

Urine proteomic signatures predicting the progression from premalignancy to malignant gastric cancer



Hua Fan,^{a,g} Xue Li,^{a,g} Zhong-Wu Li,^b Nai-Ren Zheng,^c Li-Hua Cao,^d Zong-Chao Liu,^a Ming-Wei Liu,^c Kai Li,^c Wen-Hui Wu,^a Zhe-Xuan Li,^a Tong Zhou,^a Yang Zhang,^a Wei-Dong Liu,^e Lan-Fu Zhang,^f Wei-Cheng You,^a Yi Wang,^c Jianmin Wu,^{d,****} Kai-Feng Pan,^{a,***} Jun Qin,^{c,**} and Wen-Qing Li^{a,*}



^aKey Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Cancer Epidemiology, Peking University Cancer Hospital & Institute, Beijing 100142, China

^bDepartment of Pathology, Peking University Cancer Hospital & Institute, Beijing 100142, China

^cState Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, Beijing 102206, China

^dKey Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Center for Bioinformatics, Peking University Cancer Hospital & Institute, Beijing 100142, China

^eLinqu County Public Health Bureau, Shandong 262600, China

^fLinqu County People's Hospital, Shandong 262600, China

Summary

Background Early detection of gastric cancer (GC) remains challenging. We aimed to examine urine proteomic signatures and identify protein biomarkers that predict the progression of gastric lesions and risk of GC.

Methods A case–control study was initially designed, covering subjects with GC and gastric lesions of different stages. Subjects were aged 40–69 years, without prior diagnosis of renal or urological diseases. We enrolled a total of 255 subjects, with 123 in the discovery stage from Linqu, China, a high-risk area for GC and 132 in the validation stage from Linqu and Beijing. A prospective study was further designed for a subset of 60 subjects with gastric lesions, which were followed for 297–857 days.

Findings We identified 43 differentially expressed urine proteins in subjects with GC vs. mild or advanced gastric lesions. Baseline urinary levels of ANXA11, CDC42, NAPA and SLC25A4 were further positively associated with risk of gastric lesion progression. Three of them, except for SLC25A4, also had higher expression in GC than non-GC tissues. Integrating these four proteins showed outstanding performance in predicting the progression of gastric lesions (AUC (95% CI): 0.92 (0.83–1.00)) and risk of GC (AUC (95% CI): 0.81 (0.73–0.89) and 0.84 (0.77–0.92) for GC vs. mild or advanced gastric lesions respectively).

Interpretation This study revealed distinct urine proteomic profiles and a panel of proteins that may predict the progression of gastric lesions and risk of GC. These biomarkers in a non-invasive approach may have translational significance for defining high-risk populations of GC and its early detection.

Funding Funders are listed in the Acknowledgement.

Copyright © 2022 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Gastric cancer; Proteomics; Urine biomarkers; Gastric lesion progression

eBioMedicine

2022;86: 104340

Published Online XXX

<https://doi.org/10.1016/j.ebiom.2022.104340>

1016/j.ebiom.2022.104340

*Corresponding author. Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Cancer Epidemiology, Peking University Cancer Hospital & Institute, 52 Fu-cheng Road, Haidian District, Beijing 100142, China.

**Corresponding author. State Key Laboratory of Proteomics, Beijing Proteome Research Center, National Center for Protein Sciences (Beijing), Beijing Institute of Lifeomics, 38 Lifescience Park Road, Changping District, Beijing 102206, China.

***Corresponding author. Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Cancer Epidemiology, Peking University Cancer Hospital & Institute, 52 Fu-cheng Road, Haidian District, Beijing 100142, China.

****Corresponding author. Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Center for Bioinformatics, Peking University Cancer Hospital & Institute, 52 Fu-cheng Road, Haidian District, Beijing 100142, China.

E-mail addresses: wenqing_li@bjmu.edu.cn (W.-Q. Li), jqin1965@126.com (J. Qin), pan-kf@263.net (K.-F. Pan), wujm@bjmu.edu.cn (J. Wu).

[§]Hua Fan and Xue Li contributed equally to this paper.

Research in context**Evidence before this study**

Gastric cancer remains a major public health threat worldwide. Gastric cancer develops through a long-term process in a stepwise manner. Molecular biomarkers aiding the identification of high-risk population and early detection of gastric cancer are highly needed. Urine contains circulating molecules that carry information from abnormal cells and sensitively provides a globe view of individual's health status, thus may provide a non-invasive approach for exploring biomarkers for gastric cancer.

Added value of this study

Our two-stage study with a prospective follow-up for a subset of subjects revealed distinct urine proteomic profiles and a panel of urine protein biomarkers (ANXA11, CDC42, NAPA and SLC25A4) that may predict the progression of gastric

lesions and risk of GC. Integrating the protein score of four urine proteins materially improved the ability to predict the risk of gastric lesion progression to malignancies. Among them, ANXA11, CDC42, and NAPA had higher expression in gastric cancer than non-cancer tissues.

Implications of all the available evidence

This study portrayed urine proteomic landscape and explored urine proteomic signatures for different gastric lesions along the progression continuum to gastric cancer. The non-invasive detection of urine biomarkers may have translational significance for fine discrimination of high-risk population subgroups for the progression of gastric lesions to gastric cancer and its early detection, advancing the approaches to its precision prevention and management.

Introduction

Gastric cancer (GC) is the fourth leading cause of cancer-related deaths worldwide.¹ Majority of GC patients are diagnosed at advanced stages with poor survival.² Identifying populations with markedly increased risk for developing GC, as well as detecting GC at an early stage, is crucial to increase the chance of survival and improve overall quality of life.^{3,4} Although endoscopic screening has shown its benefit for the early detection of GC,^{5,6} the screening procedure is invasive, expensive, time-intensive, and difficult for gross medical institutions.⁷

GC, particularly of the prevalent intestinal subtype, develops through a long-term process in a stepwise manner.⁸ Studies on subjects spanning the cascade of premalignant and malignant GC, particularly with prospective follow-up on the progression of gastric lesions, are warranted to uncover prediagnostic molecular signatures underlying the development of GC, which is of critical needs for population risk stratification and early diagnosis of GC, enabling efficient prevention and management endeavors.

Compared with other biospecimens, the collection of urine is non-invasive and easily accessible in large quantities, allowing for biomarker detection in large-scale population screening.⁹ Circulating molecules in urine can sensitively provide a global view of individual's health status. Indeed, the malignant, apoptotic and necrotic cells can release proteins into the microenvironment and blood, which are then filtered and reabsorbed by kidneys and eventually enter the urine. Urine contains stable peptides and some low abundant but functionally important proteins that can be reliably measured.^{10,11} Especially, about 9% of identified urine proteins from health individuals are involved in immune response.¹² These may be linked

with observed immune response and evasion in premalignant lesions, which were associated with key somatic genomic alterations.^{13,14} Together, changes in urinary proteins may reflect early biological events during GC carcinogenesis. Detecting urine proteins may serve as a promising approach for exploring biomarkers for progression of gastric lesions and risk of early GC.

Using urine as a liquid biopsy, we conducted proteomic profiling in two stages among a total of 255 individuals involving gastric lesions of different stages and GC. We portrayed the urine proteomic fingerprint and developed urine proteomic signatures associated with the progression of gastric lesions and risk of GC. We also deciphered the immune contexture from proteomic profiles and examined the highlighted proteins integrating our available data on tissue proteomics.

Methods**Study population**

A case-control study was initially designed with a total of 255 subjects in two stages, which defined subjects with GC as the case group and had subjects with the well-recognized mild (superficial gastritis (SG) or chronic atrophic gastritis (CAG)) and advanced gastric lesions (intestinal metaplasia (IM) or low-grade intraepithelial neoplasia (LGIN)) as two control groups. All controls in two stages and cases of the discovery stage were enrolled from Linqu County, Shandong Province of China, a rural area that has one of the highest GC mortality worldwide,¹⁵ based on residents attending national Upper Gastrointestinal Cancer Early Detection (UGCED) Program. For selection of controls, among those eligible Linqu residents invited through the household registration system, the respondents agreed

to undertake the first gastroscopy in the UGCED program and provided urine samples; a response rate of 72.4% has been reached.

Details of the UGCED program have been published elsewhere.^{6,16} Briefly, UGCED program investigators randomly select villages in Linqu and eligible village residents aged 40–69 years undergo endoscopic screening for early diagnosis of GC free of charge. Information on age and sex was documented based on the identity card of China. Eligibility exclusions include refusal to provide informed consent, ages outside the range of 40–69 years, previous diagnosis of cancer (except nonmelanoma skin cancer), bleeding disorder, heart failure, renal or urological disease, liver disease, emphysema, or other life-threatening illness. In the program, HGIN and invasive GC are collectively defined as GC. In addition to those undergoing the first endoscopy, individuals diagnosed with severe CAG, IM, or LGIN would be invited to take repeated examination the following year.

The discovery stage had 123 subjects, including 109 subjects with gastric lesions at different stages (“controls”) and 14 with GC (“cases”, 5 high-grade intraepithelial neoplasias (HGINs) and 9 invasive GCs). From November 22, 2018 to December 07, 2018, 231 agreed to undertake the first gastroscopy in the UGCED program. Along with the gastric biopsies for pathological diagnosis, urine sample and an extra biopsy from lesser curvature of antrum was additionally collected for each participant. Of them, 111 participants were deemed as eligible for the gastric tissue proteomics assay (with the findings published in 2021¹⁶), which had the most severe gastric histology at lesser curvature of antrum, so that any systemic effect on proteomic profiles caused by severer gastric lesions at other gastric mucosa sites can be minimized. Among 111 individuals, 109 provided sufficient mid-stream urine samples in the morning and were included as controls in the discovery stage of present study (51 with SG/CAG and 58 with IM/LGIN, two control groups). As we did not identify GCs among 231 residents, we further included all newly diagnosed GCs ($n = 14$) among those attending repeated endoscopic examination in the UGCED program between November and December of 2018, and defined them as cases in the discovery stage. For each case, 15 ml of mid-stream urine in the morning was collected.

For validation purposes, an independent validation set involved 132 subjects, with the same inclusion and exclusion criteria as the discovery set. To test the robustness of findings and improve the potential extrapolation to other populations, we sought to enroll the validation set subjects from Beijing (with a relatively low risk for GC) and Linqu simultaneously. The controls were randomly selected from participants who attended the UGCED program and provided sufficient mid-stream urine samples in the morning between

October, 2020 and March, 2021, including 114 subjects with precancerous gastric lesions at different stages. Cases were enrolled from Peking University Cancer Hospital clinics, including 18 subjects with GC (5 HGINs and 13 invasive GCs) (Fig. 1; Table S1).

Applying a normal approximation method¹⁷ for power calculation in the discovery and validation set respectively, at $\alpha = 0.05$, we found the current sample size would lead to the statistical power greater than 80% to detect an OR of 1.90 per 2-SD change of the protein expression levels for all comparisons between GC and gastric lesions (SG/CAG or IM/LGIN). The sample size also meets the empirical requirements of current proteomic techniques to detect enough number of proteins.

We prospectively followed the evolution trajectories of gastric lesions for a subset of 60 subjects with precancerous gastric lesions who volunteered to attend the endoscopic follow-up (297–857 days). Subjects with GC (HGIN and invasive GC) were not prospectively followed.

All gastroendoscopies were conducted by board-certified gastroenterologists using video endoscopes (Olympus), with detailed photographic documentation throughout the examinations. Biopsies were taken at five standardized mucosa sites. All biopsies involved in the present study were assessed independently by two experienced pathologists (initially by Dr. Xiu-Zhen Wu in Linqu County’s Hospital and then reviewed by Dr. Zhou-Wu Li in Peking University Cancer Hospital). The final diagnoses were determined after discussion for any disputed histopathologic classification, following the criteria proposed by the Updated Sydney System¹⁸ and Chinese Association of Gastric Cancer.¹⁹ Each subject was given a global diagnosis of normal, SG, CAG, IM, LGIN, HGIN (pathologically including severe dysplasia and GC in situ) or invasive GC, defined as the most severe gastric histopathology among all biopsies. Representative images of these gastric lesions are shown in Fig. S1. Gastroendoscopy was conducted again at the endpoint and the evolution of gastric lesions during follow-up was assessed for each subject. A subject was classified to have progression of gastric lesions if the endpoint severity score of gastric lesions was higher than the baseline.

Ethics

The study was approved by the Institutional Review Boards of Peking University Cancer Hospital (approval No. 2018KT117). All participants provided written informed consent.

Sample preparation, proteomic profiling, data processing and protein quantification

For all subjects, collection, transportation and storage of urine samples followed standard protocol to minimize

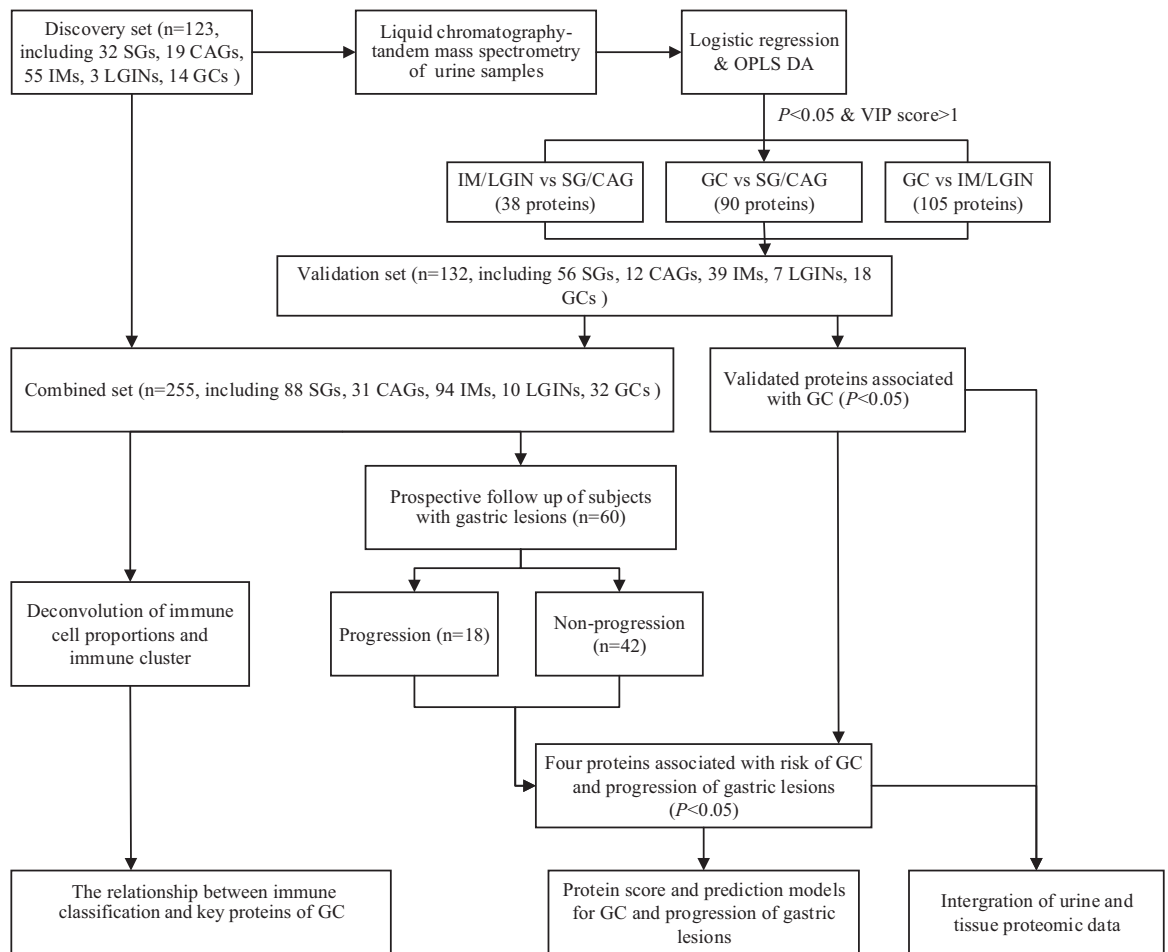


Fig. 1: The schematic diagram of study design and work flow. CAG, chronic atrophic gastritis; GC, gastric cancer; HGIN, high-grade intraepithelial neoplasia; IM, intestinal metaplasia; LGIN, low-grade intraepithelial neoplasia; OPLS-DA, Orthogonal partial least squares discrimination analysis; SG, superficial gastritis.

preanalytical variations. Approximately 15 ml of mid-stream urine was collected in the morning before endoscopy for each subject. Urine samples were immediately dispensed into 5 ml cryogenic vials containing boric acid tablets after collection and stored at -80°C until testing. The urine sample was centrifuged at 200,000g for 70 min. After ultracentrifugation, pellet was reduced with DTT to remove uromodulin.

After heating at 65°C for 30 min, the pellet was washed with wash buffer (10 mM TEA, 100 mM NaCl, pH 7.4) twice and ultra-centrifuged for 30 min. The pellet was dissolved in SDS buffer (1% SDS, 50 mM Tris, pH 8.5) and resolved on an SDS-PAGE gel. Gel was cut into six pieces and then subjected to in-gel trypsin digestion. The tryptic peptide was separated on a Homemade capillary column filled with C18 particles and analyzed by the Thermo Fisher Orbitrap mass spectrometer combined with the online Easy-nLC 1000 nano-HPLC system (Thermo Fisher Scientific). To

ensure the reproducibility of LC-MS/MS runs, tryptic digests of 293T cell lysates were routinely used as quality control (QC) samples and a total of 18 QC samples were during our sample measurement timespan for proteomic profiling assays.²⁰

LC-MS/MS data were processed in Proteome Discoverer 1.4 (Thermo Fisher Scientific), and searched against the NCBI Human Refseq protein database (released on July 4, 2013) using Mascot search engine (Matrix Science 2.3). Peptides with up to 2 missed cleavages and a false discovery rate of less than 1% were considered acceptable. The peptides had average length of 13 amino acids (range 7–66). For the screening of high-confidence proteins, a peptide that only exist in one protein in the complete theoretical digested peptide library of human proteome is defined as a unique peptide, and a peptide with Mascot ion scores of 20 or more is defined as a strict peptide. We only used proteins that were detected with at least 1 unique strict peptide and

2 strict peptides or at least 3 strict peptides in the analysis, which were deemed to have high reliability ($n = 3540$ proteins). This QC policy ensures that we map the right protein by identified peptides. A label-free intensity based absolute quantification (iBAQ) algorithm was used to quantify protein abundance.²¹ For comparisons between batches, the iBAQ was converted to intensity-based fraction of total (iFOT), which was calculated as the iBAQ value of each protein divided by the sum of the iBAQ values of all proteins in the sample, and then multiplied by 10^5 to ease the visualization of low abundant proteins.²²

Bioinformatics and statistical analysis

All analyses were conducted using R (V.4.0.2) unless otherwise noted. Considering the histopathology nature of gastric lesions and sample size, subjects were categorized to three groups (mild gastric lesions (SG or CAG), advanced gastric lesions (IM or LGIN), and GC). All analyses were restricted to proteins that were identified in at least 1/2 samples of each comparison group. Missing values were imputed by using 1/10 of the lowest iFOT value of all proteins in each sample. Pairwise Spearman's correlation coefficients were calculated for the proteomic data of all 18 QC runs to evaluate the stability and reliability of experimental method. The normalized and \log_{10} transformed iFOTs were plotted for all samples to display the stability of data quality.

Partial least squares discrimination analysis (PLS-DA) was used to display an overview of the urine proteomic profiles in the groups of mild gastric lesions, advanced gastric lesions and GC. At the individual protein level, orthogonal partial least squares discrimination analysis (OPLS-DA) was used to calculate variable importance for the projection (VIP) score between the comparison groups. Proteins with VIP >1 in the discovery set were further analyzed using OPLS-DA in the validation set. We exerted additional measures to minimize the concern on multiple testing and only focused on proteins with VIP >1 in the comparison between GC and gastric lesions in both discovery and validation sets for below individual protein-level analyses. We fitted the changing trajectories from mild gastric lesions (SG or CAG) to advanced gastric lesions (IM or LGIN) and then to GC through locally estimated scatterplot smoothing (LOESS) regression. Proteins with similar trajectories were divided into groups through hierarchical clustering. KEGG pathway enrichment analysis was then applied to explore the biological functions of proteins, with FDR <0.05 considered as significant pathways.²³

To calculate the odds ratios (ORs) and 95% confidence intervals (CIs) for the associations of each individual protein (the independent variable) with the risk of GC, two binary logistic regression models were conducted (GC vs. SG/CAG and GC vs. IM/LGIN respectively as the dependent variable), adjusting for age

and sex.²⁴ We were not able to adjust for *Helicobacter pylori* infection and smoking status,²⁵ as the information on these variables were not available for GC cases. For proteins with VIP >1, those satisfying logistic regression $P < 0.05$ in the discovery set and $P < 0.05$ in the validation set between comparison groups (GC vs. SG/CAG or GC vs. IM/LGIN) were defined as differentially expressed proteins (DEPs). For identified DEPs, protein-protein interactions (PPI) enrichment analysis was carried out with STRING database. CytoHubba plug-in was applied to identify densely connected network components and find the top 15 core hub proteins of PPI network based on the Maximal Clique Centrality algorithm through Cytoscape software.²⁶

Among DEPs defined above, we further identified proteins associated with the progression of gastric lesions (progression vs. non-progression) using logistic regression analysis adjusting for age, sex and baseline gastric lesions. Urine proteins with $P < 0.05$ were considered statistically significant. Based on the derived four key proteins (ANXA11, CDC42, NAPA and SLC25A4), a protein score was developed by summing the weighted iFOT values of proteins, calculated as the standardized expression of each protein multiplied by the regression coefficient (β) obtained from logistic regression models containing all highlighted proteins, using the below function:

$$\text{Protein score} = \beta_1 * \text{ANXA11} + \beta_2 * \text{CDC42} + \beta_3 * \text{NAPA} + \beta_4 * \text{SLC25A4}.$$

After calculating the score for each subject, we evaluated the association between the protein score and risk of gastric lesion progression and GC.

We had recently completed the gastric tissue proteomic profiling for subjects of the discovery set, as published recently.¹⁶ Integrating the tissue proteomic profiling data, we explored the association of the tissue expression of key individual proteins with the risk of GC and progression of gastric lesions using logistic regression models.

We constructed prediction models for the risk of GC and gastric lesion progression using extreme gradient boosting (XGBoost) algorithm,²⁷ integrating protein signatures with age, sex, and baseline gastric histology (for gastric lesion progression only). To fully capture the protein features, the prediction model integrated the protein score of ANXA11, CDC42, NAPA and SLC25A4 as a continuous feature and also utilized the pairwise second-order interaction features of four proteins to train the XGBoost models.²⁸ The performance of prediction models was assessed by leave-one-out cross validation and displayed by area under the receiver operating characteristic (ROC) curve (AUC). DeLong's test was used to compare the performance of the prediction model integrating urinary protein signatures vs. the model only including age, sex, and baseline gastric histology (for gastric lesion progression).

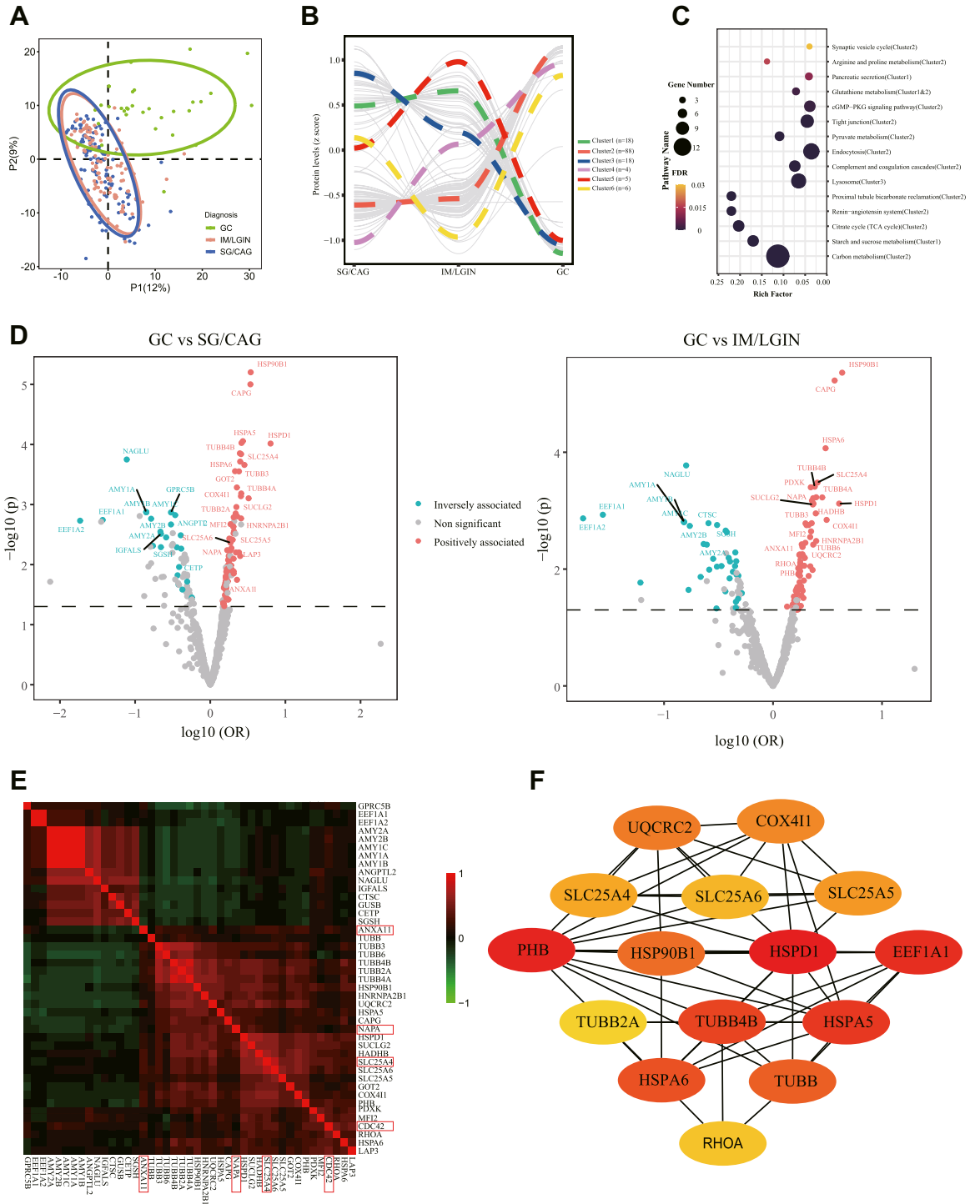


Fig. 2: Urine proteomic profiles and differentially expressed proteins for GC and precancerous gastric lesions. a. PLS-DA plot for the overview of urine proteomic profile. b. Protein clusters with similar trajectories from mild (SG/CAG) to advanced (IM/LGIN) gastric lesions and then to GC. For 139 proteins with VIP score >1 in the comparison between GC and gastric lesions in both discovery and validation sets, the changing trajectories of proteins were analyzed through locally estimated scatterplot smoothing (LOESS) regression. Proteins with similar trajectories were divided into groups through hierarchical clustering. c. Top 15 enriched pathway of proteins. KEGG pathway enrichment analysis was conducted for with VIP score >1 in the comparison between GC and gastric lesions in both discovery and validation sets. Rich factor is the ratio of the differentially expressed gene number (e.g., the gene that we submitted for the KEGG pathway analysis) to the total gene number in a

Deconvolution of immune cell proportions and immune subtyping using urine proteomic profiles

Recent studies have highlighted that urinary protein composition could be an appropriate mirror of global immune status.^{29,30} Although we did not measure peripheral blood leukocytes directly in our study, we sought to decipher the global immune contexture by deconvolution analysis using urine proteomic profiles. Our analysis took advantage of a protein signature matrix for 26 human immune cell types developed by CIBERSORT,³¹ using a reference proteome dataset for individual cell type from Rieckmann et al.,³² in which peripheral blood of four healthy donors were sorted by flow cytometry followed by MS profiling. As unquantified immune cell markers had a substantial impact on the performance of deconvolution analysis using proteomic profiles, we developed an in-house analysis pipeline NUWams (manuscript in submission) to address this issue. NUWams was applied to the urine proteomic profiles to infer the abundance of missing immune markers in the protein signature matrix. The NUWams-inferred proteomic profiles was then applied to estimate the relative proportions of 26 major immune cell types by CIBERSORT (v1.04) analysis using the developed proteomic signature. The estimated relative proportions was further grouped into ten major immune cell types (B, B.plasma, CD4⁺ T, CD8⁺ T, NK, monocytes/macrophages, dendritic cells, neutrophils, basophils and eosinophils cells) for clustering analysis.

Based on the estimated immune cell fractions of 255 urine samples, R package cola (1.6.0)³³ was employed to identify immune clusters of samples using the following parameters: spherical k-means clustering (skmeans), minK = 2, maxK = 6, partition repeat = 50, p_sampling = 1.

Spearman's and Pearson's correlation analyses were performed for the association between the estimated relative abundance of immune cells and key individual urine proteins. Wilcoxon rank-sum test was used for the differences of immune clusters and immune cell subsets across subjects with gastric lesions and GC. The immune clusters associated with highlighted individual proteins were also examined using Wilcoxon rank-sum test.

Role of funders

The funders had no role in study design, data collection and analysis, results interpretation, decision to publish, or preparation of the manuscript.

Results

Subject characteristics and urine protein detection

The study schematic diagram is shown in Fig. 1. A total of 3540 highly reliable urinary proteins were identified (Fig. S2a), of which 688 proteins were detected in more than 50% of samples in each gastric lesion and GC. The Spearman's correlation coefficients of quality control samples ranged from 0.80 to 0.91, indicating high batch-to-batch instrument reproducibility of LC-MS/MS runs (Fig. S2b). In addition, all samples showed good consistency in quantification of urine proteome, supporting its stability and reliability (Fig. S2c).

Urine differentially expressed proteins (DEPs) between GC and gastric lesions

Distinct urinary protein profiles were observed between subjects with GC and gastric lesions in partial least squares discrimination analysis (Fig. 2a). We identified 139 proteins with VIP score >1 comparing GC and gastric lesions (SG/CAG or IM/LGIN) in both discovery and validation sets, but none were identified for the comparisons between mild and advanced gastric lesions. To explore the changing trajectory of urinary protein levels among different gastric lesions and GC, six clusters of these 139 urine proteins that displayed dynamic changes from precancerous lesions to GC were defined through unsupervised hierarchical clustering (Fig. 2b). Significantly enriched pathways for 139 proteins and their aggregated clusters are shown in Fig. 2c. The top 2 enriched pathways were aggregated in Cluster-2 (carbon metabolism pathway, FDR = 1.33×10^{-12}) and Cluster-1 (starch and sucrose metabolism pathway, FDR = 1.16×10^{-6}) respectively. Cluster-2 had the largest protein assembly, the protein expression in which did not change apparently in mild and advanced gastric lesions but had soared in GC tremendously. In contrast, cluster 1 proteins had declined expression in GC. The proteins of cluster-3 showed consistent decrease with the progression of gastric lesions to GC, of which 6 proteins were enriched in the lysosomal pathway. Other protein clusters were modest in size.

The ORs for the associations between 139 proteins with VIP >1 and risk of GC were calculated using logistic regression models. Of them, 82 proteins were significantly associated with the risk of GC ($P < 0.05$, logistic regression analysis) in the discovery stage, with 43 of them replicated as DEPs

certain KEGG-defined pathway. d. A total of 43 differentially expressed proteins (DEPs) significantly associated with GC compared with SG/CAG (left) or IM/LGIN (right) as the reference respectively. *P* values and ORs were calculated based on the combined datasets. Proteins positively associated with GC (OR > 1) are shown in red and proteins inversely associated with GC (OR < 1) are shown in blue. e. Pearson correlation of 43 DEPs between GC and gastric lesions. For proteins with VIP >1 in the discovery and validation sets, those satisfying logistic regression $P < 0.05$ in both sets between comparison groups (GC vs. SG/CAG or GC vs. IM/LGIN) were defined as DEPs. f. PPI network of DEPs between GC and gastric lesions. Among 43 DEPs, the top 15 core hub proteins were defined based on the Maximal Clique Centrality algorithm using Cytoscape. CAG, chronic atrophic gastritis; GC, gastric cancer; IM, intestinal metaplasia; LGIN, low-grade intraepithelial neoplasia; SG, superficial gastritis.

for the risk of GC in the validation set (28 up-regulated and 15 down-regulated) (Table S2), involving 32 significant proteins for GC vs. SG/CAG and 30 proteins for GC vs. IM/LGIN (Fig. 2d). The four tissue protein markers that we recently reported, including DDT, HPX, PGC, and APOA1BP,¹⁶ were not significantly associated with GC risk at urine level (Table S3).

We visualized the protein clusters and network for these 43 DEPs. Many DEPs were positively correlated (Fig. 2e). The protein–protein interactions (PPI) network contains 43 nodes and 87 edges with an average node degree of 4.05 and a PPI enrichment $P < 1.0 \times 10^{-16}$. The top 15 potential core proteins were obtained (Fig. 2f).

Key proteins associated with the progression of gastric lesions and risk of GC

Using prospectively followed subjects ($n = 60$), we further examined whether 43 DEPs identified above were associated with the progression of gastric lesions. During follow-up, 18 subjects had their gastric lesions progressing to severer levels, with 2 of them developing GC. Four up-regulated urine proteins in GC across the discovery and validation sets were further significantly associated with an increased risk for gastric lesion progression ($P < 0.05$, logistic regression analysis), including ANXA11, CDC42, NAPA and SLC25A4 (Fig. 3a and b; Table 1). All four proteins were aggregated in cluster-2 and enriched in the pathways of endocytosis (CDC42), cGMP-PKG signaling (SLC25A4) and synaptic vesicle cycle (NAPA) (Fig. 2c). SLC25A4 was also displayed as a core protein in PPI network analysis (Fig. 2f). Stratified analysis found consistently higher level of these four proteins in progressed subjects either in the baseline mild (SG/CAG) or advanced gastric lesion group (IM/LGIN).

Protein scores and prediction models integrating protein biomarkers

We calculated a protein score integrating four key urine proteins. Given the correlations of highlighted proteins. The protein score was independently associated with the risk of GC compared with mild (SG/CAG) or advanced gastric lesions (IM/LGIN) as the reference, with ORs (95% CI) of 2.13 (1.38–3.28) and 2.43 (1.50–3.95) respectively, per one SD increase of the score. In addition, the protein score was significantly associated with the risk of gastric lesion progression (OR = 3.63, 95% CI: 1.49–8.81) (Fig. 3c).

Compared with the model including age, sex and baseline pathology (for the prediction of progression risk), integrating the protein score significantly improved the ability to predict the risk of GC (AUC (95% CI): 0.81 (0.73–0.89) vs. 0.63 (0.51–0.76), Delong's

$P = 0.0024$ for GC vs. SG/CAG; 0.84 (0.77–0.92) vs. 0.63 (0.51–0.76), Delong's $P = 0.0007$ for GC vs. IM/LGIN) and progression of gastric lesions (AUC (95% CI): 0.92 (0.83–1) vs. 0.62 (0.46–0.77), Delong's $P = 1.99 \times 10^{-10}$) (Fig. 3d–f).

DEPs at tissue level and risk of GC and gastric lesion progression

Taking advantage of the tissue proteomics dataset of our group, we found that 34 out of 43 urine DEPs as identified above were also detected in gastric tissue samples. We then compared their tissue expression between GC and non-GC tissues. Of them, the association for 21 tissue proteins went to the same direction with that at the urine level, although statistically significant findings were yielded for 13 tissue proteins only (Fig. 4). For four key urine proteins associated with the risk of gastric lesion progression and GC, we found significantly elevated level of three tissue proteins (ANXA11, CDC42 and NAPA) in GC, in line with the findings based on urine samples. These three proteins also had higher expression in gastric tissues at baseline for progressed than non-progressed subjects, although a limited sample size restricted the statistical power to yield significant associations (Table S4).

Immunocyte classification and immune subtyping based on urinary proteomic data

Urine proteomic profiles were utilized to estimate relative proportions of 10 major immune cell types for each subject, using the NUWA pipeline we developed for proteomics deconvolution analysis. Spearman correlation analysis showed the urine levels of two key proteins (CDC42, SLC25A4) and the protein score were statistically significantly associated with CD8⁺ T and NK cell proportions, but the correlation coefficients only demonstrated weak-to-moderate correlations (Fig. 5a, Table S5). Pearson correlation analysis reached similar findings (Table S5).

Based on the estimated proportions of immune cells for each subject, three immune clusters (ICs) were then derived (Fig. S3a). Statistically significant difference was found for all 10 major immune cell abundance across three immune clusters, with neutrophil having the lowest P value (Fig. S3b, Wilcoxon rank-sum test). We observed distinguished characteristics in immune cell types in three immune clusters, with IC1 featured high proportions of CD4⁺ T, and dendritic cells, IC2 featured abundant eosinophil and neutrophils, and IC3 featured abundant CD8⁺ T, NK cells and B cells (Fig. 5b, Fig. S3b). Furthermore, subjects with GC tended to be enriched in IC3 (Fig. 5c, $P = 0.013$, Wilcoxon rank-sum test) with up-regulated CD8⁺ T and NK cells (Fig. 5d), which can be further divided into six cell subsets, leading to identification of significantly up-regulated

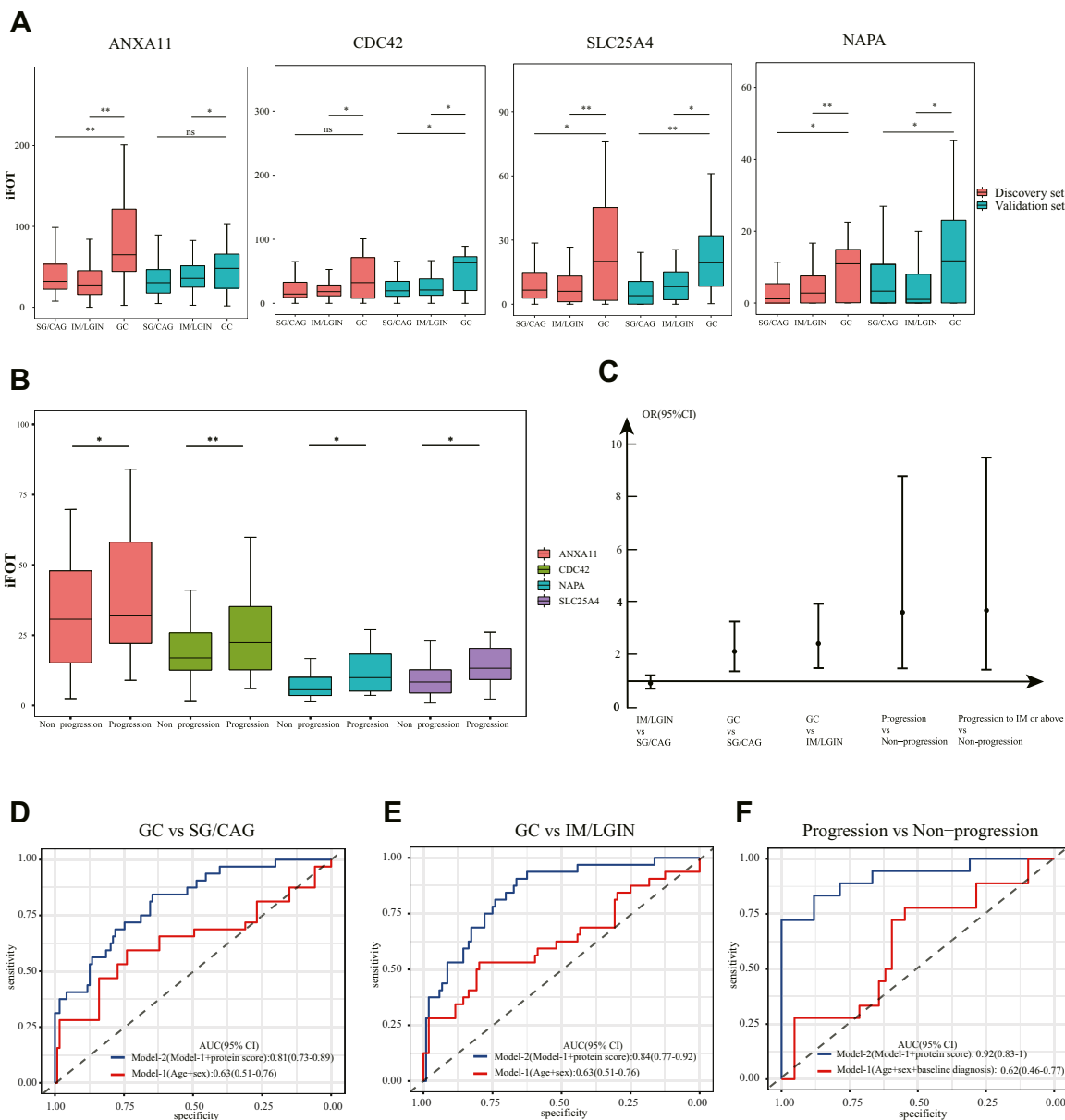


Fig. 3: Proteomic signatures associated with the progression of gastric lesion and risk of GC. **a.** Relative abundance of four highlighted proteins for subjects with SG/CAG, IM/LGIN and GC in the discovery and validation sets. **b.** Expression of four highlighted proteins at baseline for progressed (n = 18) and non-progressed (n = 42) subjects during follow-up. **c.** Associations of protein score with the risk of gastric lesion progression and GC occurrence. **d.** ROC curve of prediction model for the risk of GC occurrence compared with mild gastric lesions (SG/CAG). Model-1 includes age and sex. Model-2 includes variables in model-1 and protein score of four proteins. **e.** ROC curve of prediction model for the risk of GC occurrence compared with advanced gastric lesions (IM/LGIN). Model-1 includes age and sex. Model-2 includes variables in model-1 and protein signatures of four proteins. **f.** ROC curve of prediction model for the risk of gastric lesion progression. Model-1 includes age, sex and baseline gastric histopathology. Model-2 includes variables in model-1 and protein signatures of four proteins. AUC, area under the curve; CAG, chronic atrophic gastritis; CI, confidence interval; GC, gastric cancer; IM, intestinal metaplasia; LGIN, low-grade intraepithelial neoplasia; SG, superficial gastritis. *P < 0.05, **P < 0.01, ***P < 0.001 by logistic regression.

CD8⁺ effector memory cells re-expressing CD45RA (CD8⁺ T^{EMRA}), CD56^{dim} and CD56^{bright} NK cells in GC than subjects of precancerous gastric lesions (Fig. 5e).

Discussion

In this urine proteomics study with subjects along a spectrum ranging from premalignancy to malignant

Proteins	Discovery set			Validation set			Combined sets			Prospective follow-up of subjects						
	GC (n = 14) vs. SG/CAG (n = 51)	GC (n = 14) vs. IM/LGIN (n = 58)	GC (n = 18) vs. SG/CAG (n = 68)	GC (n = 18) vs. IM/LGIN (n = 46)	GC (n = 32) vs. SG/CAG (n = 109)	GC (n = 32) vs. IM/LGIN (n = 114)	Progression (n = 18) vs. Non-progression (n = 42)									
	OR (95% CI)	P	VIP	OR (95% CI)	P	VIP	OR (95% CI)	P	OR (95% CI)	P	OR (95% CI)	P				
ANXA11	2.32 (1.24-4.34)	0.009	1.78	2.18 (1.22-3.90)	0.008	1.98	1.29 (0.82-2.04)	0.18	1.81 (1.03-3.18)	0.042	1.60 (1.08-2.42)	0.019	1.86 (1.22-2.91)	0.004	2.22 (1.07-6.55)	0.039
CDC42	1.57 (0.87-2.82)	0.13	1.13	1.85 (1.04-3.27)	0.035	1.62	1.99 (1.21-3.28)	0.012	1.67 (1.02-2.48)	0.042	1.63 (1.11-2.46)	0.013	1.74 (1.18-2.65)	0.006	4.70 (1.74-17.34)	0.004
NAPA	2.40 (1.20-4.77)	0.013	1.15	2.32 (1.26-4.27)	0.007	1.54	1.58 (1.03-2.67)	0.046	2.18 (1.14-4.18)	0.024	1.77 (1.18-2.74)	0.007	2.33 (1.48-3.92)	0.001	1.78 (1.03-3.97)	0.045
SLC25A4	2.06 (1.10-3.87)	0.024	1.91	2.94 (1.32-6.56)	0.008	2.41	4.12 (1.98-8.60)	0.001	2.16 (1.26-3.69)	0.009	2.56 (1.63-4.35)	1.45 × 10 ⁻⁴	2.55 (1.57-4.42)	3.31 × 10 ⁻⁴	2.08 (1.04-4.51)	0.028

CAG, chronic atrophic gastritis; GC, gastric cancer; IM, intestinal metaplasia; LGIN, low-grade intraepithelial neoplasia; OR, odds ratio; SG, superficial gastritis; VIP, variable importance for the projection. ^aProteins with P < 0.05 & VIP score >1 associated with GC in the discovery set, and P < 0.05 associated with GC in validation stage and P < 0.05 associated with gastric lesion progression during prospective follow-up are shown here. Analyses were conducted using Logistic regression adjusting for age, sex, and baseline histology (for the analysis of gastric lesion progression only).

Table 1: Urine proteins significantly associated with risk of gastric cancer and gastric lesion progression during the follow-up.^a

GC, a distinct urine proteomic profile was found for GC from gastric lesions. A total of 43 urine proteins were validated for the risk of GC. Four proteins, including ANXA11, CDC42, NAPA and SLC25A4, were further positively associated with the risk of gastric lesion progression based on prospective follow-up of a subset of subjects. Among them, ANXA11, CDC42, and NAPA had higher expression in GC than non-GC tissues. Based on deconvolution analysis of urine proteomic profiles, subjects with GC were aggregated in an immune cluster of subjects characterized by abundant CD8⁺ T and NK cells.

Precision prevention and management of GC, requiring high-risk population identification and its early detection, remains quite challenging all over the world. Several blood biomarkers of GC, such as pepsinogen, gastrin-17, and *H. pylori* antibodies, have been proposed, but controversy remains either for single molecule or their combinations in the real-world settings.⁷ Development of biomarkers denoting the progression of gastric lesions to GC is highly needed, particularly in a non-invasive approach.

No past evidence on NAPA and GC is available. In contrast, there have been reports linking other three proteins with carcinogenesis and biological processes of GC. Annexin A11 (ANXA11) is a member of the annexins family involved in cell division, differentiation, apoptosis, signal transduction, and vesicle trafficking.³⁴ The down-regulation of ANXA11 was shown to inhibit the proliferation, invasion and migration of GC cells through the AKT/GSK-3β pathway.³⁵ Cell division control protein 42 (CDC42) is a member of the Rho small GTPase family which regulates the cell differentiation, progression and metastasis of cancer cells,³⁶⁻³⁸ and was found to be involved in immune escape of cancer.³⁹ Knockdown of CDC42 gene significantly inhibited the migration and invasion of GC cells.⁴⁰ SLC25A4 (ADP/ATP translocase 1) is an ATP/ADP carrier that acts as a master regulator of mitochondrial energy output by maintaining a delicate balance between ATP production and thermogenesis.⁴¹ The overexpression of SLC25A4 reflects mitochondrial dysfunction, which may play an important role in cancer development.^{42,43} SLC25A4 involves in the cGMP/PKG pathway, the activation of which is crucial to promote proliferation, metastasis, and chemoresistance of GC cells.⁴⁴ In our study, these four urine proteins stood out to be significantly associated with gastric lesion progression and risk of GC, supporting them as possible biomarkers for defining high-risk populations and early detection of GC.

In parallel with what we observed in urine samples, three intracellular proteins, including ANXA11 (located in nucleoplasm and cytosol), CDC42 (located in microtubules) and NAPA (located in cytosol) had consistently higher levels in GC than non-GC tissues. SLC25A4, one membrane protein located in mitochondria, did not present such expression pattern. Urinary proteins are

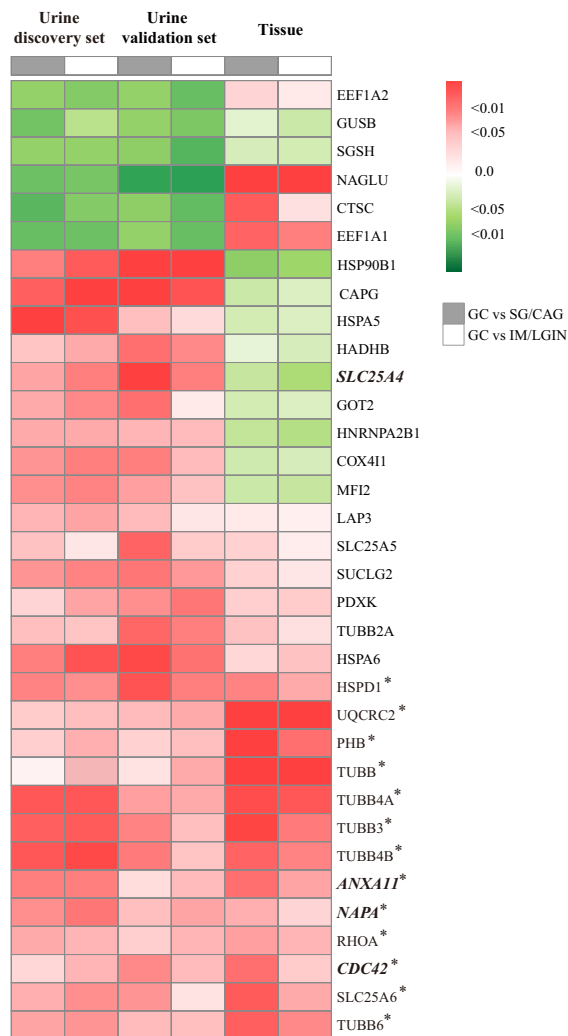


Fig. 4: Hierarchical clustering of 34 selected protein expression in urine and tissue level between GC and precancerous gastric lesion. Hierarchical analysis was conducted for the clustering of proteins. Among 43 urine DEPs significantly associated with risk of GC ($P < 0.05$ by logistic regression and VIP > 1 in both discovery and validation sets), 34 proteins were detected in gastric tissue samples. The associations for GC compared with mild (SG/CAG) and advanced (IM/LGIN) gastric lesions are shown for the urine (discovery and validation set) and tissue expression of each protein respectively. Proteins positively associated with GC (OR > 1) are shown in red and proteins inversely associated with GC (OR < 1) are shown in green. Tissue proteins significantly associated with risk of GC are labeled with *. The four proteins highlighted in the current study (ANXA11, CDC42, NAPA and SLC25A4) are italicized. CAG, chronic atrophic gastritis; GC, gastric cancer; IM, intestinal metaplasia; LGIN, low-grade intraepithelial neoplasia; SG, superficial gastritis.

primarily derived from plasma filtration and urinary tract system secretion.⁴⁵ While tissue proteins directly imply the state and function of gastric mucosa, urine proteins could well indicate global health status for cases without primary renal or urological diseases.⁴⁶ As cancer

is a systemic disease, subtle changes in protein expression during carcinogenesis process may not be observed in tissues but can be captured in urine, accommodating small and early changes that may demonstrate the systemic status of human body. Thus, it is conceivable that changes revealed in urine and tissue proteomics may provide complementary information on systemic changes during GC progression.

Indeed, none of the four recently reported GC-associated tissue proteins (DDT, HPX, PGC, and APOA1BP) were differentially expressed in urine between GC and precancerous gastric lesions. Tissue proteins undergo a series of physiological processes before entering the blood and only a small proportion of most proteins can reach the blood circulation. For example, only 1% of PGC remains in a stable form through gastric mucosal capillaries.⁴⁷ Proteins then pass through the glomerular filtration barrier, and finally get excreted in urine. The consequences of all these physiological processes may be complicated for GC cases and subjects with gastric lesions, partly helping explaining the non-differential levels of these four proteins in urine between subjects with GC and precancerous gastric lesions.

We conducted an exploratory analysis utilizing proteomic profiles to infer the global immune contexture, as urinary protein composition may be a mirror of general immunity.^{29,30} Based on the deconvolution analysis, we attempted to establish a connection between urine protein levels and the estimated immune activities and characterized an immune cluster featuring high abundance of CD8⁺ T and NK cells. Particularly, significant up-regulation of CD8⁺ T^{EMRA} cells, CD56^{dim} and CD56^{bright} NK cells was found in GC subjects. CD8⁺ T and NK cells are both cytotoxic and can kill tumor cells. Previous studies have found significantly higher proportions of NK cells in GC serums than those of healthy controls, consistent with our results in urine samples.⁴⁸ Changes in the proportion of CD8⁺ T cells in GC patients compared with gastric lesions patients or health subjects have not been reported. As CD8⁺ T^{EMRA} cells are effector memory T cells carrying the largest amount of perforin and appear only late during the immune response,⁴⁹ the estimated upregulation of CD8⁺ T^{EMRA} cells in GC may indicate an active immune surveillance within these subjects, which were possibly escaped during the progression of gastric lesions by various means. We also found that score of the highlighted proteins was moderately positively associated with CD8⁺ T and NK cell abundance, implying a potential link between highlighted proteins and immune activities. The results demonstrated that some proteins may be directly or indirectly involved in the role that immunity played in GC development,⁵⁰ which warrants further investigation of functional mechanisms.

While the current study reflected our endeavors to strengthen the precision prevention and control of GC,

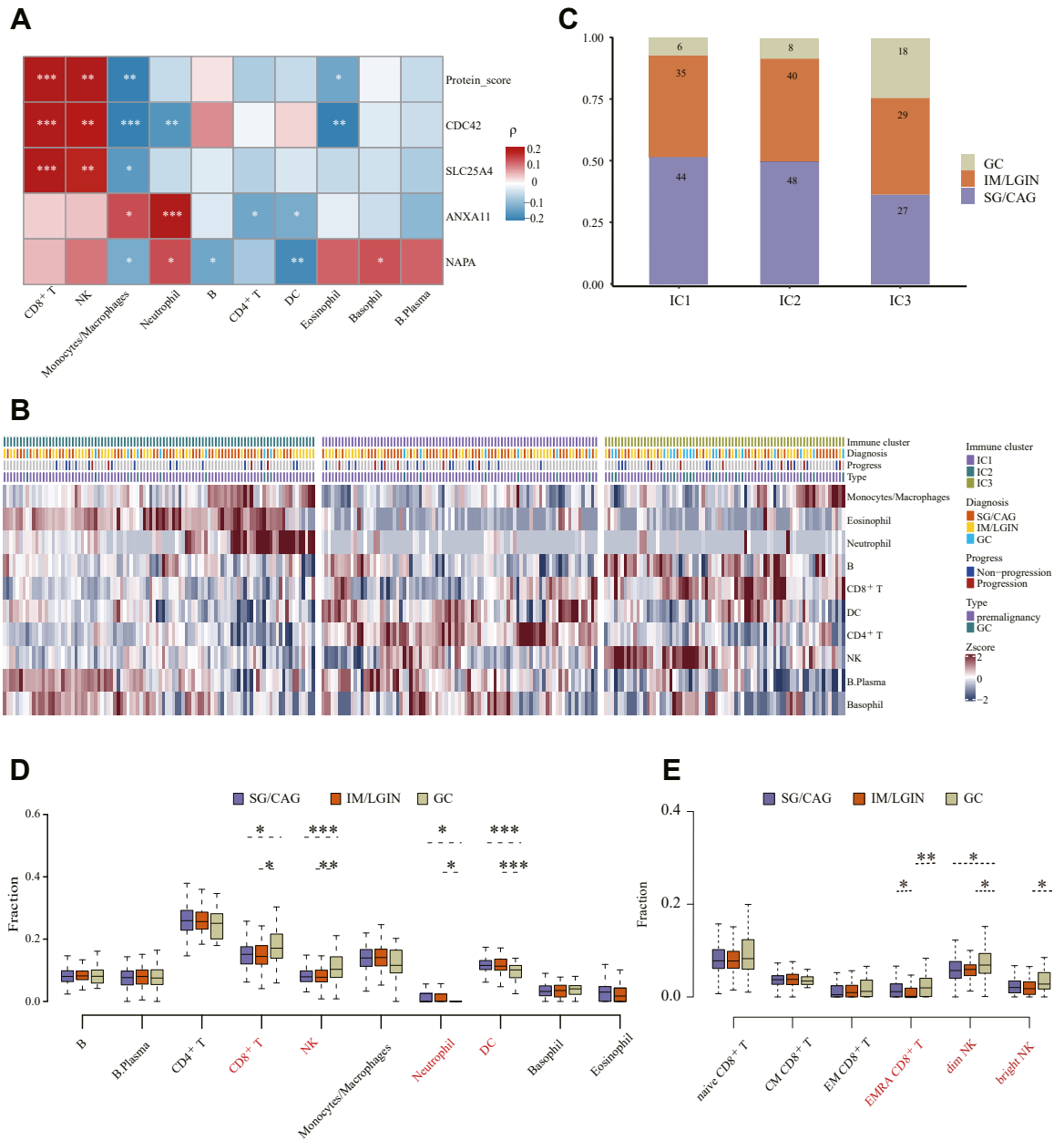


Fig. 5: Immune clustering based on urine proteomic data and association between immune cell and highlighted proteins. a. Association profiles of highlighted proteins with the abundance of immune cells as annotated by proteomics data. Red represents Spearman's correlation coefficients greater than 0, and blue represents Spearman's correlation coefficients less than 0. b. Heatmap of immune clusters identified through spherical k-means clustering. The abundance of 10 immune cells in each immune cluster is shown. c. Bar plot of different gastric lesions and GC in each immune cluster. d. The fraction of 10 immune cells for subjects with SG/CAG, IM/LGIN and GC. e. The fraction of CD8⁺ T and NK cell subsets for subjects with SG/CAG, IM/LGIN and GC; CAG, chronic atrophic gastritis; DC, dendritic cell; GC, gastric cancer; IM, intestinal metaplasia; LGIN, low-grade intraepithelial neoplasia; NK, natural killer; ns, non-significant; SG, superficial gastritis, *P < 0.05, **P < 0.01, ***P < 0.001 by Spearman's correlation analysis (panel a) or Wilcoxon rank-sum test (panels d and e).

we lacked wet lab-based explorations of mechanistic insights underlying the associations and cannot necessarily speak to cause and effect. Our primary goal was to identify non-invasive protein signatures associated with

the progression of gastric lesions and risk of early GC. Priority was therefore placed on yielding robust associations between urine protein biomarkers and gastric lesion progression. Urine is a body fluid that is excreted

after filtering by kidney, and its composition reflects the overall state of all tissues and organs of the entire body. With the current approach, it is then difficult to pinpoint which tissue/organ is the main source of changes in protein abundance. Nevertheless, we tried to integrate with the tissue proteomics data and found significantly elevated level of three tissue proteins (ANXA11, CDC42 and NAPA) in GC, in line with the findings based on urine samples. Despite so, with a proper model, mechanistic study should be a future direction in revealing the roles of these proteins in GC progression.

Our study involved in-depth urine proteomics in two stages with reasonable sample size of different gastric lesions and GC. We attempted to conduct a prospective study of urine proteomics for GC and were able to integrate the analysis of individual proteins at urine and gastric tissue samples. However, only part of the research participants had longitudinal follow-up. A large-scale prospective study with prolonged follow-up, ideally involving external multi-centers, is warranted to further validate the prediction of identified urine proteins for the progression of gastric lesions to GC. We are aware of other limitations. First, we did not collect urine samples during the endoscopic follow-up so were not able to investigate the dynamic changes of urine protein levels with the evolution of gastric lesions. Second, as the study priority was on the urine proteomic signatures along the cascade of gastric lesion progression to GC, we did not enroll GC cases at different clinical stages and were underpowered to examine the associations of highlighted proteins with therapeutic options or GC prognosis. Third, we attempted to infer immune activities based on deconvolution analysis of urine proteomic profiles in an exploratory analysis. The estimated immune profiles by deconvolution analysis could only to some extent reflect the global immune contexture, which should be directly measured by blood assays. Therefore, interpretation of the results should be cautious and further mechanistic studies are required. Fourth, the validation stage used a mixed subject set from Beijing (low-risk area) and Linqu (high-risk area). However, all GC cases in the validation stage were from Beijing and we were able to replicate key urine proteins associated with GC risk, corroborating the robustness of our findings. Despite so, further large-scale validation studies are warranted before translation of potential biomarkers in the real-world setting.

In conclusion, our in-depth urine proteomics study sheds lights on widely divergent proteomic landscape between precancerous gastric lesions and GC and established a panel of urine proteins that may potentially predict the progression of gastric lesions and risk of GC occurrence. Detection of biomarkers based on a non-invasive approach may have translational significance for fine discrimination of high-risk exposure phenotypes and concentration of high-risk population subgroups for the progression of gastric lesions to GC, advancing the

transition to precision prevention mode for GC in the future.

Contributors

W.Q.-L. supervised and designed the research; W.Q.-L., J.Q., K.-F.P. and J.W. had verified the underlying data and had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. X.L., N.-R.Z., L.H.C. and Y.W. adapted algorithms and software for data analyses; X.L., Z.-X.L., T.Z., Y.Z., W.-H.W., W.-D.L., L.-F.Z. contributed to subject recruitment and sample collection; Z.-W.L. completed histological diagnoses; M.-W.L. and K.L. carried out sample preparation and mass spectrometry analyses; H.F. wrote the draft of the manuscript; W.-Q.L., J.Q., K.-F.P., W.-C.Y. and J.W. revised the manuscript. All authors read and approved the submitted version.

Data sharing statement

The MS proteomics data have been deposited to the ProteomeXchange Consortium via the iProX partner repository and accession number will be provided after acceptance. The data is available for review purposes (please refer to <https://www.iprox.cn/page/SSV024.html?url=1662706036243NCx7>, password: uTnd, link is valid for 360 days).

Declaration of interests

W.-Q.L., J.Q., K.-F.P. and H.F. have a pending patent entitled “The identification and translation of urinary protein signatures associated with progression of premalignant to malignant gastric cancer” (Application No. 202210733363.0). The other authors have declared that no conflict of interest exists.

Acknowledgements

We sincerely thank all participants of the National Upper Gastrointestinal Cancer Early Detection program. This study was supported by grants from the National Natural Science Foundation of China (82273704), Capital's Funds for Health Improvement and Research (CFH 2020-2-1026), Michigan Medicine-PKUHSJ Joint Institute for Translational and Clinical Research (BMU2020J1004), Beijing Talents Foundation (2018000021223ZK01) and PKU-Baidu Fund (No. 2020BD034).

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.ebiom.2022.104340>.

References

- 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):209–249.
- 2 Allemani C, Matsuda T, Di Carlo V, et al. Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. *Lancet*. 2018;391(10125):1023–1075.
- 3 Kim H, Hwang Y, Sung H, et al. Effectiveness of gastric cancer screening on gastric cancer incidence and mortality in a community-based prospective cohort. *Cancer Res Treat*. 2018;50(2):582–589.
- 4 Jun JK, Choi KS, Lee HY, et al. Effectiveness of the Korean National Cancer screening program in reducing gastric cancer mortality. *Gastroenterology*. 2017;152(6):1319–1328.e7.
- 5 Chen R, Liu Y, Song G, et al. Effectiveness of one-time endoscopic screening programme in prevention of upper gastrointestinal cancer in China: a multicentre population-based cohort study. *Gut*. 2021;70(2):251–260.
- 6 Li WQ, Qin XX, Li ZX, et al. Beneficial effects of endoscopic screening on gastric cancer and optimal screening interval: a population-based study. *Endoscopy*. 2022;54(9):848–858.

- 7 Fan X, Qin X, Zhang Y, et al. Screening for gastric cancer in China: advances, challenges and visions. *Chin J Cancer Res*. 2021;33(2):168–180.
- 8 Correa P. Human gastric carcinogenesis: a multistep and multifactorial process—first American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. *Cancer Res*. 1992;52(24):6735–6740.
- 9 Di Meo A, Bartlett J, Cheng Y, Pasic MD, Yousef GM. Liquid biopsy: a step forward towards precision medicine in urologic malignancies. *Mol Cancer*. 2017;16(1):80.
- 10 Jakobsen KR, Paulsen BS, Baek R, Varming K, Sorensen BS, Jorgensen MM. Exosomal proteins as potential diagnostic markers in advanced non-small cell lung carcinoma. *J Extracell Vesicles*. 2015;4:26659.
- 11 Gao Y. Urine—an untapped goldmine for biomarker discovery? *Sci China Life Sci*. 2013;56(12):1145–1146.
- 12 Marimuthu A, O’Meally RN, Chaerkady R, et al. A comprehensive map of the human urinary proteome. *J Proteome Res*. 2011;10(6):2734–2743.
- 13 William Jr WN, Zhao X, Bianchi JJ, et al. Immune evasion in HPV (-) head and neck precancer-cancer transition is driven by an aneuploid switch involving chromosome 9p loss. *Proc Natl Acad Sci U S A*. 2021;118(19):e2022655118.
- 14 Krysan K, Tran LM, Grimes BS, et al. The immune contexture associates with the genomic landscape in lung adenomatous premalignancy. *Cancer Res*. 2019;79(19):5022–5033.
- 15 You WC, Zhao L, Chang YS, Blot WJ, Fraumeni Jr JF. Progression of precancerous gastric lesions. *Lancet*. 1995;345(8953):866–867.
- 16 Li X, Zheng NR, Wang LH, et al. Proteomic profiling identifies signatures associated with progression of precancerous gastric lesions and risk of early gastric cancer. *EBioMedicine*. 2021;74:103714.
- 17 Hsieh FY, Bloch DA, Larsen MD. A simple method of sample size calculation for linear and logistic regression. *Stat Med*. 1998;17(14):1623–1634.
- 18 Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *Am J Surg Pathol*. 1996;20(10):1161–1181.
- 19 You WC, Blot WJ, Li JY, et al. Precancerous gastric lesions in a population at high risk of stomach cancer. *Cancer Res*. 1993;53(6):1317–1321.
- 20 Zhang C, Leng W, Sun C, et al. Urine proteome profiling predicts lung cancer from control cases and other tumors. *EBioMedicine*. 2018;30:120–128.
- 21 Schwanhauss B, Busse D, Li N, et al. Corrigendum: global quantification of mammalian gene expression control. *Nature*. 2013;495(7439):126–127.
- 22 Liu Q, Ding C, Liu W, et al. In-depth proteomic characterization of endogenous nuclear receptors in mouse liver. *Mol Cell Proteomics*. 2013;12(2):473–484.
- 23 Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res*. 2017;45(D1):D353–D361.
- 24 Thrift AP, El-Serag HB. Burden of gastric cancer. *Clin Gastroenterol Hepatol*. 2020;18(3):534–542.
- 25 Guo Y, Li ZX, Zhang JY, et al. Association between lifestyle factors, vitamin and garlic supplementation, and gastric cancer outcomes: a secondary analysis of a randomized clinical trial. *JAMA Netw Open*. 2020;3(6):e206628.
- 26 Szklarczyk D, Gable AL, Lyon D, et al. STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res*. 2019;47(D1):D607–D613.
- 27 Liu P, Fu B, Yang SX, Deng L, Zhong X, Zheng H. Optimizing survival analysis of XGBoost for ties to predict disease progression of breast cancer. *IEEE Trans Biomed Eng*. 2021;68(1):148–160.
- 28 Bruzelius M, Bottai M, Sabater-Lleal M, et al. Predicting venous thrombosis in women using a combination of genetic markers and clinical risk factors. *J Thromb Haemost*. 2015;13(2):219–227.
- 29 Tian W, Zhang N, Jin R, et al. Immune suppression in the early stage of COVID-19 disease. *Nat Commun*. 2020;11(1):5859.
- 30 Bi X, Liu W, Ding X, et al. Proteomic and metabolomic profiling of urine uncovers immune responses in patients with COVID-19. *Cell Rep*. 2022;38(3):110271.
- 31 Newman AM, Liu CL, Green MR, et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat Methods*. 2015;12(5):453–457.
- 32 Rieckmann JC, Geiger R, Hornburg D, et al. Social network architecture of human immune cells unveiled by quantitative proteomics. *Nat Immunol*. 2017;18(5):583–593.
- 33 Gu Z, Schlesner M, Hubschmann D. cola: an R/Bioconductor package for consensus partitioning through a general framework. *Nucleic Acids Res*. 2021;49(3):e15.
- 34 Shibata H, Kanadome T, Sugiura H, et al. A new role for annexin A11 in the early secretory pathway via stabilizing Sec31A protein at the endoplasmic reticulum exit sites (ERES). *J Biol Chem*. 2015;290(8):4981–4993.
- 35 Hua K, Li Y, Zhao Q, Fan L, Tan B, Gu J. Downregulation of Annexin A11 (ANXA11) inhibits cell proliferation, invasion, and migration via the AKT/GSK-3beta pathway in gastric cancer. *Med Sci Monit*. 2018;24:149–160.
- 36 Zhang Y, Li J, Lai XN, Jiao XQ, Xiong JP, Xiong LX. Focus on Cdc42 in breast cancer: new insights, target therapy development and non-coding RNAs. *Cells*. 2019;8(2):146.
- 37 Maldonado MDM, Dharmawardhane S. Targeting Rac and Cdc42 GTPases in cancer. *Cancer Res*. 2018;78(12):3101–3111.
- 38 Stengel K, Zheng Y. Cdc42 in oncogenic transformation, invasion, and tumorigenesis. *Cell Signal*. 2011;23(9):1415–1423.
- 39 Marques CA, Hahnel PS, Wolfel C, et al. An immune escape screen reveals Cdc42 as regulator of cancer susceptibility to lymphocyte-mediated tumor suppression. *Blood*. 2008;111(3):1413–1419.
- 40 Du DS, Yang XZ, Wang Q, et al. Effects of CDC42 on the proliferation and invasion of gastric cancer cells. *Mol Med Rep*. 2016;13(1):550–554.
- 41 Namba T, Doczi J, Pinson A, et al. Human-specific ARHGAP11B acts in mitochondria to expand neocortical progenitors by glutaminolysis. *Neuron*. 2020;105(5):867–881.e9.
- 42 Clemencin B, Babot M, Trezeguet V. The mitochondrial ADP/ATP carrier (SLC25 family): pathological implications of its dysfunction. *Mol Aspects Med*. 2013;34(2-3):485–493.
- 43 Moreno-Sanchez R, Rodriguez-Enriquez S, Marin-Hernandez A, Saavedra E. Energy metabolism in tumor cells. *FEBS J*. 2007;274(6):1393–1418.
- 44 Xiang T, Yuan C, Guo X, et al. The novel ZEB1-upregulated protein PRTG induced by Helicobacter pylori infection promotes gastric carcinogenesis through the cGMP/PKG signaling pathway. *Cell Death Dis*. 2021;12(2):150.
- 45 Zhao M, Li M, Yang Y, et al. A comprehensive analysis and annotation of human normal urinary proteome. *Sci Rep*. 2017;7(1):3024.
- 46 Wu J, Gao Y. Physiological conditions can be reflected in human urine proteome and metabolome. *Expert Rev Proteomics*. 2015;12(6):623–636.
- 47 Shen S, Jiang J, Yuan Y. Pepsinogen C expression, regulation and its relationship with cancer. *Cancer Cell Int*. 2017;17:57.
- 48 Chen YY, Feng Y, Mao QS, et al. Diagnostic and prognostic value of the peripheral natural killer cell levels in gastric cancer. *Exp Ther Med*. 2020;20(4):3816–3822.
- 49 Sallusto F, Geginat J, Lanzavecchia A. Central memory and effector memory T cell subsets: function, generation, and maintenance. *Annu Rev Immunol*. 2004;22:745–763.
- 50 Lazar DC, Avram MF, Romosan I, Cornianu M, Taban S, Goldis A. Prognostic significance of tumor immune microenvironment and immunotherapy: novel insights and future perspectives in gastric cancer. *World J Gastroenterol*. 2018;24(32):3583–3616.