



# Identification of the consistently differential expressed hub mRNAs and proteins in lung adenocarcinoma and construction of the prognostic signature: a multidimensional analysis

Yiran Liu, MSc<sup>a</sup>, Zhenyu Li, MSc<sup>a</sup>, Qianqiao Meng, MSc<sup>c</sup>, Anhui Ning, MSc<sup>a</sup>, Shenxuan Zhou, MSc<sup>a</sup>, Siqi Li, MSc<sup>a</sup>, Xiaobo Tao, MSc<sup>a</sup>, Yutong Wu, MSc<sup>a</sup>, Qiong Chen, MSc<sup>a</sup>, Tian Tian, MD, PhD<sup>a</sup>, Lei Zhang, MD, PhD<sup>a</sup>, Jiahua Cui, MD, PhD<sup>a,\*</sup>, Liping Mao, MSc<sup>b,\*</sup>, Minjie Chu, MD, PhD<sup>a,\*</sup>

**Background:** This study aimed to elucidate the consistency of differentially expressed hub mRNAs and proteins in lung adenocarcinoma (LUAD) across populations and to construct a comprehensive LUAD prognostic signature.

**Methods:** The transcriptomic and proteomics data from different populations were standardized and analyzed using the same criteria to identify the consistently differential expressed mRNAs and proteins across genders and races. We then integrated prognosis-related mRNAs with clinical, pathological, and EGFR (epidermal growth factor receptor) mutation data to construct a survival model, subsequently validating it across populations. Through plasma proteomics, plasma proteins that consistently differential expressed with LUAD tissues were screened and validated, with their associations discerned by measuring expressions in tumor tissues and tumor vascular normalization.

**Results:** The consistency rate of differentially expressed mRNAs and proteins was ~20–40%, with ethnic factors leading to about 40–60% consistency of differentially expressed mRNA or protein across populations. The survival model based on the identified eight hub mRNAs as well as stage, smoking status, and EGFR mutations, demonstrated good prognostic prediction capabilities in both Western and East Asian populations, with a higher number of unfavorable variables indicating poorer LUAD prognosis. Notably, GPI expression in tumor tissues was inversely correlated with vascular normalization and positively correlated with plasma GPI expression.

**Conclusion:** Our study underscores the significance of integrating transcriptomics and proteomics data, emphasizing the need to account for genetic diversity among ethnic groups. The developed survival model may offer a holistic perspective on LUAD progression, enhancing prognosis and therapeutic strategies.

**Keywords:** consistency rate, GPI, lung adenocarcinoma, prognosis, race

## Introduction

Lung cancer remains the leading cause of cancer-related mortality worldwide<sup>[1]</sup>, with non-small cell lung cancer (NSCLC) accounting for 85% of cases<sup>[2]</sup>, and lung adenocarcinoma (LUAD) stands out as the primary subtype within NSCLC. The incidence of LUAD has been steadily increasing over the years<sup>[3,4]</sup>, and there exist gender disparities in its incidence,

possibly linked to sex-specific genetic differences<sup>[5,6]</sup>. While genetic differences lay the foundation, mRNA and protein variations offer valuable information on tumor behavior, biomarkers, and therapeutic targets. Focusing on genders and races, as well as mRNA and protein differences may help uncover the complex molecular mechanisms underlying LUAD development and progression, enabling more targeted and effective treatment strategies.

<sup>a</sup>Department of Epidemiology, School of Public Health, Nantong University, <sup>b</sup>Department of Oncology, Affiliated Nantong Hospital of Shanghai University (The Sixth People's Hospital of Nantong), Nantong, Jiangsu, People's Republic of China and <sup>c</sup>Department of Global Health and Population, School of Public Health, Harvard University, Boston, USA  
Yiran Liu, Zhenyu Li, and Qianqiao Meng contributed equally to this article.

Sponsorships or competing interests that may be relevant to content are disclosed at the end of this article.

\*Corresponding author. Address: Department of Epidemiology, School of Public Health, Nantong University, Nantong, Jiangsu 226019, People's Republic of China.

Tel.: +86 13914487601; fax: +86 0513 85012910. E-mail: chuminjie@ntu.edu.cn (M. Chu), and Tel.: +86 15951731309; fax: +86 0513 85012910. E-mail: cuijiahua@ntu.edu.cn (J. Cui); Department of Oncology, Affiliated Nantong Hospital of Shanghai University (The Sixth People's Hospital of Nantong), Nantong, Jiangsu 226019, People's Republic of China. Tel.: +86 18921606619; fax: +86 0513 80886111. E-mail: yanmao7471@126.com (L. Mao).

Copyright © 2023 The Author(s). Published by Wolters Kluwer Health, Inc. This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

International Journal of Surgery (2024) 110:1052–1067

Received 29 August 2023; Accepted 12 November 2023

Supplemental Digital Content is available for this article. Direct URL citations are provided in the HTML and PDF versions of this article on the journal's website, [www.com/international-journal-of-surgery](http://www.com/international-journal-of-surgery).

Published online 27 November 2023

<http://dx.doi.org/10.1097/JS9.0000000000000943>

The central dogma of molecular biology<sup>[7]</sup> governs the flow of genetic information from DNA to mRNA to protein. While the central dogma implies a faithful translation of genetic information, disparities often arise between mRNA and protein expression levels. The correlation between mRNA and protein expression, known as mRNA–protein correlation, is a topic of great interest in cancer research. Traditionally, it was assumed that mRNA and protein levels exhibit a strong positive correlation. However, emerging evidence suggests that this correlation is often moderate or even weak<sup>[8,9]</sup>. Several factors affecting gene expression regulation contribute to the disparity between mRNA and protein expression, including post-transcriptional modifications, translational efficiency, protein stability, and degradation processes.

One major factor contributing to mRNA–protein discordance is post-transcriptional regulation. Variations in mRNA stability, transport, and translation efficiency can lead to differences between mRNA abundance and corresponding protein levels<sup>[10]</sup>. Dysregulation of RNA-binding proteins, microRNAs, and other regulatory molecules can influence mRNA turnover rates, resulting in altered protein expression<sup>[11–14]</sup>. Additionally, alternative splicing, which generates different protein isoforms by including or excluding specific exons, can further contribute to protein diversity<sup>[15,16]</sup>, complicating the mRNA–protein relationship. Post-translational modifications, such as phosphorylation, acetylation, and glycosylation, also play a role in modulating protein activity, stability, and subcellular localization<sup>[17]</sup>. Perturbations in these modifications can lead to variations between mRNA and protein expression patterns, ultimately affecting cancer-related processes<sup>[18,19]</sup>, including cell proliferation, apoptosis<sup>[20]</sup>, angiogenesis, and metastasis.

Differential expression of mRNA and proteins in LUAD can have significant implications. Altered mRNA expression may not always be reflected in corresponding changes at the protein level due to regulatory mechanisms occurring post-transcriptionally or during translation. This discordance can impact our understanding of LUAD carcinogenic mechanisms, biomarker discovery, and therapeutic targeting. Therefore, studying the differences between mRNA and protein expression levels in LUAD is crucial. While some studies have addressed this issue<sup>[21,22]</sup>, few have specifically focused on this aspect, particularly in the context of comprehensive analysis across races and genders.

Many existing survival models for LUAD have been developed with a focus on specific aspects. For instance, some models emphasize the role of pathology, while others delve into genetic mutations or rely heavily on clinical information<sup>[23–25]</sup>. These models, while valuable, often provide a narrow perspective, potentially overlooking the multifaceted nature of LUAD progression and prognosis. Recognizing this limitation, our study aims to construct a comprehensive survival model for LUAD. By integrating diverse factors such as pathology, genetics, and clinical data, we aspire to offer a more holistic understanding of LUAD prognosis, thereby enhancing the precision and applicability of survival predictions.

This study aims to analyze the consistency rates of differentially expressed LUAD-related mRNAs and proteins in the Western Population and the Chinese Population. Additionally, it seeks to examine the consistency rate of differentially expressed LUAD-related mRNAs and the consistency rate of differentially expressed LUAD-related proteins across different races.

## HIGHLIGHTS

- Consistency rates between differentially expressed mRNAs and proteins were around 20–40%.
- The constructed prognostic model may predict the survival of LUAD (lung adenocarcinoma) patients.
- The integrated survival model encompasses diverse data may assist individualized therapeutic strategies.

Furthermore, the study investigates the consistency rate of differential mRNA expression and differential protein expression between genders in the Western Population and Chinese Population. Finally, the identification of a group of mRNAs associated with LUAD prognosis that consistently exhibits differential expression between races can provide valuable insights for future LUAD studies, aiding in biomarker discovery and the identification of therapeutic targets.

## Methods

### Study design

The study design involved comparing the consistency rates of differentially expressed mRNAs and differentially expressed proteins in tumor tissues among patients with different races (Western Population and Chinese Population) and genders (males and females). Additionally, the consistency rate of differential expression between mRNA and protein was calculated in LUAD in the Western Population and the Chinese Population. We identified prognostic mRNAs that overlapped between Western and Chinese populations and underwent functional analysis. Utilizing these mRNAs, we constructed a survival model, integrating both clinical, pathological, and EGFR mutation data. We conducted plasma proteomics on samples from 10 LUAD cases and 10 healthy controls. Subsequently, we expanded our sample size and validated candidate proteins in 102 LUAD cases and 102 healthy controls using enzyme-linked immunosorbent assay (ELISA). We then quantified the expression levels of these validated proteins in tumor tissues to evaluate the correlation between their expression in plasma and tissue with vascular normalization. All the work has been reported in line with the REMARK criteria<sup>[26]</sup> (Supplemental Digital Content 1, <http://links.lww.com/J9/B413>).

### Study population of plasma proteomics

The 10 LUAD cases in the screening phase were obtained between August 2021 and November 2021 at Hai'an People's Hospital, Jiangsu Province, with 10 healthy controls matched by age. The 102 LUAD cases in the validation phase were obtained between December 2021 and June 2022 at Hai'an People's Hospital and Affiliated Nantong Hospital of Shanghai University, with 102 healthy controls matched by age and gender. Informed consent was obtained from all patients. The study was reviewed and approved by the ethics committee of Nantong University (Approval No. 2022-2).

### Data sources

The RNA-seq raw data and clinical data for LUAD tissues were obtained from the TCGA (The Cancer Genome Atlas) dataset and

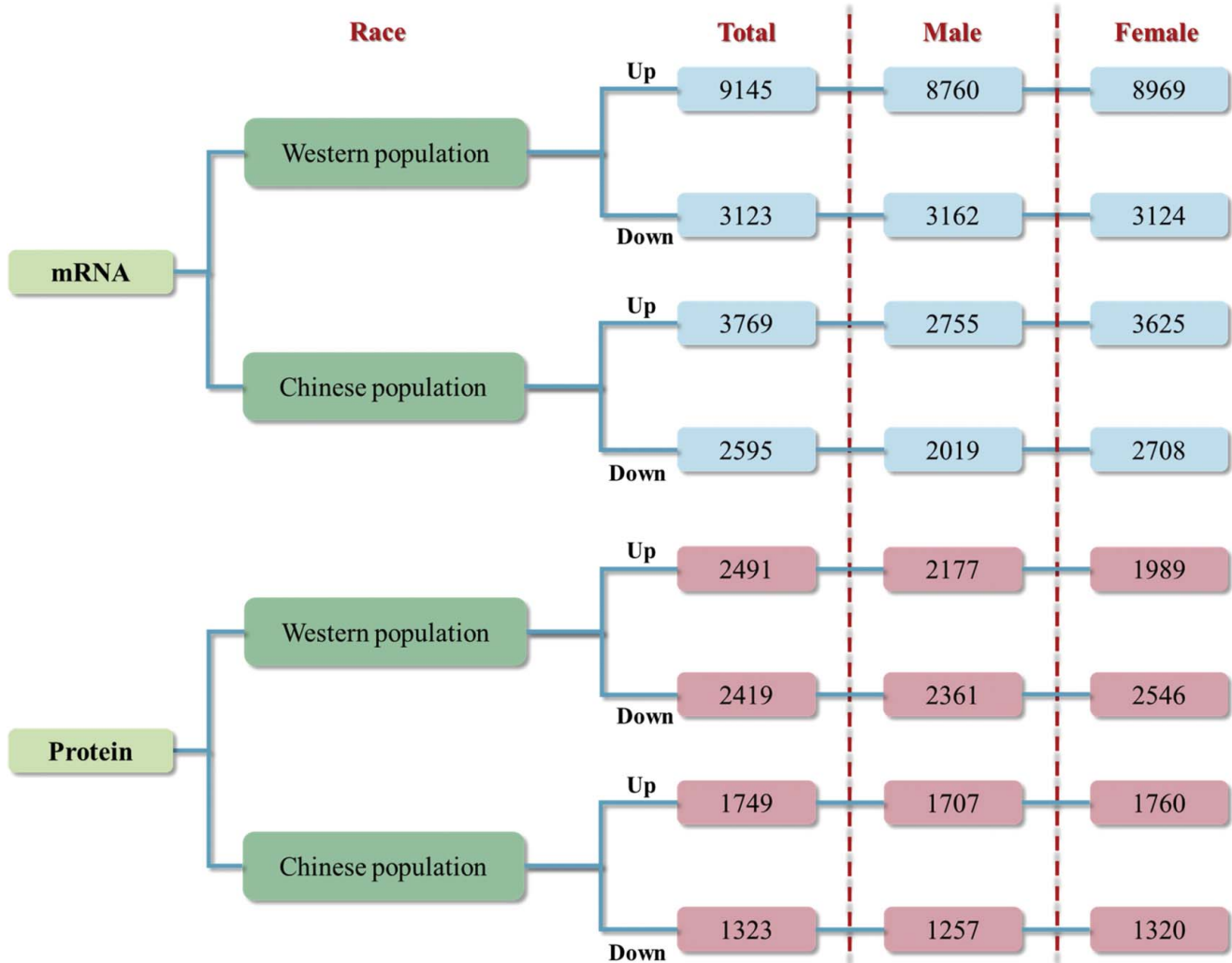


Figure 1 . Differentially expressed mRNAs (proteins) in each database.

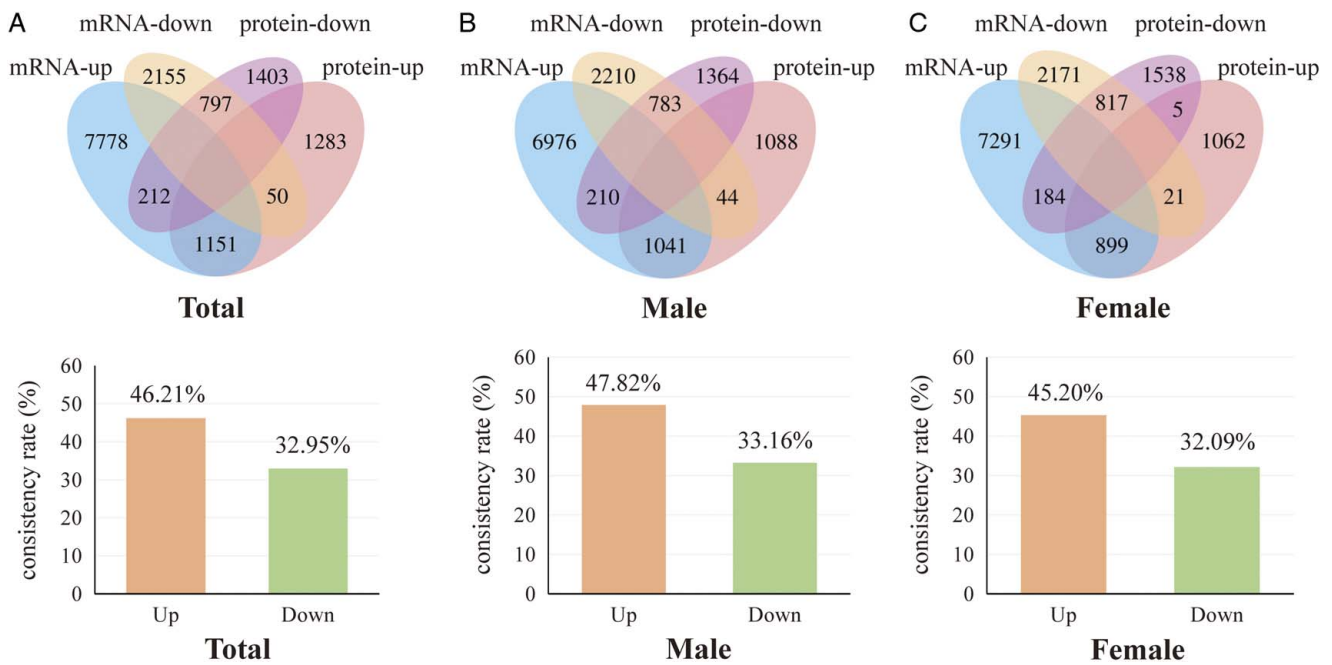
downloaded from UCSC Xena (<https://xenabrowser.net/data/pages/>). We excluded samples with missing racial information and Asians from the TCGA and used the remaining sample as the Western population. The proteomic data of protein expression from LUAD tumor tissues and adjacent nontumor tissues in the Western Population was obtained through the CPTAC (Clinical Proteomic Tumor Analysis Consortium) data portal with the subproject ID PDC000153<sup>[22]</sup>. Similarly, we included the proteomics data of samples from PDC000153, whose original region is Western, as the Western population proteomics data. The RNA-seq data of mRNA expression in the Chinese population were downloaded from the Gene Expression Omnibus (GEO) database with accession number GSE140343<sup>[27]</sup>. The proteomic data of protein expression from LUAD tumor tissues and adjacent nontumor tissues from the same Chinese population were downloaded from the iProX consortium with the subproject ID IPX0001804000<sup>[27]</sup>. We sourced mRNA expression datasets and corresponding clinical survival data for Japanese<sup>[28]</sup> and American<sup>[29]</sup> cohort studies from the GEO database (GSE31210 and GSE72094). After excluding samples with incomplete clinical information, the Japanese cohort included 226 samples, while the American cohort included 331 samples.

#### Kaplan–Meier survival analysis

Cox regression analysis was performed to assess the relationship between mRNA expression and survival time in LUAD patients. The Kaplan–Meier method was used to evaluate differences between high-risk and low-risk groups based on the median expression, employing R packages ‘survminer’ and ‘survival’.

#### ‘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 (liquid chromatography–mass spectrometry/mass spectrometry) 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 the BCA kit according to the instructions. After trypsin digestion of each sample protein into peptide fragments, 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



**Figure 2.** Consistency rate of differential mRNA and protein expression in the Western population. (A) Differentially upregulated and downregulated mRNAs and proteins in the Western population and consistency rates of upregulated and downregulated mRNAs and proteins. (B) Differentially upregulated and downregulated mRNAs and proteins in the male Western population and consistency rates of upregulated and downregulated mRNAs and proteins in males. (C) Differentially upregulated and downregulated mRNAs and proteins in the female Western population and consistency rates of upregulated and downregulated mRNAs and proteins in females.

processed using the MaxQuant search engine (v.1.6.15.0). False discovery rate (FDR) was adjusted to less than 1%.

#### Enzyme-linked immunosorbent assay (ELISA)

ELISA kits allow for the determination of GPI and GAPDH in plasma samples in this study. Collected plasma was sampled after centrifugation at 2000 rpm for 20 min at 4°C and stored at -80°C until analysis. Commercial ELISA kits purchased from Meimian Biotechnology (Yancheng, Jiangsu, China) were used to measure protein levels in human plasma samples according to the manufacturer's recommendations. Absorbances were measured in a microplate reader (Sunrise, Tecan, Austria) at 450 nm.

#### Immunohistochemical staining and immunofluorescent staining

For immunohistochemistry stainings, the slides of LUAD tumor tissue specimens were incubated with GPI (CST, 94068T) antibody solution. The images were scanned with a 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  $\alpha$ -SMA (Servicebio, GB111364) antibody solution.

#### Statistical analysis

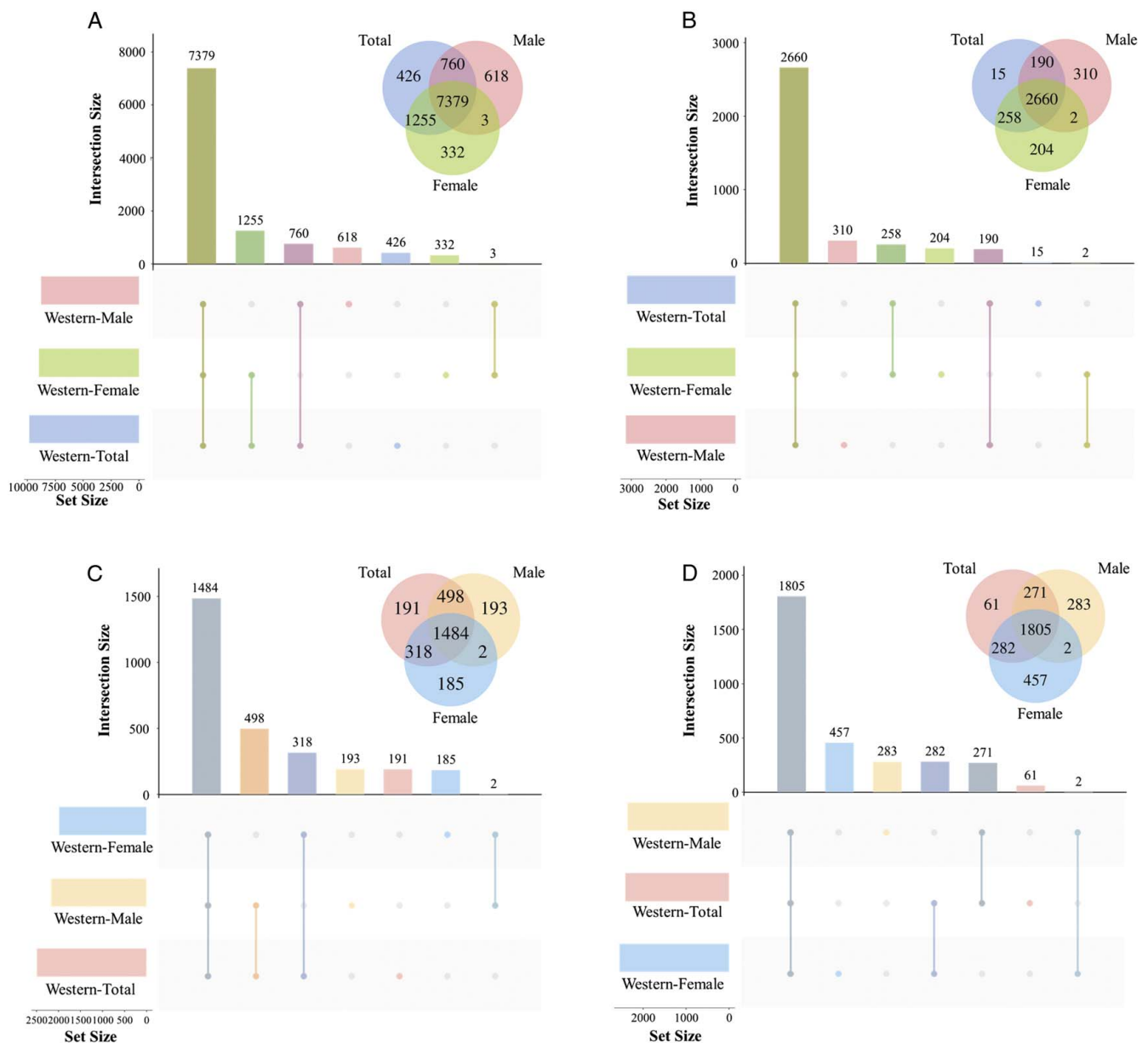
After downloading the initial data from the databases, mRNAs, or proteins with a sample response rate higher than 75% were retained for analysis. Since the data in the TCGA database were  $\log_2$  (norm\_count + 1) transformed, an adjustment of  $2^{\text{count}} - 1$  was performed on the RNA-seq raw data from the TCGA-LUAD datasets to maintain data consistency. The data in the PDC000153

database were  $\log_2$  norm\_count transformed, and an adjustment of  $2^{\text{count}}$  was performed on the proteomics raw data from PDC000153 to maintain data consistency. The Wilcoxon Rank-Sum Test was used to determine the differential expression of mRNAs or proteins between tumor tissues and adjacent nontumor tissues in LUAD patients. mRNA (proteins) with a fold change (FC) greater than 3/2 or less than 2/3 (FC = Average (tumor tissues)/Average (adjacent nontumor tissues)) and FDR adjusted  $P < 0.05$  were considered significantly differentially expressed. R 4.1.1 was employed for statistical analysis.

#### Results

##### Consistency rate of LUAD-related differential mRNA and protein expression in the Western Population

We first compared the differential mRNA expression in LUAD patients utilizing the TCGA database. Compared with adjacent nontumor tissues, 9145 upregulated mRNAs and 3123 downregulated mRNAs are identified in LUAD tumor tissues (Fig. 1). Then, we compared the differential protein expression in LUAD patients by PDC000153, with 2491 proteins upregulated and 2419 proteins downregulated. We then screened for consistently differentially expressed mRNAs and proteins in LUAD tumor tissues to obtain a consistent rate of differential mRNA and protein expression in LUAD. In differential expressed proteins, 46.21% (1151/2491) proteins were consistently upregulated, and 32.95% (797/2419) proteins were consistently downregulated with differentially expressed mRNAs (Fig. 2A). The consistency rate remained relatively stable in different genders, in male differential expressed proteins, 47.82% (1041/2177) proteins were consistently



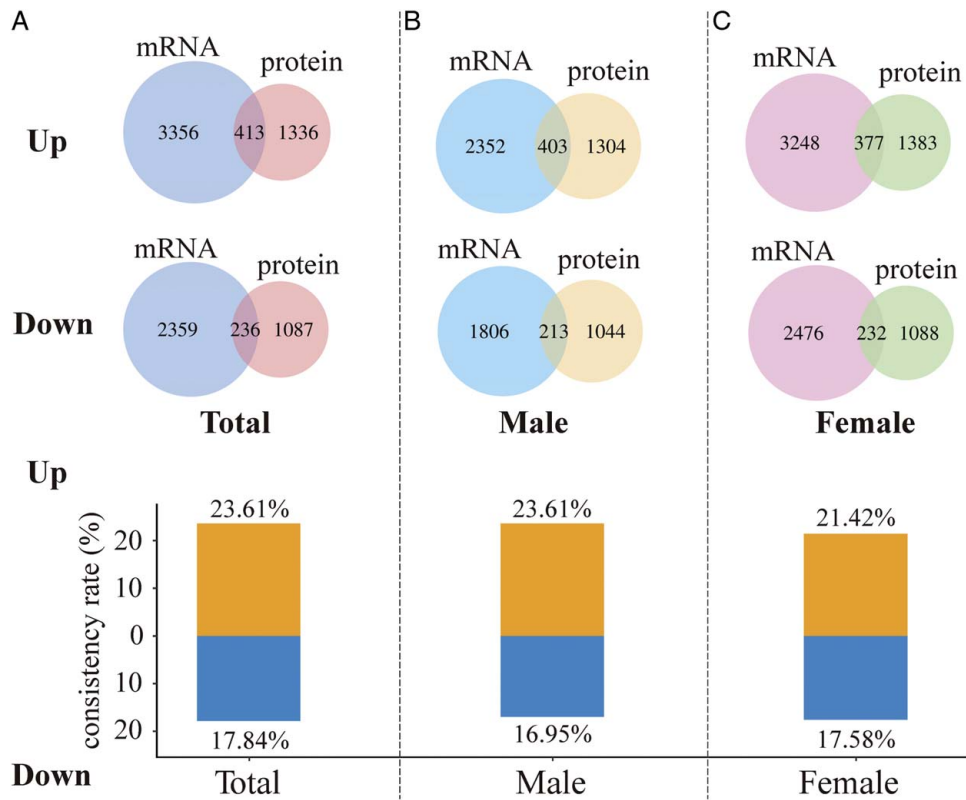
**Figure 3 .** Consistency rate of differential mRNA expression and differential protein expression between genders in the Western population. (A) Differentially upregulated mRNAs in the Western population between genders. (B) Differentially downregulated mRNAs in the Western population between genders. (C) Differentially upregulated proteins in the Western population between genders. (D) Differentially downregulated proteins in the Western population between genders.

upregulated, and 33.16% (783/2361) proteins were consistently downregulated with differentially expressed mRNAs (Fig. 2B). In the differential expressed proteins in female LUAD, 45.20% (899/1989) proteins were consistently upregulated, and 32.09% (817/2546) proteins were consistently downregulated with differential expressed mRNAs (Fig. 2C).

#### **Consistency rate of LUAD-related differential mRNAs expression and differential protein expression among genders in the Western population**

We then analyzed the distribution of differential mRNA expression between genders. In the differential expressed mRNAs in

LUAD tumor tissues, there were 8760 mRNA upregulated in males and 8969 mRNA upregulated in females, while 3162 mRNA downregulated in males and 3124 mRNA downregulated in females (Fig. 1). In the differential expressed proteins in LUAD tumor tissues, 7379 mRNAs were consistently upregulated, and 2660 mRNAs were consistently downregulated in both males and females. Compared with females, the upregulation and downregulation consistency rates were 84.24% (7379/8760) and 84.12% (2660/3162) in males, respectively (Fig. 3A, B). In the differential expressed proteins in LUAD tumor tissues, there were 2177 proteins upregulated in males and 1989 proteins upregulated in females, while 2361 proteins downregulated in males and



**Figure 4 .** Consistency rate of differential mRNA and protein expression in the Chinese population. (A) Differentially upregulated and downregulated mRNAs and proteins in the Chinese population and consistency rates of upregulated and downregulated mRNAs and proteins. (B) Differentially upregulated and downregulated mRNAs and proteins in the male Chinese population and consistency rates of upregulated and downregulated mRNAs and proteins in males. (C) Differentially upregulated and downregulated mRNAs and proteins in the female Chinese population and consistency rates of upregulated and downregulated mRNAs and proteins in females.

2546 proteins downregulated in females (Fig. 1). Among them, 1484 proteins were consistently upregulated and 1805 mRNAs were consistently downregulated in both males and females. Compared with females, the upregulation and downregulation consistency rates in males were 68.17% (1484/2177) and 76.45% (1805/2361), respectively (Fig. 3C, D).

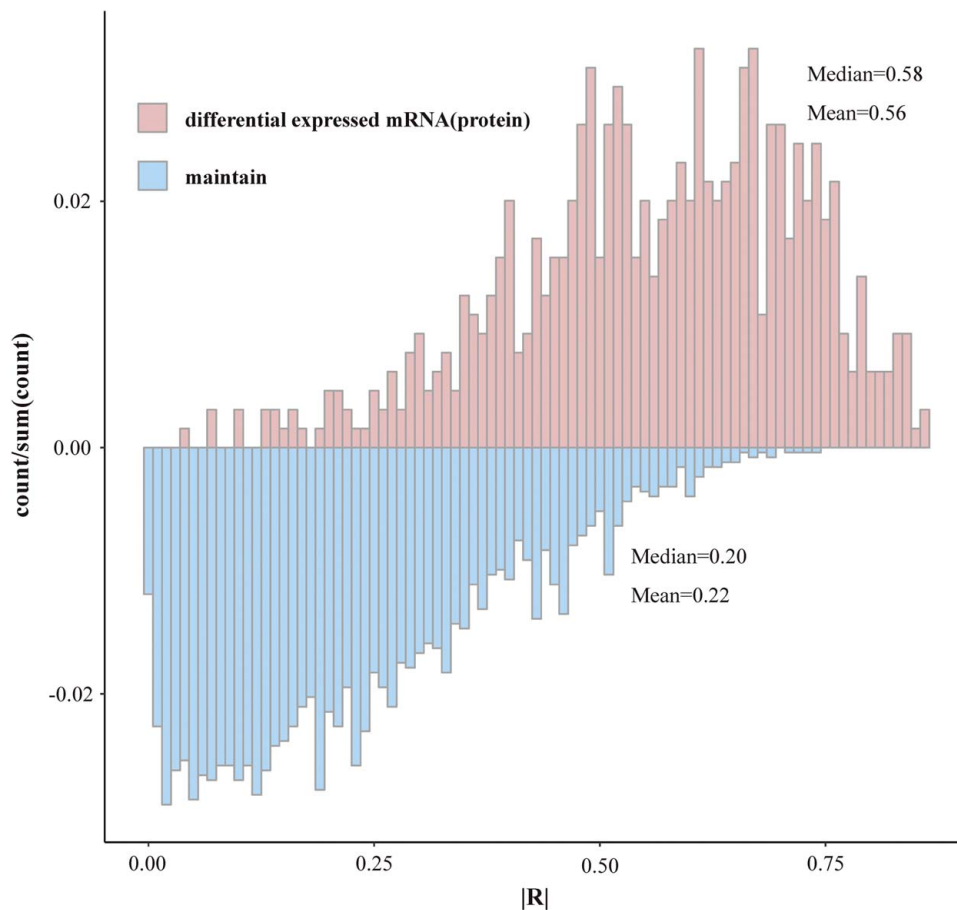
**Consistency rate of LUAD-related differential mRNA and protein expression in the Chinese population**

Having compared the consistency rate of differential mRNA and protein expression in the Western population, we further explored this among Chinese populations. We obtained 3769 upregulated mRNAs and 2595 downregulated mRNAs in Chinese LUAD tumor tissues from GSE140343 (Fig. 1). Meanwhile, we obtained proteomic data of LUAD in the Chinese population from IPX0001804000s. Among them, 1749 proteins upregulated expressed and 1323 proteins downregulated expressed in LUAD tumors tissues (Fig. 1). With these differential expressed proteins, 23.61% (413/1749) mRNAs (proteins) were consistently upregulated and 17.84% (236/1323) mRNAs (proteins) were consistently downregulated (Fig. 4A). Similarly, the consistency rate of differential mRNA and protein expression in Chinese population was stable in males and females, with 23.61% (403/1707) consistent upregulation and 16.95% (213/1257) consistent downregulation in males (Fig. 4B) and 21.42% (377/1760) consistent upregulation and 17.58% (232/1320) consistent downregulation in females (Fig. 4C), respectively.

Since the transcriptomics and proteomics data in the Chinese population came from the same sample, we were able to do correlation analyses between the expression of mRNA and protein. We categorized the mRNA and protein pairs into differential expression and non-differential expression groups and observed that the median correlation coefficient for the differential expression group was 0.58, notably higher than the corresponding figure of 0.20 for the non-differential expression group (Fig. 5).

**Consistency rate of LUAD-related differential mRNAs expression and differential protein expression among genders in the Chinese population**

By analyzing the RNA-seq data of LUAD tissues in the Chinese population between genders, 2755 mRNAs were upregulated in males and 3625 mRNAs were upregulated in females, 2019 mRNAs were downregulated in males and 2708 mRNAs were downregulated in females (Fig. 1). Within the differentially expressed mRNAs in male LUAD tumor tissues, 2177 mRNAs were co-upregulated with females, with an upregulation consistency rate of 79.02% (2177/2755), and 1695 mRNAs were co-downregulated with females, with a downregulation consistency rate of 83.95% (1695/2019) (Fig. 6A). When it comes to differential expressed LUAD-related proteins between genders, 1707 proteins were upregulated in males and 1760 proteins were upregulated in females (Fig. 1), of which 1438 proteins



**Figure 5 .** Correlation between mRNA and protein expression in the Chinese population.

overlapped and were co-upregulated in males and females, accounting for 84.24% (1438/1707) consistency rate of the male upregulated proteins (Fig. 6B). And 1257 proteins were downregulated in males and 1320 proteins were downregulated in females (Fig. 1), with 1116 proteins overlapping and co-downregulated in males and females, accounting for 88.78% (1116/1257) consistency rate of male downregulated proteins (Fig. 6B).

#### **Consistency rate of LUAD-related differential expressed mRNAs between races**

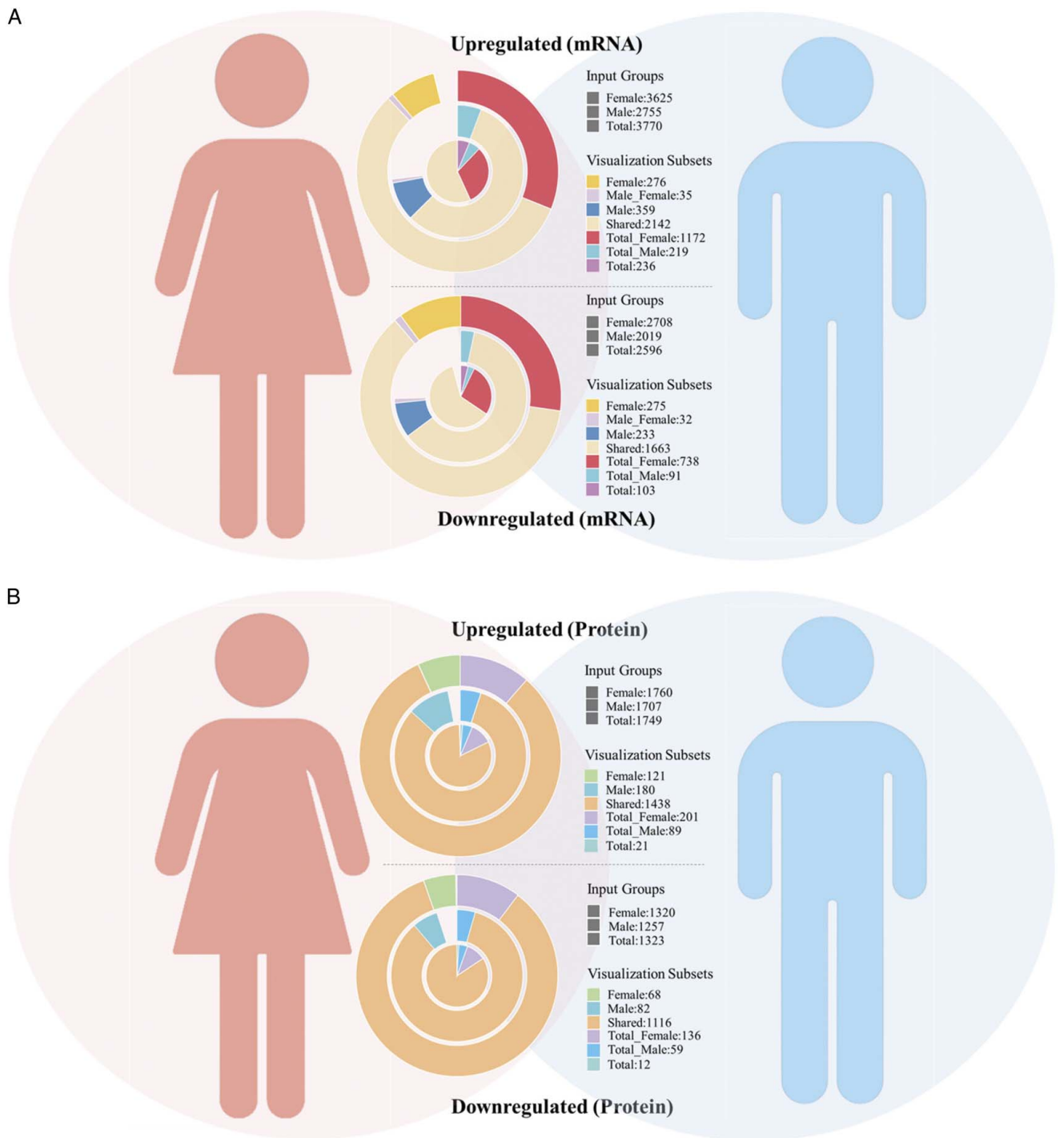
By analyzing the RNA-seq data of LUAD tissues in the Western Population and the Chinese Population and retaining the consistently differential expressed mRNAs, leaving 2192 consistently upregulated and 1087 consistently downregulated mRNAs (Fig. 7A). In the LUAD tumor tissues of the Chinese population, there was a 58.16% (2192/3769) consistent upregulation and 41.89% (1087/2595) consistent downregulation of mRNA with that from the Western population. Similarly, the consistency rate of differential mRNA expression between the Chinese and Western populations was stable in males and females, with 59.93% (1651/2755) consistent upregulation and 40.22% (812/2019) consistent downregulation in males and 57.08% (2069/3625) consistent upregulation and 41.10% (1113/2708) consistent downregulation in females, respectively.

#### **Consistency rate of LUAD-related differential expressed proteins between races**

We further analyzed the proteomic data of LUAD tissues in the Western Population and the Chinese Population to get the consistency rate of differential proteins between races. In the differential expressed proteins, 809 proteins were co-upregulated and 747 proteins were co-downregulated in both the Western population and the Chinese population (Fig. 7B). In the LUAD tumor tissues of the Chinese population, there was a 46.26% (809/1749) consistent upregulation and 56.46% (747/1323) consistent downregulation of protein with that from the Western population. Meanwhile, in both Western population and Chinese populations, 42.65% (728/1707) proteins consistently upregulated in males and 40.68% (716/1760) proteins consistently upregulated in females, 56.09% (705/1257) proteins consistently downregulated in male and 59.70% (788/1320) proteins consistently downregulated in female.

#### **Screening for prognosis-related mRNAs (protein) and constructing a survival model**

Among the 1948 mRNAs consistently differential expressed with proteins in the Western population, 457 mRNAs were identified as prognosis-related to LUAD. In the Chinese population, 74 mRNAs were identified as prognosis-related to LUAD. Of these prognosis-related mRNAs, we obtained eight overlapped

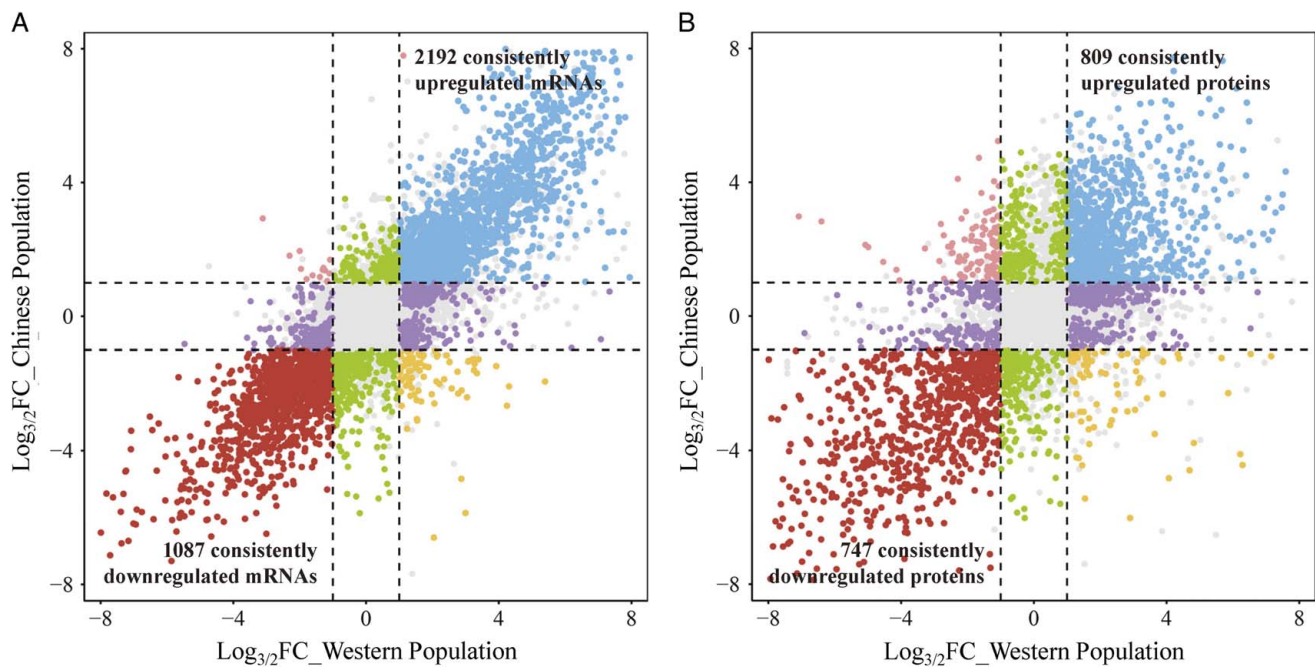


**Figure 6 .** Consistency rate of differential mRNA expression and differential protein expression between genders in the Chinese population. (A) Consistency rate of differential mRNA expression between genders in the Chinese population. (B) Consistency rate of differential protein expression between genders in the Chinese population.

mRNAs (Fig. 8A). The specific screening process is as shown in Figure 9. We assigned scores based on the potential prognostic implications of these mRNAs: mRNAs considered unfavorable were assigned a score of 1, whereas those deemed favorable received a score of 0. Specifically, if the expression of an up-regulated mRNA was above the median, it was considered

unfavorable and given a score of 1. Conversely, if the expression of a downregulated mRNA was equal to or below the median, it was also considered unfavorable and assigned a score of 1. We performed survival analysis in both the Western population and the Chinese population; the KM (Kaplan–Meier) plots were present in Figure 10A.





**Figure 7 .** Consistently differential expressed mRNAs and proteins between races. (A) Consistently differential expressed mRNAs between races. (B) Consistently differential expressed proteins between races.

#### Validate hub mRNAs and construct survival models

We utilized survival data from Japanese and U.S. populations to validate the prognostic relevance of these eight hub mRNAs on LUAD. Our analysis revealed that, in the Japanese population, seven out of the eight mRNAs demonstrated a significant association with LUAD prognosis. All eight mRNAs in the U.S. population were found to be significantly associated with LUAD prognosis (Fig. 8B).

Given that the prognostic relevance of the majority of mRNAs was consistently validated across both the Japanese and U.S. populations, we could employ these eight hub mRNAs to construct a comprehensive prognostic survival model and validate it in different populations. Recognizing the significance of clinical and pathological information in prognostic model construction, we expanded our model to include smoking status and tumor staging in addition to the eight hub mRNAs. Subjects with a history of smoking were classified as unfavorable, receiving a score of 1. Pathological stages III/IV were deemed unfavorable, with an assigned score of 1, with the specific assignment methodology detailed in the provided table (Supplementary Table 1, Supplemental Digital Content 2, <http://links.lww.com/JS9/B414>). We discovered that there are significant prognostic differences among subjects who possess less than 4 unfavorable variables, 4–7 unfavorable variables, and more than 7 unfavorable variables. The more unfavorable variables a subject carries, the poorer their prognosis (Fig. 10B). In parallel, we sought to validate the robustness of our survival model by assessing its performance in both the Japanese and American populations. The predictive accuracy of the model was found to be consistent with its performance observed in Chinese and Western populations (Fig. 10C), underscoring its potential as a reliable prognostic tool across diverse ethnic groups.

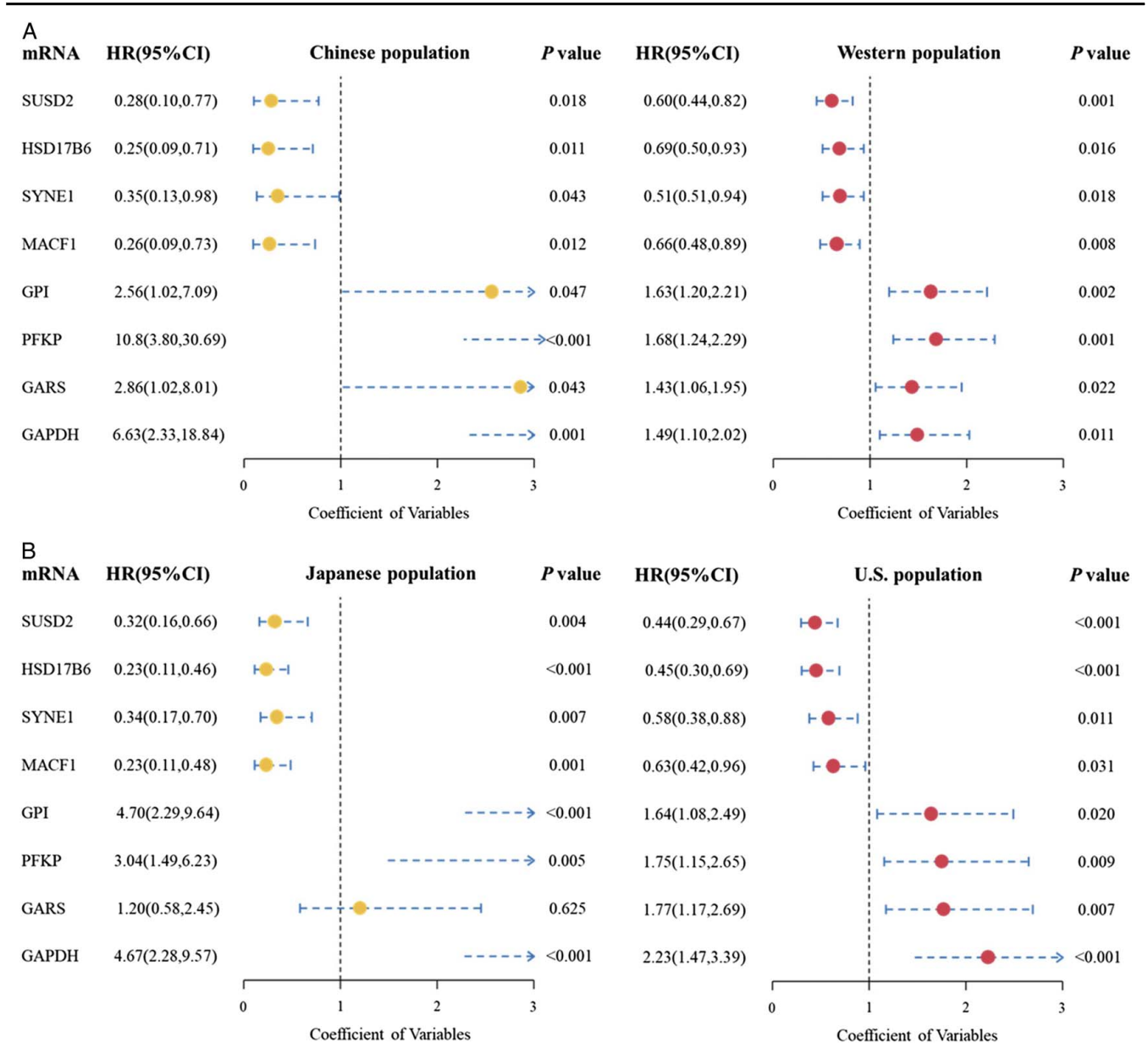
Utilizing the mutation data available to us, we incorporated the mutational status of EGFR into the survival models for the Chinese and Japanese populations. Subjects without EGFR mutations were assigned a score of 1. This decision was informed by our survival analysis, which highlighted the positive prognostic implications of EGFR mutation about LUAD prognosis (Figure S1, Supplemental Digital Content 3, <http://links.lww.com/JS9/B415>). We observed that upon incorporating the mutational status of EGFR, the predictive accuracy of the survival model remained consistent and robust (Fig. 10D).

#### Functional analysis of prognosis-related mRNAs

Based on the above identified eight hub mRNAs, the enrichment of the GO and KEGG pathways was carried out by DAVID Bioinformatics Resources. In the GO analysis, upregulated and downregulated mRNAs were enriched in eight distinct GO categories. Notably, ‘extracellular exosome’ and ‘membrane’ were the two categories with the highest mRNA involvement (Fig. 11A). Both of these categories play pivotal roles in intercellular communication. In the KEGG analysis, ‘glutathione metabolism’ emerged as the most significant pathway, while ‘metabolic pathways’ had the highest gene ratio (Fig. 11B).

#### Validate hub mRNAs in plasma and explore related mechanisms

We conducted plasma proteomics analysis on 10 LUAD patients and 10 healthy controls to assess the expression of proteins encoded by 8 hub mRNAs in the plasma of LUAD patients. Our findings indicated a differential upregulation of plasma proteins GPI and GAPDH in the LUAD. To validate these initial observations, we expanded our sample size and examined the expression of these proteins in plasma from 102 LUAD patients and 102



**Figure 8.** Forest plot of hazard ratio (HR) of prognosis-related mRNA expression in the different populations. (A) Survival results of eight hub mRNAs in Chinese and Western populations. (B) Survival results of eight hub mRNAs in Japanese and the U.S. population.

healthy controls using the ELISA assay. Consistently, our results confirmed an upregulated expression of GPI in the plasma of LUAD patients (Fig. 12A).

We determined the expression of GPI in the corresponding tumor tissue from 10 LUAD cases in which plasma was previously used for plasma proteomics analysis (Fig. 12B). To identify the role of GPI 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 (Fig. 12C). Significant positive correlations were detected between the expression of GPI in tissues and plasma ( $R = 0.93, P = 1.2 \times 10^{-4}$ ) (Fig. 12D). Tumor cells and its related substances breaking through the vascular basement

membrane and entering the vasculature are the prerequisites for blood biopsy. The immature and unstable vasculature of the tumor microenvironment provides a convenience for primary tumor cells and its related substances to enter the vasculature. Correlation analysis revealed a negative correlation between GPI expression in tissues and the normalized level of tumor vessels ( $R = -0.66, P = 0.036$ ) (Fig. 12D), 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 GPI ( $R = -0.71, P = 0.021$ ) in plasma (Fig. 12D), indicating that abnormalities in tumor vessels may facilitate the infiltration of substances from tumor tissue into the bloodstream.

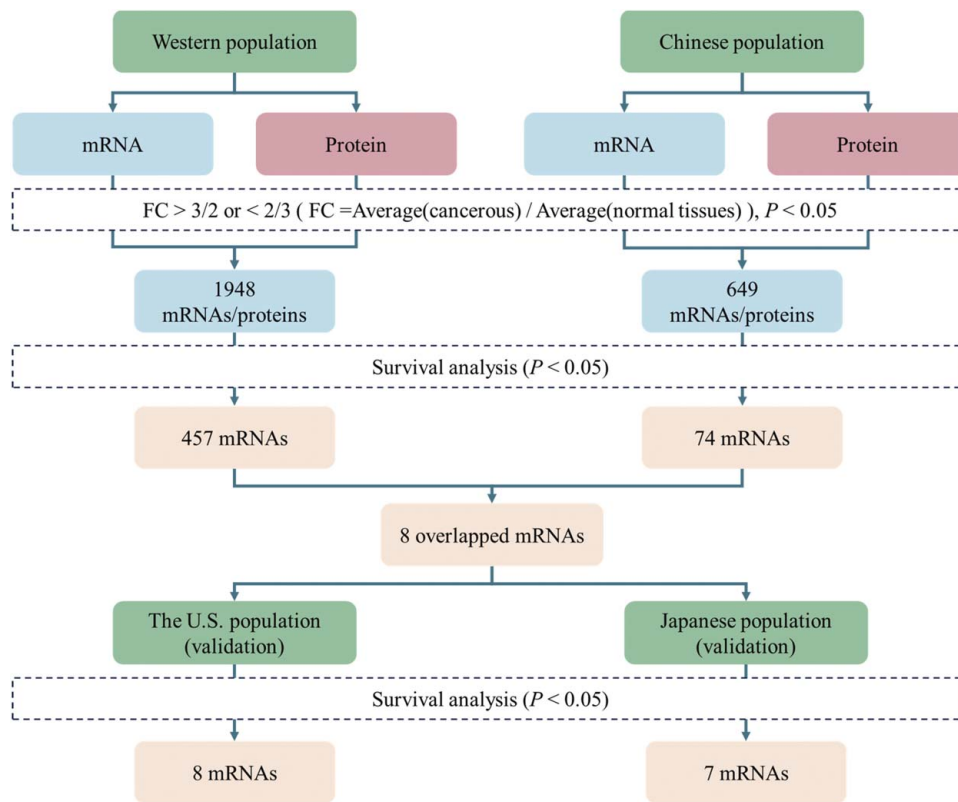


Figure 9 . Screening flowchart for prognostic-related hub mRNAs.

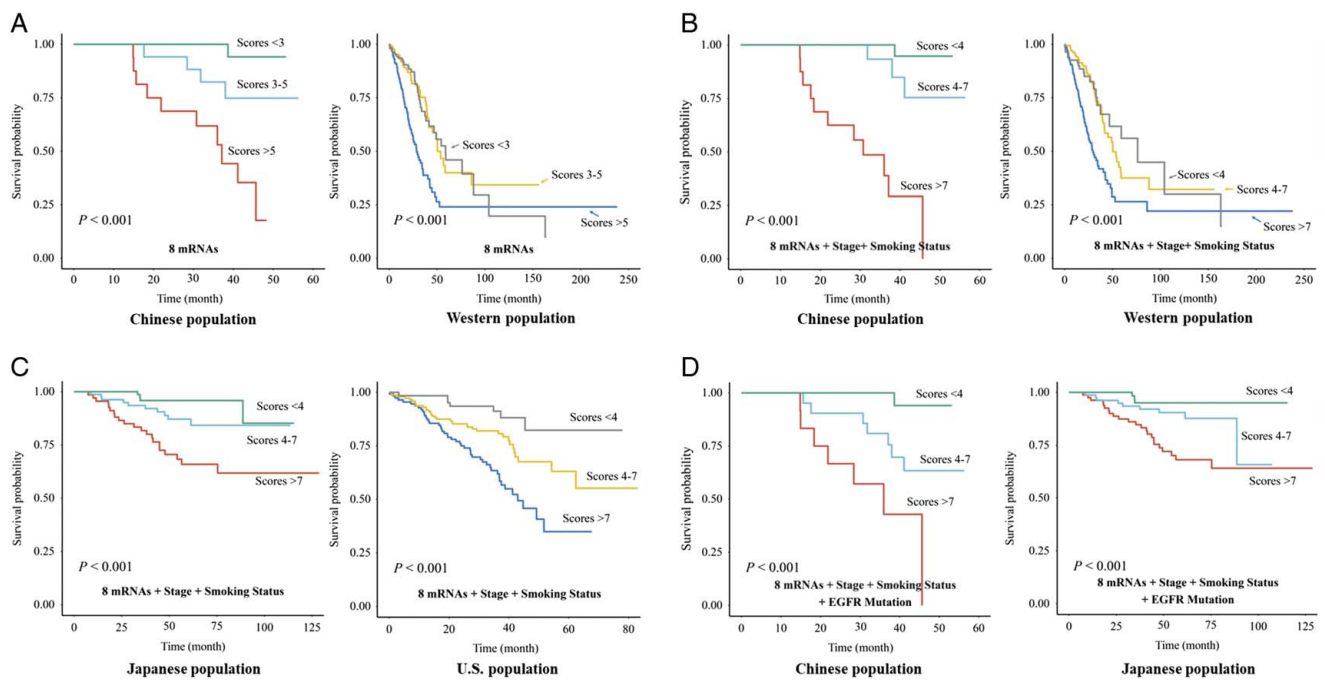
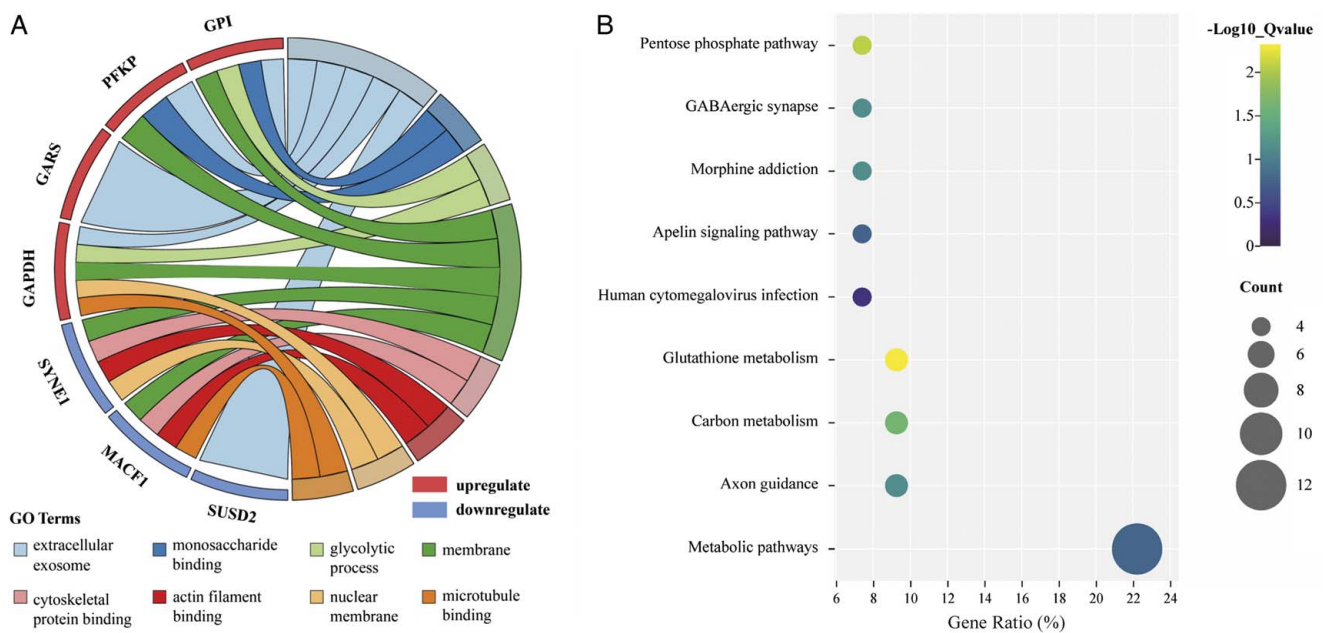


Figure 10 . Survival model results. (A) Survival model from eight hub mRNAs in Chinese and Western populations. (B) Survival model from eight hub mRNAs, stage, and smoking status in Chinese and Western population. (C) Validation of survival model from eight hub mRNAs, stage, and smoking status in Japanese and the U.S. population. (D) Validation of survival model from eight hub mRNAs, stage, smoking status, and EGFR mutation in Chinese and Japanese population.



**Figure 11 .** GO and KEGG pathway analysis of the prognosis-related mRNAs. (A) GO chord plot the eight most significant GO terms of the prognosis-related mRNAs. (B) Scatter plot of enriched KEGG pathways statistics. The color and size of the dots represent the range of the *P*-value and gene ratios mapped to the indicated pathways, respectively.

## Discussion

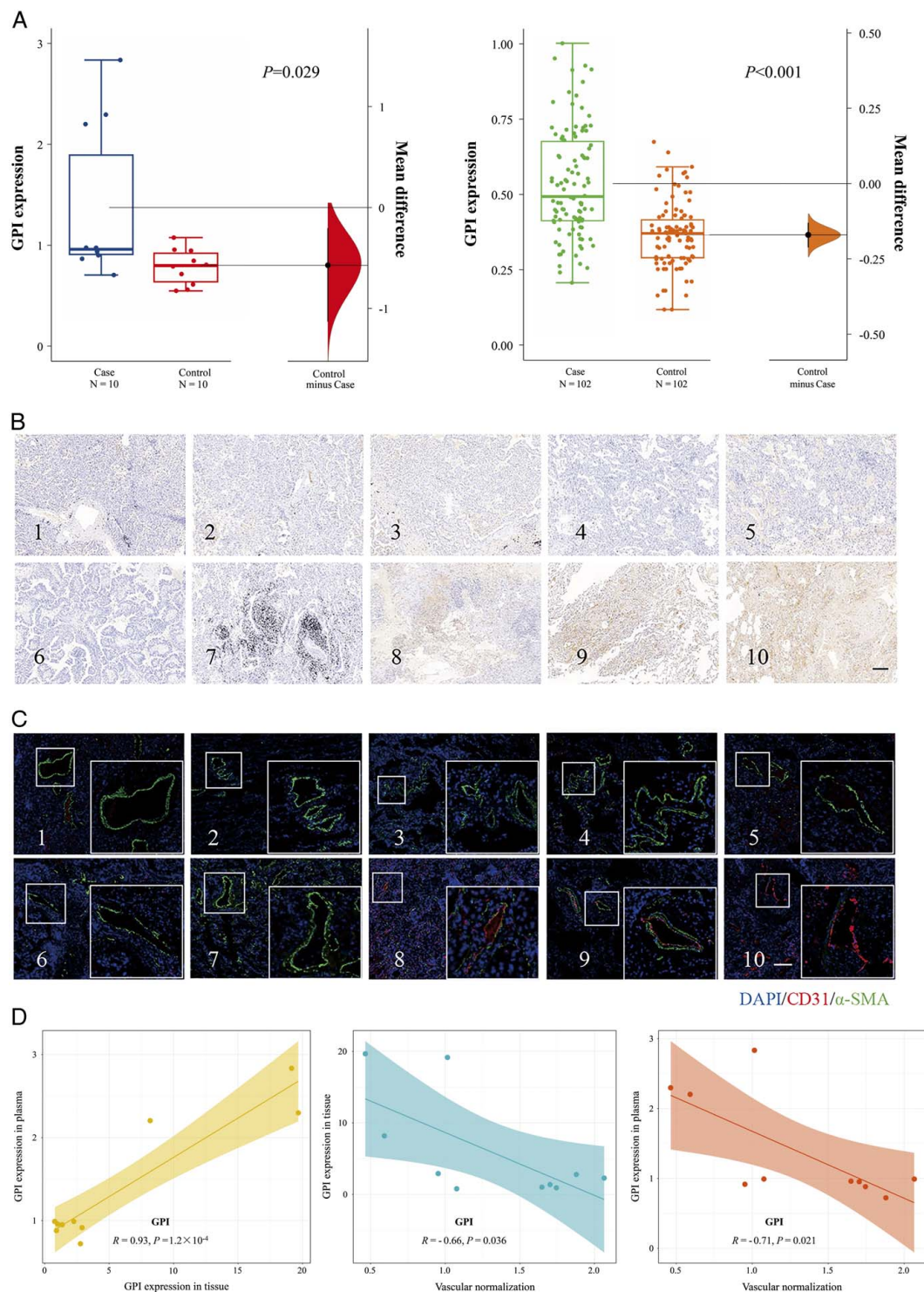
Differences in mRNA and protein expression among LUAD tumor tissues vary by race, potentially influencing our understanding of LUAD development, biomarker discovery, and therapeutic targets. By comparing mRNA and protein expression across different databases, we aim to quantify these variations. Furthermore, our integrated survival model, which encompasses diverse data, not only offers enhanced prognostic predictions for LUAD but also paves the way for individualized therapeutic strategies. Impressively, this model has demonstrated robust predictive capabilities across both Western and East Asian populations, underscoring its universal applicability.

Our study found a consistency rate of the differential expression between mRNA and proteins in Western populations was ~40%, and we observed a lower consistency rate of around 20% in the Chinese population, while the correlation between mRNA and protein expression was not consistent in some previous<sup>[22,27,30]</sup>. This discrepancy can be attributed to genetic diversity among populations, which may contribute to variations in mRNA–protein expression correlations. Differences in genetic backgrounds, including single nucleotide polymorphisms (SNPs) and genetic variations, can impact post-transcriptional processes and protein synthesis efficiency. Meanwhile, the choice of analytical methods and platforms used for mRNA and protein quantification may differ between studies and populations. Variances in experimental techniques, such as RNA-seq and mass spectrometry-based proteomics, can introduce technical biases that affect the observed consistency rates.

Previous studies on the correlation analysis between mRNA and protein expression mainly encompassed all detected mRNAs and proteins or categorized them into LUAD and normal groups<sup>[21,22]</sup>. When conducting a correlation analysis of mRNA and protein expression in the Chinese population with LUAD, it

was observed that the median absolute correlation coefficients of mRNA and protein pairs within the consistently differentially expressed group were higher than those within the non-differentially expressed group. This suggests a stronger correlation between mRNA and protein within the consistently differentially expressed group, indicating that many mRNA-level alterations may undergo modifications during translation, potentially reducing their impact. Consequently, further analysis and exploration of these consistently differentially expressed mRNA and protein pairs hold promise for in-depth investigation.

Sex-specific genetic differences in LUAD are widely recognized, and some studies have suggested that gender-biased molecular differences in LUAD are ethnically influenced<sup>[6]</sup>, roughly around 80%. There is substantial evidence of genetic differences between Asian and Western populations with LUAD, such as a significantly higher frequency of EGFR somatic mutations in Asian LUADs than in Western populations<sup>[31]</sup>. In this study, a great difference was discovered in differentially expressed mRNA between Chinese and Western populations, with only about 40–60% consistency, suggesting that ethnic factors may play a significant role in LUAD development. The 40–60% consistency rate of differentially expressed proteins between races signifies that a substantial portion of the protein-level alterations in LUAD differs between Western and Chinese populations. This observation underscores the complex interplay of genetic, environmental, and molecular factors in shaping the proteomic landscape of this cancer type. One possible explanation for this discrepancy lies in genetic variations between populations. Different ethnic groups often exhibit genetic polymorphisms, single nucleotide variations, and allelic frequencies that influence gene expression at multiple levels. These genetic differences can impact the transcription and translation processes that ultimately yield proteins. As a result, we observe larger disparities in protein



**Figure 12 .** Correlation between vascular normalization and GPI expression in LUAD plasma and tissues. (A) GPI expression in 'Blood + ' screening and validation phases. (B) Immunohistochemical staining results of GPI in tumor tissues. Scale: 200  $\mu$ M. (C) Immunofluorescence photographs of LUAD tumor tissue with CD31 and  $\alpha$ -SMA staining. Scale: 200  $\mu$ M. (D) Correlation analysis between the GPI expression in tumor tissues and plasma; Correlation analysis between the GPI expression in tumor tissues and vascular normalization; Correlation analysis between the GPI expression in plasma and vascular normalization.

expression profiles, even when the underlying genomic variations are relatively subtle.

The construction of a comprehensive LUAD survival model offers several advantages. It provides clinicians with a nuanced

tool to predict patient outcomes, enabling personalized therapeutic strategies and more informed clinical decisions. Our model's strength lies in its integration of both molecular markers and clinical parameters, offering a more holistic perspective of

disease progression. While many existing models might lean heavily on either genetic markers or clinical data, our approach harmoniously melds the two, potentially improving its predictive precision. The inclusion of EGFR mutation status, a pivotal marker in LUAD, further underscores our model's contemporary relevance. Furthermore, by comparing datasets from diverse demographic backgrounds, we have ensured the model's broad applicability, a vital aspect in today's interconnected medical landscape.

GPI, a dimeric enzyme with a molecular weight of ~132 000, is responsible for catalyzing the conversion between D-glucose-6-phosphate and D-fructose-6-phosphate<sup>[32]</sup>. Its role in glycolysis is evident, as suppression of GPI has been shown to hinder glycolysis in cancer cells without affecting their overall viability<sup>[33]</sup>. Beyond its glycolytic function, GPI has been implicated in various cancer-related processes<sup>[34,35]</sup>. For instance, it influences metastasis in colorectal cancer and has been identified as a prognostic marker for hepatocellular carcinoma<sup>[36,37]</sup>. Our research has further highlighted the significance of GPI in LUAD. We found that high GPI expression in tissues correlates with a poorer prognosis for LUAD. This observation was consistent across both Western and East Asian populations. Additionally, an upregulation of GPI was detected in the plasma of LUAD patients. Given these findings, along with GPI's impact on the tumor immune microenvironment<sup>[38]</sup>, it is evident that GPI holds both diagnostic and prognostic value for LUAD.

Tumor abnormal vasculature is characterized by discontinuous endothelial lining and defective basement membrane, which facilitate the egress of tumor cells and associated molecules into the systemic circulation. In our investigation, we discerned that elevated GPI expression in tumor tissues is associated with these vascular anomalies. Such vascular defects can enhance the intravasation of neoplastic cells and their derivatives, potentiating metastatic spread. This assertion is buttressed by the inverse relationship observed between GPI levels in plasma and tumor vessel normalization, suggesting GPI leakage from the tumor milieu via its heterogeneous vasculature. Tumor cells are highly glycolytic<sup>[39]</sup>, while glycolysis inhibits the vascular support function of pericytes<sup>[40]</sup>. Targeting tumor cell glycolysis induces normalization of tumor vasculature, thereby reducing invasion, endocytosis, and dissemination of cancer cells<sup>[39]</sup>. Therefore, as key enzyme proteins of glycolysis, monitoring the concentration of GPI in plasma can reflect the normalized level of blood vessels within the tumor and infer the risk and malignancy of tumor metastasis.

Our study innovatively explores the consistency of differentially expressed mRNAs and proteins. We highlighted the differences in mRNA and protein expression, as well as differences in mRNA expression in LUAD across races and differences in protein expression across races. Given that mRNA cannot directly perform biological functions, it is vital to consider the role played by proteomics alongside mRNA expression. Consequently, we assessed the consistency between differentially expressed mRNAs and proteins in LUAD, identifying a set of hub mRNAs (proteins) associated with LUAD prognosis. By integrating these mRNAs with clinical and pathological data, we constructed a survival model. We anticipate that these findings will serve as a valuable reference for subsequent LUAD research. Concurrently, through plasma proteomics screening and validation, we identified the plasma protein GPI and endeavored to elucidate

its potential mechanism in relation to LUAD prognosis. However, this study has limitations. The transcriptomics data and proteomics data included in this study are not from the same sample, and the experimental methods of the different studies may have biased the results, which also led to the fact that we were not able to do the correlation analyses of differentially expressed mRNAs and proteins in the Western populations.

## Conclusion

The consistency rate between differentially expressed mRNAs and proteins in LUAD patients was about 40% in the Western population and 20% in the Chinese population. When comparing Western and Chinese LUAD populations, the consistency rate for differentially expressed mRNAs was around 40–60%. Moreover, the consistency rate for differentially expressed proteins was around 40–60% between races. These findings underscore the imperative need to account for differences between transcriptomics and proteomics in future LUAD research endeavors. We have developed a survival model for LUAD, wherein samples with a higher number of unfavorable variables exhibit poorer prognoses. Subsequently, we selected and validated the plasma protein GPI, monitoring its expression in plasma could potentially infer the metastatic risk and malignancy grade in LUAD patients.

## Ethical approval

The study was reviewed and approved by the ethics committee of Nantong University (Approval No. 2022-2).

## Consent

Written informed consent was obtained from the patient for the publication of this case report and accompanying images. A copy of the written consent is available for review by the Editor-in-Chief of this journal on request.

## Sources of funding

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).

## Author contribution

Y.L.: conceptualization, investigation, and writing – original draft, review, and editing; Z.L.: investigation, methodology, and writing – review and editing; Q.M.: conceptualization, investigation, resources, and writing – original draft; A.N.: validation, visualization, and writing – review and editing; S.Z.: data curation and methodology; S.L.: formal analysis, software, and writing – review and editing; X.T.: funding acquisition and writing – review and editing; Y.W.: conceptualization and writing – review and editing; Q.C.: formal analysis and software; T.T.: data curation and investigation; L.Z.: resources and methodology; J.C.: methodology and writing – review and editing; L.M.: funding acquisition,

investigation, and supervision; M.C.: funding acquisition, investigation, methodology, supervision, validation, visualization, and project administration.

### Conflicts of interest disclosure

The authors declare that they have no conflicts of interest.

### Research registration unique identifying number (UIN)

1. Name of the registry: researchregistry.
2. Unique identifying number or registration ID: researchregistry8875.
3. Hyperlink to your specific registration (must be publicly accessible and will be checked): <https://www.researchregistry.com/registernow#home/registrationdetails/6441fc18c390230027033630/>.

### Guarantor

Minjie Chu.

### Data availability statement

The datasets generated during and/or analyzed during the current study are not publicly available but are available from the corresponding author on reasonable request.

### Provenance and peer review

Not commissioned, externally peer-reviewed.

### Acknowledgements

Assistance with the study: none.

Presentation: none.

### References

- [1] Siegel RL, Miller KD, Wagle NS, *et al.* Cancer statistics, 2023. *CA Cancer J Clin* 2023;73:17–48.
- [2] Duma N, Santana-Davila R, Molina JR. Non-small cell lung cancer: epidemiology, screening, diagnosis, and treatment. *Mayo Clin Proc* 2019; 94:1623–40.
- [3] Yu XQ, Yap ML, Cheng ES, *et al.* Evaluating prognostic factors for sex differences in lung cancer survival: findings from a large Australian cohort. *J Thorac Oncol* 2022;17:688–99.
- [4] Tamasi L, Horvath K, Kiss Z, *et al.* Age and gender specific lung cancer incidence and mortality in Hungary: trends from 2011 through 2016. *Pathol Oncol Res* 2021;27:598862.
- [5] Fan T, Li C, He J. Prognostic value of immune-related genes and comparative analysis of immune cell infiltration in lung adenocarcinoma: sex differences. *Biol Sex Differ* 2021;12:64.
- [6] Li X, Wei S, Deng L, *et al.* Sex-biased molecular differences in lung adenocarcinoma are ethnic and smoking specific. *BMC Pulm Med* 2023; 23:99.
- [7] Crick F. Central dogma of molecular biology. *Nature* 1970;227:561–3.
- [8] de Sousa Abreu R, Penalva LO, Marcotte EM, *et al.* Global signatures of protein and mRNA expression levels. *Mol Biosyst* 2009;5: 1512–26.
- [9] Wang H, Wang Q, Pape UJ, *et al.* Systematic investigation of global coordination among mRNA and protein in cellular society. *BMC Genom* 2010;11:364.
- [10] Sonneveld S, Verhagen BMP, Tanenbaum ME. Heterogeneity in mRNA translation. *Trends Cell Biol* 2020;30:606–18.
- [11] Zhao BS, Roundtree IA, He C. Post-transcriptional gene regulation by mRNA modifications. *Nat Rev Mol Cell Biol* 2017;18:31–42.
- [12] Machnicka MA, Milanowska K, Osman Oglou O, *et al.* MODOMICS: a database of RNA modification pathways–2013 update. *Nucleic Acids Res* 2013;41:D262–7.
- [13] Statello L, Guo CJ, Chen LL, *et al.* Gene regulation by long non-coding RNAs and its biological functions. *Nat Rev Mol Cell Biol* 2021;22: 96–118.
- [14] Guo L, Louis IV, Bohjanen PR. Post-transcriptional regulation of cytokine expression and signaling. *Curr Trends Immunol* 2018;19:33–40.
- [15] Ule J, Blencowe BJ. Alternative splicing regulatory networks: functions, mechanisms, and evolution. *Mol Cell* 2019;76:329–45.
- [16] Wright CJ, Smith CWJ, Jiggins CD. Alternative splicing as a source of phenotypic diversity. *Nat Rev Genet* 2022;23:697–710.
- [17] Czuba LC, Hillgren KM, Swaan PW. Post-translational modifications of transporters. *Pharmacol Ther* 2018;192:88–99.
- [18] Pan S, Chen R. Pathological implication of protein post-translational modifications in cancer. *Mol Aspects Med* 2022;86:101097.
- [19] Wu Z, Huang R, Yuan L. Crosstalk of intracellular post-translational modifications in cancer. *Arch Biochem Biophys* 2019;676:108138.
- [20] Liu X, Shi F, Li Y, *et al.* Post-translational modifications as key regulators of TNF-induced necroptosis. *Cell Death Dis* 2016;7:e2293.
- [21] Soltis AR, Bateman NW, Liu J, *et al.* Proteogenomic analysis of lung adenocarcinoma reveals tumor heterogeneity, survival determinants, and therapeutically relevant pathways. *Cell Rep Med* 2022;3:100819.
- [22] Gillette MA, Satpathy S, Cao S, *et al.* Proteogenomic characterization reveals therapeutic vulnerabilities in lung adenocarcinoma. *Cell* 2020; 182:200–225.e35.
- [23] Wu L, Wen Z, Song Y, *et al.* A novel autophagy-related lncRNA survival model for lung adenocarcinoma. *J Cell Mol Med* 2021;25:5681–90.
- [24] Wang X, Yao S, Xiao Z, *et al.* Development and validation of a survival model for lung adenocarcinoma based on autophagy-associated genes. *J Transl Med* 2020;18:149.
- [25] Hou S, Xu H, Liu S, *et al.* Integrated bioinformatics analysis identifies a new stemness index-related survival model for prognostic prediction in lung adenocarcinoma. *Front Genet* 2022;13:860268.
- [26] McShane LM, Altman DG, Sauerbrei W, *et al.* Reporting recommendations for tumour MARKer prognostic studies (REMARK). *Br J Cancer* 2005;93:387–91.
- [27] Xu JY, Zhang C, Wang X, *et al.* Integrative proteomic characterization of human lung adenocarcinoma. *Cell* 2020;182:245–261.e17.
- [28] Okayama H, Kohno T, Ishii Y, *et al.* Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* 2012;72:100–11.
- [29] Schabath MB, Welsh EA, Fulp WJ, *et al.* Differential association of STK11 and TP53 with KRAS mutation-associated gene expression, proliferation and immune surveillance in lung adenocarcinoma. *Oncogene* 2016;35:3209–16.
- [30] Chen YJ, Roumeliotis TI, Chang YH, *et al.* Proteogenomics of non-smoking lung cancer in East Asia delineates molecular signatures of pathogenesis and progression. *Cell* 2020;182:226–244.e17.
- [31] Chen J, Yang H, Teo ASM, *et al.* Genomic landscape of lung adenocarcinoma in East Asians. *Nat Genet* 2020;52:177–86.
- [32] Achari A, Marshall SE, Muirhead H, *et al.* Glucose-6-phosphate isomerase. *Philos Trans R Soc Lond B Biol Sci* 1981;293:145–57.
- [33] Mazziro 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: 469–97.
- [34] 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.
- [35] 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.
- [36] Tsutsumi S, Fukasawa T, Yamauchi H, *et al.* Phosphoglucose isomerase enhances colorectal cancer metastasis. *Int J Oncol* 2009;35:1117–21.
- [37] 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;40:698–704.

- [38] 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.
- [39] 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;30:968–85.
- [40] 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;12:6011.