

Plasma GPI and PGD are associated with vascular normalization and may serve as novel prognostic biomarkers for lung adenocarcinoma: Multi-omics and multi-dimensional analysis

Yiran Liu^{1,*}, Yanchi Wang^{3,*}, Liping Mao^{3,*}, Yang Hu², Rongrong Zhao⁴, Wendi Zhang¹, Huiwen Xu¹, Buyun Xu¹, Yutong Wu¹, Qiong Chen¹, Xiaobo Tao¹, Shufan Xu¹, Xun Zhuang¹, Yulong Lian¹, Yihua Lu¹, Lei Zhang¹, Tian Tian¹, Guangyu Tian^{4,#}, Jiahua Cui^{1,#}, Minjie Chu^{1,#}

¹Department of Epidemiology, School of Public Health, Nantong University, Nantong, Jiangsu, China

²Affiliated Nantong Hospital of Shanghai University (The Sixth People's Hospital of Nantong), Nantong, Jiangsu, China

³Department of nutrition, Hai 'an City People's Hospital, Nantong, Jiangsu, China

⁴Department of Oncology, Jiangdu People's Hospital of Yangzhou, Yangzhou, Jiangsu, China

*These authors contributed equally to this work.

#Correspondence:

Minjie Chu

Department of Epidemiology, School of Public Health, Nantong University, Nantong, Jiangsu, China, 226019

E-mail: chuminjie@ntu.edu.cn

Jiahua Cui

Department of Epidemiology, School of Public Health, Nantong University, Nantong, Jiangsu, China, 226019

E-mail: cuijiahua@ntu.edu.cn

Guangyu Tian

Department of Oncology, Jiangdu People's Hospital of Yangzhou, Yangzhou, Jiangsu, China, 225202

E-mail: jdryzlzx@hotmail.com

WITHDRAWN
see manuscript DOI for details

Abstract

Background: The aim of this study was to explore potentially novel plasma protein biomarkers for lung adenocarcinoma (LUAD).

Methods: A plasma proteomics analysis was carried out in LUAD cases and healthy controls. Candidate protein biomarkers were validated in 102 LUAD cases and 102 matched healthy controls. The same LUAD tumor tissues were detected to explore the correlation between the expression of candidate proteins in tissues and plasma and vascular normalization. A LUAD active metastasis mice model was constructed to explore the role of candidate proteins for lung metastasis.

Results: In the screening phase, 5 candidate proteins were selected. In the validation phase, GPI and PGD were verified to be upregulated in plasma from LUAD patients. The expression of GPI in tumor tissue was positively correlated with the expression of GPI in plasma and negatively correlated with the normalization of tumor blood vessels. Meanwhile, a negative correlation between the expression of GPI and PGD in plasma and tumor vascular normalization was discovered. In the LUAD active metastasis model, the lowest levels of vascular normalization and the highest expression of GPI and PGD were found in mice with lung metastases.

Conclusion: This study found that GPI and PGD may be potential diagnostic plasma biomarkers for LUAD, providing new clues to assist in LUAD early screening. Meanwhile, monitoring the concentration of GPI and PGD in plasma may reflect the level of vascular normalization within the tumor and infer the risk of metastasis and malignancy of the tumor.

Keywords: Glucose-6-phosphate isomerase (GPI); 6-phosphogluconate dehydrogenase (PGD); Biomarker; Proteomics; Lung adenocarcinoma.

1. Introduction

Lung cancer still ranks the leading cause of cancer-related mortality, and 1.8 million patients died of which in 2020 worldwide¹. Non-small cell lung cancer (NSCLC) accounts for about 85% of all lung cancer cases², while lung adenocarcinoma (LUAD) accounted for nearly 60% of all NSLSC cases, presenting the primary histologic subtype³. What's more, the incidence characteristics of LUAD present differently between genders, with the incidence of female LUAD patients increasing significantly in recent years^{4,5}. Owing to LUAD is often initially asymptomatic, lacks effective screening programs, and has a greater metastasis tendency⁶, thus always leads many patients to be diagnosed at an advanced stage with poor survival. However, the fact that LUAD patients diagnosed with early-stage or have undergone surgery are more likely to get a better survival⁶, indicating the urgent need for an effective screening and diagnostic strategy for detecting LUAD at an earlier time.

Although tissue biopsy and tissue biomarkers are commonly recommended and still irreplaceable in clinical application, however, the clinical significance of tissue biopsies is mainly applied to distinguishing of benign and malignant tumors and sometimes can hardly be performed for early cancer screening and diagnosis. Besides, tissue biomarkers could not be sampled until after the tumor was detected. Meanwhile, tissue sampling is invasive, time-consuming, challenging to obtain, and harmful to patients^{7,8}. Sometimes, in order to obtain enough samples, patients need repeated biopsies, and some patients are unable to undergo tissue sampling due to their wishes or other medical reasons⁹. The advent of liquid biopsy may make up for these limitations and may be more appropriate for patients who unwilling or unable to undergo a biopsy.

Samples for liquid biopsy involve blood, mucosa, pleural effusions, urine, and cerebrospinal fluid¹⁰. As blood is more accessible, efficient, convenient, and affordable than other body fluids, blood has become the critical fluid for liquid biopsy analysis. It is a valuable diagnostic tool to support differential diagnosis, estimate prognosis, and early detection of residual or progressive disease¹¹. Analytes for blood biopsy include circulating tumor cells, circulating nucleic acids, the tumor-derived

fraction of cell-free DNA, cell-free RNAs, extracellular vesicles, tumor-educated platelets, proteins, and metabolites¹². One particular concern is that plasma contains an extensive and relatively comprehensive array of proteins, and the high stability of plasma proteins and the wealth of biological information they carry make them available as protein biomarkers for cancer. A previous study performed integrative proteomics analysis of 103 paired LUAD tumor tissues and adjacent non-tumor tissues from the Chinese population and identified differentially expressed proteins in the tissues. The most highly and commonly expressed protein with prognostic ability in tumor tissues was validated in plasma, resulting in HSP90AB1 as a potential prognostic biomarker¹³. Meanwhile, other proteins, like SAA2¹⁴, CXCL17¹⁵, CEACAM5, VEGFR2, and ERBB3¹⁶, have been identified as plasma biomarkers for LUAD via validating proteins that are differentially expressed in tissues. However, previous studies usually simply validated LUAD-related proteins in plasma rather than systematically performing high-throughput comprehensive proteomic analysis of plasma proteins and lacked integration of proteins consistently expressed in both tissues and plasma for analysis.

Considering its specific protein composition structure, it makes sense that there was rarely a comprehensive and systematic analysis of plasma proteins. Just one single protein, albumin, accounts for about 50% of all proteins; the top 22 proteins comprise 99% of all proteins. "Blood+" high-depth blood proteomics analysis can filter out interfering signals of high-abundance proteins and increase the detection depth of blood over 1000 while still maintaining good quantitative parallelism. Potential biomarkers for gastric cancer and type 2 diabetes have been identified by this high-depth proteomics technique^{17 18}, however, this technique has not yet been applied to exploring of LUAD protein biomarkers.

In this study, to explore potential plasma protein biomarkers for LUAD, we first performed "Blood+" high-depth plasma proteomics analysis on LUAD patients and healthy controls. Proteomics data of LUAD tumor tissues and adjacent non-tumor tissues were then employed to select proteins that were consistently differentially expressed in both plasma and tissues. Meanwhile, to further maintain the consistency

of differential protein expression on both transcriptome and proteomics levels, we introduced RNA-seq data of LUAD tumor tissues and adjacent non-tumor tissues in the previous step for further screening for plasma proteins consistent with mRNA and protein expression in tissues. After that, we expanded the sample and validated the candidate plasma protein biomarkers. Finally, we performed functional experiments on cancer metastasis in mice to explore the possible mechanisms for the differences in plasma protein expression, further supporting the value of plasma protein biomarkers to diagnose and predict the prognosis of LUAD.

2. Materials and methods

2.1 Study design

In the first phase, proteomic analysis was performed on plasma collected from 10 LUAD cases and 10 matched healthy controls to identify differential expressed proteins. To select proteins consistently expressed in plasma and tumor tissues, we combined proteomic data from LUAD tumor tissues and adjacent non-tumor tissues from the Chinese population. Proteins in tissues with fold change (FC) $>3/2$ or $<2/3$ were retained from raw proteomics data uploaded in the iProX Consortium with the subproject ID IPX0001804000, and proteins among them that overlapped with the differential plasma proteins in the first phase were selected. Considering that only about 40% of protein expression differences were consistent with changes in transcript levels, we cited RNA-seq data collected in GSE140343 and selected mRNAs with FC $>3/2$ or $<2/3$ that were consistently differentially expressed at the mRNA and protein levels in tissues as candidate plasma protein biomarkers (**Figure 1**). To validate the differential expression of candidate proteins biomarkers in plasma, we included 62 female LUAD cases and 62 age-matched healthy controls in the validation phase. For exploring whether the expression of candidate protein biomarkers is consistently expressed in male LUAD patients, 40 male LUAD cases and 40 age-matched healthy controls were included for validation of the candidate proteins.

2.2 Study Population

The 10 female LUAD cases in the first phase were obtained between 2021.08 and

2021.11 at Hai'an City People's Hospital, Jiangsu Province, with 10 healthy controls matched by age. The population in the validation phase was obtained from two parts: a portion of 94 LUAD cases and 94 healthy controls matched according to gender and age from Hai'an City People's Hospital and the Sixth People's Hospital of Nantong between 2021.12 and 2022.06; The other 8 LUAD cases are from the inclusion phase of a cohort with more than two thousand participants in Jiangdu, Jiangsu Province which were discovered with LUAD at recruitment, and eight healthy controls were matched by gender and age in the cohort. The study was reviewed and approved by the ethics committee of Nantong University (Approval No. 2022-2). Informed consent was obtained from all patients.

2.3 "Blood+" high depth blood proteomics analysis

Proteomics analysis was conducted by Jingjie PTM BioLabs (Hangzhou, China). Primary experimental procedures for "Blood+" high depth blood proteomics analysis included protein extraction, trypsin digestion, LC-MS/MS analysis, and data analysis. First, the plasma samples were centrifuged at 4 °C, 12,000 g for 10 min to remove cell debris. Then, the protein concentration was determined with BCA kit according to the instructions. After trypsin digestion of each sample protein into peptide fragments, then the tryptic peptides were dissolved in solvent A (0.1% formic acid, 2% acetonitrile/ in water) and separated using an EASY-nLC 1200 UPLC system (ThermoFisher Scientific). Finally, the resulting raw data were processed using MaxQuant search engine (v.1.6.15.0). FDR was adjusted to < 1%. For the determination of differentially expressed proteins, the fold change is calculated as the average of the comparisons between biological replicates.

2.4 Sample size calculation

The sample size for the validation phase is determined by the following formula, where n = sample size, α =significant level, δ =1- power of test, and p = sensitivity or specificity. It is assumed that the sensitivity and specificity for LUAD diagnostic screening were both 80%, setting significant level as 0.05 and power of test as 0.90, thus giving an estimated sample size of 61.

$$n = [\mu_{\alpha}/(\delta/\sqrt{p(1-p)})]^2, \alpha = 0.05, \delta = 0.10$$

2.5 Data sources

The proteomic and transcriptomic data of LUAD tumor tissues and adjacent non-tumor tissues in the Chinese population were obtained from the research published in *cell* by Xu *et al*¹³. Their proteomic data were downloaded on the iProx consortium with the subproject ID IPX0001804000. And the RNA-seq data were downloaded from the Gene Expression Omnibus (GEO) database with an accession number: GSE140343.

2.6 Enzyme-linked immunosorbent assay (ELISA)

ELISA kits allow for the determination of *ASS1*, *PGD*, *GAPDH*, *GPI* in plasma samples in this study. Collected plasma was sampled after centrifugation at 2,000 rpm for 20 min at 4°C and stored at -80°C until analysis. Commercial enzyme-linked immunosorbent assay (ELISA) kits purchased from Meimian Biotechnology (Yancheng, Jiangsu, China) were used to measure protein level in human plasma samples according to the manufacturer's recommendations. Absorbances were measured in a microplate reader (Sunrise, Tecan, Austria) at 450 nm.

2.7 Evaluation of candidate plasma protein biomarkers for LUAD diagnosis

To further evaluate the ability of candidate plasma protein biomarkers as potential biomarkers for LUAD diagnosis, receiver operating characteristic (ROC) curves were constructed, and the area under the ROC curves (AUC) were calculated based on samples originating from the validation phase.

2.8 Kaplan-Meier Survival Analysis

The Kaplan-Meier Plotter website (<https://kmplot.com/analysis/>) was used to analyze the relationship between mRNA expression and overall survival of LUAD patients. The proteomics data and clinical data were obtained from iProx Consortium with the subproject ID IPX0001804000 (<https://www.iprox.org/page/PSV023.html?url=15717295820259eyj>) for further analysis. The relationship between the expression of proteins with survival time of

LUAD patients was assessed via Cox regression analysis. Differences between high and low risk groups were evaluated using the Kaplan-Meier method, based on the best separation of expression, employing R packages of survminer and survival.

2.9 Functional analysis of GPI, PGD, and related genes

To identify enriched functional classes of *GPI* and *PGD* and related genes, the functional relationships of genes associated with *GPI* and *PGD* were revealed using STRING (<https://string-db.org/>), and Cytoscape was used to construct the PPI (protein-protein interaction) network of the top 50 related genes. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) (<https://david.ncifcrf.gov/>). Following the instructions of the DAVID manual, *GPI* and *PGD* and the top 50 related genes were uploaded, and the function charts were generated. The 10 groups with more than two genes and the most significant *P*-values in each category were retained for study.

2.10 Immunohistochemical staining and immunofluorescent staining

LUAD tumor tissue specimens were fixed in Bouin's fixative for 24 h, then embedded in paraffin and sectioned by paraffin microtome (Thermo HM340E). After deparaffinization and rehydration, the slides were immersed in sodium citrate buffer (pH 6.0) for antigen retrieval and maintained at sub-boiling temperature for 15 min. Subsequently, the slides were incubated with 10 % normal goat serum for 30 min at room temperature to block non-specific staining. For immunohistochemistry stainings, the slides were incubated with GPI (CST, 94068T) or PGD (Proteintech, 14718-1-AP) antibody solution for 24 h at 4 °C with gentle shaking. The images were scanned with pathology section scanner (Pannoramic MIDI) after DAB staining. For dual immunofluorescence staining, the slides were simultaneously incubated with mouse-derived CD31 (Servicebio, GB12063) and rabbit-derived α -SMA (Servicebio, GB111364) antibody solution for 24 h at 4 °C with gentle shaking. The images were scanned with pathology section scanner (Pannoramic MIDI) after incubation with the mouse-derived (red) and rabbit-derived (green) fluorescent secondary antibody. The staining intensity was analyzed with Image J.

2.11 LUAD active metastasis model

All care and treatment of experimental mice were approved by Nantong University Animal Care and Use Committee guidelines. The 6 weeks old C57BL/6 male mice were purchased from SLAC Laboratory Animal Center (Shanghai, China) and maintained in SPF facilities. Murine-derived LUDA LLC cells (5×10^5) were subcutaneously injected into the right axilla of C57BL/6 mice. When the subcutaneous tumor volume of the mice reached 1500 mm^3 , the subcutaneous tumor tissue was removed by aseptic surgery. After 4 weeks, the mice were euthanized to observe the occurrence of lung metastasis. Tumor tissues from mice without lung metastases and mice with lung metastases were taken, and the vascular marker CD31 was co-stained with the perivascular endothelial marker α -SMA by immunofluorescence. Take tumor tissues of mice without lung metastasis, the tumor tissues of mice with lung metastasis and the corresponding lung metastasis tumor tissues, and evaluate the expression of GPI, PGD and perivascular endothelial marker α -SMA by Western Blot.

2.12 Statistical analysis

Differences in the distribution of demographic characteristics between the case and control groups were expressed as medians and percentages, calculated using the two-sided chi-square tests. T-test and Wilcoxon signed-rank test were performed to compare the expression difference of each detectable protein between the case group and control groups. Protein intensities were normalized before Wilcoxon signed-rank test were performed. Before performing the t-test, the normalized protein intensities were \log_2 transformed to ensure the normality. Proteins with a fold change $>3/2$ or $<2/3$ ($\text{FC} = \text{Average (tumor tissues or LUAD)} / \text{Average (adjacent non-tumor tissues or healthy controls)}$) and $P < 0.05$ are considered as significantly differentially expressed. R 4.1.1 and SPSS 23.0 were used for the statistical analysis.

3. Result

3.1 Characteristics of the study population

The characteristics of the participants in the case and control groups are detailed in **Table 1**. In the screening phase, comparable ages existed between the case and control groups ($P=0.79$), and the proportion of cases with stage I LUAD reached 80%. In the validation phase, the age distribution was not different in females between the case and control groups ($P=0.14$), and 91% of the female cases were stage I LUAD. Meanwhile, there was no difference in male age distribution between the case and control groups ($P=0.95$), with stage I LUAD accounting for 75% of the male LUAD cases.

3.2 “Blood+” proteomics analysis of plasma samples

A total of 1,331 proteins were identified by “Blood+” high-depth blood proteomics analysis, of which 1,113 could be quantitatively analyzed. On this basis, analyzed at $FC>3/2$ or $<2/3$ level, 415 proteins are selected, with 333 proteins upregulated and 82 proteins downregulated in the case group. Among these, 102 proteins were upregulated and 28 proteins were downregulated by the significance test at $P<0.05$. The details are mentioned in **Figure 2A-2B**.

3.3 Selection of differentially expressed proteins in both plasma and tissues

We subsequently analyzed the expression of above 415 candidate plasma proteins in 103 paired LUAD tumor tissues and adjacent non-tumor tissues, selecting proteins with consistent direction of altered expression in both plasma and tissues. According to the criteria of $FC>3/2$ or $<2/3$, 4,854 proteins were upregulated and 3,045 proteins were downregulated in tissues. Among which, 147 proteins were differentially expressed and altered in the same direction in both tissues and plasma, including 100 upregulated proteins and 47 downregulated proteins.

3.4 Selection of differentially expressed proteins in both proteome and transcriptome

We expect to identify plasma proteins that are consistent with the direction of differential mRNA and protein expression in tumor tissues. Therefore, we introduced GSE140343, which including RNA-seq data from 51 LUAD tumor tissues and 49 adjacent non-tumor tissues, with 5,269 upregulated mRNAs ($FC>3/2$) and 2,929

downregulated mRNAs ($FC < 2/3$) in LUAD tumors. Of the previous 147 proteins differentially expressed in both tissues and plasma, 14 upregulated proteins were also expressed as upregulated in the mRNA level in tissues (**Table 2**). Wilcoxon signed-rank test and t-test were performed separately for these 14 proteins to improve the test efficacy, and five overlapped proteins exhibited expression differences at the criterion of $P < 0.05$ (**Table 3**). Among them, HSP90AB1 has been completely validated as a potential prognostic marker in previous study, so this protein was not repeated in this study and was excluded in the next phase of validation.

3.5 Validation of candidate plasma proteins

To verify the difference in expression of *ASS1*, *GAPDH*, *PGD*, and *GPI* between LUAD patients and the healthy controls, we expanded samples and assessed the expression levels of these proteins in the plasma from female and male LUAD cases and matched healthy controls using ELISA. Statistically significant differences in the expression of *GPI* and *PGD* were found between the 62 female LUAD cases and 62 healthy controls. Compared to the control group, *PGD* and *GPI* expression levels were both significantly higher in the female case group, with P -value were 2.17×10^{-15} and 4.12×10^{-14} , respectively (**Figure 2C**, **Figure 2D**). Further, the expression of *GPI* and *PGD* is also found upregulated in the plasma of male LUAD cases with P values of 1.22×10^{-9} and 3.83×10^{-7} . After combining male and female cases, *GPI* ($P = 1.2 \times 10^{-14}$) and *PGD* ($P = 4.1 \times 10^{-13}$) were both highly expressed in plasma of the combined LUAD cases (**Figure 2C**, **Figure 2D**). However, the differences in the expression levels of *ASS1* and *GAPDH* in the case and healthy control groups were not statistically significant ($P > 0.05$), thus they will be excluded from future research. As a result, *GPI* and *PGD* proteins are consistently upregulated in plasma and tumor tissues of LUAD cases (**Figure 3A-3B**, **Figure 3D-3E**). And *GPI* and *PGD* proteins and mRNA are consistently upregulated in tumor tissues of LUAD cases (**Figure 3B-3C**, **Figure 3E-3F**).

3.6 Evaluation of *PGD* and *GPI* for LUAD diagnoses

To investigate the potential of *GPI* and *PGD* as plasma biomarker in LUAD diagnosis, the ROC curves and the AUC were calculated using subjects from validation phase,

including 102 LUAD cases and 102 healthy controls. The ROC curves showed strong distinction between the LUAD cases and healthy controls, with the AUC of 0.795 (95%CI: 0.733-0.856) for GPI (**Figure 2E**) and 0.839 (95%CI: 0.785-0.893) for PGD (**Figure 2F**). The sensitivity and specificity of GPI for LUAD diagnosis screening were 77.5% and 74.5%, respectively, and the sensitivity and specificity of PGD for LUAD diagnosis screening were 87.3% and 72.5%, respectively. By combining GPI and PGD, we generated a ROC curve with an AUC of 0.873 (95%CI: 0.824-0.922), sensitivity of 93.1% and specificity of 72.5% (**Figure 2G**).

3.7 Survival analysis of PGD and GPI in LUAD

Using the Kaplan-Meier Plotter website, the association between the mRNA expression levels of GPI and PGD in LUAD patients and their prognosis was investigated by the Kaplan-Meier analysis. As illustrated in **Figure 4A-4B**, expression of GPI mRNA and PGD mRNA were significantly correlated with the prognosis in LUAD, with overexpression of both GPI and PGD leading to a worse prognosis ($P=0.024$). In addition, the results were validated at the protein level. The group with high protein expression of GPI and PGD had a considerably worse prognosis than group with low expression in LUAD patients ($P=0.048$, $P=0.034$) (**Figure 4C- 4D**).

3.8 Gene ontology (GO) classifications and KEGG mapping

The enrichment of the GO and KEGG pathways was carried out by DAVID Bioinformatics Resources. 106 GO terms, including 48 biological processes, 16 cellular components, and 42 molecular function terms, were assigned to GPI, PGD, and the related top 50 genes. In the molecular function, "identical protein binding" was the most significantly enriched GO-term. GO cellular compartment analysis showed that GPI and related genes were highly enriched in "cytosol". The most prevalent biological processes assignment was "glycolytic process" (**Figure 5A-5B**). The pathways in the KEGG database serve as representations of molecular interactions and reaction networks. To identify signaling pathways involved in GPI and PGD, we had mapped the KEGG database and discovered that the relevant genes were significantly enriched in 22 KEGG pathways. The relevant genes were highly clustered in several signaling pathways, such as "Metabolic

pathways”, “Glycolysis/Gluconeogenesis”, “Carbon metabolism” and “Pentose phosphate pathway” (**Figure 5C-5D**).

3.9 Association between GPI and PGD expression and tumor vascular normalization

It is unclear whether the expression of GPI and PGD in plasma correlates with the expression of GPI and PGD in tissue, even though both GPI and PGD are upregulated in tissue and plasma. For that, we determined the expression of GPI and PGD in the corresponding tumor tissue from 10 LUAD cases in which plasma was previously used for plasma proteomics analysis. The immunohistochemical staining results of GPI and PGD in tumor tissue are shown in **Figure 6A** and **Figure 6B**, respectively. A correlation analysis was performed to assess the relationship between the expression of GPI and PGD in plasma and tumor tissue. Significant positive correlations were detected between the expression of GPI in tissues and plasma ($R=0.93$, $P=1.2\times 10^{-4}$), while the expression of PGD did not show a correlation in tissues and plasma ($R=0.13$, $P=0.710$) (**Figure 6C**).

Tumor cells and its related substances breaking through the vascular basement membrane and entering the vasculature are the prerequisites for blood biopsy. And the immature and unstable vasculature of the tumor microenvironment provides a convenience for primary tumor cells and its related substances to enter the vasculature. To identify the role of GPI and PGD in tumor vascular normalization, we performed immunofluorescent staining on the corresponding tumor vessels from 10 LUAD cases in which plasma was previously used for plasma proteomics analysis (**Figure 6D**). Correlation analysis revealed a negative correlation between GPI expression in tissues and the normalized level of tumor vessels ($R=-0.66$, $P=0.036$) (**Figure 6E**), suggesting that high expression of GPI in tumor tissues may lead to abnormal tumor vascular development. The degree of tumor vascular normalization was inversely correlated with the expression of both GPI ($R=-0.71$, $P=0.021$) and PGD ($R=-0.64$, $P=0.045$) in plasma (**Figure 6F**), indicating that abnormalities in tumor vessels may facilitate the infiltration of substances from tumor tissue into the bloodstream.

3.10 LUAD active metastasis model

We constructed a LUAD active metastasis model to explore the association between the expression of GPI and PGD in tumor tissue and lung metastasis (**Figure 7A**). The blood vessels of the tumor tissue in situ without lung metastasis (a) and the tumor tissue in situ with lung metastasis (b) were performed immunofluorescent staining assay (**Figure 7B**). The results showed that the tumor tissue in situ with lung metastasis presented abnormal vascular morphology (**Figure 7C**), demonstrating that abnormalities in tumor vessels increase the risk of lung metastasis.

Western-blot assay showed that the expression of GPI and PGD was highest in the tumor tissues of mice with lung metastasis. The expression of GPI and PGD was higher in the tumor tissues of mice with lung metastasis (c) than those without lung metastasis (a) (**Figure 7D, Figure 7E**), suggesting that high expression of GPI and PGD in tumor tissues may promote lung metastasis in mice. The Western-blot results of α -SMA, representing the level of vascular normalization, were consistent with the immunofluorescent staining results of tumor vessel. The expression of α -SMA in tumor tissues of mice with lung metastases (b) was lower than that in mice without lung metastases (a), and the expression of α -SMA was lowest in lung tumors with lung metastasis (c) (**Figure 7E**), indicating that abnormalization of blood vessels may increase the risk of lung metastasis.

4. Discussion

Screening and diagnosis of LUAD remains dependent on imaging nowadays, lacking more economical and efficient strategies to detect LUAD at the beginning of tumor development. The adoption of plasma proteins as biomarkers is less invasive and more concise and efficient than previous tissue tests, while may also help in dynamic monitoring of LUAD progression. After several phases of screening and validation, transcriptomic and proteomic analyses of LUAD revealed that GPI and PGD are highly expressed in both plasma and tumor tissues of LUAD patients. ROC curve analysis demonstrates that GPI and PGD can distinguish well between LUAD and healthy controls, showing that they may be used for the early diagnosis of LAUD. The results of immunohistochemistry, immunofluorescence staining and mice models

showed a negative correlation between GPI and PGD expression in plasma and vascular normalization, suggesting that high expression of GPI and PGD may indicate a risk of tumor metastasis, leading to a poor prognosis.

This study initially identified differential protein expression in the plasma of LUAD patients, and then referenced data from proteomic analysis in LUAD tumor tissues, that also from Chinese patients, to maintain the consistency of protein expression in both tissues and plasma. Gene expression can be divided into two primary phases: mRNA and protein expression, which are regulated by complex mechanisms. Although, on average, the correlation between protein and mRNA concentration was weaker in multicellular organisms¹⁹, we could find five proteins that were consistently upregulated in mRNA and protein levels by integrating the data of multi-omics. The expression of these five proteins is consistent in tissues and plasma, improving the candidate proteins' reliability as biomarkers for LUAD. At the same time, the five proteins are co-upregulated at the mRNA and protein levels in tissues, which may provide clues for subsequent exploration of possible regulatory pathways.

Among female lung cancer patients, the proportion of LUAD patients keeps increasing, from 54.2% in 1993-1999 to 62.3% in 2010-2015²⁰. Therefore, in the first phase of this study we included only female LUAD cases and female healthy controls for proteomic analysis, with the aim of focusing on plasma protein biomarkers for female LUAD. After expansion of the sample, ELASA was performed separately by gender grouping, GPI and PGD showed upregulation in the plasma of both male and female LUAD cases. Although proteomic analysis included only female cases and healthy controls at the screening phase, differences in expression of GPI and PGD were validated in both female and male, suggesting that the above two potential protein biomarkers may be universal between male and female LUAD patients, despite the gender-specific differences that exist in the incidence of LUAD.

In this study, analysis of the KEGG pathway for genes related to GPI and PGD revealed that these genes are enriched in carbon metabolism, glycolytic and pentose phosphate pathways (PPP), all associated with energy metabolism. One of the

"hallmarks of cancer" is an altered energy metabolism, characterized by a preferential reliance on glycolysis for energy production in a manner that is not oxygen-dependent^{21 22}. Cancer cells exhibit enhanced glucose absorption and glycolysis, and glycolytic genes are often overexpressed in 24 types of cancer, including lung cancer²³. PPP is critical for cancer cell survival as well as ribonucleotide and lipid production, and accumulating evidence suggests that tumorigenic lesions in cancer cells regulate the flux of PPP either directly or indirectly²⁴. GPI (EC 5. 3. 1. 9), a dimeric enzyme with a molecular mass of 132,000, catalyzes the interconversion of D-glucose-6-phosphate and D-fructose-6-phosphate²⁵. Knockdown of GPI has been shown to severely impair glycolysis in cancer cells without affecting overall cell survival²⁶. And regulation of glycolytic proteins like GPI has anti-tumor effects and may be a potential target for tumor treatment^{27 28}. As well, GPI is implicated in the prognosis of cancer. GPI promotes metastasis in colorectal cancer²⁹ and is also a prognostic predictor for hepatocellular carcinoma³⁰. In addition to glycolytic mechanisms, GPI functions in the development of the tumor immune microenvironment and LUAD³¹. Most importantly, it has recently been demonstrated that GPI is informative in the diagnosis and prognostic assessment of LUAD³¹.

PGD is the third enzyme in the PPP, has been reported to contribute to tumor development in cancer cell metabolism³², lipogenesis³³, redox homeostasis³⁴, cell metastasis and proliferation³⁵, and radiation resistance³⁶. PGD is linked to tumor cell migration, and by lowering down PGD, tumor cell proliferation can be reduced^{33 34}. PGD was reported to be upregulated in lung cancer, and upregulation of PGD expression plays an important role in lung cancer cell proliferation and tumor growth³⁷. Hyperexpression of PGD gene was associated with reduced survival in ovarian and lung cancer patients³⁸, consistent with the results of this study. Blockade of PGD can enhance anti-tumor function by generating CD8+ effector T cells, and accordingly was screened as a potential immunotherapeutic biomarker for LUAD³⁹. PGD expression was markedly upregulated in cisplatin-resistant cancer cell lines, suggesting that PGD may represent a fresh avenue for overcoming the drug's resistance³⁸.

Tumor abnormal vasculature were characterized by discontinuous endothelial lining and defective basement membrane, which Promotes tumor cells and its related substances into the body circulation. Then, plasma protein biomarkers derived from tumor cells may be more sensitive and specific in recognizing LUAD, so it is desirable to identify the source of plasma proteins. In this study, we found that high expression of GPI in tumor tissues may lead to abnormalities in tumor vasculature, and the abnormalities in tumor vasculature may promote the entry of cancer cells and related substances into the circulation, thus leading to tumor metastasis. This is supported by the negative correlation between GPI and PGD expression in plasma and normalization of tumor vessels, indicating that GPI and PGD in plasma are leaking into the blood through heterogeneous vessels in tumor tissue. Tumor cells are highly glycolytic⁴⁰, while glycolysis inhibits the vascular support function of pericytes⁴¹. Targeting tumor cell glycolysis induces normalization of tumor vasculature, thereby reducing invasion, endocytosis, and dissemination of cancer cell⁴⁰. Therefore, as key enzyme proteins of glycolysis, monitoring the concentration of GPI and PGD in plasma can reflect the normalized level of blood vessels within the tumor and infer the risk and malignancy of tumor metastasis (**Figure 7F, Figure 7G**).

This study has several advantages. Firstly, we initially performed proteomic analysis of plasma from LUAD cases, and then selected candidate plasma proteins that were also differentially expressed in tumor tissues with the same direction. The selection of plasma proteins with the consistent direction of differential expression in both plasma and tissues enhances the reliability for subsequent biomarker evaluation and provides higher feasibility for subsequent screening of high-risk groups for LUAD. Secondly, we maintained the consistency of differential protein expression at the transcriptomic and proteomic levels and screened for plasma candidate proteins consistent with mRNA and protein expression in tissues. Maintaining consistent expression of mRNA and protein provides clues for subsequent in-depth exploration of the relevant pathways and mechanisms regulating the development of LUAD. Thirdly, plasma from 10 LUAD cases was selected for proteomic analysis, and tumor tissues from the same 10 LUAD cases were also collected for immunohistochemical and

immunofluorescence staining. As a result, information on plasma GPI and PGD expression, tissue GPI and PGD expression, and vascular normalization levels were obtained from the same individuals. Determining the expression of GPI and PGD in both tissues and plasma, and observing the corresponding vascular normalization, which are all from the same individual, for correlation analysis is more convincing and credible, avoiding bias from individual differences. Last, this study identified two potential protein biomarkers for LUAD from the population and conducted in vivo animal experiments and a tumor-bearing mice model to attempt to explore possible mechanisms for the high expression of GPI and PGD in plasma. By starting from a proteomic analysis, then validating it in plasma, and finally rationalizing the results via functional experiments, thus provides a comprehensive illustration that GPI and PGD may be potential biomarkers in plasma for LUAD.

5. Conclusion

In summary, we identified GPI and PGD as potential novel diagnostic and prognostic biomarkers for LUAD. PGD and GPI can be used as diagnostic biomarkers in combination with other available strategies to assist in the screening and diagnosis of LUAD, and as prognostic biomarkers aid in predict the risk of tumor metastasis and malignancy in patients with LUAD.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (82273715, 82203771), the National Key Research and Development Program of China (2022YFC2503202), the Science and Technology Program of Nantong City (MS22022062, JC22022002), the Talent Training Climbing Action Plan Project of the Sixth People's Hospital of Nantong, and the Postgraduate Research and Practice Innovation Program of Jiangsu Province (KYCX21-3126, KYCX20-2852).

Author contributions

Yiran Liu: Conceptualization, Investigation, Writing – original draft, Writing – review & editing. **Yang Hu:** Conceptualization, Investigation, Resources, Writing – original draft; **Yanchi Wang:** Funding acquisition, Writing – review & editing. **Liping Mao:** Conceptualization, Writing – review & editing. **Rongrong Zhao:** Data curation, Investigation. **Wendi Zhang:** Formal analysis, Software. **Huiwen Xu:** Data curation, Investigation. **Buyun Xu:** Data curation **Yutong Wu:** Data curation, Formal analysis, Investigation. **Qiong Chen:** Software, Investigation. **Xiaobo Tao:** Data curation, Software. **Shufan Xu:** Formal analysis, Data curation. **Xun Zhuang:** Writing – review & editing. **Yulong Lian:** Investigation, Formal analysis. **Yihua Lu:** Methodology, Supervision. **Lei Zhang:** Resources, Methodology. **Tian Tian:** Investigation, Writing – review & editing. **Guangyu Tian:** Resources, Supervision. **Jiahua Cui:** Methodology, Writing – review & editing. **Minjie Chu:** Funding acquisition, Investigation, Methodology, Supervision, Validation, Visualization, Project administration.

Reference

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin* 2021;71(3):209-49. doi: 10.3322/caac.21660 [published Online First: 2021/02/05]
2. Peters S, Adjei AA, Gridelli C, et al. Metastatic non-small-cell lung cancer (NSCLC): ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol* 2012;23 Suppl 7:vii56-64. doi: 10.1093/annonc/mds226 [published Online First: 2012/11/20]
3. Flores R, Patel P, Alpert N, et al. Association of Stage Shift and Population Mortality Among Patients With Non-Small Cell Lung Cancer. *JAMA Netw Open* 2021;4(12):e2137508. doi: 10.1001/jamanetworkopen.2021.37508 [published Online First: 2021/12/18]
4. Li D, Shi J, Dong X, et al. Epidemiological characteristics and risk factors of lung adenocarcinoma: A retrospective observational study from North China. *Front Oncol* 2022;12:892571. doi: 10.3389/fonc.2022.892571 [published Online First: 2022/08/23]
5. Radkiewicz C, Dickman PW, Johansson ALV, et al. Sex and survival in non-small cell lung cancer: A nationwide cohort study. *PLoS One* 2019;14(6):e0219206. doi: 10.1371/journal.pone.0219206 [published Online First: 2019/06/28]
6. Wang BY, Huang JY, Chen HC, et al. The comparison between adenocarcinoma and squamous cell carcinoma in lung cancer patients. *J Cancer Res Clin Oncol* 2020;146(1):43-52. doi: 10.1007/s00432-019-03079-8 [published Online First: 2019/11/11]
7. Lokhandwala T, Bittoni MA, Dann RA, et al. Costs of Diagnostic Assessment for

- Lung Cancer: A Medicare Claims Analysis. *Clin Lung Cancer* 2017;18(1):e27-e34. doi: 10.1016/j.clcc.2016.07.006 [published Online First: 2016/08/18]
8. Lim C, Tsao MS, Le LW, et al. Biomarker testing and time to treatment decision in patients with advanced nonsmall-cell lung cancer. *Ann Oncol* 2015;26(7):1415-21. doi: 10.1093/annonc/mdv208 [published Online First: 2015/04/30]
 9. Gutierrez ME, Choi K, Lanman RB, et al. Genomic Profiling of Advanced Non-Small Cell Lung Cancer in Community Settings: Gaps and Opportunities. *Clin Lung Cancer* 2017;18(6):651-59. doi: 10.1016/j.clcc.2017.04.004 [published Online First: 2017/05/10]
 10. Lone SN, Nisar S, Masoodi T, et al. Liquid biopsy: a step closer to transform diagnosis, prognosis and future of cancer treatments. *Mol Cancer* 2022;21(1):79. doi: 10.1186/s12943-022-01543-7 [published Online First: 2022/03/20]
 11. Holdenrieder S. Biomarkers along the continuum of care in lung cancer. *Scand J Clin Lab Invest Suppl* 2016;245:S40-5. doi: 10.1080/00365513.2016.1208446 [published Online First: 2016/08/20]
 12. Ignatiadis M, Sledge GW, Jeffrey SS. Liquid biopsy enters the clinic - implementation issues and future challenges. *Nat Rev Clin Oncol* 2021;18(5):297-312. doi: 10.1038/s41571-020-00457-x [published Online First: 2021/01/22]
 13. Xu JY, Zhang C, Wang X, et al. Integrative Proteomic Characterization of Human Lung Adenocarcinoma. *Cell* 2020;182(1):245-61 e17. doi: 10.1016/j.cell.2020.05.043 [published Online First: 2020/07/11]
 14. Zhang FF, Han B, Xu RH, et al. Identification of plasma SAA2 as a candidate biomarker for the detection and surveillance of non-small cell lung cancer. *Neoplasma* 2021;68(6):1301-09. doi: 10.4149/neo_2021_210228N263 [published Online First: 2021/10/15]
 15. Fang W, You J, Xu Q, et al. Plasma Exosomal CXCL7 is a Potential Biomarker for Lung Adenocarcinoma. *Clin Lab* 2022;68(10) doi: 10.7754/Clin.Lab.2022.220128 [published Online First: 2022/10/18]
 16. Djureinovic D, Ponten V, Landelius P, et al. Multiplex plasma protein profiling identifies novel markers to discriminate patients with adenocarcinoma of the lung. *BMC Cancer* 2019;19(1):741. doi: 10.1186/s12885-019-5943-3 [published Online First: 2019/07/31]
 17. Song Y, Wang J, Sun J, et al. Screening of Potential Biomarkers for Gastric Cancer with Diagnostic Value Using Label-free Global Proteome Analysis. *Genomics Proteomics Bioinformatics* 2020;18(6):679-95. doi: 10.1016/j.gpb.2020.06.012 [published Online First: 2021/02/20]
 18. Yu H, Liu Y, He T, et al. Platelet biomarkers identifying mild cognitive impairment in type 2 diabetes patients. *Aging Cell* 2021;20(10):e13469. doi: 10.1111/ace1.13469 [published Online First: 2021/09/17]
 19. de Sousa Abreu R, Penalva LO, Marcotte EM, et al. Global signatures of protein and mRNA expression levels. *Mol Biosyst* 2009;5(12):1512-26. doi: 10.1039/b908315d [published Online First: 2009/12/22]
 20. Nguyen PT, Katanoda K, Saito E, et al. Trends in lung cancer incidence by gender, histological type and stage at diagnosis in Japan, 1993 to 2015: A multiple

- imputation approach. *Int J Cancer* 2022;151(1):20-32. doi: 10.1002/ijc.33962 [published Online First: 2022/02/10]
21. Ganapathy-Kanniappan S, Geschwind JF. Tumor glycolysis as a target for cancer therapy: progress and prospects. *Mol Cancer* 2013;12:152. doi: 10.1186/1476-4598-12-152 [published Online First: 2013/12/05]
 22. Abbaszadeh Z, Cesmeli S, Biray Avci C. Crucial players in glycolysis: Cancer progress. *Gene* 2020;726:144158. doi: 10.1016/j.gene.2019.144158 [published Online First: 2019/10/21]
 23. Altenberg B, Greulich KO. Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics* 2004;84(6):1014-20. doi: 10.1016/j.ygeno.2004.08.010 [published Online First: 2004/11/10]
 24. Patra KC, Hay N. The pentose phosphate pathway and cancer. *Trends Biochem Sci* 2014;39(8):347-54. doi: 10.1016/j.tibs.2014.06.005 [published Online First: 2014/07/20]
 25. Achari A, Marshall SE, Muirhead H, et al. Glucose-6-phosphate isomerase. *Philos Trans R Soc Lond B Biol Sci* 1981;293(1063):145-57. doi: 10.1098/rstb.1981.0068 [published Online First: 1981/06/26]
 26. Mazzio E, Badisa R, Mack N, et al. Whole-transcriptome Analysis of Fully Viable Energy Efficient Glycolytic-null Cancer Cells Established by Double Genetic Knockout of Lactate Dehydrogenase A/B or Glucose-6-Phosphate Isomerase. *Cancer Genomics Proteomics* 2020;17(5):469-97. doi: 10.21873/cgp.20205 [published Online First: 2020/08/30]
 27. Wu ST, Liu B, Ai ZZ, et al. Esculetin Inhibits Cancer Cell Glycolysis by Binding Tumor PGK2, GPD2, and GPI. *Front Pharmacol* 2020;11:379. doi: 10.3389/fphar.2020.00379 [published Online First: 2020/04/16]
 28. Wang C, Zhou Q, Wu ST. Scopolin obtained from Smilax china L. against hepatocellular carcinoma by inhibiting glycolysis: A network pharmacology and experimental study. *J Ethnopharmacol* 2022;296:115469. doi: 10.1016/j.jep.2022.115469 [published Online First: 2022/06/20]
 29. Tsutsumi S, Fukasawa T, Yamauchi H, et al. Phosphoglucose isomerase enhances colorectal cancer metastasis. *Int J Oncol* 2009;35(5):1117-21. doi: 10.3892/ijo_00000427 [published Online First: 2009/09/30]
 30. Lyu Z, Chen Y, Guo X, et al. Genetic variants in glucose-6-phosphate isomerase gene as prognosis predictors in hepatocellular carcinoma. *Clin Res Hepatol Gastroenterol* 2016;40(6):698-704. doi: 10.1016/j.clinre.2016.05.001 [published Online First: 2016/06/12]
 31. Han J, Deng X, Sun R, et al. GPI Is a Prognostic Biomarker and Correlates With Immune Infiltrates in Lung Adenocarcinoma. *Front Oncol* 2021;11:752642. doi: 10.3389/fonc.2021.752642 [published Online First: 2021/12/17]
 32. Hitosugi T, Zhou L, Elf S, et al. Phosphoglycerate mutase 1 coordinates glycolysis and biosynthesis to promote tumor growth. *Cancer Cell* 2012;22(5):585-600. doi: 10.1016/j.ccr.2012.09.020 [published Online First: 2012/11/17]
 33. Lin R, Elf S, Shan C, et al. 6-Phosphogluconate dehydrogenase links oxidative PPP, lipogenesis and tumour growth by inhibiting LKB1-AMPK signalling. *Nat Cell Biol* 2015;17(11):1484-96. doi: 10.1038/ncb3255 [published Online First: 2015/10/20]
 34. Chan B, VanderLaan PA, Sukhatme VP. 6-Phosphogluconate dehydrogenase regulates tumor cell migration in vitro by regulating receptor tyrosine kinase

- c-Met. *Biochem Biophys Res Commun* 2013;439(2):247-51. doi: 10.1016/j.bbrc.2013.08.048 [published Online First: 2013/08/27]
35. Sukhatme VP, Chan B. Glycolytic cancer cells lacking 6-phosphogluconate dehydrogenase metabolize glucose to induce senescence. *FEBS Lett* 2012;586(16):2389-95. doi: 10.1016/j.febslet.2012.05.052 [published Online First: 2012/06/09]
 36. Liu R, Li W, Tao B, et al. Tyrosine phosphorylation activates 6-phosphogluconate dehydrogenase and promotes tumor growth and radiation resistance. *Nat Commun* 2019;10(1):991. doi: 10.1038/s41467-019-08921-8 [published Online First: 2019/03/03]
 37. Sheng H, Li Z, Su S, et al. YTH domain family 2 promotes lung cancer cell growth by facilitating 6-phosphogluconate dehydrogenase mRNA translation. *Carcinogenesis* 2020;41(5):541-50. doi: 10.1093/carcin/bgz152 [published Online First: 2019/09/11]
 38. Zheng W, Feng Q, Liu J, et al. Inhibition of 6-phosphogluconate Dehydrogenase Reverses Cisplatin Resistance in Ovarian and Lung Cancer. *Front Pharmacol* 2017;8:421. doi: 10.3389/fphar.2017.00421 [published Online First: 2017/07/18]
 39. Daneshmandi S, Cassel T, Lin P, et al. Blockade of 6-phosphogluconate dehydrogenase generates CD8(+) effector T cells with enhanced anti-tumor function. *Cell Rep* 2021;34(10):108831. doi: 10.1016/j.celrep.2021.108831 [published Online First: 2021/03/11]
 40. Cantelmo AR, Conradi LC, Brajic A, et al. Inhibition of the Glycolytic Activator PFKFB3 in Endothelium Induces Tumor Vessel Normalization, Impairs Metastasis, and Improves Chemotherapy. *Cancer Cell* 2016;30(6):968-85. doi: 10.1016/j.ccell.2016.10.006 [published Online First: 2016/11/22]
 41. Meng YM, Jiang X, Zhao X, et al. Hexokinase 2-driven glycolysis in pericytes activates their contractility leading to tumor blood vessel abnormalities. *Nat Commun* 2021;12(1):6011. doi: 10.1038/s41467-021-26259-y [published Online First: 2021/10/16]

Figure Legends:

Figure 1: Schematic of candidate protein biomarkers screening.

Figure 2: Screening and validation of candidate plasma proteins.

(A) 1,331 identified proteins in plasma, CP: cytoplasm; ER: endoplasmic reticulum; EC: extracellular; MC: mitochondria; PM: plasma membrane.

(B) The differentially expressed proteins between the case and control groups are described using volcano plots. The differentially expressed proteins on the statistical criteria of fold change $<2/3$, $>3/2$ (\log_2 scaled), and $P < 0.05$ ($-\log_{10}$ scaled).

(C) Differential expression of GPI between 102 LUAD cases and 102 healthy controls in the validation phase. (D) Differential expression of PGD between 102 LUAD cases

and 102 healthy controls in the validation phase. Navy blue, total LUAD cases, red, total healthy controls; green, female LUAD cases, light blue, female healthy controls; purple, male LUAD cases, orange, male healthy controls. (E) ROC curve for GPI expression levels of LUAD cases versus healthy controls. (F) ROC curve for PGD expression levels of LUAD cases versus healthy controls. (G) ROC curve for combined GPI and PGD expression levels of LUAD cases versus healthy controls.

Figure 3: GPI and PGD are consistently differential expressed in both plasma and tissues. (A) GPI protein are differential expressed in plasma. (B) GPI protein are differential expressed in tissue. (C) GPI mRNA are differential expressed in tissue. (D) PGD protein are differential expressed in plasma. (E) PGD protein are differential expressed in tissue. (F) PGD mRNA are differential expressed in tissue.

Figure 4: Kaplan-Meier estimates the overall survival probability of LUAD patients. (A-B) Kaplan-Meier estimates of the overall survival probability of LUAD patients in Kaplan-Meier Plotter. (C-D) Kaplan-Meier survival curve analysis of the prognostic significance of a high and a low protein expression of GPI and PGD.

Figure 5: GO functional classification and KEGG pathway mapping of GPI, PGD, and relevant genes. (A) The STRING database was used to analyze the top 50 relevant genes and Cytoscape was used to display the protein-protein interaction (PPI) network. (B) The distributions were summarized in three main categories: biological process, molecular function (MF), and cellular component (CC). (C) Scatter plot of enriched KEGG pathways statistics. The color and size of the dots represent the range of the P-value and the number of genes mapped to the indicated pathways, respectively. Top 10 enriched pathways are shown in the figure. (D) The map of “glycolysis” was modified from the KEGG map.

Figure 6 Association between expression of GPI and PGD and vascular normalization. (A) Immunochemical staining results of GPI in tumor tissues. Scale: 200μM. (B) Immunochemical staining results of PGD in tumor tissues. Scale: 200μM. (C) Correlation analysis between the expression of GPI and PGD in tumor tissues and plasma. (D) Immunofluorescence photographs of LUAD tumor tissue with CD31 and α-SMA staining. Scale: 200μM (E) Correlation analysis between the expression of

GPI and PGD in tumor tissues and vascular normalization. (F) Correlation analysis between the expression of GPI and PGD in plasma and vascular normalization.

Figure 7: Lung adenocarcinoma active metastasis model. (A) Schematic of active metastasis tumor-bearing mice model, tumor tissues of mice without lung metastasis (a), the tumor tissues of mice with lung metastasis (b) and the corresponding lung metastasis tumor tissues (c). (B) Immunofluorescence photographs of tumor tissue of mice without lung metastasis (a), the tumor tissues of mice with lung metastasis (b) with CD31 and α -SMA staining. Scale: 200 μ M. (C) Differential expression of α -SMA in tumor tissue of mice without lung metastasis (a), the tumor tissues of mice with lung metastasis (b). (D) The expression of α -SMA, GPI, PGD in tumor tissues of mice without lung metastasis(a), the tumor tissues of mice with lung metastasis (b) and the corresponding lung metastasis tumor tissues (c). (E) Differential expression of α -SMA, GPI, and PGD in tumor tissue of mice without lung metastasis(a), the tumor tissues of mice with lung metastasis (b) and the corresponding lung metastasis tumor tissues (c). (* $P < 0.05$; *** $P < 0.001$).

Table 1 Characteristics of the subjects enrolled in this study

Variables	Screening ("blood+")			Validation (ELISA)					
	Case (Female) (N=10)	Control (Female) (N=10)	<i>P</i>	Case (Female) (N=62)	Control (Female) (N=62)	<i>P</i>	Case (Male) (N=40)	Control (Male) (N=40)	<i>P</i>
Age, years (mean±SD)	62.30±13.33	63.70±9.64	0.79	62.27±9.35	65.06±4.73	0.10	66.80±9.25	66.82±4.05	0.99
Stage, N (100%)									
I	8 (80%)			57 (91.9%)			30 (75.0%)		
II	0 (0%)			2 (3.2%)			2 (5.0%)		
III	2 (20%)			1 (1.6%)			4 (10.0%)		
IV	0 (0%)			2 (3.2%)			4 (10.0%)		

Table 2 Results of the 14 proteins(mRNAs) that consistently differentially expressed in both plasmas and tissues in the screening phase

Number	Accession	Protein description	Name	Regulated Type	FC ^a	FC ^b	FC ^c
1	P00966	Argininosuccinate synthase	ASS1	Up	4.651	2.86	1.777
2	Q04760	Lactoylglutathione lyase	GLO1	Up	3.454	2.838	1.657
3	P52209	6-phosphogluconate dehydrogenase, decarboxylating	PGD	Up	3.186	4.252	1.72
4	P08238	Heat shock protein HSP 90-beta	HSP90AB1	Up	3.156	2.774	1.927
5	B8ZZK2	Gamma-glutamylcyclotransferase	GGCT	Up	2.935	9.913	2
6	P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	Up	2.588	2.674	3.19
7	P13646	Keratin, type I cytoskeletal 13	KRT13	Up	2.458	22.245	1.787
8	A0A075B6I0	Immunoglobulin lambda variable 8-61	IGLV8-61	Up	2.405	2.084	20.228
9	P23528	Cofilin-1	CFL1	Up	2.032	2.211	1.565
10	Q01518	Adenylyl cyclase-associated protein 1	CAP1	Up	1.871	1.534	1.673
11	P06744	Glucose-6-phosphate isomerase	GPI	Up	1.715	3.935	1.942
12	P19012	Keratin, type I cytoskeletal 15	KRT15	Up	1.654	3.172	6.817
13	Q14574	Desmocollin-3	DSC3	Up	1.636	4.483	3.647
14	Q9BSE5	Agmatinase, mitochondrial	AGMAT	Up	1.511	3.378	4.512

^a: the fold change of protein in plasma

^b: the fold change of protein in tissue

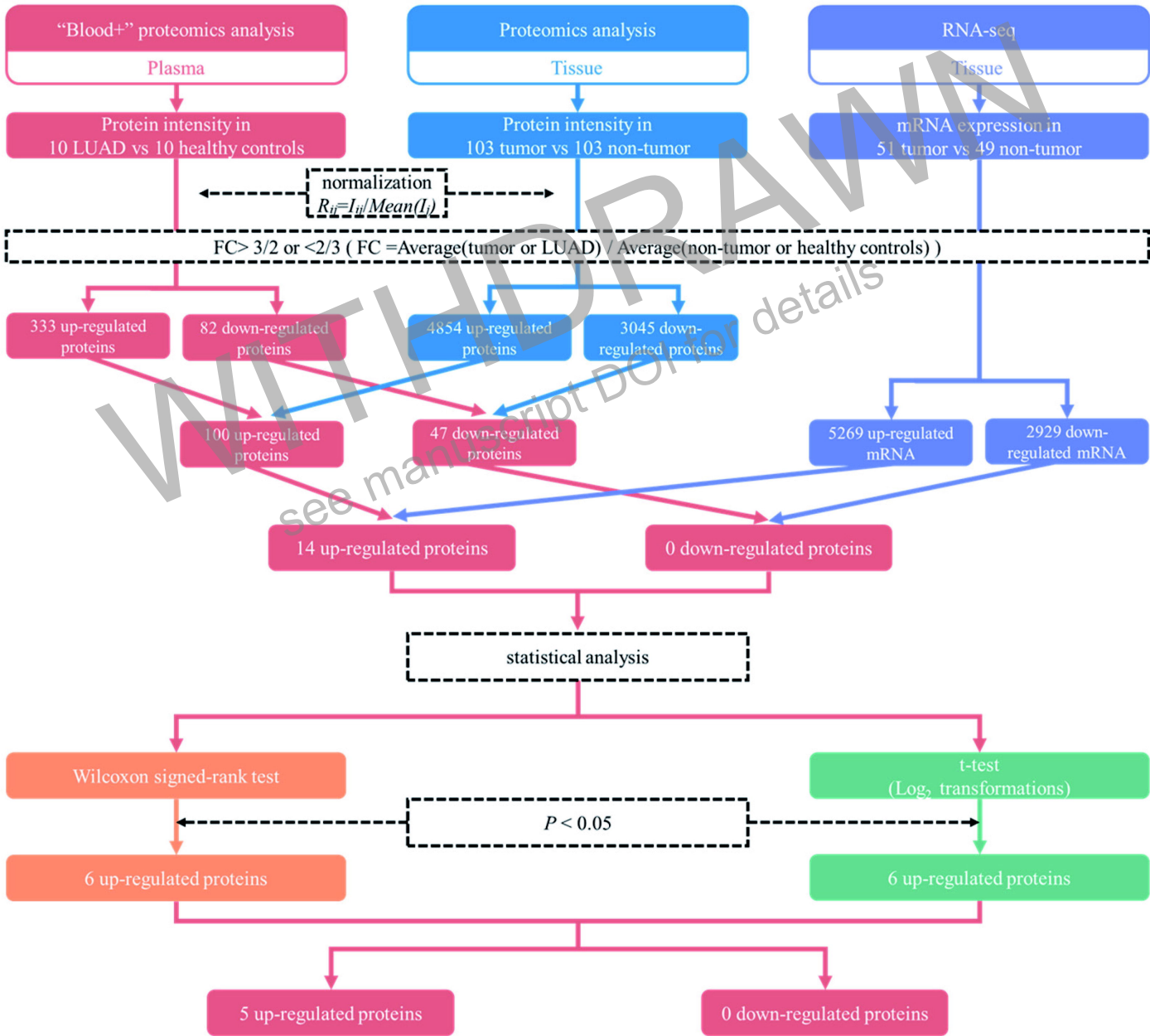
^c: the fold change of mRNA in tissue

Table 3 Results of the 5 proteins differentially expressed in plasma of LUAD patients

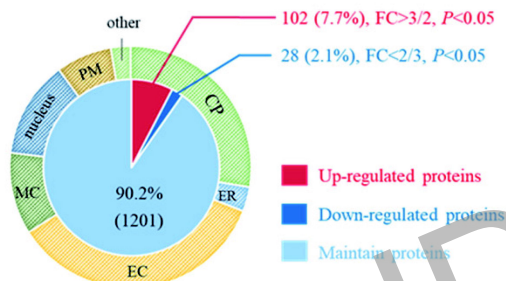
Number	Accession	Protein description	Name	Regulated Type	P_w	P_t
1	P00966	Arginosuccinate synthase	ASS1	Up	0.023	0.015
2	P52209	6-phosphogluconate dehydrogenase, decarboxylating	PGD	Up	0.010	0.011
3	P04406	Glyceraldehyde-3-phosphate dehydrogenase	GAPDH	Up	<0.001	<0.001
4	P06744	Glucose-6-phosphate isomerase	GPI	Up	0.028	0.017
5	P08238	Heat shock protein HSP 90-beta	HSP90AB1	Up	0.007	0.009

P_t : the p-value of t-test

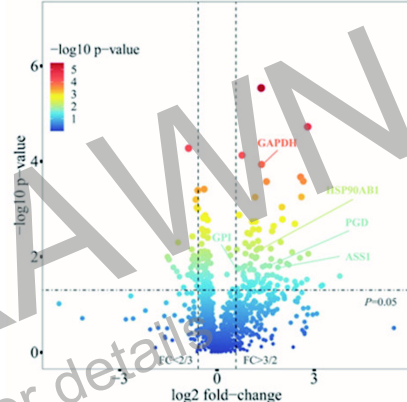
P_w : the p-value of Wilcoxon signed-rank test



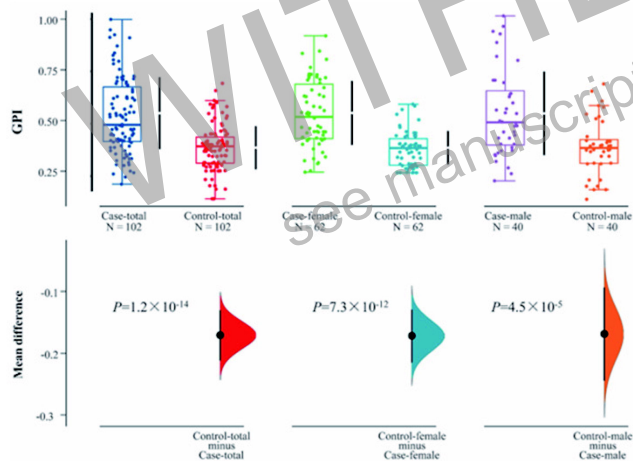
A



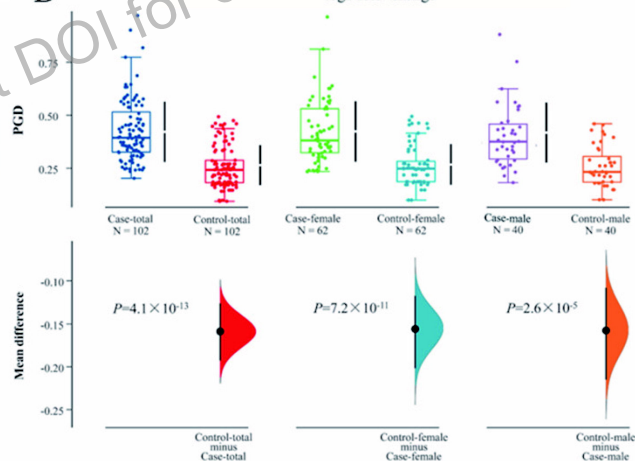
B



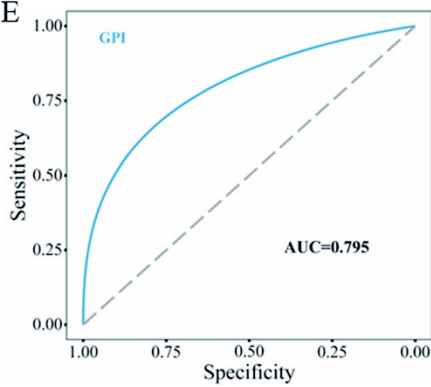
C



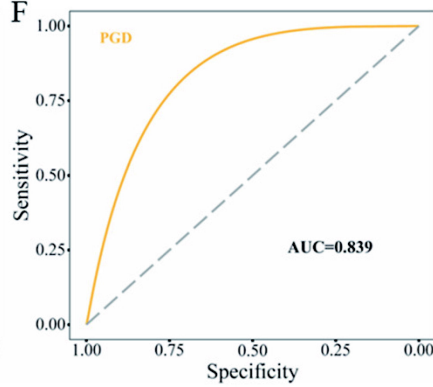
D



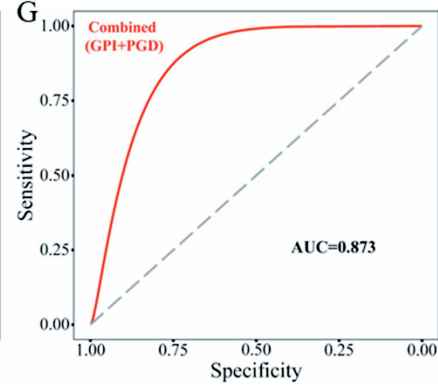
E

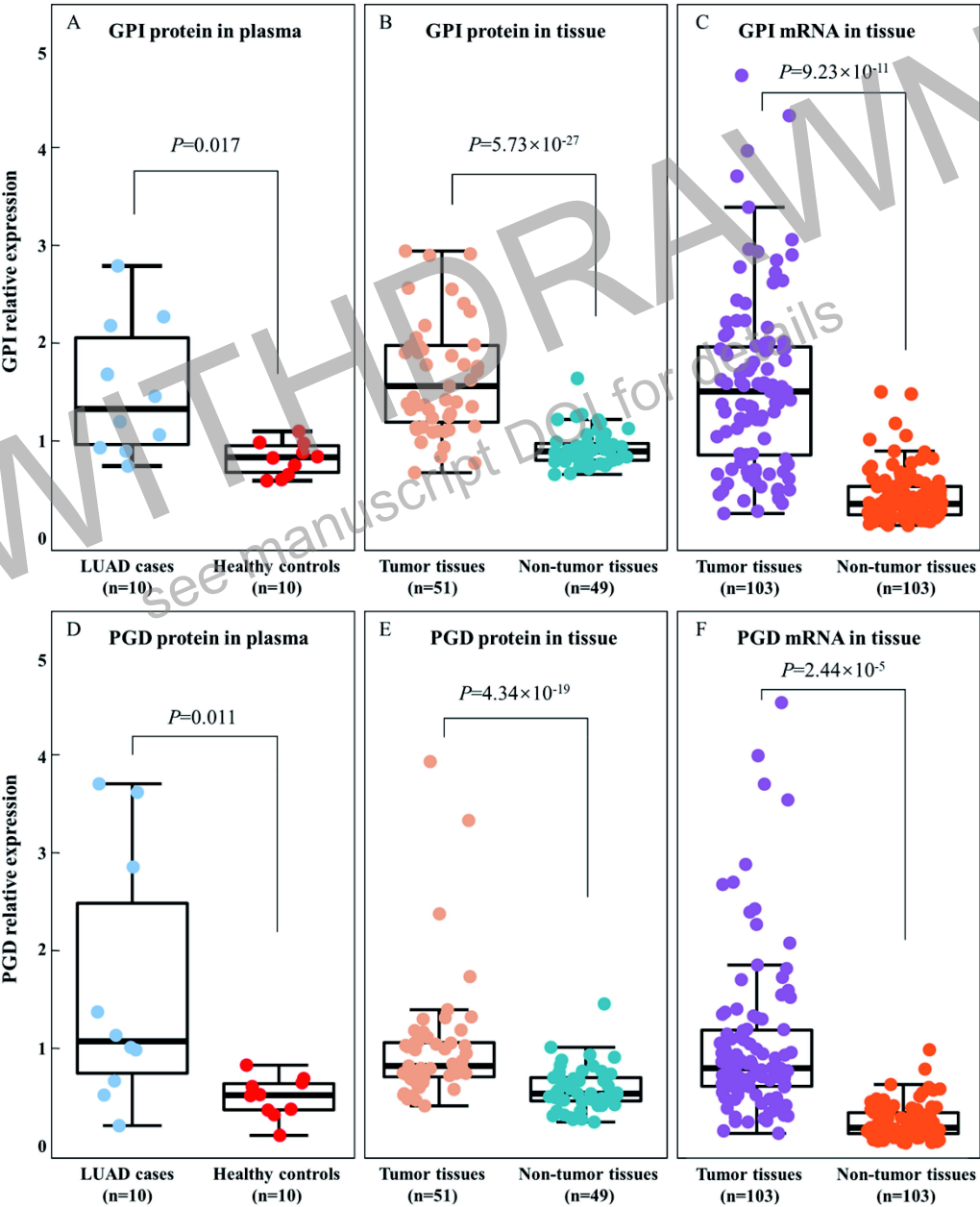


F



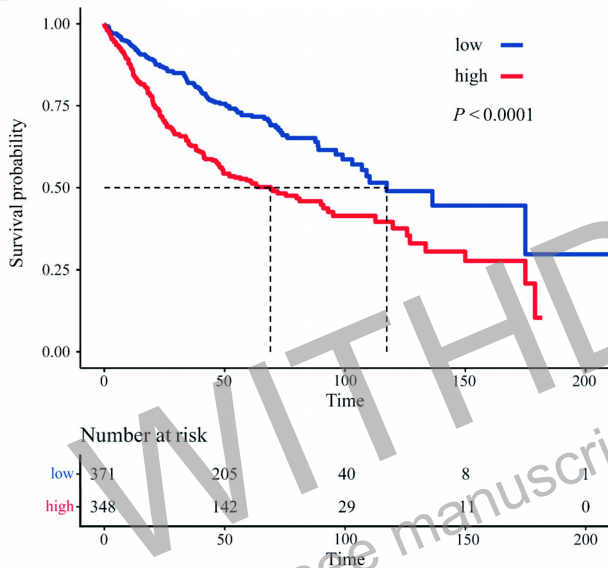
G





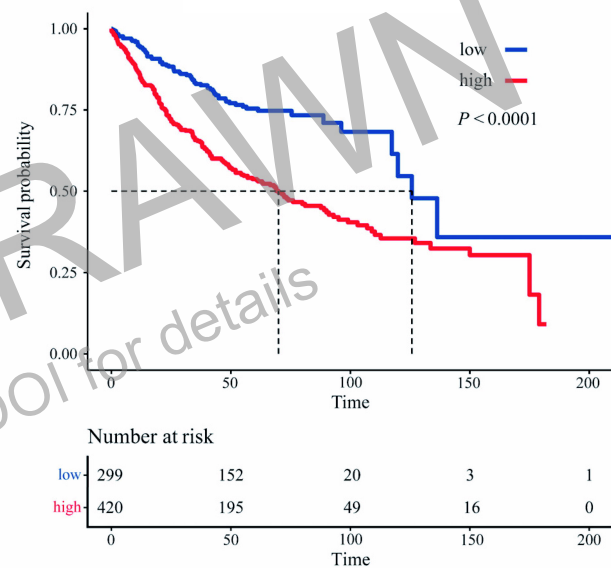
A

GPI mRNA



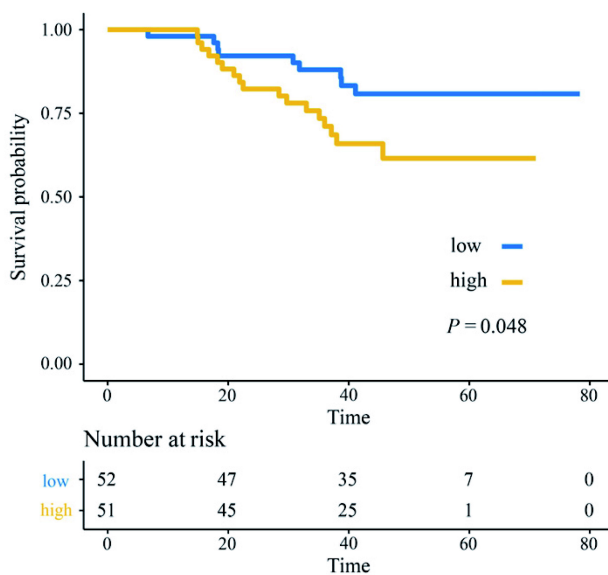
B

PGD mRNA



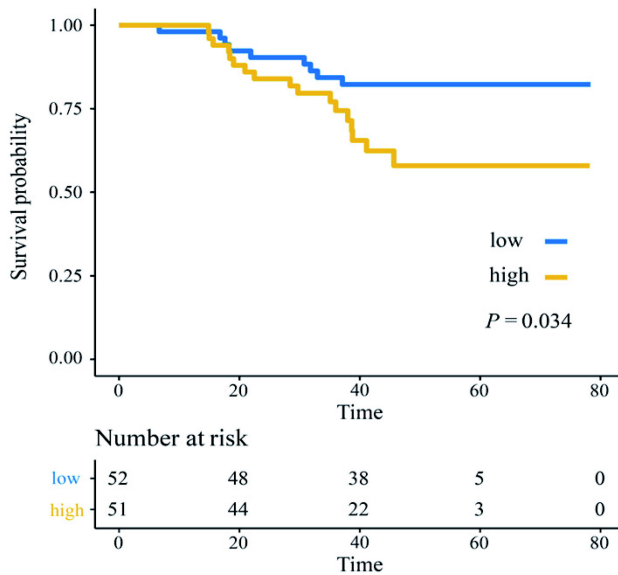
C

GPI protein

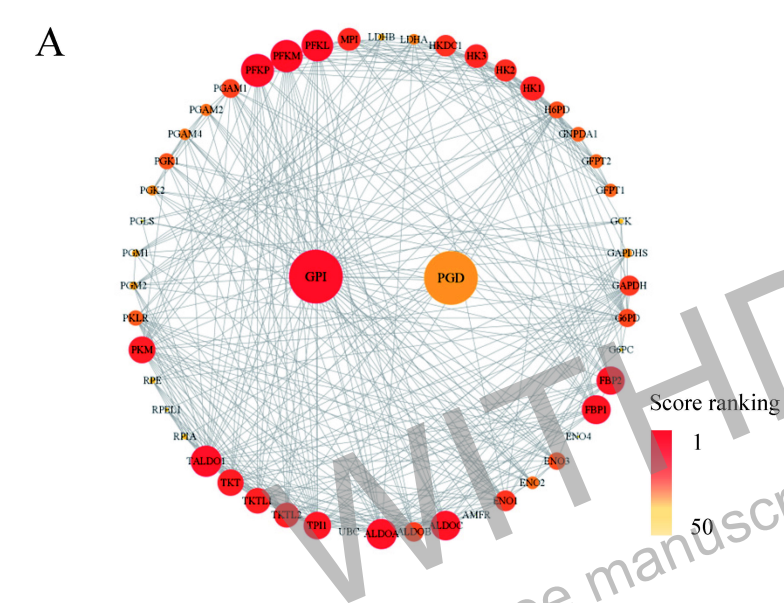


D

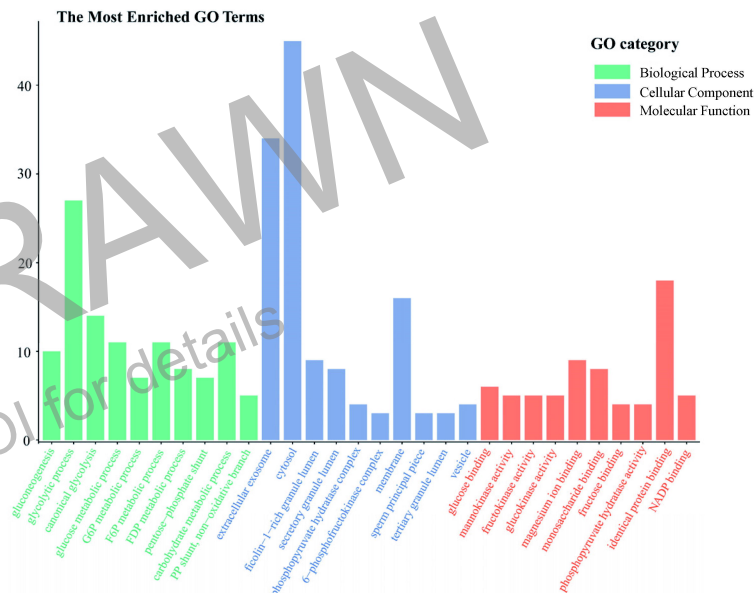
PGD protein



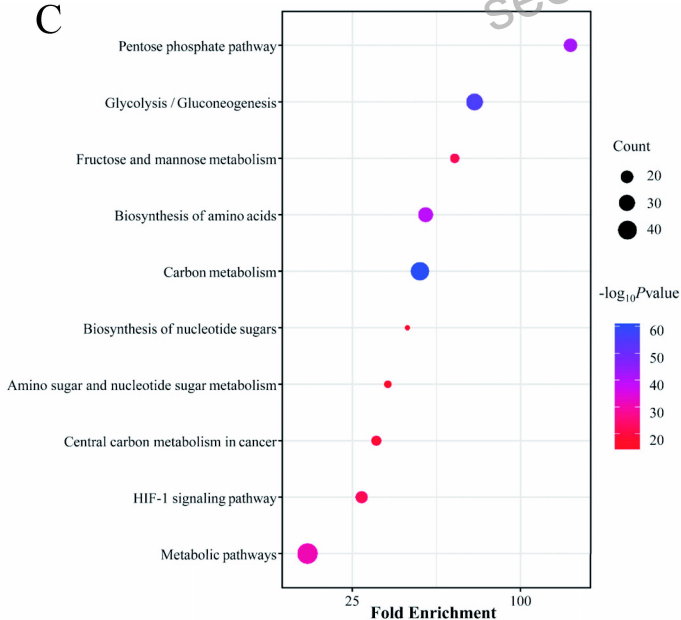
A



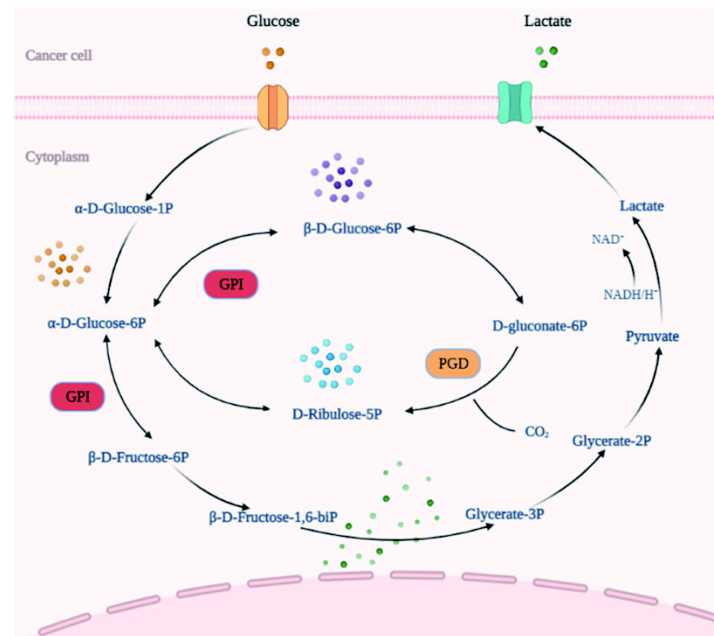
B



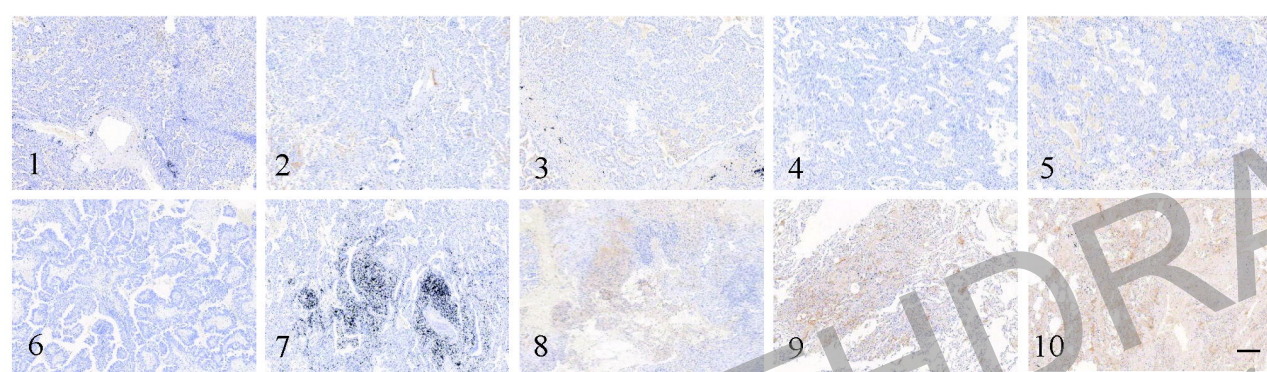
C



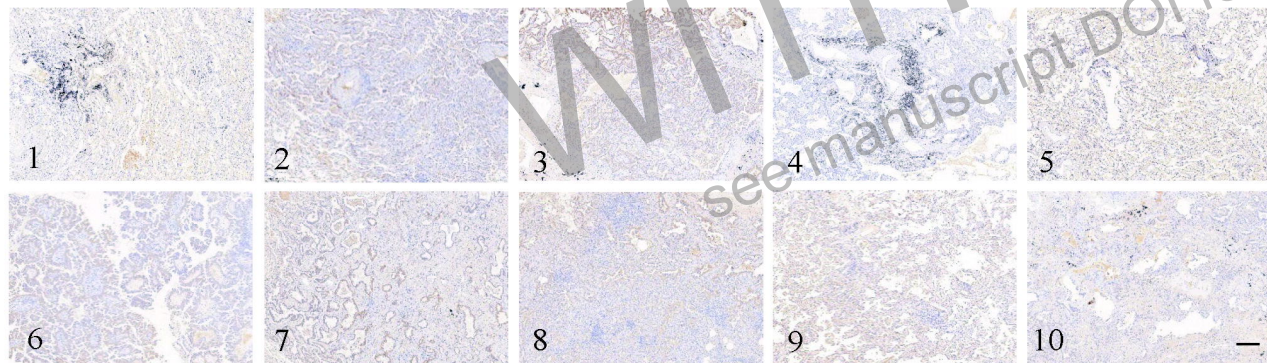
D



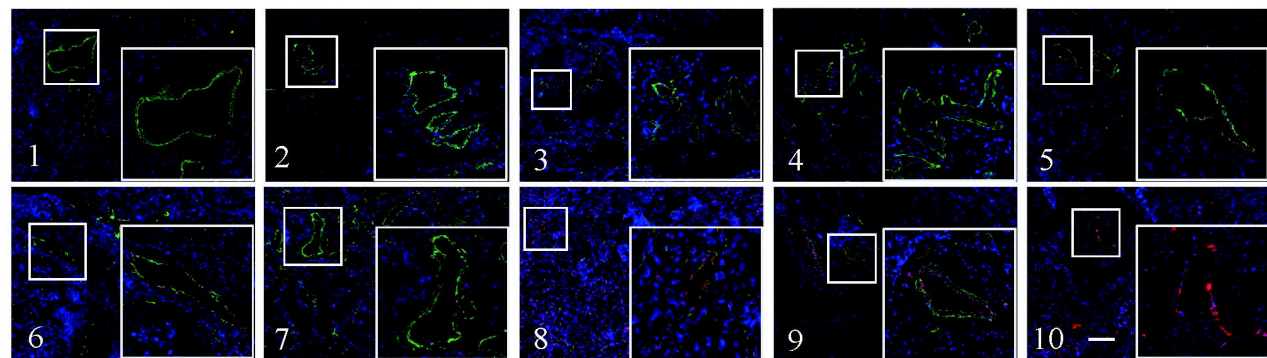
A



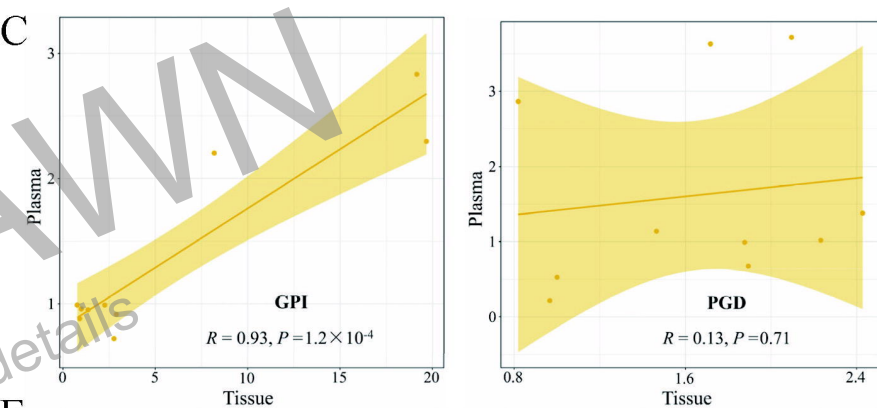
B



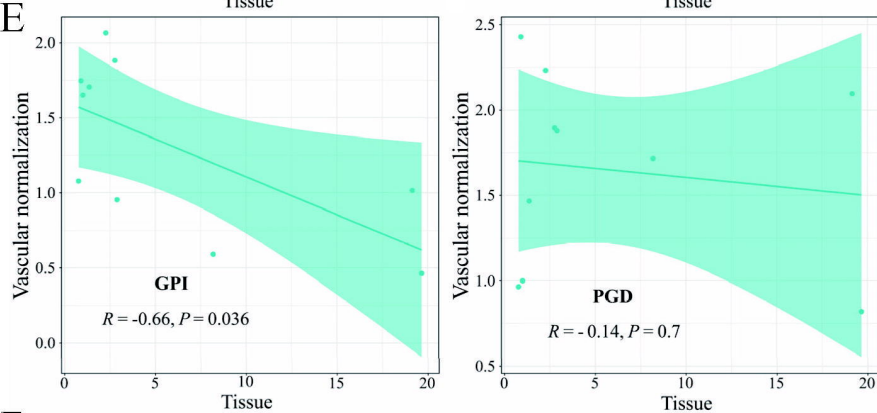
D

DAPI/CD31/ α -SMA

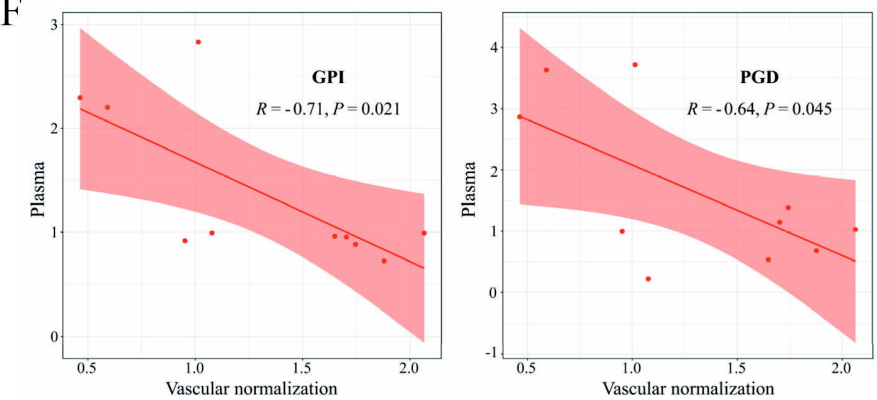
C



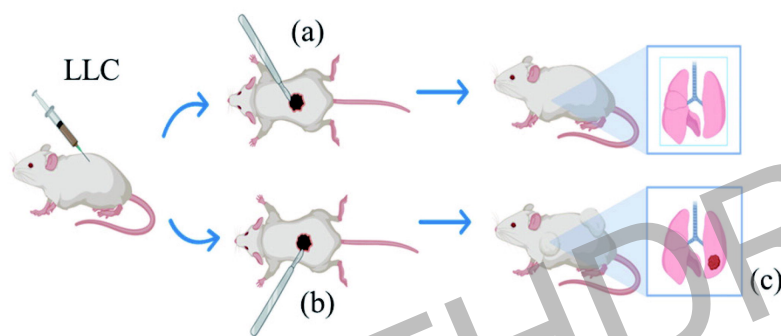
E



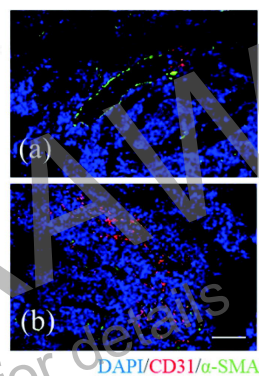
F



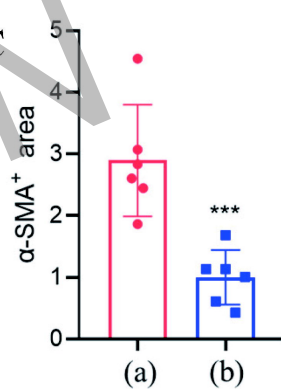
A



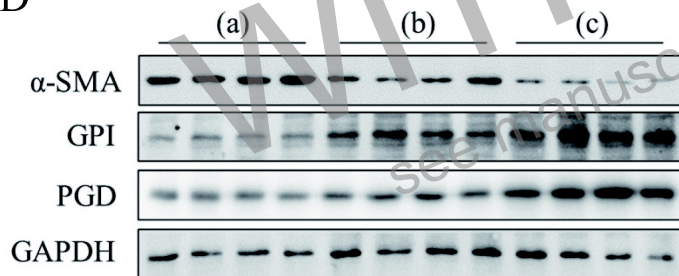
B



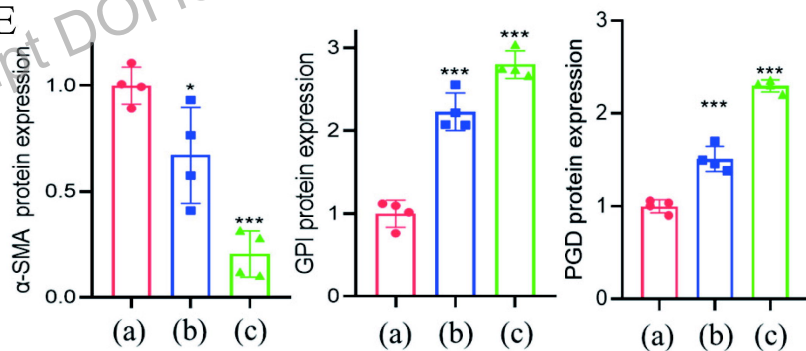
C



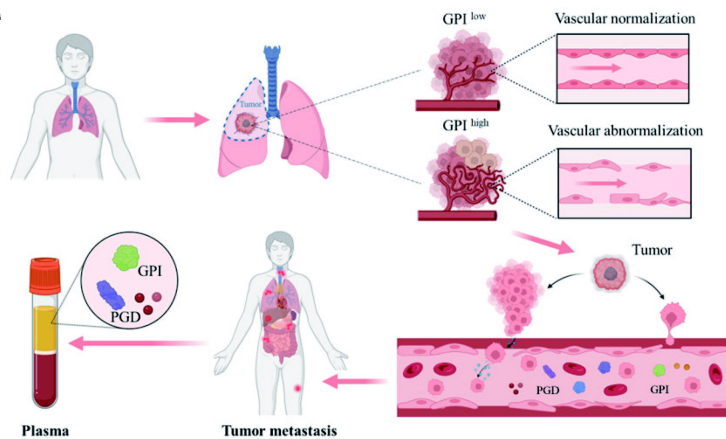
D



E



F



G

