Open camera or QR reader and
scan code to access this article
and other resources online.

Urinary Proteomics for Discovery of Gastric Cancer Biomarkers to Enable Precision Clinical Oncology

Neha Joshi,[1–3] Firdous Bhat,[3] Anikha Bellad,[1,2] Gajanan Sathe,[1,2] Anu Jain,[3] Sandip Chavan,[3] Ravi Sirdeshmukh,[1,2] and Akhilesh Pandey[2–4]

Abstract

For precision in clinical oncology practice, detection of tumor-derived peptides and proteins in urine offers an attractive and noninvasive alternative for diagnostic or screening purposes. In this study, we report comparative quantitative proteomic profiling of urine samples from patients with gastric cancer and healthy controls using tandem mass tags-based multiplexed mass spectrometry approach. We identified 1504 proteins, of which 246 were differentially expressed in gastric cancer cases. Notably, ephrin A1 (EFNA1), pepsinogen A3 (PGA3), sortilin 1 (SORT1), and vitronectin (VTN) were among the upregulated proteins, which are known to play crucial roles in the progression of gastric cancer. We also found other overexpressed proteins, including shisa family member 5 (SHISA5), mucin like 1 (MUCL1), and leukocyte cell derived chemotaxin 2 (LECT2), which had not previously been linked to gastric cancer. Using a novel approach for targeted proteomics, SureQuant, we validated changes in abundance of a subset of proteins discovered in this study. We confirmed the over-expression of vitronectin and sortilin 1 in an independent set of urine samples. Altogether, this study provides molecular candidates for biomarker development in gastric cancer, and the findings also support the promise of urinary proteomics for noninvasive diagnostics and personalized/precision medicine in the oncology clinic.

Keywords: gastric cancer, proteomics, urine biomarkers, personalized medicine, cancer research, precision oncology, SureQuant

Introduction

GASTRIC CANCER is one of the salient life-threatening diseases with the annual incidence of over 1 million (Sung et al., 2021). It is the fifth most common malignancy and fourth leading cause of cancer-related deaths worldwide (Sung et al., 2021). The incidence of gastric cancer is higher in Asian countries like China, India, and Japan (Sung et al., 2021). Gastric cancer is influenced by various risk factors, including diet, age, smoking, alcohol consumption, family history of gastric cancer, exposure to cement, and mineral dust (Yusefi et al., 2018).

Detecting gastric cancer at an early stage is challenging as it is asymptomatic and often detected at the advanced stages.

Among the many approaches to diagnose gastric cancer, endoscopy with pathological interpretation from biopsy is the current gold standard detection test. It is not only invasive but also unsuitable for mass screening. In such a case, therefore, it is necessary to develop a noninvasive diagnostic test for gastric cancer, which can be used for mass screening in a larger population.

Biofluids such as serum, plasma, and gastric juice have been used so far to identify the potential biomarkers for detection of gastric cancer (Huang et al., 2021; Subbannayya et al., 2015; Zhou et al., 2020). Numerous diagnostic and prognostic biomarkers in gastric cancer have been proposed and the blood-based markers that are commonly used in clinical setup include carbohydrate antigen (CA)19-9,

[1]Institute of Bioinformatics, International Technology Park, Bangalore, India.
[2]Manipal Academy of Higher Education (MAHE), Manipal, India.
[3]Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester, Minnesota, USA.
[4]Center for Individualized Medicine, Mayo Clinic, Rochester, Minnesota, USA.

carcinoembryonic antigens, and CA72-4 (Herrera-Pariente et al., 2021; Ishigami et al., 2001; Matsuoka and Yashiro, 2018; Yin et al., 2015). Other protein biomarkers like trefoil factor 3 (TFF3), vascular endothelial growth factor (VEGF), metalloproteinase domain-containing protein 8 (ADAM8), and serum pepsinogen I and II (PGI, PGII) have demonstrated the potential for screening gastric cancer (Aikou et al., 2011; Miki, 2006; Tong et al., 2016).

Although serum and plasma have emerged as a good source for promising novel and minimally invasive markers for various diseases, they have complex proteomes and require laborious sample preparation techniques (Luque-Garcia and Neubert, 2007). In addition to serum and plasma, urine is another biospecimen that has been used to study the biomarkers for the diagnosis of diseases, including cancer (Arnaudova and Romansky, 1989; Fernandez et al., 2005; Han et al., 2005). Urine as a noninvasive source, is advantageous for disease marker discovery due to its easy accessibility and potential for repeat sampling in unlimited volumes. With its less complex protein content than serum or plasma, urine is a convenient specimen for clinical research (Kalantari et al., 2015).

Our group has previously carried out a comprehensive study on human urinary proteome and reported 1823 proteins in normal human urine (Marimuthu et al., 2011). Urinary proteomics has been performed in many disease conditions like diabetic nephropathy, kidney disorders, and urinary tract infections (Lee and Choi, 2015; Vitko et al., 2020). Furthermore, there are Food and Drug Administration (FDA)-approved urine-based biomarker assays for the diagnosis and surveillance of bladder cancer, which include nuclear matrix protein 22 (NMP22) and bladder tumor antigen assay (BTA) (Sugeeta et al., 2021). Urinary protein biomarkers have also been reported in kidney cancer, endometrial cancer, and prostate cancer, among others (Morrissey et al., 2010; Njoku et al., 2020; Schiffer et al., 2012).

Mass spectrometry has been used to identify differentially expressed proteins from tumor tissues, serum, plasma, and gastric juice in gastric cancer. Li et al. (2021) demonstrated the association of four proteins pepsinogen C, hemopexin, NAD(P)H-hydrate epimerase, and D-dopachrome tautomerase with the risk of gastric lesion progression using mass spectrometry and validated their findings based on immunohistochemistry and The Cancer Genome Atlas data. In another study, the authors reported increased expression of proteins like integrin alpha-4 (ITGA4), aromatic-L-amino-acid decarboxylase (DDC), and carnitine O-palmitoyltransferase 1 (CPT1A), which are strongly associated with the development and progression of gastric cancer and are potential diagnostic markers (Dhondrup et al., 2022).

Recently, our team employed mass spectrometry-based approach to demonstrate distinctive proteomic signatures, which included gremlin1 (GREM1), bcl-2-associated athanogene 2 (BAG2), olfactomedin 4 (OLFM4), thyroid hormone receptor-interacting protein 6 (TRIP6), and melanoma-associated antigen 9 (MAGE-A9), between diffuse and intestinal subtypes of gastric cancer and were further validated using IHC (Singh et al., 2021). Most of these studies have used tumor tissue samples and/or serum for the proteomics analysis. To the best of our knowledge, there are limited studies that have elucidated protein markers associated with gastric cancer in urine.

In this study, we carried out a mass spectrometry-based proteomic analysis of urine from patients with gastric cancer for identification of novel noninvasive biomarkers. We performed tandem mass tags (TMT) 10-plex labeling for multiplexing of urine samples from control and gastric cancer groups. This comparative analysis of urine samples revealed significant changes at the proteomic level. We validated a subset of proteins by deploying SureQuant, a recently developed targeted mass spectrometry approach. We believe that identifying urine-based proteomic signatures associated with gastric cancer will provide significant boost in screening general population for gastric cancer.

Materials and Methods

Samples

The urine samples from gastric cancer patients ($n = 5$) and apparently healthy controls ($n = 5$) were recruited from the study site at the Kidwai Memorial Institute of Oncology, Bangalore. The study was approved by the Institutional Review Board and Medical Ethics Committee of Kidwai Memorial Institute of Oncology, Bangalore, India (KMIO/MEC/021/24). A written informed consent was obtained from all individuals. The study was conducted according to the guidelines of the Declaration of Helsinki. The diagnosis of gastric cancer and the healthy control status were ascertained through patient history, basic laboratory testing, and clinical examination by a physician at Kidwai Memorial Institute of Oncology, Bangalore, India.

The discovery set included five gastric cancer patients (diagnosed with grade 3 adenocarcinoma) and five healthy control individuals with age ranging from 40 to 60 years. The validation set included 19 urine samples from gastric cancer patients (diagnosed with grade 2 or grade 3 adenocarcinoma) and 12 urine samples from healthy controls with age ranging between 30 and 60 years. The urine samples were collected from patients with gastric cancer before any chemotherapy. The demographic details of patients and controls included in the study are shown in Supplementary Table S1.

Thirty to 50 mL of urine was collected from each patient as well as healthy individuals. Samples were immediately kept on ice and transported to the laboratory for further processing. These urine samples were then centrifuged at 1500 g for 10 min at 4°C. The supernatant was filtered using 0.22 μm filter (Millipore, Billerica, MA). Thirty milliliters of filtered urine was concentrated using 3 KDa cutoff filters (Millipore, Billerica, MA). The retentate (concentrated urine) was stored immediately at −80°C until further use.

Sample preparation for quantitative proteomics

The protein concentration of concentrated urine samples was estimated using bicinchoninic acid protein estimation kit (Thermo Scientific Pierce). An aliquot of 250 μg of protein from each sample was taken for further proteomic sample preparation. Samples were then reduced using 5 mM dithiothreitol at 60°C for 30 min followed by alkylation using 10 mM iodoacetamide for 30 min in dark at room temperature. Digestion of proteins was carried out by treating samples with Promega Lys-C (Mass Spec Grade) in the ratio of 1:50 (enzyme: protein) and samples were incubated at 37°C for 4 h. This was followed by digestion with Sequencing

Grade Trypsin (Promega) in the ratio of 1:20 (enzyme: protein) and incubating the samples at 37°C overnight. Samples were then acidified by 1% formic acid (FA) and desalted using Sep-Pak C18 Cartridges (Waters). Eluted samples were dried under vacuum and then stored at −20°C until further use.

All dried samples were reconstituted in 100 μL of 100 mM TEAB (Triethyl Ammonium Bicarbonate) and the peptide amount was estimated using Pierce Quantitative Colorimetric Peptide Assay (Thermo Scientific) kit. Equal peptide amount ($\sim$100 μg) from each sample was labeled using TMT 10-plex (TMT) as per the manufacturer's instructions (Catalog no. 90110; Thermo Fisher Scientific). Briefly, the equal amounts of peptides from each sample were aliquoted and labeled separately with the TMT 10-plex labels in the ratio of 1:2 (peptide: label).

The labeling reaction mixture was incubated at room temperature for an hour and quenched using 8 μL of 5% hydroxylamine solution with incubation for 15 min at room temperature before pooling. The pooled labeled sample was dried using speed vac and fractionated into 24 fractions using basic pH reversed-phase liquid chromatography. The fractions obtained were desalted using C18 stage tips and dried under vacuum. Samples were stored at −80°C until further analysis.

Liquid chromatography-mass spectrometry (MS)/MS analysis

Proteomic data acquisition for 24 fractions of urine samples was carried out on Thermo Scientific Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher Scientific, Bremen, Germany) interfaced with Thermo Scientific EASY-nLC 1000 (Thermo Fisher Scientific). Peptides were reconstituted in 0.1% FA and loaded on to a trap column (Thermo Scientific Acclaim PepMap 100 C18 LC Column, 75 μm × 2 cm). Peptides were then resolved on an analytical column at a flow rate of 300 nL/min using an optimized linear gradient of 8–90% acetonitrile over 90 min.

Mass spectrometry analysis was carried out in a data-dependent mode from 350 to 1600 m/z range with full scans acquired using orbitrap mass analyzer at a mass resolution of 120,000 at 200 m/z. The automatic gain control (AGC) target for precursor ion acquisition was set as 4×10^5 and ion filling time set to 50 ms. Top 15 most intense precursor ions from a survey scan were selected for fragmentation using higher-energy collision dissociation with 34% normalized collision energy and detected at a mass resolution of 50,000 at m/z 200. The AGC target for mass spectrometry (MS)/MS was set as 1×10^5 and ion filling time set to 100 ms. The ions selected for fragmentation were excluded for 30 sec.

Data analysis

The MS/MS raw data were analyzed using Sequest search algorithm in Proteome Discoverer 2.1 (Thermo Fisher Scientific) against Human RefSeq protein database (Ver. 89). The workflow used for the searches included spectrum selector, SEQUEST search nodes, and peptide validator. The search parameters included carbamidomethylation at cysteine residues (+57.021 Da), TMT 10-plex (+229.163 Da) modification at N-terminus of peptide, and C-terminus of lysine as fixed modifications, while oxidation of methionine (+15.995 Da) was set as a dynamic modification.

MS and MS/MS mass tolerances were set to 10 ppm and 0.05 Da, respectively. Trypsin was specified as protease and a maximum of two missed cleavage were allowed. A false discovery rate (FDR) of 1% was applied as a cutoff value for reporting identified peptides. To determine the differentially expressed proteins in gastric cancer, we carried out two sample t-test using Perseus software at the significance level of ≤0.05. The fold change cutoff of 1.5 was applied to the proteomic data. Significantly dysregulated proteins were considered for further data analysis.

Bioinformatics analysis

To assess the functional profile of differentially expressed proteins, gene ontology and functional enrichment analysis was performed using DAVID (https://david.ncifcrf.gov/) online tool. Various biological processes and cellular components were enriched in the analysis. To evaluate the translocation of proteins, we carried out the signal peptide and transmembrane domain analysis for the dysregulated proteins using SignalP and TransMembrane prediction tools using Hidden Markov Models, respectively.

Peptide synthesis for targeted analysis

After analyzing the discovery data set, 19 potential proteins were selected for validation, which showed significant overexpression in gastric cancer in our TMT-10 plex experiment. Thirty isotope-labeled peptides were synthesized for use as internal standards (ISs) corresponding to 19 proteins for targeted quantitative analysis. These peptides were synthesized using fluorenylmethoxycarbonyl protecting group (FMOC) solid-phase technology with the following specifications: Crude purity and synthetic isotope-labeled c-terminal lysine (K) or arginine (R). The crude peptides after synthesis were dissolved in 0.1% FA in 30% (v/v) acetonitrile/water and stored at −20°C. Each heavy isotope-labeled peptide was resuspended in 0.1% FA and subjected to nano-liquid chromatography (LC)/MS-MS analysis in data-dependent mode to determine the intensity response of 30 peptides. A final heavy peptide mixture was prepared by increasing the concentration of the low-intensity heavy peptides in the mixture before injecting and spiking into the samples.

Targeted mass spectrometry of selected peptides

Targeted mass spectrometry analysis in SureQuant mode was carried out for 31 urine samples (19 gastric cancer and 12 control samples), which were processed in similar manner up to trypsin digestion, as described above for data-dependent acquisition analysis. These samples were analyzed on an Orbitrap Eclipse mass spectrometer (Thermo Scientific) coupled to UltiMate 3000 RSLC Nano LC system (Dionex).

An injection of 1 μg peptide digest of urine sample spiked in with heavy isotope-labeled peptides was loaded onto a 15 cm column (Pepsep) with a column oven temperature of 50°C. Peptides were eluted at a flow rate of 500 nL/min across a linear gradient consisting of 0.1% FA (buffer A) and 100% acetonitrile in 0.1% FA (buffer B). The gradient was as follows: 5–35% B from 4 to 40 min, 35–90% B from 40–45 min and maintained at 90% for the next 5 min, and changed to 3% B for column equilibration for 10 min.
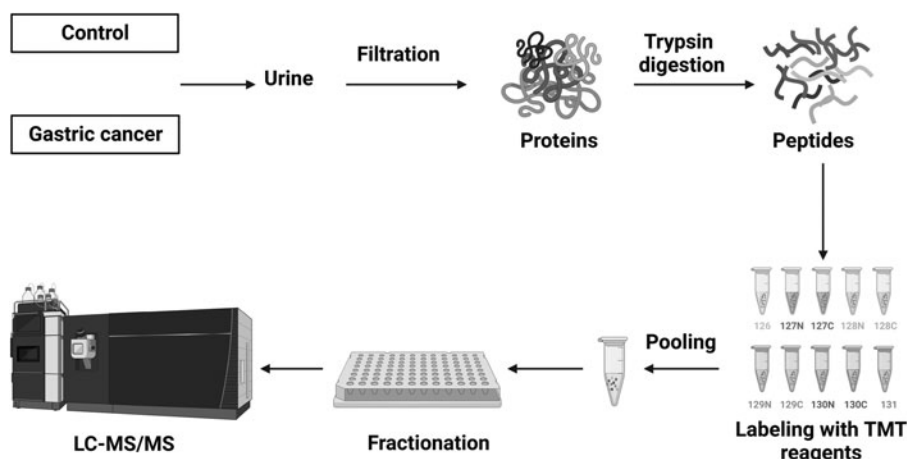
FIG. 1. Schematics of proteomics workflow deployed in the quantitative analysis of gastric cancer urinary proteins. Urine sample from control individuals ($n=5$) and gastric cancer patients ($n=5$) was centrifuged and proteins were extracted using 3 KDa centrifugal filter units. Proteins were then enzymatically digested, and peptides were labeled with TMT 10-plex, and fractions were analyzed on mass spectrometer. TMT, tandem mass tags.

Peak area of endogenous peptides and corresponding heavy IS peptides was exported from Skyline, and peptides were filtered according to the following criteria: First, only IS peptides with an integrated peak area >0 for $n \geq 5$ product ions were considered. Of these remaining tar-

gets, only endogenous targets with an integrated peak area >0 for $n \geq 3$ product ions were considered. For quantification, the peak area values of the three product ions present for both the light/heavy peptides were summed, and the ratio of endogenous to heavy IS peptide signal was
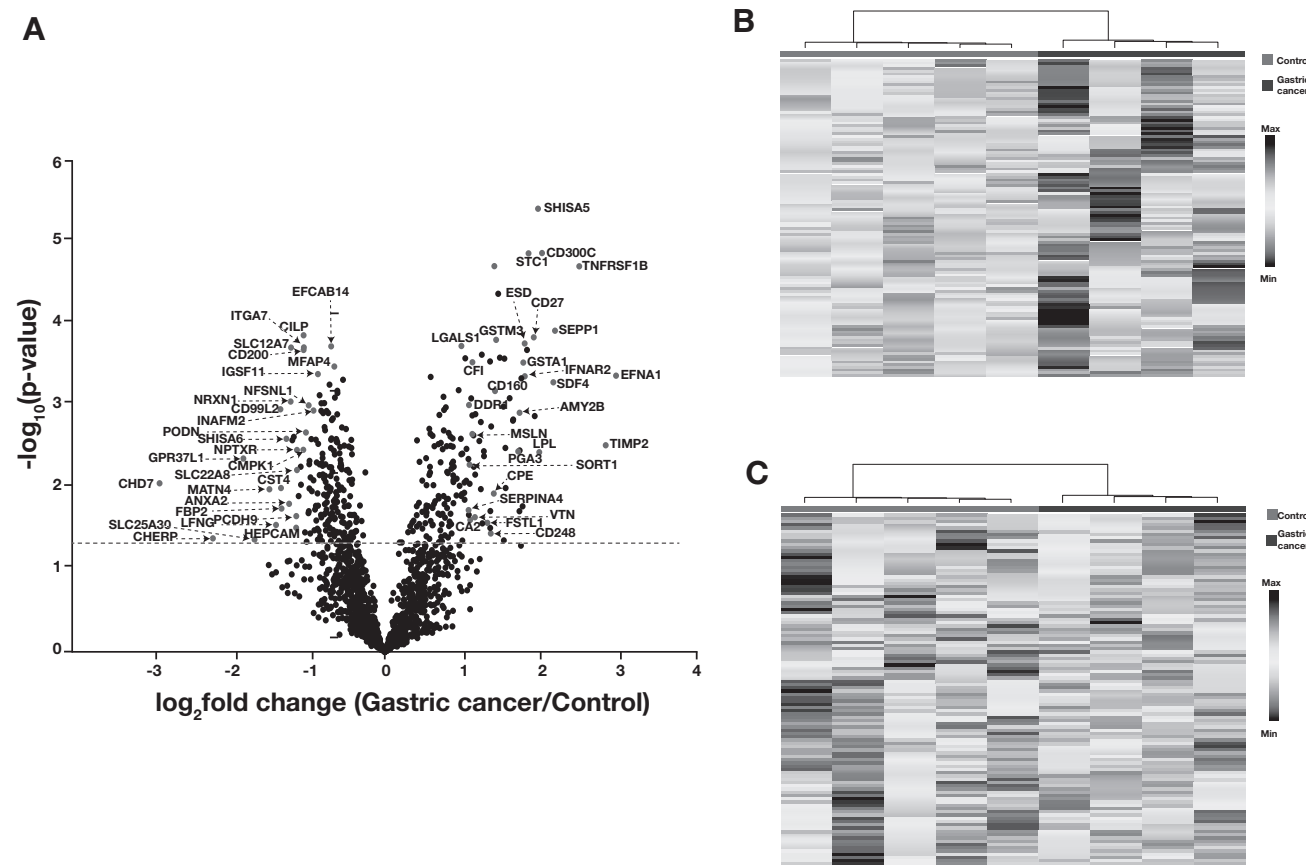

FIG. 2. Summary of proteomics data. **(A)** Distribution of proteomics data by volcano plot. Proteins annotated on the left side of the volcano plot are significantly overexpressed in gastric cancer compared to the control. Proteins downregulated in gastric cancer are annotated and present at the right side of the volcano plot. **(B)** Heat map representing the proteomic distribution for proteins that are significantly overexpressed in urine from gastric cancer patients (p-value ≤ 0.05). **(C)** Heat map representing the proteomic distribution for proteins that are significantly downregulated in urine from gastric cancer patients (p-value ≤ 0.05).

taken across samples. Student t-test was done to calculate the p-value of proteins in SureQuant analysis.

Performance assessment using receiver operating characteristic analysis

Receiver operating characteristic (ROC) curve was generated to assess the performance of the potential protein candidate(s) from SureQuant analysis. Individual and paired ROC curves were plotted and area under the curve (AUC) was calculated using the pROC package in R (Robin et al., 2011).

Data availability

The mass spectrometry proteomics data have been deposited to the Proteome Xchange Consortium using the PRIDE partner repository with the dataset identifier: PXD040956.

Results and Discussion

We carried out TMT-based quantitative proteomic profiling to compare the expression of proteins in the urine from gastric cancer patients and age- and sex-matched control individuals without a diagnosis of any cancer. The workflow deployed for proteomic analysis is illustrated in Figure 1.
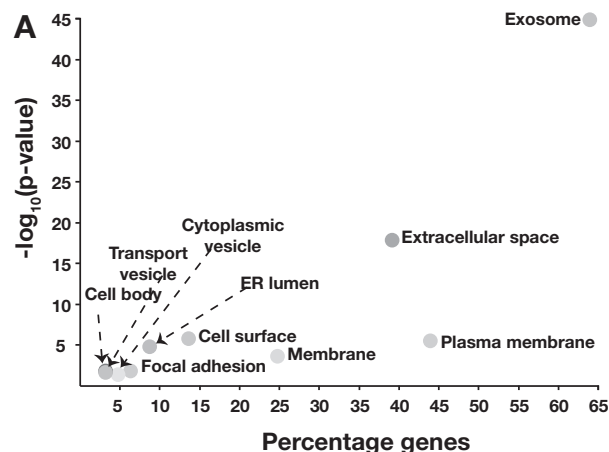
Quantitative proteomic profiling of urine samples

We performed a multiplexed quantitative proteomics analysis using TMT 10-plex labeling coupled with LC-MS/MS to profile urinary proteins in gastric cancer (Fig. 1). Urine samples were collected, centrifuged, and filtered through a $0.22 \mu m$ filter. Proteins from urine were concentrated using molecular weight cutoff filters (molecular weight cutoff 3 KDa), filtered, and then used for further sample processing. The digestion of extracted urinary proteins was carried out using two different enzymes, lys-C and trypsin, for efficient digestion. The peptides were then labeled with TMT 10-plex reagents, mixed, and fractionated into 24 fractions. Each of these 24 fractions was analyzed on a high-resolution mass spectrometer. The raw data acquired from mass spectrometry analysis were searched against the human RefSeq database_81 using Proteome Discoverer 2.1. Proteomic analysis resulted in the identification of 1504 proteins, out of which 1430 were quantified (Supplementary Table S2).

Differential expression of proteins in urine of gastric cancer patients

We carried out student's t-test to calculate the significance of proteins that were quantified in the TMT experiment. The
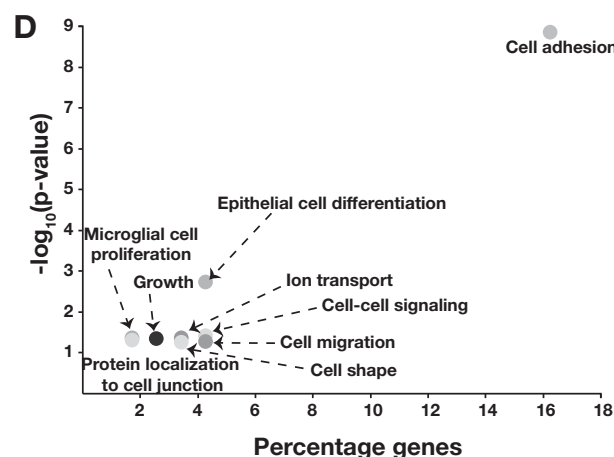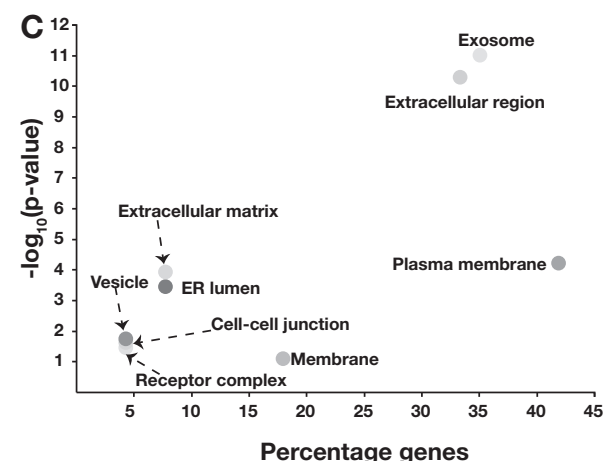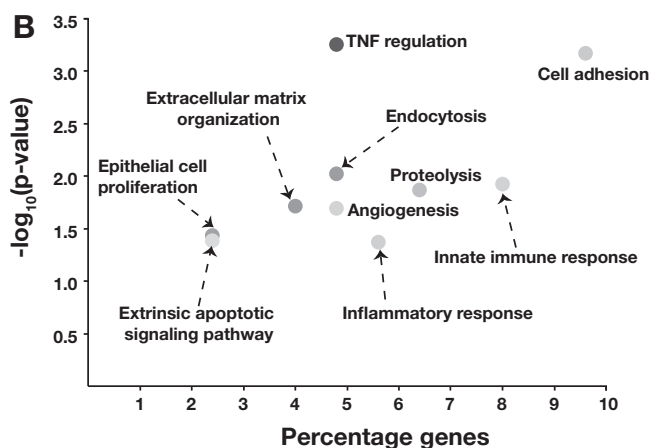

FIG. 3. Gene ontology analysis of overexpressed proteins. (**A**) Cellular components enriched for overexpressed proteins. (**B**) Biological processes enriched for overexpressed proteins. Similarly, gene ontology analysis of downregulated proteins showed (**C**) cellular components enriched for downregulated proteins and (**D**) biological processes enriched for downregulated proteins.

distribution of proteomic data is depicted as a volcano plot in Figure 2A. This analysis revealed 246 significantly dysregulated proteins (p-value ≤0.05) in gastric cancer compared to the control group. Out of these, 128 proteins were found to be significantly higher in abundance and 118 proteins to be significantly lower in abundance in gastric cancer after applying a fold change cutoff of 1.5 (Supplementary Table S3).

Proteins, including ephrin A1 (EFNA1), metalloproteinase inhibitor 2 (TIMP2), shisa family member 5 (SHISA5), pepsinogen A-3 (PGA3), vitronectin (VTN), and sortilin 1 (SORT1), were significantly overexpressed in gastric cancer patients. Similarly, proteins, including chromodomain-helicase-DNA-binding protein 7 (CHD7), podocin (PODN), shisa-6 (SHISA6), matrilin-4 (MATN4), and annexin A2 (ANXA2), were downregulated in gastric cancer. These proteins are tissue specific and were also detected in urine. Figure 2B shows a heat map for significantly upregulated proteins in gastric cancer compared to controls, while Figure 2C depicts significantly downregulated proteins in gastric cancer.

Gene ontology analysis of differentially expressed proteins

To identify the cellular components and biological functions that are impacted in gastric cancer, gene ontology analysis of dysregulated proteins was carried out using DAVID, a web-based tool (https://david.ncifcrf.gov/). As expected, extracellular matrix or extracellular space was significantly enriched (p-value 3.59×10^{-43}) for differentially expressed proteins in gastric cancer (Fig. 3A).

In addition, cellular components such as plasma membrane, cytoplasmic vesicles, and Golgi lumen were also significantly enriched for overexpressed proteins (Fig. 3A). Similarly, the downregulated proteins were localized in different cellular components such as endoplasmic reticulum, plasma membrane, vesicles along with the extracellular matrix, and cell-cell junctions (Fig. 3C). Proteins overexpressed in gastric cancer were predominantly involved in biological processes, including cell adhesion, angiogenesis, endocytosis, regulation of necrotic factors, and cell proliferation (Fig. 3B). Along with cell adhesion, many biological processes like dephosphorylation, cell differentiation, and regulation of ERK cascade were significantly enriched for the downregulated proteins (Fig. 3D).

We carried out SignalP analysis to evaluate the nature/translocation of proteins that are differentially expressed in our dataset. Signal peptides are short (usually 16–30 amino acid long), hydrophobic sorting, or recognition signals present at the N-termini of proteins and direct the protein through the classical secretory pathway (Owji et al., 2018). SignalP analysis predicted 150 proteins with N-terminal signal peptides, which could potentially be secreted out of the cell and excreted in urine (Supplementary Table S4).

For proteins spanning the membrane, we carried out TransMembrane prediction using Hidden Markov Models (TMHMM) analysis. Transmembrane helices are present in

TABLE 1. PANEL OF PROTEINS AND CORRESPONDING PEPTIDES SELECTED FOR VALIDATION IN SUREQUANT ASSAY

Gene symbol	Protein name	Fold change	Shortlisted peptide sequence
EFNA1	Ephrin-A1	7.8	WQCNRPSAK
			FTPFTLGK
SHISA5	Shisa family member 5	3.9	FVWSEER
CD27	CD27 antigen isoform X1	3.7	HYWAQGK
			HCNSGLLVR
IFNAR2	Interferon alpha/beta receptor 2 isoform A	3.5	SILSWELK
GSTA1	Glutathione S-transferase A1 isoform 1	3.4	YFPAFEK
			ISNLPTVK
PGA3	Pepsin A-3 preproprotein	3.3	QYFTVFDR
GSTM3	Glutathione S-transferase Mu 3	2.7	ITQSNAILR
			IAAYLQSDQFCK
CD248	Endosialin precursor	2.6	GASLLCVK
			VDSLVGAGPASR
CPE	Carboxypeptidase E preproprotein	2.6	NSLISYLEQIHR
MMP9	Matrix metalloproteinase 9	2.6	EYSTCTSEGR
FSTL1	Follistatin-related protein 1 precursor	2.5	ICANVFCGAGR
			LDSSEFLK
CDH16	Cadherin-16 isoform 1 precursor	2.5	LGALALSPK
MSLN	Mesothelin	2.2	LLGPHVEGLK
			TDAVLPLTVAEVQK
VTN	Vitronectin	2.2	AVRPGYPK
			FEDGVLDPDYPR
SORT1	Sortilin 1	2.1	SAPGEDEECGR
SERPINA4	Kallistatin isoform 1	2.1	VGSALFLSHNLK
			LGFTDLFSK
CFI	Complement factor I	2.1	VFSLQWGEVK
			SLECLHPGTK
CA2	Carbonic anhydrase 2	2.1	VVDVLDSIK
C7	Complement component C7 precursor	2.0	VLFYVDSEK
			LTPLYELVK

proteins, which are membrane bound. These are integral proteins that are implicated in numerous biological functions, including cell-cell communication, protein folding, transportation, and attachment. The variations in these proteins result in membrane disassembly and could lead to a disease phenotype (Ng et al., 2012). In this study, 34 proteins were predicted to have a transmembrane helix and hence could be the membrane-bound proteins (Supplementary Table S4).

Development of a targeted approach using SureQuant

To further investigate our findings in the discovery experiment, we selected a panel of proteins, which were significantly dysregulated in urine samples from gastric cancer patients. For developing a panel of candidate proteins, we considered proteins that were reported earlier to have correlation with gastric cancer prognosis and/or other types of
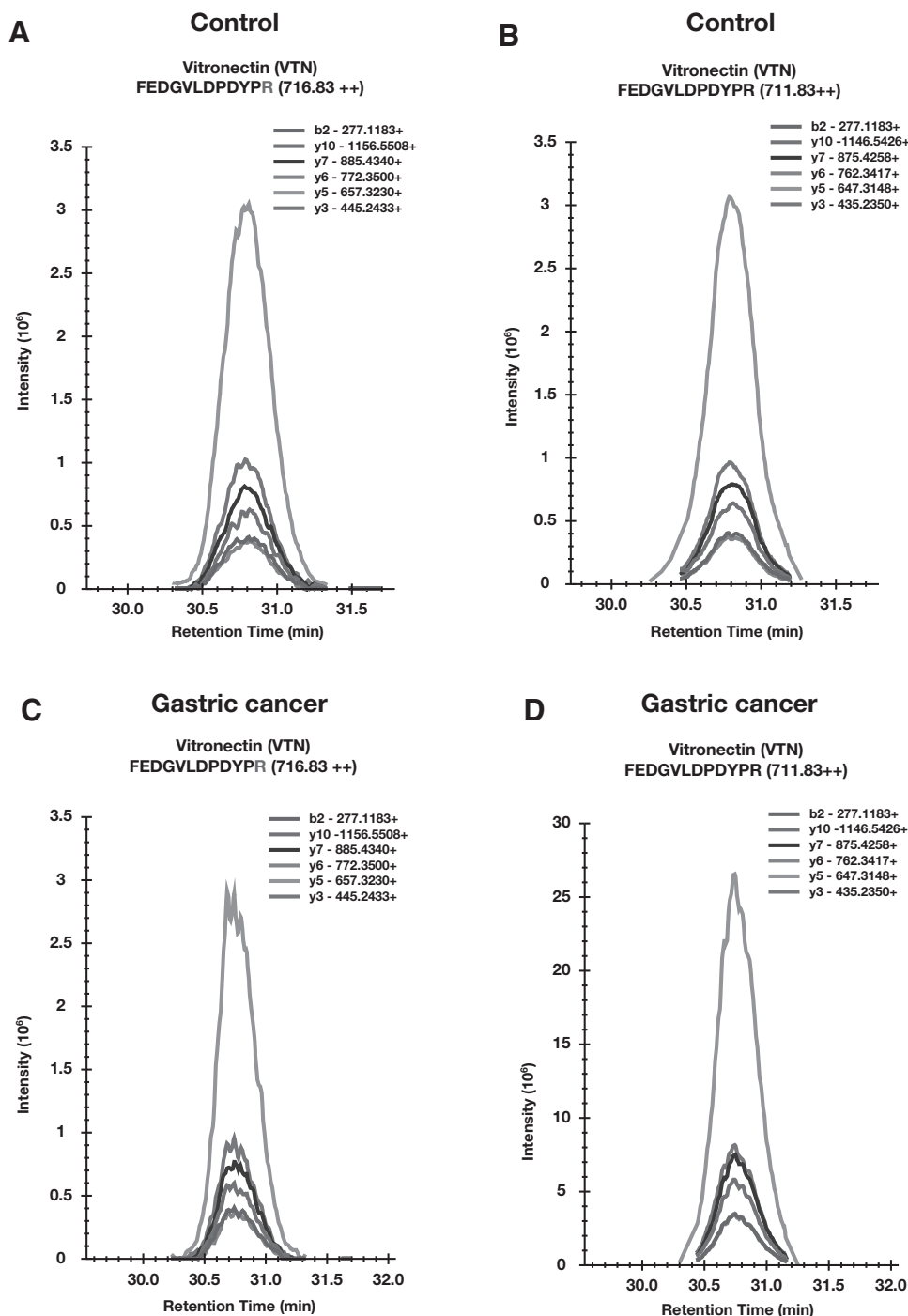

FIG. 4. Optimization of SureQuant assay in gastric cancer. **(A)** Transitions for heavy isotope-labeled peptide FEDG-VLDPDYPR (m/z 716.83) and **(B)** transitions for endogenous peptide FEDGVLDPDYPR (m/z 711.83) in urine from control individuals detected in SureQuant analysis. Similarly, **(C, D)** represents the transitions from both heavy labeled FEDGVLDPDYPR (m/z 716.83) and endogenous peptide FEDGVLDPDYPR (m/z 711.83) in urine from gastric cancer patient, respectively, detected in the SureQuant assay.

cancers. Table 1 lists the proteins and their corresponding peptides that were selected for validation. We chose a novel targeted mass spectrometry-based approach, SureQuant, for validation studies. SureQuant is a triggered targeted mass spectrometry assay used for quantification of peptides. This approach consists of two different acquisition modes—a low-resolution scan in which continuous monitoring of a heavy IS peptide is carried out and upon detection, acquisition of the endogenous peptide is triggered in which MS/MS scan at a defined mass offset is acquired at a high resolution (Stopfer et al., 2021; van Bentum and Selbach, 2021).

We deployed the SureQuant targeted assay to analyze 30 peptides from 19 proteins, which were significantly over-expressed in urine from gastric cancer patients. The criteria for selecting peptides included peptide length (8–15 amino acids), peptide uniqueness, and a peptide with significant intensity ($\sim 10^6$ or more). We considered only fully tryptic

peptides and avoided peptides with any missed cleavages. These 30 peptides were synthesized and labeled at the C-terminus with a stable heavy isotope of arginine or lysine.

For the validation set, we processed 31 urine samples, 19 from gastric cancer and 12 from the healthy control group in which IS peptides were spiked before analysis. We monitored 30 peptides for which we observed good intense signal from mass spectrometer for IS peptides. Figure 4 shows a signal pattern for SureQuant analysis for one representative peptide. The raw data were analyzed using Skyline software and peak area of IS peptide and endogenous peptides was calculated, which was used to generate the ratio for
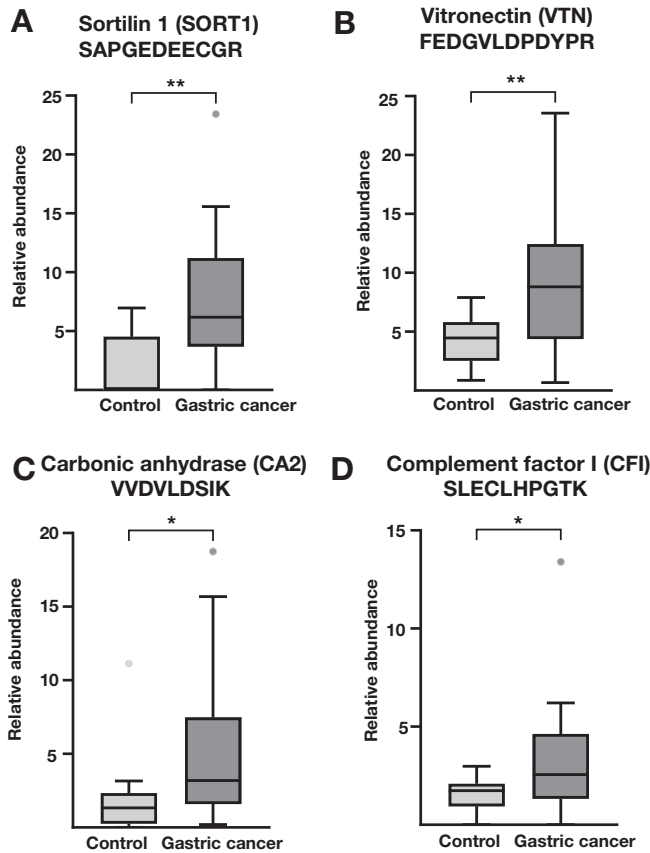

FIG. 5. Box plots for peptides detected in gastric cancer. **(A)** Ratio of endogenous to heavy peptide calculated from SureQuant data and represented for a peptide SAPGEDE-ECGR corresponding to SORT1. Ratio of endogenous to heavy peptides was calculated in urine from both control and gastric cancer individuals and plotted as a box plot for the peptides **(B)** FEDGVLDPDYPR corresponding to VTN. **(C)** Ratio of endogenous to heavy peptide calculated from SureQuant data and represented for a peptide VVDVLDSIK corresponding to CA2 protein, **(D)** ratio of endogenous to heavy peptide calculated from SureQuant data and represented for a peptide SLECLHPGTK corresponding to CFI *p*-value ranging from 0.01 to 0.05 is denoted by * and *p*-value ≤0.01 is denoted by **.
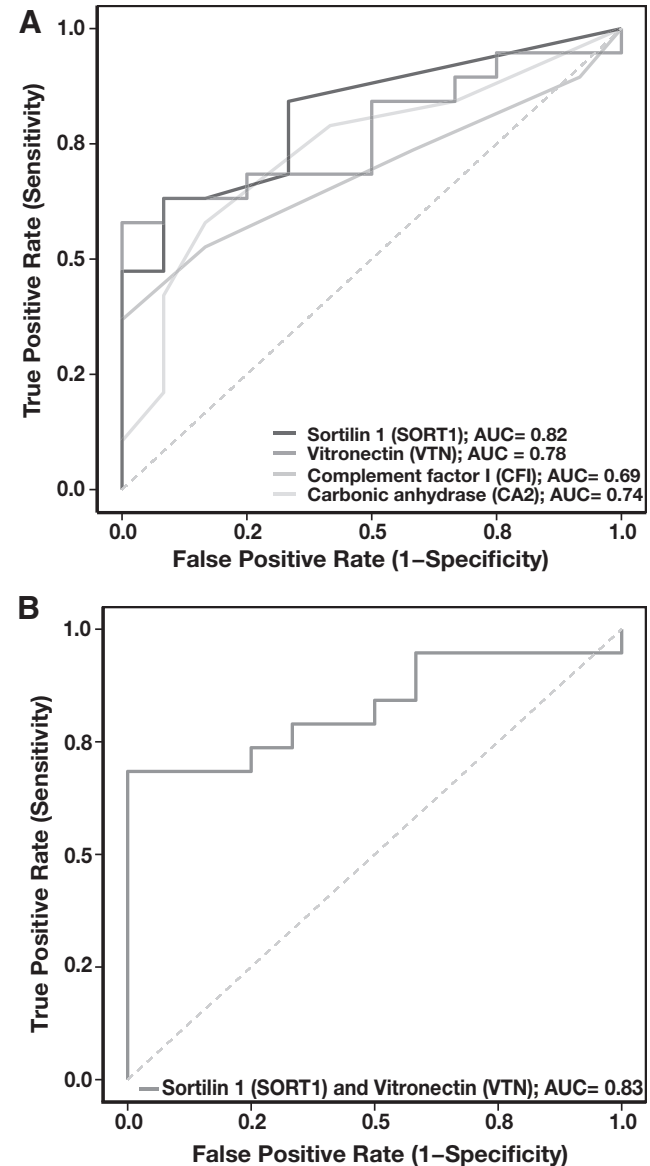

FIG. 6. ROC curve. **(A)** The AUC was calculated and ROC plots were generated for four peptides—VVDVLDSIK (CA2), SLECLHPGTK (CFI), SAPGEDEECGR (SORT1), and FEDGVLDPDYPR (VTN). **(B)** Cumulative AUC of two peptides SAPGEDEECGR (SORT1) and FEDGVLDPDYPR (VTN) was plotted in ROC. AUC, area under the curve; CA2, carbonic anhydrase 2; CFI, complement factor I; ROC, receiver operating characteristic; SORT1, sortilin 1; VTN, vitronectin.

TABLE 2. AREA UNDER THE RECEIVER OPERATING CHARACTERISTIC CURVE FOR PAIRS
OF POTENTIAL URINARY PROTEIN MARKERS FOR GASTRIC CANCER

Peptide sequence (protein name)	SLECLHPGTK (CFI)	VVDVLDSIK (CA2)	SAPGEDEECGR (SORT1)	FEDGVLDPDYPR (VTN)
SLECLHPGTK (CFI)	0.69	0.77	0.84	0.78
VVDVLDSIK (CA2)		0.74	0.83	0.78
SAPGEDEECGR (SORT1)			0.82	0.83
FEDGVLDPDYPR (VTN)				0.78

endogenous to heavy peptide. For calculating the ratio, we selected only the peptides for which endogenous peptide was identified in more than 50% of the total number of samples. We observed 20 peptides corresponding to 12 proteins overexpressed in gastric cancer samples compared to control samples correlating with our discovery proteomic analysis.

The panel of 12 proteins include SORT1, EFNA1, kallistatin isoform 1 (SERPINA4), complement factor I (CFI), complement component C7 precursor (C7), CD27 antigen isoform X1 (CD27), glutathione S-transferase A1 isoform 1 (GSTA1), glutathione S-transferase Mu 3 (GSTM3), interferon alpha/beta receptor 2 isoform A (IFNAR2), carbonic anhydrase 2 (CA2), mesothelin (MSLN), and VTN. Student's *t*-test was performed to examine the difference in the abundance of proteins in the two groups. Box plots were generated for four proteins, which were significantly different in their expression in gastric cancer samples (Fig. 5).

These proteins have been reported earlier in gastric or other cancers. Sortilin 1 (SORT1) functions as a sorting receptor in Golgi apparatus as well as acting as a clearance receptor on cell membrane. SORT1 is required for the protein transport from Golgi apparatus to lysosomes and in the transport of proteins from the Golgi body to endosome. Sortilin is a developmental protein involved in differentiation, endocytosis, and osteogenesis along with the protein transport. Upregulation of SORT1 by circular RNA was shown to promote the progression of gastric cancer (Liang et al., 2021). Overexpression of SORT1 in urine from gastric cancer patient was observed in the validation (Fig. 5A).

VTN is a cell adhesion protein and differential expression of this protein can promote cell migration. Lian et al. (2019) showed overexpression of VTN by immunohistochemistry in gastric cancer tissue samples, suggesting its probable role in regulating cell growth and motility. In another study, the elevated level of VTN was identified in cancer-associated fibroblasts in gastric cancer, leading to metastasis (Yang et al., 2020; Zhang et al., 2022). Our data also showed the overexpression of VTN in gastric cancer samples (Fig. 5B).

CA2, an isoform of carbonic anhydrase, was overexpressed in this study (Fig. 5C). CA2 has been reported to be overexpressed in urothelial carcinomas and associated with tumor grade, invasiveness, and progression of cancer (Tachibana et al., 2017). In gastrointestinal stromal tumors, CA2 has been shown to be overexpressed using immunohistochemistry and western blotting (Parkkila et al., 2010).

CFI is part of complement system, which is involved in body's immune system. Changes in the expression of proteins that are part of complement system have been shown to be involved in the development and progression of cancers

(Revel et al., 2020). Overexpression of CFI in cutaneous squamous carcinoma cells promotes proliferation and upregulates expression of matrix metallopeptidase 2 and 13, which stimulate invasion of these cells (Rahmati Nezhad et al., 2021). It has also been shown to correlate with poor prognosis and overall survival in glioma patients (Cai et al., 2020). Our data also revealed overexpression of CFI in gastric cancer (Fig. 5D). However, Liu et al. (2007) reported downregulation of CFI in gastric cancer to be associated with TNM staging. This needs to be further evaluated to confirm the role of CFI in the development of gastric cancer.

The panel of four proteins—SORT1, VTN, CA2, and CFI—was assessed for their specificity to distinguish gastric cancer from control samples. ROC analysis for these candidates revealed the peptide SAPGEDEECGR derived from SORT1 to be the best performing peptide with AUC of 0.82. Peptides, including FEDGVLDPDPYPR from VTN and VVDVLDSIK from CA2, showed suboptimal performance individually with AUC of 0.78 and 0.74, respectively (Fig. 6A). However, paired ROC analysis of SAPGEDEECGR from SORT1 and FEDGVLDPDPYPR from VTN further showed slight improvement in the performance with an AUC of 0.83 (Fig. 6B). Table 2 represents the AUC calculated for each peptide individually as well as in pairs.

Conclusions

Urine is a commonly used noninvasive biospecimen for diagnostic and prognostic assays because it has proteins derived from the systemic circulation as well as the excretory system. Mass spectrometry-based proteomics of urine provides vital information not only for renal disorders but also for identification of biomarkers for a variety of other diseases. Our discovery studies led us to identification of a panel of candidate proteins, which were subsequently validated by deploying SureQuant, a novel targeted mass spectrometry method. This study demonstrates the potential of a urinary proteomics approach for the detection of gastric cancer and could be extended to other cancers as well.

Authors' Contributions

Conceptualization, N.J., G.S., R.S., and A.P.; data curation, N.J. and F.B.; formal analysis, N.J., F.B., and A.B.; funding acquisition, A.P.; investigation, N.J., F.B., and G.S.;

methodology, N.J., G.S., and A.P.; project administration, A.P.; resources, N.J., G.S., and A.P.; software, N.J., F.B., and S.C; supervision, R.S. and A.P.; validation, N.J., F.B., A.B., and S.C.; visualization, N.J. and A.P.; writing—original draft, N.J.; writing—review and editing, N.J., F.B., A.B., G.S., S.C., A.J., R.S., and A.P. All authors have read and agreed to the published version of the article.

Author Disclosure Statement

Funding Information

Supplementary Material

Supplementary Table S1
Supplementary Table S2
Supplementary Table S3
Supplementary Table S4

References

Aikou S, Ohmoto Y, Gunji T, et al. Tests for serum levels of trefoil factor family proteins can improve gastric cancer screening. Gastroenterology 2011;141(3):837–845.e831–e837; doi: 10.1053/j.gastro.2011.05.040

Arnaudova V, Romansky K. Postoperative hematomas following 505 operations for intracranial meningiomas. Zentralbl Neurochir 1989;50(2):99–100.

Cai X, Qiu W, Qian M, et al. A candidate prognostic biomarker complement factor I promotes malignant progression in glioma. Front Cell Dev Biol 2020;8:615970; doi: 10.3389/fcell.2020.615970

Dhondrup R, Zhang X, Feng X, et al. Proteomic analysis reveals molecular differences in the development of gastric cancer. Evid Based Complement Alternat Med 2022;2022:8266544; doi: 10.1155/2022/8266544

Fernandez CA, Yan L, Louis G, et al. The matrix metalloproteinase-9/neutrophil gelatinase-associated lipocalin complex plays a role in breast tumor growth and is present in the urine of breast cancer patients. Clin Cancer Res 2005;11(15):5390–5395; doi: 10.1158/1078-0432.CCR-04-2391

Han WK, Alinani A, Wu CL, et al. Human kidney injury molecule-1 is a tissue and urinary tumor marker of renal cell carcinoma. J Am Soc Nephrol 2005;16(4):1126–1134; doi: 10.1681/ASN.2004070530

Herrera-Pariente C, Montori S, Llach J, et al. Biomarkers for gastric cancer screening and early diagnosis. Biomedicines 2021;9(10):1448; doi: 10.3390/biomedicines9101448

Huang ZB, Zhang HT, Yu B, et al. Cell-free DNA as a liquid biopsy for early detection of gastric cancer. Oncol Lett 2021;21(1):3; doi: 10.3892/ol.2020.12264

Ishigami S, Natsugoe S, Hokita S, et al. Clinical importance of preoperative carcinoembryonic antigen and carbohydrate antigen 19-9 levels in gastric cancer. J Clin Gastroenterol 2001;32(1):41–44; doi: 10.1097/00004836-200101000-00010

Kalantari S, Jafari A, Moradpoor R, et al. Human urine proteomics: Analytical techniques and clinical applications in renal diseases. Int J Proteomics 2015;2015:782798; doi: 10.1155/2015/782798

Lee SY, Choi ME. Urinary biomarkers for early diabetic nephropathy: Beyond albuminuria. Pediatr Nephrol 2015;30(7):1063–1075; doi: 10.1007/s00467-014-2888-2

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; doi: 10.1016/j.ebiom.2021.103714

Lian L, Li XL, Xu MD, et al. VEGFR2 promotes tumorigenesis and metastasis in a pro-angiogenic-independent way in gastric cancer. BMC Cancer 2019;19(1):183; doi: 10.1186/s12885-019-5322-0

Liang M, Yao W, Shi B, et al. Circular RNA hsa_circ_0110389 promotes gastric cancer progression through upregulating SORT1 via sponging miR-127-5p and miR-136-5p. Cell Death Dis 2021;12(7):639; doi: 10.1038/s41419-021-03903-5

Liu W, Liu B, Xin L, et al. Down-regulated expression of complement factor I: A potential suppressive protein for gastric cancer identified by serum proteome analysis. Clin Chim Acta 2007;377(1–2):119–126; doi: 10.1016/j.cca.2006.09.005

Luque-Garcia JL, Neubert TA. Sample preparation for serum/plasma profiling and biomarker identification by mass spectrometry. J Chromatogr A 2007;1153(1–2):259–276; doi: 10.1016/j.chroma.2006.11.054

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; doi: 10.1021/pr2003038

Matsuoka T, Yashiro M. Biomarkers of gastric cancer: Current topics and future perspective. World J Gastroenterol 2018;24(26):2818–2832; doi: 10.3748/wjg.v24.i26.2818

Miki K. Gastric cancer screening using the serum pepsinogen test method. Gastric Cancer 2006;9(4):245–253; doi: 10.1007/s10120-006-0397-0

Morrissey JJ, London AN, Luo J, et al. Urinary biomarkers for the early diagnosis of kidney cancer. Mayo Clin Proc 2010;85(5):413–421; doi: 10.4065/mcp.2009.0709

Ng DP, Poulsen BE, Deber CM. Membrane protein misassembly in disease. Biochim Biophys Acta 2012;1818(4):1115–1122; doi: 10.1016/j.bbamem.2011.07.046

Njoku K, Chiasserini D, Jones ER, et al. Urinary biomarkers and their potential for the non-invasive detection of endometrial cancer. Front Oncol 2020;10:559016; doi: 10.3389/fonc.2020.559016

Owji H, Nezafat N, Negahdaripour M, et al. A comprehensive review of signal peptides: Structure, roles, and applications. Eur J Cell Biol 2018;97(6):422–441; doi: 10.1016/j.ejcb.2018.06.003

Parkkila S, Lasota J, Fletcher JA, et al. Carbonic anhydrase II. A novel biomarker for gastrointestinal stromal tumors. Mod Pathol 2010;23(5):743–750; doi: 10.1038/modpathol.2009.189

Rahmati Nezhad P, Riihila P, Piipponen M, et al. Complement factor I upregulates expression of matrix metalloproteinase-13 and -2 and promotes invasion of cutaneous squamous carcinoma cells. Exp Dermatol 2021;30(11):1631–1641; doi: 10.1111/exd.14349

Revel M, Daugan MV, Sautes-Fridman C, et al. Complement system: Promoter or suppressor of cancer progression? Antibodies (Basel) 2020;9(4):57; doi: 10.3390/antib9040057

Robin X, Turck N, Hainard A, et al. pROC: An open-source package for R and S+ to analyze and compare ROC curves. BMC Bioinform 2011;12:77; doi: 10.1186/1471-2105-12-77

Schiffer E, Bick C, Grizelj B, et al. Urinary proteome analysis for prostate cancer diagnosis: Cost-effective application in routine clinical practice in Germany. Int J Urol 2012;19(2):118–125; doi: 10.1111/j.1442-2042.2011.02901.x

Singh S, Bhat MY, Sathe G, et al. Proteomic signatures of diffuse and intestinal subtypes of gastric cancer. Cancers (Basel) 2021;13(23):5930; doi: 10.3390/cancers13235930

Stopfer LE, Flower CT, Gajadhar AS, et al. High-density, targeted monitoring of tyrosine phosphorylation reveals activated signaling networks in human tumors. Cancer Res 2021;81(9):2495–2509; doi: 10.1158/0008-5472.CAN-20-3804

Subbannayya Y, Mir SA, Renuse S, et al. Identification of differentially expressed serum proteins in gastric adenocarcinoma. J Proteomics 2015;127(Pt A):80–88; doi: 10.1016/j.jprot.2015.04.021

Sugeeta SS, Sharma A, Ng K, et al. Biomarkers in bladder cancer surveillance. Front Surg 2021;8:735868; doi: 10.3389/fsurg.2021.735868

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; doi: 10.3322/caac.21660

Tachibana H, Gi M, Kato M, et al. Carbonic anhydrase 2 is a novel invasion-associated factor in urinary bladder cancers. Cancer Sci 2017;108(3):331–337; doi: 10.1111/cas.13143

Tong W, Ye F, He L, et al. Serum biomarker panels for diagnosis of gastric cancer. Onco Targets Ther 2016;9:2455–2463; doi: 10.2147/OTT.S86139

van Bentum M, Selbach M. An introduction to advanced targeted acquisition methods. Mol Cell Proteomics 2021;20:100165; doi: 10.1016/j.mcpro.2021.100165

Vitko D, Cho PS, Kostel SA, et al. Characterizing patients with recurrent urinary tract infections in vesicoureteral reflux: A pilot study of the urinary proteome. Mol Cell Proteomics 2020;19(3):456–466; doi: 10.1074/mcp.RA119.001873

Yang X, Chen L, Mao Y, et al. Progressive and prognostic performance of an extracellular matrix-receptor interaction signature in gastric cancer. Dis Markers 2020;2020:8816070; doi: 10.1155/2020/8816070

Yin LK, Sun XQ, Mou DZ. Value of combined detection of serum CEA, CA72-4, CA19-9 and TSGF in the diagnosis of gastric cancer. Asian Pac J Cancer Prev 2015;16(9):3867–3870; doi: 10.7314/apjcp.2015.16.9.3867

Yusefi AR, Bagheri Lankarani K, Bastani P, et al. Risk factors for gastric cancer: A systematic review. Asian Pac J Cancer Prev 2018;19(3):591–603; doi: 10.22034/APJCP.2018.19.3.591

Zhang J, Zhang N, Fu X, et al. Bioinformatic analysis of cancer-associated fibroblast related gene signature as a predictive model in clinical outcomes and immune characteristics of gastric cancer. Ann Transl Med 2022;10(12):698; doi: 10.21037/atm-22-2810

Zhou B, Zhou Z, Chen Y, et al. Plasma proteomics-based identification of novel biomarkers in early gastric cancer. Clin Biochem 2020;76:5–10; doi: 10.1016/j.clinbiochem.2019.11.001

Address correspondence to:
Akhilesh Pandey, MD, PhD
Department of Laboratory Medicine and Pathology
Mayo Clinic
200 First Street SW
Rochester, MN 55905
USA

E-mail: pandey.akhilesh@mayo.edu

Abbreviations Used

AGC = automatic gain control
ANXA2 = annexin A2
AUC = area under the curve
BAG2 = bcl-2-associated athanogene 2
BCA = bicinchoninic acid
BTA = bladder tumor antigen assay
C7 = complement component C7 precursor
CA = carbohydrate antigen
CA2 = carbonic anhydrase 2
CD27 = CD27 antigen isoform X1
CFI = complement factor I
CHD7 = chromodomain-helicase-DNA-binding protein 7
CPT1A = Carnitine O-palmitoyltransferase 1
DDC = aromatic-L-amino-acid decarboxylase
DTT = dithiothreitol
EFNA1 = ephrin A1
FA = formic acid
GREM1 = gremlin1
GSTA1 = glutathione S-transferase A1 isoform 1
GSTM3 = glutathione S-transferase Mu 3
IAA = iodoacetamide
IFNAR2 = interferon alpha/beta receptor 2 isoform A
IS = internal standard
ITGA4 = integrin alpha-4
LECT2 = leukocyte cell-derived chemotaxin 2
MAGE-A9 = melanoma-associated antigen 9
MATN4 = matrilin-4
MSLN = mesothelin
MUCL1 = mucin like 1
NMP22 = nuclear matrix protein 22
OLFM4 = olfactomedin 4
PGI, PGII = serum pepsinogen I and II
PODN = Podocan
ROC = receiver operating characteristic
SERPINA4 = kallistatin isoform 1
SHISA5 = shisa family member 5
SHISA6 = shisa family member 6
SORT1 = sortilin 1
TFF3 = trefoil factor 3
TIMP2 = metalloproteinase inhibitor 2
TMT = tandem mass tags
TRIP6 = thyroid hormone receptor interacting protein 6
VEGF = vascular endothelial growth factor
VTN = vitronectin