples. Data were analyzed using a paired Student's t-test. *p < 0.05.

4. Discussion

By integrating multi-omics analysis of public databases with experimental validation in an independent clinical cohort, this study reveals a novel regulatory cascade in the lung adenocarcinoma (LUAD) tumor

microenvironment, linking specific microbial dysbiosis, constitutive mTORC1 pathway activation, and the aberrant overexpression of the glycosyltransferase B3GNT3. The core finding is twofold: not only B3GNT3 is an independent biomarker for poor prognosis in LUAD, but its high expression is also highly synchronous with the systemic activation of the mTORC1 signaling pathway. We propose that this phenomenon is likely driven by the enrichment of opportunistic pathogens within the tumor, such as *Porphyromonas* and *Caulobacter*, which in turn endows tumor cells with enhanced glycosylation capabilities and a more malignant phenotype.

As a key glycosyltransferase that catalyzes the extension of poly-N-acetyllactosamine (poly-LacNAc) chains, the abnormal expression of B3GNT3 has been gradually reported in various malignancies(Gao et al. 2018). Aberrant glycosylation is one of the hallmarks of cancer and is considered an important molecular basis for tumor cells to acquire invasive, metastatic, and therapy-resistant capabilities(Pinho and Reis 2015; Munkley and Elliott 2016). This study, for the first time in LUAD, systematically confirms that high B3GNT3 expression is independently associated with shorter overall survival. This clinical disadvantage can be mechanistically explained by well-documented literature, aberrant glycosylation is known to directly regulate the signaling intensity of key growth factor receptors (e.g., EGFR, TGF-βR), promote tumor cell-matrix adhesion, and mediate an immunosuppressive microenvironment(Mereiter et al. 2019). Therefore, the aberrant glycan structures mediated by B3GNT3 likely constitute a crucial functional effector in LUAD progression.

The most significant novelty of this study lies in placing tumor microbiome dysbiosis at the top of the B3GNT3-mTORC1 regulatory axis, proposing a complete "microbe-metabolism-signal-glycosylation" cascade hypothesis. Interestingly, although our initial beta-diversity analysis indicated that the overall microbial community structure remained macroscopically stable, further taxonomic examination revealed a state of severe baseline dysbiosis overwhelmingly dominated by the phylum Proteobacteria. Against this uniformly dysbiotic backdrop, higher-resolution profiling uncovered a decisive genus-level remodeling, notably the significant enrichment of periodontal pathogen-associated genera, such as Porphyromonas, in the tumor-bearing lung lobe. As a central hub for sensing cellular nutrients and energy, the overactivation of mTORC1 in the vast majority of solid tumors is widely recognized as a core driving event of tumor metabolic reprogramming(Mossmann et al. 2018). Our GSEA results showed that the Hallmark mTORC1 signaling gene set was significantly enriched in the B3GNT3-high expression group, suggesting that B3GNT3 is likely a downstream transcriptional target of mTORC1. Previous studies have confirmed that such microorganisms can directly activate the PI3K/Akt/mTORC1 pathway(subsequently triggering the phosphorylation of its downstream target S6), either through their metabolic products like short-chain fatty acids (e.g., butyrate), polyamines (e.g., putrescine, spermine), and lipopolysaccharides (LPS), or via the TLR4/MyD88 pathway(Rubinstein et al. 2013; Iraporda et al. 2015). These findings strongly support our hypothesis that the abnormal colonization of opportunistic pathogens in the TME remodels the local metabolic profile, continuously "feeding" and activating the tumor cell's mTORC1 pathway. This, in turn, upregulates a series of glycosylation and biosynthesis-related genes, including B3GNT3, ultimately reinforcing the tumor's malignant glycophenotype.

The aforementioned findings offer multi-level potential strategies for the precision treatment of LUAD. (1) B3GNT3 itself could serve as a novel therapeutic target. The development of its specific small-molecule inhibitors or nanobodies targeting its catalytic products may block invasion and metastasis driven by aberrant glycosylation(Gloster 2012). (2) Given the central role of mTORC1 in this axis, mTORC1 inhibitors (e.g., everolimus, rapamycin analogs) may achieve better efficacy in the B3GNT3-high LUAD subpopulation, for which our study provides a clear biomarker for patient stratification. (3) The most promising for translation are intervention strategies targeting the tumor microbiome. These include specific antibiotics, engineered bacteriophages, or fecal microbiota transplantation (FMT) combined with immune checkpoint inhibitors, which could potentially block mTORC1 activation from the top of the cascade and reverse the tumor glycophenotype, an approach that has already shown synergistic anti-tumor effects in models like colorectal cancer(Sethi et al. 2018).

Although this study provides relatively complete correlational evidence and a mechanistic hypothesis, several limitations exist. First, the main conclusions are derived from retrospective bioinformatics analysis and experimental validation with a small sample size. The causal relationship of the "microbe-mTORC1-B3GNT3" axis needs to be further confirmed through in vitro co-culture, germ-free animal models, or microbiota transplantation experiments. Second, the clinical validation cohort was limited in size (n = 10), and the prognostic value of B3GNT3 and the mTORC1 activation status need to be validated in larger, multi-center cohorts. Finally, this study has not yet identified the specific microbial metabolites driving mTORC1 activation. Future work combining metabolomics and spatial transcriptomics is required to deeply resolve the key signaling molecules.

Declarations

Ethics Statement

This study was conducted in accordance with the Declaration of Helsinki and was approved by the Ethics Committee of the 900th Hospital of Joint Logistics Support Force (Approval No. 2026-055). Written informed consent was obtained from all participants prior to their enrollment.

Conflicts of Interest

The authors declare no conflicts of interest.

Author Contribution

Weitao Song analyzed the data, drafted the manuscript, and performed the experiments; Jinhe Xu collected the tissue specimens; Jianqing Lian performed the preliminary processing of the specimens; Zhangyi Bao provided constructive suggestions; Shuaicai Lin assisted with the data analysis; Jun Lu

guided the overall design of the manuscript; Duo Zhang contributed to the experimental design and revised the manuscript.

Acknowledgement

The authors would like to thank all the individuals who participated in this study. We thank Dr. Ranxun Lin of Fuzhou University. This study was supported by the Research Programs Foundation of Fujian Medical University(2022QH1337).

Data Availability

The datasets analyzed during the current study are available in public repositories. The 16S rRNA gene amplicon sequencing data can be accessed in the NCBI Sequence Read Archive (SRA) under the BioProject accession number PRJNA303190. The transcriptomic and clinical datasets for lung adenocarcinoma (LUAD) are available in The Cancer Genome Atlas (TCGA) and can be accessed via the cBioPortal database (study ID: luad_tcga_pan_can_atlas_2018). All other original experimental data generated during this study (including qRT-PCR and Western blot results) are available from the corresponding author upon reasonable request.

References

1. Bolyen E, Rideout JR, Dillon MR et al (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. Nat Biotechnol 37:852–857. https://doi.org/10.1038/s41587-019-0209-9
2. Bray F, Laversanne M, Sung H et al (2024) Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 74:229–263. https://doi.org/10.3322/caac.21834
3. Callahan BJ, McMurdie PJ, Rosen MJ et al (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nat Methods 13:581–583. https://doi.org/10.1038/nmeth.3869
4. Chepkemoi L, Salifu D, Murungi LK, Tonnang HEZ (2024) Species JDAIP 12:210–231. https://doi.org/10.4236/jdaip.2024.122012. High Dimension Multivariate Data Analysis for Small Group Samples of Chemical Volatile Profiles of African Nightshade
5. Gao L, Zhang H, Zhang B et al (2018) B3GNT3 overexpression is associated with unfavourable survival in non-small cell lung cancer. J Clin Pathol 71:642–647. https://doi.org/10.1136/jclinpath-2017-204860
6. Gloster TM (2012) Development of inhibitors as research tools for carbohydrate-processing enzymes. Biochem Soc Trans 40:913–928. https://doi.org/10.1042/BST20120201
7. Hanahan D (2022) Hallmarks of Cancer: New Dimensions. Cancer Discov 12:31–46. https://doi.org/10.1158/2159-8290.CD-21-1059

8. Hanahan D, Weinberg RA (2011) Hallmarks of Cancer: The Next Generation. Cell 144:646–674. https://doi.org/10.1016/j.cell.2011.02.013
9. Herbst RS, Morgensztern D, Boshoff C (2018) The biology and management of non-small cell lung cancer. Nature 553:446–454. https://doi.org/10.1038/nature25183
10. Iraporda C, Errea A, Romanin DE et al (2015) Lactate and short chain fatty acids produced by microbial fermentation downregulate proinflammatory responses in intestinal epithelial cells and myeloid cells. Immunobiology 220:1161–1169. https://doi.org/10.1016/j.imbio.2015.06.004
11. Katoh K (2002) MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059–3066. https://doi.org/10.1093/nar/gkf436
12. Korotkevich G, Sukhov V, Budin N et al (2016) Fast gene set enrichment analysis
13. Li R, Li J, Zhou X (2024) Lung microbiome: new insights into the pathogenesis of respiratory diseases. Sig Transduct Target Ther 9:19. https://doi.org/10.1038/s41392-023-01722-y
14. Liberzon A, Birger C, Thorvaldsdóttir H et al (2015) The Molecular Signatures Database Hallmark Gene Set Collection. Cell Syst 1:417–425. https://doi.org/10.1016/j.cels.2015.12.004
15. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550. https://doi.org/10.1186/s13059-014-0550-8
16. McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLoS ONE 8:e61217. https://doi.org/10.1371/journal.pone.0061217
17. Mereiter S, Balmaña M, Campos D et al (2019) Glycosylation in the Era of Cancer-Targeted Therapy: Where Are We Heading? Cancer Cell 36:6–16. https://doi.org/10.1016/j.ccell.2019.06.006
18. Mossmann D, Park S, Hall MN (2018) mTOR signalling and cellular metabolism are mutual determinants in cancer. Nat Rev Cancer 18:744–757. https://doi.org/10.1038/s41568-018-0074-8
19. Munkley J, Elliott DJ (2016) Hallmarks of glycosylation in cancer. Oncotarget 7:35478–35489. https://doi.org/10.18632/oncotarget.8155
20. Pinho SS, Reis CA (2015) Glycosylation in cancer: mechanisms and clinical implications. Nat Rev Cancer 15:540–555. https://doi.org/10.1038/nrc3982
21. Quast C, Pruesse E, Yilmaz P et al (2012) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219
22. Rubinstein MR, Wang X, Liu W et al (2013) Fusobacterium nucleatum Promotes Colorectal Carcinogenesis by Modulating E-Cadherin/β-Catenin Signaling via its FadA Adhesin. Cell Host Microbe 14:195–206. https://doi.org/10.1016/j.chom.2013.07.012
23. Sethi V, Kurtom S, Tarique M et al (2018) Gut Microbiota Promotes Tumor Growth in Mice by Modulating Immune Response. Gastroenterology 155:33–37e6. https://doi.org/10.1053/j.gastro.2018.04.001

24. Shah T, Shah Z, Baloch Z, Cui X (2021) The role of microbiota in respiratory health and diseases, particularly in tuberculosis. Biomed Pharmacother 143:112108. https://doi.org/10.1016/j.biopha.2021.112108
25. Swart AL, Gomez-Valero L, Buchrieser C, Hilbi H (2020) Evolution and function of bacterial RCC1 repeat effectors. Cell Microbiol 22. https://doi.org/10.1111/cmi.13246
26. Thai AA, Solomon BJ, Sequist LV et al (2021) Lung cancer. Lancet 398:535–554. https://doi.org/10.1016/S0140-6736(21)00312-3
27. Thomas D, Rathinavel AK, Radhakrishnan P (2021) Altered glycosylation in cancer: A promising target for biomarkers and therapeutics. Biochimica et Biophysica Acta (BBA) -. Reviews Cancer 1875:188464. https://doi.org/10.1016/j.bbcan.2020.188464
28. Tsay J-CJ, Wu BG, Sulaiman I et al (2021) Lower Airway Dysbiosis Affects Lung Cancer Progression. Cancer Discov 11:293–307. https://doi.org/10.1158/2159-8290.CD-20-0263
29. Wu T, Hu E, Xu S et al (2021) clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov 2:100141. https://doi.org/10.1016/j.xinn.2021.100141

Figures

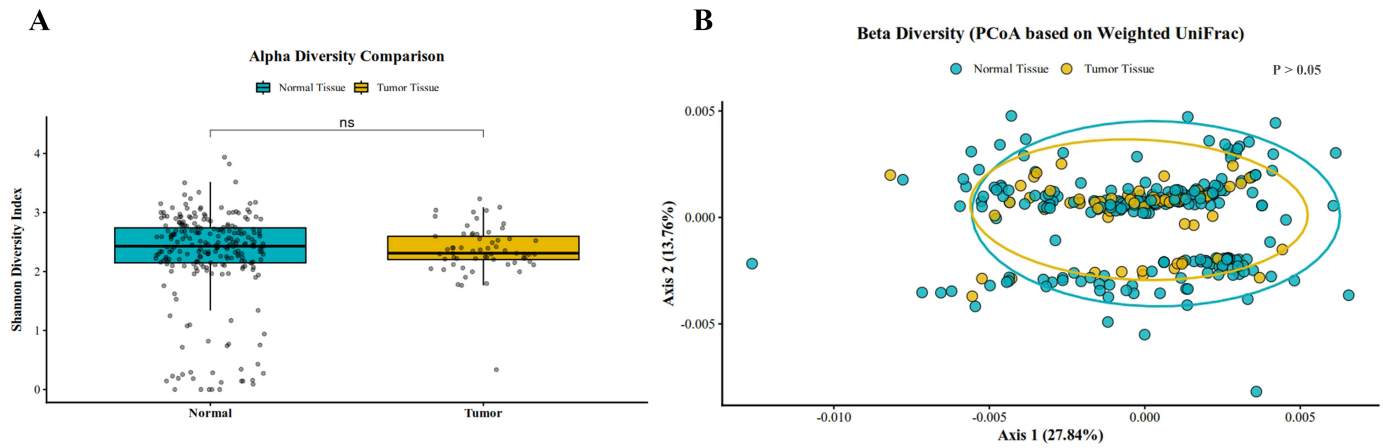


Figure 1

Diversity analysis of the microbiome in lung adenocarcinoma tumor and adjacent normal tissues.

(A) Comparison of alpha diversity of the microbiome between tumor tissues and adjacent normal tissues. The boxplot displays the distribution of the Shannon Diversity Index. "ns" indicates no statistical significance (Wilcoxon signed-rank test, $p > 0.05$).
(B) Beta diversity Principal Coordinates Analysis (PCoA) based on the weighted UniFrac distance. Each point represents an individual sample (blue: adjacent normal tissue; yellow: tumor tissue). Ellipses represent the 95% confidence intervals. The p-value ($p > 0.05$) calculated by PERMANOVA indicates no significant difference in overall community structure. The percentage in the parentheses on each axis indicates the proportion of the total variance explained by that principal coordinate.

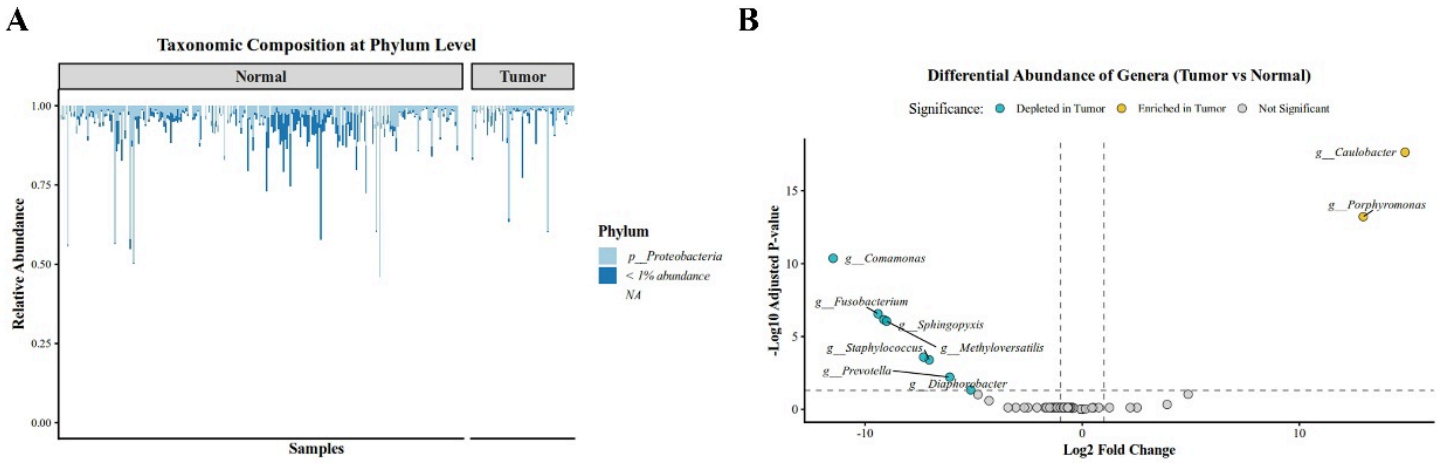


Figure 2

Taxonomic composition and differential abundance analysis of the LUAD microbiome.
(A) Stacked bar plot showing the taxonomic composition at the phylum level for tumor and adjacent normal tissues. Each vertical bar represents an individual sample, and the Y-axis represents the relative abundance.
(B) Volcano plot of differential abundance at the genus level. The X-axis represents the Log2 Fold Change of tumor versus normal groups, and the Y-axis represents the -Log10 Adjusted P-value. Yellow points indicate genera significantly enriched in tumors, blue points indicate genera significantly depleted in tumors, and grey points indicate genera with no significant difference.

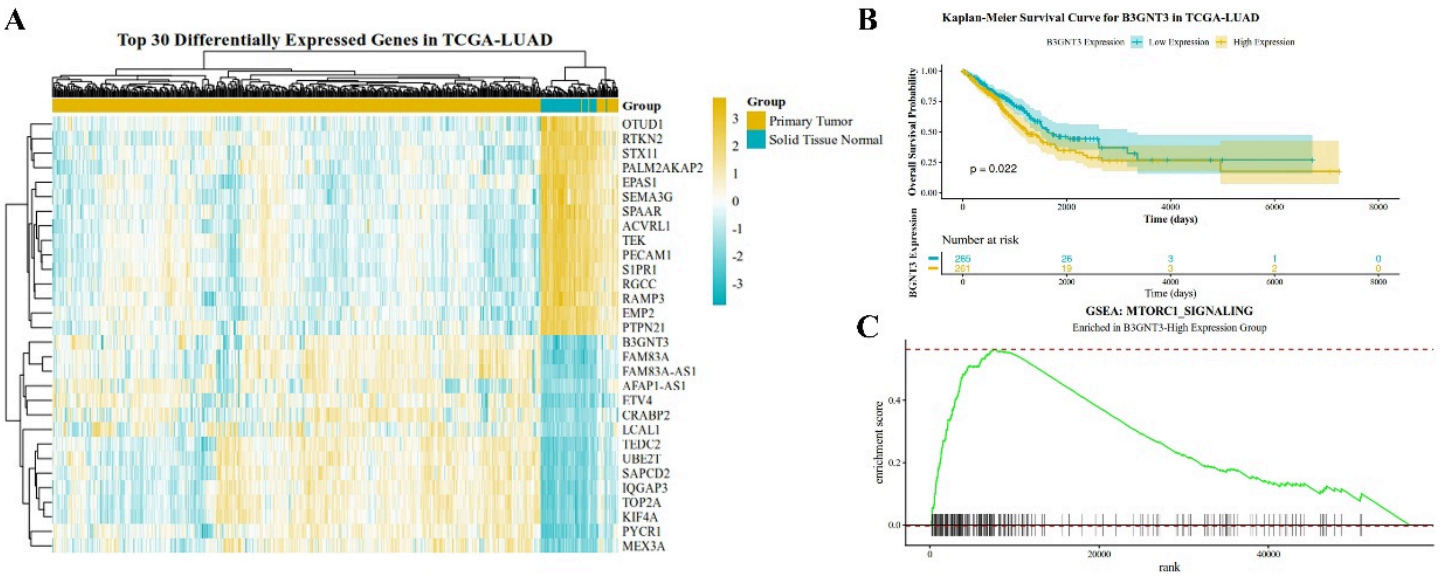


Figure 3

Analysis of B3GNT3 expression, clinical significance, and its association with mTORC1 pathway activation in LUAD.
(A) Heatmap of the top 30 most significantly differentially expressed genes in the TCGA-LUAD cohort.

Each column represents a sample, and each row represents a gene. The color scale indicates the relative gene expression level (Z-score), with yellow representing high expression and blue representing low expression.
(B) Kaplan-Meier survival curve for overall survival of LUAD patients based on B3GNT3 expression levels. Patients were divided into high-expression (yellow) and low-expression (blue) groups based on the median expression of B3GNT3. The p-value was calculated using the Log-rank test.
(C) Gene Set Enrichment Analysis (GSEA) plot showing the significant enrichment of the mTORC1 signaling pathway in the B3GNT3-high expression group.

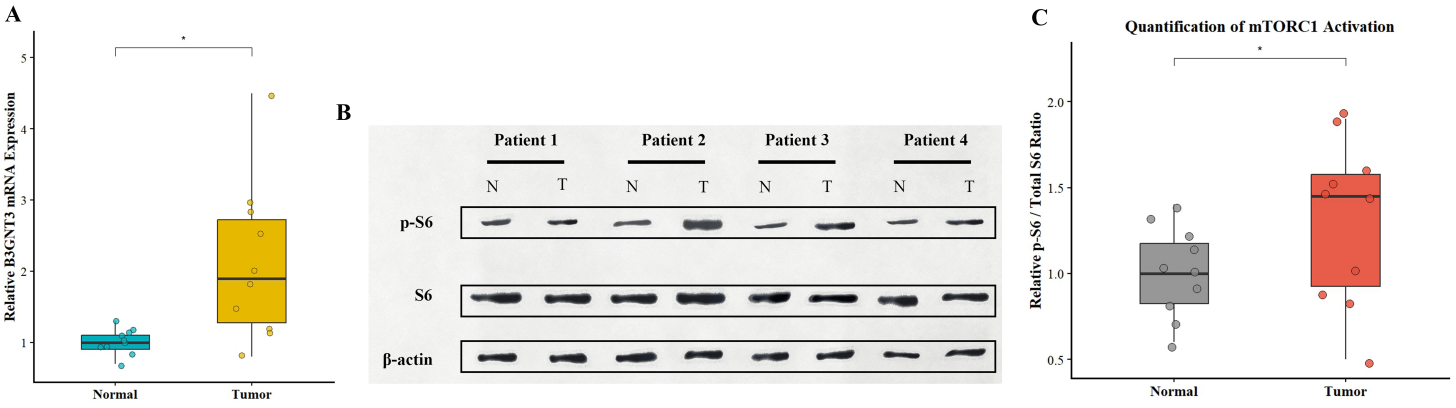


Figure 4

Experimental validation of B3GNT3 upregulation and mTORC1 pathway activation in clinical LUAD samples.
(A) The relative mRNA expression of B3GNT3 in paired LUAD tumor tissues (Tumor) and adjacent normal tissues (Normal) was measured by qRT-PCR (n = 10 pairs). The boxplot overlays jittered points representing individual samples. Statistical significance was determined using a paired Student's t-test. *p < 0.05.
(B) Representative Western blot analysis of phosphorylated S6 (p-S6), total S6, and β-actin in four paired clinical samples. β-actin served as the loading control. N, adjacent normal tissue; T, tumor tissue.
(C) Semi-quantitative analysis of the relative p-S6 to total S6 protein ratio across all 10 pairs of matched samples. Data were analyzed using a paired Student's t-test. *p < 0.05.