

Proteomic Profiling of Gastric Signet Ring Cell Carcinoma Tissues Reveals Characteristic Changes of the Complement Cascade Pathway

Authors

Yang Fan, Bin Bai, Yuting Liang, Yan Ren, Yanxia Liu, Fenli Zhou, Xiaomin Lou, Jin Zi, Guixue Hou, Fei Chen, Qingchuan Zhao, and Siqi Liu

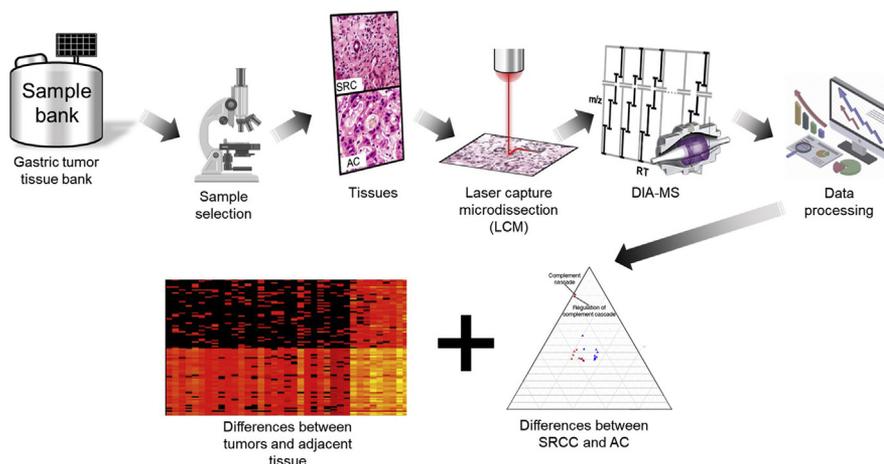
Correspondence

zhaoqc@fmmu.edu.cn;
siqiliu@genomics.cn

In Brief

This study took advantages of LCM and DIA-MS, generating a data set in the context of different subtypes of gastric tumors, globally and precisely. It was discovered for the first time that the complement cascade in SRCC tumors was specifically activated compared with AC.

Graphical Abstract



Highlights

- LCM-DIA extracted unprecedented proteomic details of gastric in different subtypes.
- Complement cascade was found to be an SRCC-specific pathway for the first time.
- Gastric cell lines were evaluated based on proteomic features for the first time.
- Re-analyzable DIA data collected provide rich opportunity for future research.

Proteomic Profiling of Gastric Signet Ring Cell Carcinoma Tissues Reveals Characteristic Changes of the Complement Cascade Pathway

Yang Fan^{1,2,3} , Bin Bai⁴, Yuting Liang^{2,3}, Yan Ren³, Yanxia Liu^{1,2}, Fenli Zhou⁴, Xiaomin Lou¹, Jin Zi³, Guixue Hou³, Fei Chen^{1,2}, Qingchuan Zhao^{4,*}, and Siqi Liu^{1,2,3,*}

Signet ring cell carcinoma (SRCC) is a histological subtype of gastric cancer with distinct features in multiple aspects compared with adenocarcinomas (ACs). The lack of a systematic molecular overview of this disease has led to slow progress in its clinical practice. In the present proteomics study, gastric tissues were collected from tumors and adjacent tissues, including 14 SRCCs and 34 ACs, and laser capture microdissection (LCM) was employed to eradicate the cellular heterogeneity of the tissues. The proteomes of tissues were profiled by data-independent acquisition (DIA) mass spectrometry (MS). Based on the over 6000 proteins quantified, univariate analysis and pathway enrichment revealed that some proteins and pathways demonstrated differences between SRCC and ACs. Importantly, the upregulation of a majority of complement-related proteins was notable for SRCC but not for ACs. A hypothesis, based on the proteomics evidence, was proposed that the complement cascade was evoked in the SRCC microenvironment upon infiltration, and the SRCC cells survived the complement cytotoxicity by secreting endogenous negative regulators. Moreover, an attempt was made to establish appropriate cell models for gastric SRCC through proteomic comparison of the 15 gastric cell lines and gastric tumors. The predictions of a supervised classifier suggested that none of these gastric cell lines qualified to mimic SRCC. This study discovered that the complement cascade is activated at a higher level in gastric SRCC than in ACs.

Gastric signet ring cell carcinoma (SRCC) is a histological subtype of gastric cancer defined by the World Health Organization (WHO) as gastric tumors composed predominantly or exclusively of signet-ring cells, which are characterized by a central optically clear globoid droplet of cytoplasmic mucin with an eccentrically placed nucleus (1). As opposed to the trend of a decreasing incidence of gastric cancer worldwide,

the SRCC incidence has continued to rise (2). The molecular features of the pathology and pharmacology relevant to SRCC are highly attractive in the frontier of studying gastric cancer.

Gastric SRCC is not only special in its histology but is also different in its clinicopathological features from other subtypes of gastric cancer. The female incidence of SRCC among all gastric cancers is approximately 50%, whereas that of non-SRCC is approximately 30%; the average incidence age of SRCC is approximately 62 years, whereas that of non-SRCC is roughly 69 years (3). Comparing SRCC to the other two main subtypes of gastric cancer, well-moderately differentiated adenocarcinoma (WMDAC) and poorly differentiated adenocarcinoma (PDAC), Chon *et al.* (4) observed that at the early stage the prognosis of SRCC was better than those of WMDAC and PDAC, whereas at later stage, the SRCC prognosis was worse than those of the other two subtypes.

During the last decade, a number of studies investigated the molecular indicators of gastric SRCC. Immunostaining revealed that all gastric SRCC and mucinous adenocarcinomas had a high abundance of trafficking kinesin protein 1 (5). The RT-PCR and IHC evidence demonstrated that the expression product of FOXP3 was significantly upregulated in gastric cancer, especially the corresponding abundance was higher in SRCC than in adenocarcinomas (79.3% versus 0%) (6). Similar to these observations, several proteins such as pyruvate kinase M1/2, glypican-3, cathepsin E, and transmembrane protein 207 were found to have abundance changes in the gastric SRCC cells or tissues. These studies investigating the SRCC-related proteins are still in the preliminary phase and are far from clinical practice. Most of those proteins were individually divulged through different approaches and laboratories and were not commonly verified. Proteomics, as a powerful means for the identification and

From the ¹CAS Key Laboratory of Genome Sciences and Information, Beijing Institute of Genomics, Chinese Academy of Sciences, Beijing, China; ²College of Life Sciences, University of the Chinese Academy of Sciences, Beijing, China; ³Clinical Laboratory of BGI Health, BGI-Shenzhen, Shenzhen, China; ⁴State Key Laboratory of Cancer Biology & Department of Surgery, Xijing Hospital of Digestive Diseases, Fourth Military Medical University, Xi'an, China

*For correspondence: Qingchuan Zhao, zhaorc@fmmu.edu.cn; Siqi Liu, sqiliu@genomics.cn.

quantification of proteins, has naturally become a main technique for the exploration of the SRCC-related proteins.

Proteomic investigation on gastric SRCC is still limited by a slow pace. There are only three published papers thus far that discussed SRCC using proteomics, but they did not reach any significant conclusion to help in the understanding of the molecular features of SRCC (7–9). What barrier hinders the relevant studies on gastric SRCC? Three factors at least, according to our view, indeed affect the study of SRCC. First, how to obtain a reasonable cohort of the SRCC samples is an obvious limitation in this area. All of the studies in the published literature regarding SRCC proteomics only dealt with less than four cases and were less convincing with regard to statistical evaluation. Second, how to excise the SRCC tissues is a key limitation in the sample preparation. Since gastric SRCC has its special histological features, the gastric tissues with dominant signet ring cells should be carefully estimated and isolated. Third, how to conduct proteomic analysis is an important technique issue so that it provides a deep and large data set for proteomic comparison, especially in a relatively large cohort.

With awareness of the three gaps, in this communication we present a comprehensive comparison of the proteomes derived from the gastric tissues of SRCC, PDAC, and WMDAC. A cohort with 48 cases including 14 SRCC, 17 PDAC, and 17 WMDAC cases was strictly selected from more than 2500 cases, and the cases were carefully evaluated on the basis of histological examination. The cancer and adjacent tissues were well isolated using laser capture microdissection (LCM) (10). We employed data-independent acquisition (DIA)-based proteomics (11) in quantitatively profiling the proteomes for all of the individuals on a large scale. For the first time, the quantitative proteomes in gastric SRCC, PDAC, and WMDAC were deeply characterized in parallel, revealing that the proteins in the complement cascade pathway were significantly upregulated in SRCC. Moreover, we performed a proteomic survey on 14 gastric cancer cell lines with the aim of classifying the cancer subtype representativeness of each cell line.

EXPERIMENTAL PROCEDURE

Experimental Design and Statistical Rationale

A total of 48 cases of gastric tumor tissues were involved, including 14 cases of SRCC and 17 cases each for PDAC and WMDAC. All the cases consisted of the paired tissues, which were defined by two criteria, 1) the paired tissues being obtained from the same patient and 2) according to the surgeon experiment, the tumor tissue as the center tissue on the visual view, while the adjacent tissue located more than 5 cm away from the tumor tissue. Univariate tests were used to probe the differential expression between disease conditions and subtypes. Using overrepresentation analysis and gene set enrichment analysis (GSEA), a running test was employed to find differential pathways. To avoid confounding factors, the samples were selected in a way to minimize cross-subtype differences of age, gender, and clinical characteristics. The batch effect was avoided by randomizing the

injection order of samples. Technical duplication was carried out to reduce missing values and errors of quantification.

Sample Collection

Frozen tissues were obtained from the specimen library of digestive diseases in Xijing Hospital, China. All of the tissue samples in the specimen library were frozen in liquid nitrogen at -196°C immediately after surgical resection. According to pathological records, 2522 cases of gastric tumors samples were stored between 2011 and 2016. Considering the relative low incidence of SRCC, we adopted a strategy to first select the SRCC cases followed by the inclusion of PDAC/WMDAC cases by matching with the SRCC cases in terms of age and gender (Fig. 1). The tissues diagnosed as SRCC by pathologists were further confirmed by microscopic recheck with hematoxylin and eosin (H&E) (BA4025, BASO, China) staining. In order to select the matched cases of PDAC and WMDAC, the cases for each subtype were selected in a case-wise matching manner to the SRCC cases as described Figure 1. Since the incidences tend to be distributed at a lower age in gastric SRCC than ACs, low-age cases with available frozen tissues and $>50\%$ cancer cells for the AC subtypes were supplemented. The intersubtype matching goodness in terms of age, gender, T and N staging was measured by the Wilcoxon test (12) (age) and χ^2 test (13) (gender, T and N staging). The paired tumor and adjacent tissues for those cases were collected for this study.

The cell lines in supplemental Table S1 were collected from multiple sources, including a gastric epithelia-derived GES-1 and 14 gastric tumor-derived lines including GCSR-1, AGS, NCI-N87, KATO-3, HGC-27, BGC-823, MKN-1, MKN-7, MKN-28, MKN-45, SNU-1, SNU-5, SNU-16, and SNU-668.

Generation of Proteomic Data

LCM was employed to purify cancer cells and adjacent epithelial cells from tissues. Trypsin digestion of the LCM tissues and cell lines were analyzed by LC-MS/MS using an Ultimate 3000 nanoLC system (Thermo Scientific) coupled to an Orbitrap Fusion Lumos mass spectrometer (Thermo Scientific) operated in either DDA or DIA acquisition mode (supplemental Information 1) at a gradient time of 120 min (supplemental Fig. S1).

Data Processing

The protein quantification data for all runs were globally normalized to an equal total, *i.e.*, the mean sum of all runs. Then, the normalized data for each of the three subtypes, SRCC, PDAC, and WMDAC, were strictly filtered, and their differentially expressed proteins (DEPs) between tumor and adjacent tissues (T/A-DEPs) were identified through the procedure depicted by supplemental Fig. S2. The data for each of the three subtypes, SRCC, PDAC, and WMDAC, were filtered, and their differentially expressed proteins between tumor and adjacent tissues (T/A-DEPs) were identified through the procedure depicted by supplemental Fig. S2. In brief, each protein in each case (tumor/adjacent pair) was classified according to its abundances in corresponding tumor and adjacent samples. Proteins with enough (more than 2/3 of all cases) full data points (cases with available data from tumor and adjacent samples, class 1 in supplemental Fig. S2) or enough tendency-agreed unilateral cases (cases with data from either tumor or adjacent samples, classes 2 and 3 in supplemental Fig. S2) plus cases whose absolute \log_2 transferred ratio >1 (classes 1.1 and 1.2 in supplemental Fig. S2) were retained while other proteins were filtered out. A protein with enough full data points was determined as T/A-DEP if its FDR-adjusted *p* value of the paired *t*-test (calculated with built-in functions in R 3.4.4) < 0.05 and absolute \log_2 transferred fold-change > 1 , and all the proteins with greater than 2/3 tendency-agreed unilateral cases plus cases whose absolute \log_2 transferred

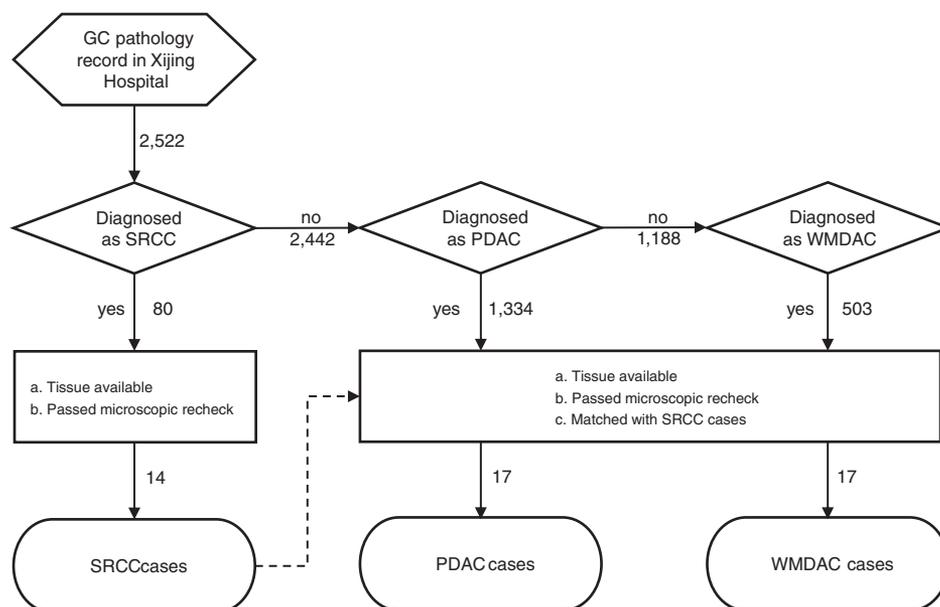


FIG. 1. The evaluation procedure to select proper tissue samples of SRCC, PDAC, and WMDAC for the proteomics study using LCM.

ratio >1 were designated as T/A-DEPs. A stricter filter ($>75\%$ cases possess at least eightfold difference) was applied to the common T/A-DEPs identified in the three subtypes to extract T/A-DEPs that infer the most obvious gastric cancer characteristics. The pathway enrichment analysis on T/A-DEPs was achieved with g:Profile (14) on Reactome (15) pathways through overrepresentation analysis (FDR adjusted p value < 0.05).

The proteins that passed the filtering as described above and presented in all subtypes were first selected, and then, the abundance ratios for all of the selected proteins were attained (in SRCC termed RS and in AC termed RA). To determine the DEPs between SRCC and AC (S/A-DEPs), two comparisons were conducted, comparison of the protein abundance (SRCC/AC) and comparison of abundance ratios (RS/RA) between SRCC and AC. A S/A-DEP was defined once its FDR-adjusted Wilcoxon test p value < 0.05 and its absolute \log_2 transferred fold-change > 1 in both comparisons.

The differential pathways between SRCC and AC were determined through GSEA (16) based on the protein abundance ratios of tumor/adjacent.

Aiming to classify cell lines into gastric cancer subtypes, a random forest model was trained on the protein abundances in tissues (supplemental Fig. S3), in which the proteins (square root n) ranked by abundance were randomly picked up 1000 times for training the classifier using the scikit-learn module (17), python 3.7. Data were used by machine learning in which cross-validation in a “leave one out” manner was used to seek the feature subset with the highest accuracy. Protein abundance data from cell lines were treated in the same manner, and then, they were fed to the trained classifier for their tissue-type likelihood prediction.

RESULTS

Collection of High-Quality Cancer Tissues and Proteomic Data

Of the 80 SRCC cases recorded in the tissue bank of Xijing hospital, only 14 tumor tissues were qualified with major tumor cells with the characteristics of signet ring type by an H&E

staining recheck (Fig. 1). In the tissue bank, 2442 cases were primarily diagnosed as PDAC and WMDAC, and 685 cases were removed after recheck, resulting in 1334 PDAC and 503 WMDAC cases. For each subtype, 14 cases were selected in a case-wise matching manner regarding the 14 SRCC cases. Then, three low-age cases with available frozen tissues and $>50\%$ cancer cells for each subtype were supplemented, resulting in the inclusion of 17 cases for PDAC and WMDAC. In total, 48 cases of gastric cancer that were well collected paired tumor and adjacent tissues were histologically classified into SRCC ($n = 14$), PDAC ($n = 17$), and WMDAC ($n = 17$). In order to set a base for cross-subtype comparison, the matching of clinicopathological features was especially considered in the selected cases. As a result, the age, gender, T and N staging of SRCC cases were not significantly different from those of PDAC or WMDAC cases (Table 1). Of the three subtypes, the mean ages ranged from 54.79 to 58.35 years, the percentages of male cases ranged from 64% to 71%, the tumors were all in advanced stages, *i.e.*, the T2, T3, and T4 stages, and the percentages of cases in the N0, N1/N2, and N3 stages ranged from 0% to 7%, 21% to 35%, and 59% to 71%, respectively.

Tumor cells are generally unevenly distributed in resected tumor tissue. To obtain tissues with high contents of tumor cells, we adopted LCM and collected the tissues with low intratumor heterogeneity for protein extraction. The typical microscopic images of the LCM-treated tissues are presented in Figure 2A, with three cases randomly selected from each subtype of gastric cancer, clearly demonstrating the “signet ring” morphology of SRCC, the dense formation of separate tumor cells of PDAC as well as the gland-like structures formed by WMDAC cells.

The LCM samples with an approximate area of 20 mm² were processed through an established method in our laboratory that was suitable for extracting peptides from micro amounts of biological samples (Experimental Procedure). A range of 1.4 to 12.5 µg of peptides was retrieved from a LCM sample, and the peptide yields ranged from 0.14 to 0.65 µg/mm² LCM sample (supplemental Table S2).

For the sake of better protein identification, we employed Preview to rapidly interrogate the occurrences of 25 chemical modifications in the samples. The assessment results in supplemental Table S3 surprisingly suggested that carbamidomethylation artifacts (+57 on N-terminus, H and K), deamidation (+1 on N and Q), and DTT addition (+152 on C) were the top three modifications, while pyroglutamate formation (-17 on N-terminal Q) and oxidation (+16 on M) were ranked at the sixth and seventh places. Therefore, the top three modifications were set as variable modifications in subsequent database searching. The spectral library required by DIA analysis was constructed by merging the DDA search results from the samples that were treated with pooling and fractionation and the publicly available pan human library (18). The library covered 10,990 proteins, 163,254 peptides, and 299,808 precursors, correspondingly.

For quality control of the proteomic data, we checked for a potential batch effect by feeding the unprocessed quantification data into principal component analysis (PCA). As visualized in supplemental Fig. S4, no obvious deviation was found among the four sequential batches, which were obtained from the continuous runs lasting a month. This implicated that the data quality was solid, and the batch effect could be ruled out. For protein identification and quantification, DIA analysis against the library generated a quantitative proteome containing 6195 proteins (supplemental Table S4) from these tissues in total, averagely 4835 proteins per sample (Fig. 2B). Among all of the quantified proteins, 62% were based on three unique peptides, while the default maximum unique peptides used for protein quantification in Spectronaut are just set at 3 (Fig. 2C). The number of unique peptides

related to each protein quantified in each injection was recorded in supplemental Table S5. These data made a solid base for further qualitative and quantitative analysis.

Proteomic Characteristics of Gastric Cancers

To obtain a glimpse of the overall pattern from all of the samples, the filtration of proteomic data was conducted through criteria described in the Experimental Procedure and resulted in 4945 proteins (supplemental Table S4). These protein abundances in the individual samples were compressed to 96 two-dimensional data points via t-SNE and visualized by a scatter plot (Fig. 3A). The figure revealed that the t-SNE-derived distances seemed not to distinguish different subtypes of gastric cancer according to the overall protein abundance patterns; meanwhile, tumors and adjacent tissues presented clearly different protein abundance patterns.

Based on the criteria and cutoffs described in the Experimental Procedure, we were able to identify the T/A-DEPs, with 574/263, 530/235, and 468/213 (downregulated/upregulated) from SRCC, PDAC, and WMDAC, respectively (supplemental Fig. S5 and supplemental Table S4). The overlap status of T/A-DEPs was assessed in supplemental Fig. S5, with 30.9% (380) T/A-DEPs shared by the three subtypes. Once the strict criteria were set, more than 75% cases had at least eightfold changes in abundance, and 48 proteins among the 380 common T/A-DEPs were further extracted (Fig. 3B). Of the 48 proteins, 11 were found in upregulated abundance in tumors, including ATP11A, ELN, FKBP10, FSTL1, LOX, MAP3K20, MFGE8, MGP, PLA2G2A, SFRP4, and TIMP3, and all of these proteins were also observed to be associated with gastric cancer in previous investigations. With regard to the eight proteins ATP11A (19), ELN (20), FKBP10 (21), FSTL1 (22), LOX (23, 24), MFGE8 (25), MGP (26), and SFRP4 (27, 28), other reports agreed with our findings. Contrary to our results, the previous studies uncovered the three proteins MAP3K20 (29), PLA2G2A (30–32), and TIMP3 (33) with downregulated abundances in gastric cancer.

TABLE 1
Statistics for tissue sample collection and pairing

Sample feature	SRCC (n = 14)	PDAC (n = 17)	(Measure of matching ^a)	WMDAC (n = 17)	(Measure of matching ^a)
Mean age	54.79	54.82	(0.95)	58.35	(0.33)
Gender					
Male	64%	65%	(0.98)	71%	(0.71)
Female	36%	35%	(0.98)	29%	(0.71)
T staging					
T1	0%	0%	(1)	0%	(1)
T2, T3, T4	100%	100%	(1)	100%	(1)
N staging					
N0	7%	0%	(0.26)	6%	(0.89)
N1, N2	21%	29%	(0.77)	35%	(0.4)
N3	71%	71%	(0.96)	59%	(0.47)

^aMeasure of matching is represented by the *p* value of the statistics test. Wilcox test for age and χ^2 test for gender, T and N staging.

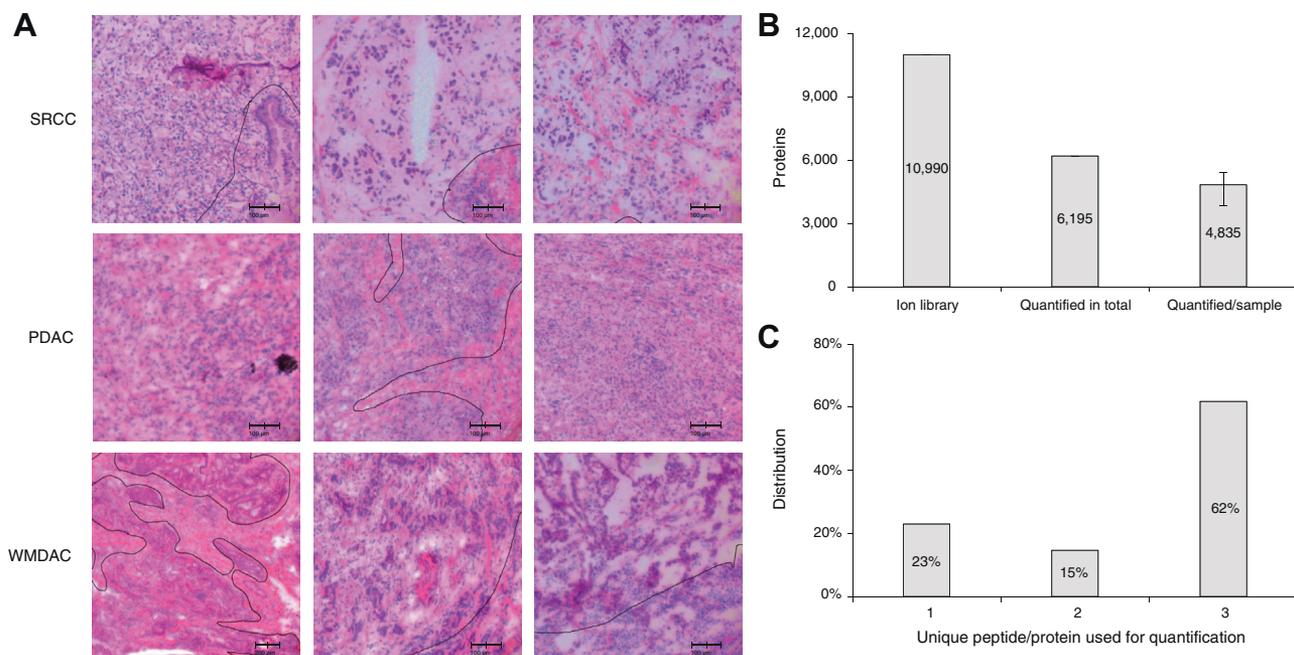


FIG. 2. **Assessment of data quality.** *A*, the H&E images for the diagnosis of SRCC, PDAC, and WMDAC. *B*, the proteins in the gastric tissues identified using a DDA and DIA approach (the error bar indicates the upper and lower bounds of proteins quantified/sample). *C*, distribution of the unique peptides in the quantified proteins.

In a query of the functions related to the 380 common T/A-DEPs, pathway enrichment analysis was performed using the Reactome pathway database (Fig. 3, C and D, supplemental Table S6). On the basis of evaluation by FDR-adjusted p values produced by the Fisher's exact tests, extracellular matrix (ECM) organization, collagen biosynthesis, and modifying enzymes as well as collagen formation were the top three pathways that were commonly upregulated in all of the subtypes. As for the enriched pathways with the down-regulated T/A-DEPs, TCA cycle, respiratory electron transport, and metabolism were the three most pronounced ones. It is a common phenomenon that the activation of ECM modification and suppression of aerobic metabolism are well-recognized hallmark behaviors of many tumors (34, 35). All of the results described above revealed that the protein characteristics of gastric cancers, both SRCC and AC, possessed the major and common differences from the adjacent tissues, in either typical biomarkers of gastric cancer or pathway dysregulations such as ECM and energy metabolism. Importantly, these molecular events commonly associated with gastric cancer were also perceived by other investigators.

It is well accepted that the tumor microenvironment (TME) plays a key role in the regulation of tumor growth and migration. Although an LCM sample mainly contains normal or abnormal epithelia, it may have several epithelia-associated cells, such as macrophages, monocytes, and so on. Thus, the proteomes derived from the tumor or adjacent tissues are reasoned with the proteomic information related with TME. Nirmal *et al.* (36, 37) claimed that over 300 proteins were

immune cell signatures in solid tumors. Based on the list, the immune cell-associated proteins were extracted from the proteome data set, and their corresponding abundances in the adjacent and tumor tissues were assessed. In total, 115 of the immune-associated proteins were found; however, they appeared to have a large portion of missing values, averaging 24%. Once the proteins with missing values were filtered out, 38 immune-associated proteins were hierarchically clustered against the GC samples. As illustrated in Figure 3E, the immune-associated proteins exhibited relatively higher abundances in tumor tissues compared with adjacent tissues, implicating that the immune cells around the tumor were enriched. Based on the expression ratio of corresponding signature genes (supplemental Fig. S6), the B cells and NK cells were absent in nearly all of the tissues tested, while macrophages, monocytes, neutrophils, plasma cells, and T cells were present in up to half of the samples. The GC tissues in general seemed to be infiltrated more by macrophages, monocytes, and neutrophils than adjacent tissues, and plasma cells presented at the same level in both tumor and adjacent tissues. Interestingly, the distribution of the T cells in the three types of GC was quite different.

Comparison of the Proteomic Characteristics Among SRCC and ACs

We further inquired to whether there was any subtype-based abundance feature. In an attempt to hierarchically cluster the 48 cases based on their tumor/adjacent protein ratios (supplemental Fig. S7), it was not easy to distinguish

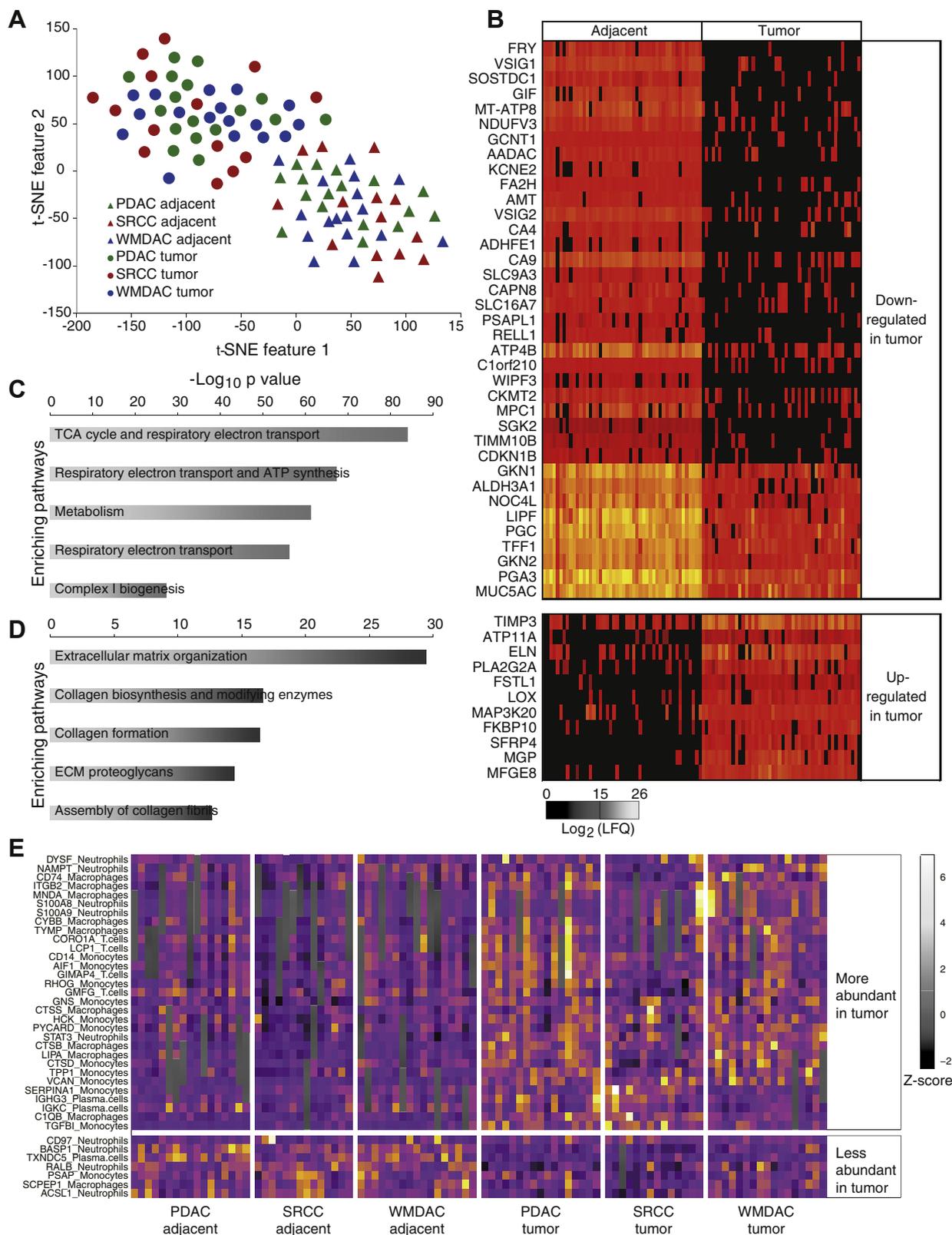


FIG. 3. Basic information of the quantitative proteomics in the gastric tissues. A, T-SNE analysis toward the protein abundance gained from the tumor and adjacent tissues in the 3 GC subtypes. B, heatmap of the 48 most obvious T/A-DEPs, as defined in the Experimental Procedure. C and D, the top five enriched pathways for the downregulated and upregulated proteins in all of the GC tissues. E, heatmap of 38 immune-related proteins that were probed in this DIA data set. The protein signals were normalized to Z-scores.

individual subtypes from each other. Then, the protein abundances were compared by linear regression, and the closeness was evaluated by Pearson's correlation coefficient (R^2). As shown in Figure 4A, PDAC and WMDAC were mutually more similar in the protein fold-change pattern, $R^2 = 0.79$, compared with PDAC to SRCC, $R^2 = 0.70$, and WMDAC to SRCC, $R^2 = 0.66$, implying that the proteomic abundances of PDAC and WMDAC were generally comparable, whereas that of SRCC was unique to some extent. Based on this overview, more investigations were conducted to pinpoint the proteins as well as pathways with different expression patterns between SRCC and ACs. The Venn diagram of the unique T/A DEPs also gave similar conclusions. As shown in supplemental Fig. S8 and supplemental Table S7, the unique

downregulated DEPs in SRCC were significantly enriched in energy metabolism pathways, and the unique upregulated ones were enriched in complement-related pathways. However, the two pathways were not significantly changed in ACs.

According to the definition of S/A-DEP, of the 4133 candidate proteins, only ten proteins matched with the criteria (supplemental Table S8), and six proteins had higher abundances in SRCC, including carcinoembryonic antigen-related cell adhesion molecule 5, matrix Gla protein, mucin-2, mucin-5B, ribosomal RNA processing protein 1 homolog B, and serpin B6, while four proteins had higher abundances in AC, including cytochrome c oxidase assembly protein COX11, mitochondrial 28S ribosomal protein S11, mitochondrial peptidyl-tRNA hydrolase 2, and selenoprotein H (supplemental

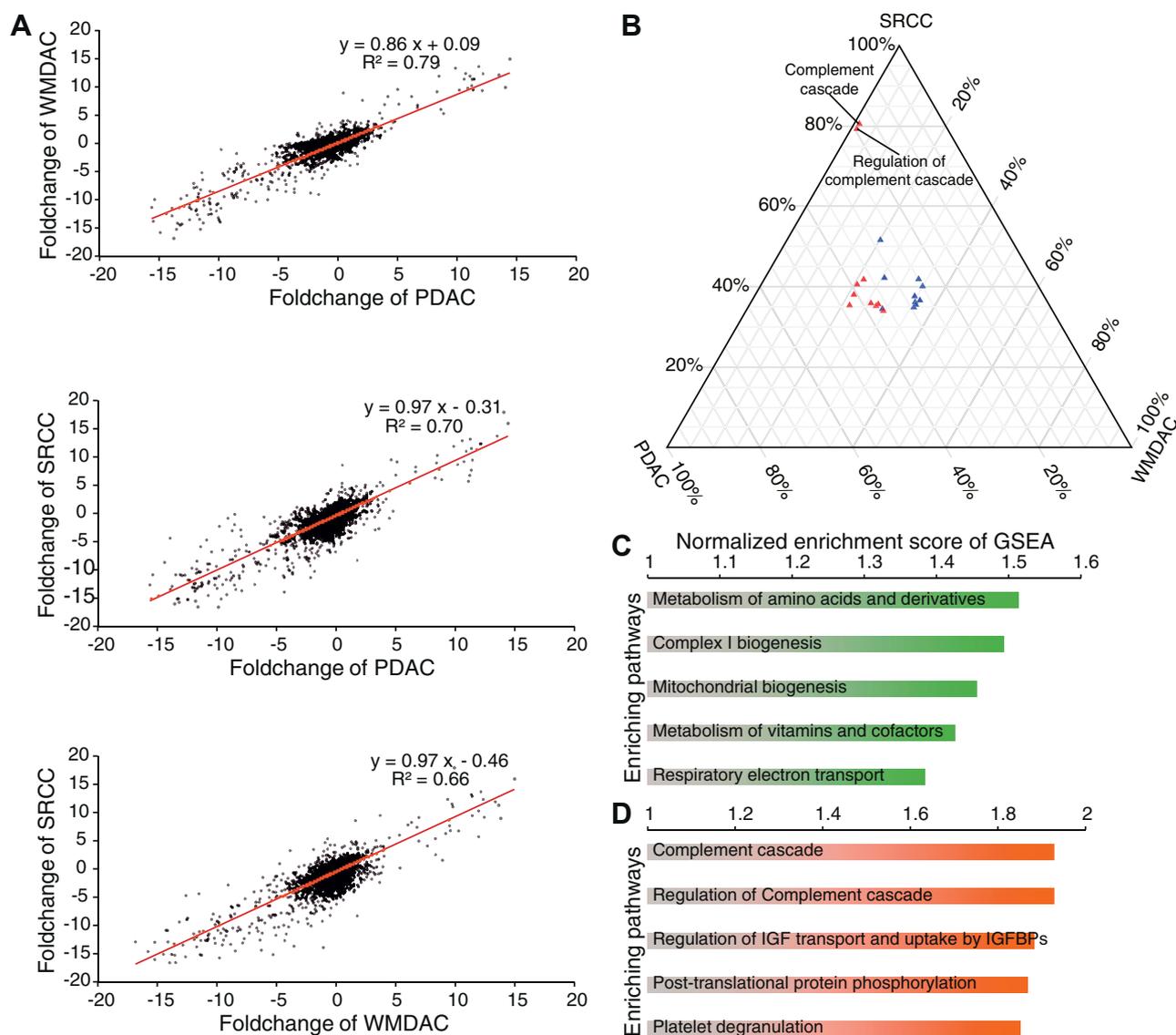


FIG. 4. **The proteomic characterization of SRCC.** A, linear correlations of the protein abundance ratios (T/A) intersubtypes. B, ternary plot indicating the subtype specificities of the top ten pathways enriching upregulated T/A-DEPs (red) and downregulated T/A-DEPs (blue). C and D, the top five AC-specific and the SRCC-specific pathways based on GSEA.

Fig. S9). In order to uncover pathways whose regulations were different among subtypes, we carried out pathway enrichment on T/A-DEPs identified from three subtypes (supplemental Table S9). The \log_{10} FDR-adjusted p values of the top enriched pathways were normalized and plotted to the ternary scale (Fig. 4B), demonstrating that the complement cascade and its regulation pathway were specifically upregulated in SRCC. Furthermore, GSEA was used to mine pathways harboring proteomic signals distinguishing SRCC from AC (supplemental Table S9). The top AC-specific pathways were mostly related to mitochondrial function. Meanwhile, SRCC-specific pathways were significantly related to extracellular reactions including the complement cascade (Fig. 4, C and D), which agreed with previous analysis. The complement cascade involves 138 proteins according to the Reactome database, of which approximately 60% (78) were quantified in the gastric tissues (Fig. 5). As over one-third of the gastric complement proteins exhibited higher abundances in SRCC and the average abundance ratios of T/A for complement proteins were twofold more than those in PDAC and WMDAC (supplemental Fig. S10), we came to a deduction that the proteins involved in the complement cascade were largely regulated in the SRCC microenvironment, while such an observation in the study of gastric cancer has not been reported yet. To conclude, despite the overall similar pattern observed among the three subtypes, a few proteins were found to be differentially expressed between SRCC and ACs. Meanwhile, pathway enrichment results indicated that the complement cascade was much more upregulated in SRCC than in ACs.

For the sake of confirming the upregulated proteins of complement cascades in SRCC, PRM was utilized to quantify these proteins in the SRCC and AC samples (the method is described in supplemental Information 1 Detailed methods). A total of 56 peptides related with 25 proteins were quantified, and the PRM results are summarized in supplemental Fig. S11 (the entire data set is listed in supplemental Table S10). Of these peptides, most peptides exhibited an increased trend in SRCC, including 23 peptides with significantly upregulated abundances. The PRM evidence, hence, endorsed the deduction from the proteomic analysis.

The Complement Relevant Proteome Events in SRCC

The complement cascade is harbored in human innate immunity, which likely consists of two events in cancer, complement activation followed by consensus amplification in the TME, and complemental regulation proteins (CRPs) function in membrane bound or secreted forms in tumor cells. The complement activation generally takes three distinct pathways, namely classical, lectin, and alternative, while all of the activated pathways finally merge into consensus amplification to exert the cascaded influence of innate immunity. As shown in Figure 5, the bulk of proteins in classical and lectin pathways were identified in SRCC with significant upregulation but not in AC tissues, except for FCN3, whereas only two proteins

of alternative pathways were perceived in all of the tissues of SRCC and ACs with insignificant changes in their protein abundances. Moreover, a large amount of immunoglobulins that might recognize the tumor-specific antigens and bind to C1 complexes in the classical pathway were identified with increased abundance (supplemental Table S4), implying the activation of the classic pathway in SRCC. Mucins (MUC2 and MUC5B) that are the secreted glycoproteins with rich N-acetylglucosamine moieties (38) and are liganded with lectins (39, 40) detected in high abundance were significantly upregulated in SRCC compared with ACs, setting the basis for the activation of the lectin pathway in SRCC (Fig. 5). In consensus amplification, complement proteins were upregulated to higher extent in SRCC than ACs, whereas the protein fold-changes in the pathway appeared less than those in classical and lectin activation (a 5.00-fold increase on average in classical, 8.32-fold in lectin, and 3.51-fold in consensus). All of the proteomic evidence thus led to a deduction that the classical and lectin pathways but not alternative pathway were activated in SRCC. The activation signals should be enlarged through consensus amplification; however, the changes of protein abundance in the consensus pathway were not fully coordinated with the complement activation. This suggested that the delivery of the activation signals was possibly attenuated in SRCC.

Tumor-derived CRPs identified in this study, either membrane bound or secreted, generally function as negative regulators to block the complement cascade. As depicted in Figure 5, many CRPs exhibited higher abundances in the SRCC tissues. These upregulated CRPs exhibited two characteristics. First, membrane cofactor protein (CD46), complement decay-accelerating factor (DAF, CD55), and CD59 are the common membrane-bound CRPs related with tumors to inhibit the complement cascade (37). Although the three proteins were identified in the gastric tissues, only CD55 was found to have increased abundance in SRCC, but not the others. Second, over ten secreted CRPs were identified to have augmented abundances in SRCC. For example, there were C4b-binding protein (C4bp) and complement factor I that bind or cleave C3/C5 convertases (41, 42), complement factor H and its related proteins (FHRs) that target and degrade C5 convertase (43), and carboxypeptidase N and clusterin that inactivate the membrane attach complex (MAC) (44, 45). Importantly, these secreted CRPs showed significantly higher abundances in SRCC compared with the corresponding adjacent tissues, while their fold-changes in SRCC were obviously larger than those in ACs (C4bp (6.75/3.95), DAF (9.28/4.22), factor I (4.39/1.41), clusterin (2.87/1.42), and carboxypeptidases N (6.03/4.22)). Hence, the proteomic evidence supported the postulation that the secreted CRPs in the tumor cells of SRCC were greatly expressed and secreted into the matrix, which might mainly be a response to complement activation in the cancer microenvironment and effectively attenuate the pathway of complement consensus amplification.

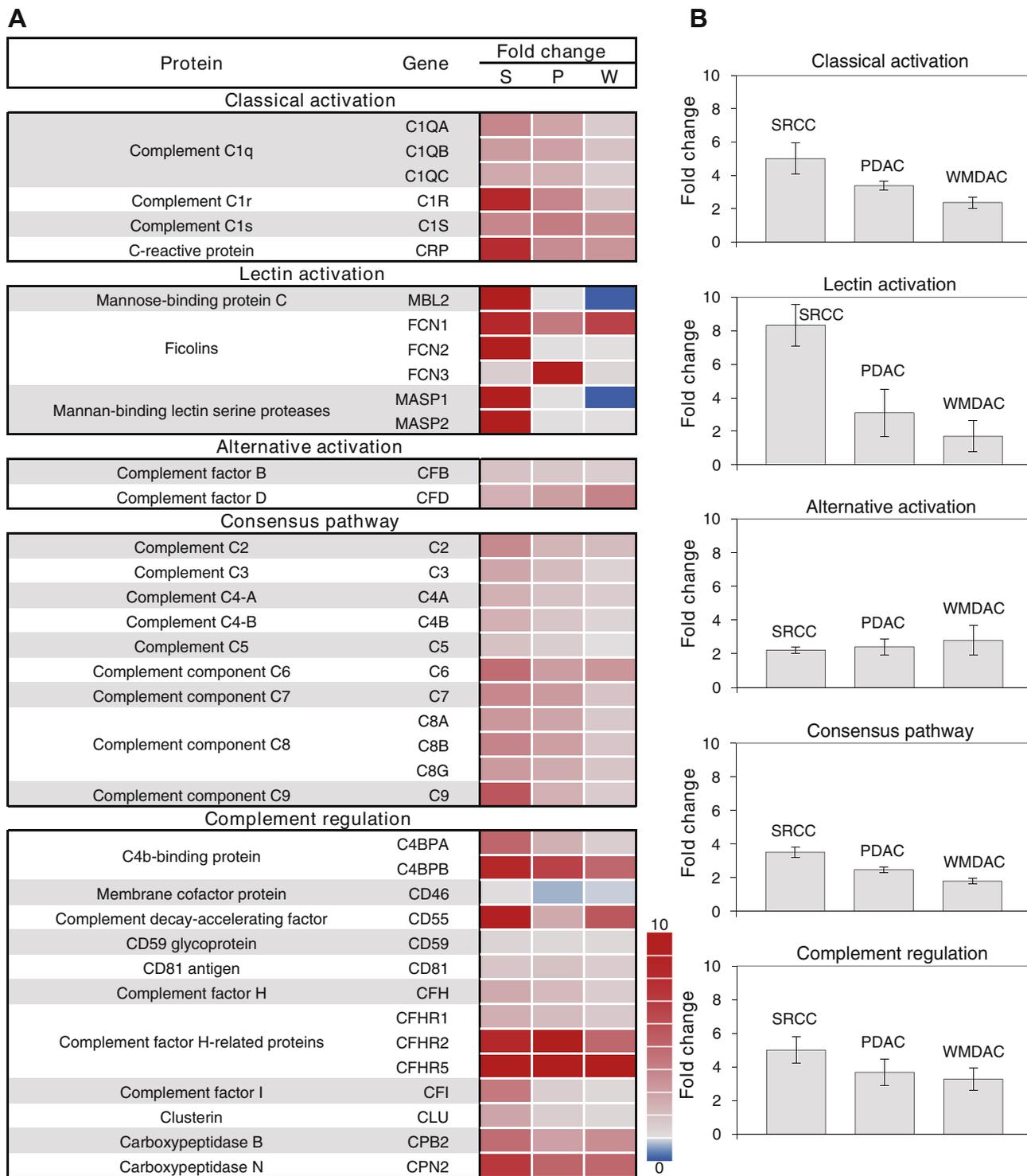


FIG. 5. Expression levels of the complement cascade in the three subtypes. Complement-related proteins quantified in the three subtypes of gastric cancer were grouped into five segments, namely classical activation, lectin activation, alternative activation, consensus pathways, and complement regulation. *A*, heatmaps indicating the protein fold-changes. *B*, corresponding bar plots describe the average fold-change for these segments.

A question is naturally raised how the complement activation coordinates with complement regulation because both events demonstrate enhanced abundances of the involved proteins.

We hypothesize a molecular scenario that in the TME, complement activation, such as classic and lectin, is triggered by the degradation products of phagocytosis, chemotaxis of

inflammatory cells, or tumor cell lysis (supplemental Fig. S12). Once the complement activation components are deposited on the tumor cell surface, the defense systems within them would be stimulated to exhibit complement-avoidance, by either DAF or a set of the secreted CRPs. Therefore, in SRCC tissues, a balance between complement activation and regulation remains so that some tumor cells escape from complement-mediated cytotoxicity (Fig. 6).

Comparison of the Proteomic Characteristics Between Gastric Cancer Tissues and Cell Lines

Many cell lines derived from the tissues of gastric cancer are widely accepted in academic investigation. After cell proliferation in many generations and the special treatment of cell immortalization, a question has remained whether those cell lines still retain the molecular characteristics of gastric cancer. In this study, we tried to seek the answer through comparison of the proteomic characteristics between tissues and cell lines. A total of 6639 proteins were quantified from all of the cell lines, and on average, 5213 proteins were perceived in an individual cell line (supplemental Table S11 and supplemental Fig. S13). The number of unique peptides related to each protein quantified in each injection is recorded in supplemental Table S12. The globally normalized protein abundance data were hierarchically clustered with an unsupervised mode, as shown in Figure 7A, demonstrating no obvious hierarchical cluster because over 50% of the proteins had relatively comparable abundances, whereas the other proteins possessed a diverse distribution of abundances.

The comparability assessment toward proteomic data was carried out in both qualitative and quantitative information. For

qualitative comparison, the Jaccard index (46) was obtained by the ratio of the overlapped proteins to the total proteins in any two samples and resulted in a Jaccard matrix. As illustrated in Figure 7B, the Jaccard index mean (0.64) for proteins between tissue and cell samples was much less than the value of 0.82 for the proteins within tissues or within cell lines, suggesting that the overall features in the tissue proteome were incomparable with those in cell lines. There were 1409 proteins uniquely identified in cell lines and 965 uniquely identified in tissues (supplemental Fig. S14A). Through the pathway enrichment analysis, the unique proteins in cell lines were significantly concentrated in 86 Reactome pathways, and those in tissues were enriched in 53 Reactome pathways (supplemental Table S13), whereas the converged pathways in tissues were completed differently from cell lines, strongly endorsing the conclusion drawn from Figure 7B and supplemental Fig. S14B. For quantitative comparison, a correlation matrix of the protein abundances (Fig. 7C) was generated from the correlation coefficients (R^2) of the co-identified proteins between tissues and cell lines. Similar to the results of the Jaccard matrix, the mean R^2 of 0.57 between tissues and cell lines was obviously smaller than the mean R^2 within tissues (0.81) or cell lines (0.80), implying that the quantification distribution of proteomes was largely different between tissues and cell lines.

Machine learning classifiers are efficient means to elucidate similar or dissimilar groups in a large data set. Among a variety of algorithms, the random forest (47) classifier is able to smartly weight and combine the intrinsic input features, thus generalizing reasonable predictions. There were three random forest classifiers that were constructed as follows: 1) the NT

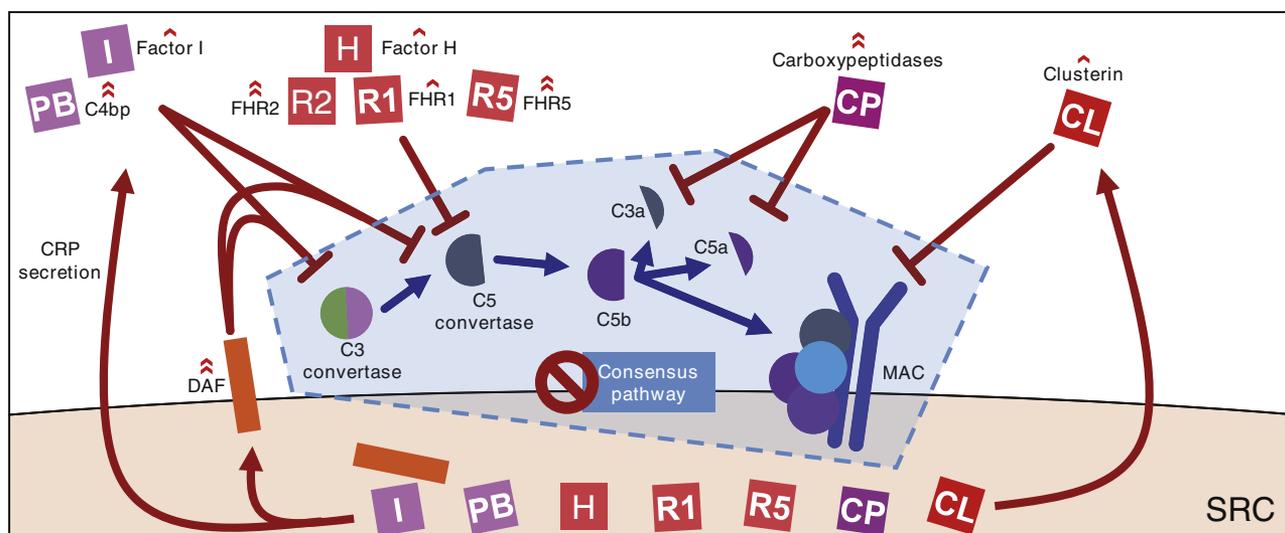


FIG. 6. Postulated complement regulation occurred in the SRCC context. To survive the complement cascade-induced cytotoxicity, SRC expressed DAF, factor 1, and C4bp to inhibit C3 and C5 convertase, factor H, FHR1, FHR2, and FHR5 to inhibit C5 convertase, carboxypeptidases to inhibit C3a and C5a, and clusterin to abolish MAC (blue arrows and shadows indicate the complement consensus pathway, while red lines indicate complement regulation. Red marks indicate the upregulation of corresponding molecules, with a single mark indicating a fold change >2 but ≤4 and double marks indicating a fold change >4).

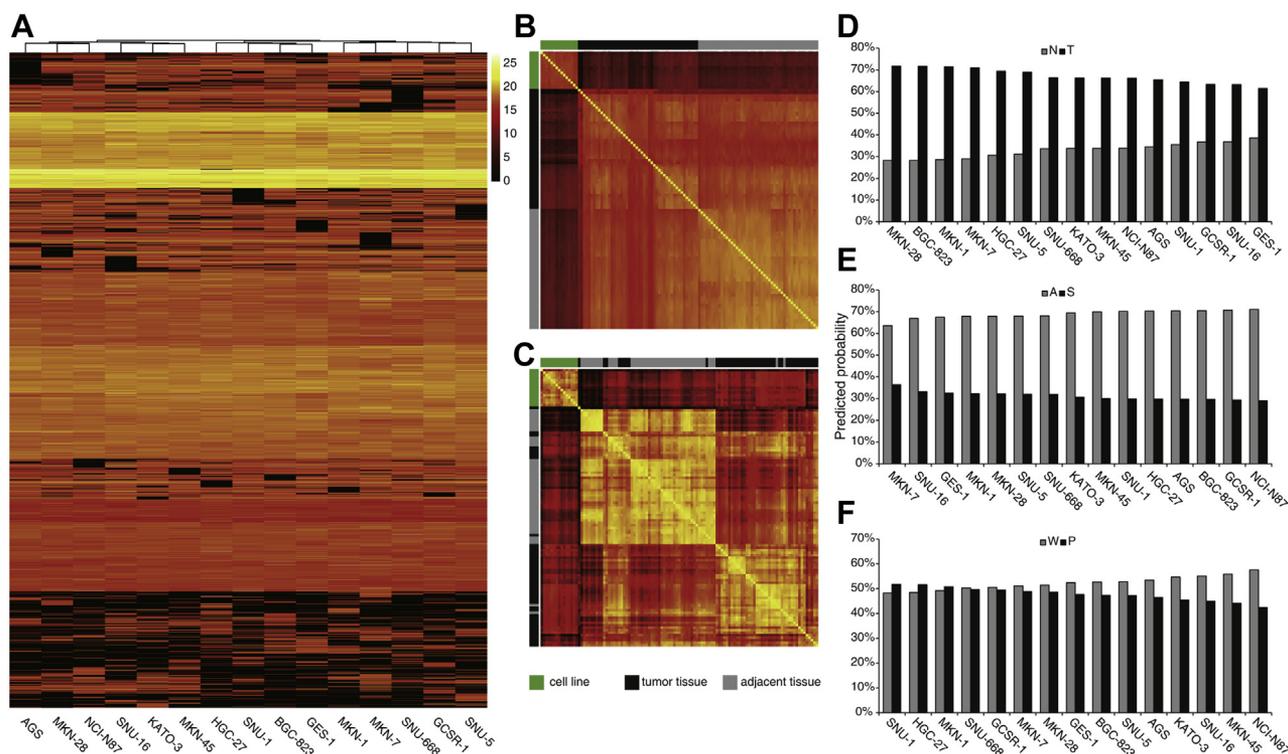


FIG. 7. Analysis of the quantitative proteomes in 15 human gastric cell lines. A, clustering of the quantified proteins in all of the cell lines. B and C, parallel abundance comparison of cell lines and tissues via the Jaccard index and correlation coefficient. D–F, similarity predictions toward GC cell lines and tissues by NT, AS, and PW classifiers.

classifier was trained on data from all of the 96 LCM samples to classify a cell line into “normal” or “tumor”, 2) the AS classifier was trained on data from 48 tumor LCM samples to classify a cell line into “SRCC” or “AC”; and 3) the PW classifier was trained on data from 34 adenocarcinoma LCM samples to classify a cell line into “PDAC” or “WMDAC”. Cross-validations were carried out to discover that all three classifiers yield acceptable accuracy with the entire data set (supplemental Fig. S15, supplemental Table S14). A probability of 50% was set as the threshold for class prediction. As a result, all of the cell lines of gastric cancer selected were classified to “tumor” with probabilities of 62%~73% by the NT classifier (Fig. 7D), and the three best representatives for tumor were MKN-28, BGC-823, and MKN-1 with probabilities over 71%. Although the GES-1 cell line was derived from normal gastric epithelia, it was also classified into “tumor” due to that its predicted probability was 62%, and GES-1 had the lowest probability to be “tumor” among all of the cell lines, implying that it was still different from tumor tissue somehow. All of the cell lines were classified to “adenocarcinoma” with probabilities of 63%~71% by the AS classifier with the top three representatives of AC being NCI-N87, GCSR-1, and BGC-823 (Fig. 7E). Surprisingly, three cell lines derived from SRCC tumors, KATO-3, GCSR-1, and SNU-668 (Table S1), were also recognized as AC. As for PW, the prediction probabilities generated by this classifier ranged from 48% to 58%

(Fig. 7F), which were too close to 50% to reach acceptable predictions, suggesting that the cell lines for PDAC or WMDAC were not well grouped through the PW classifier. Based upon these classifiers, we came to a conclusion that the 14 selected cell lines of gastric cancer appeared to have similar proteomic features with the ACs in tissues. Nevertheless, the cells currently used for the SRCC study were incomparable with the correspondent tumor tissues, at least based upon proteomic features.

DISCUSSION

Previously, in terms of the depth, the best result obtained from MS-based proteomics on gastric cancer was reported by Ge *et al.* (48), who managed to quantify ~4400 proteins per sample on average and ~9200 proteins for a total of 168 samples. This was done by feeding a large amount of peptide samples and fractions (~100 μ g in 6 fractions) into LC-MS/MS with DDA. In comparison, we applied a DIA strategy in this study to quantify ~4800 proteins per sample on average and ~6100 proteins for a total of 96 samples, achieving slightly better quantification results per sample yet much improved intersample comparability. We observed that, in spite of the similar scale and depth achieved by the present work and Ge’s work, there are differences in the DEP-enriched pathways concluded by the two works. To investigate, we

compared the pathways enriching DEPs from the three histological subtypes in our work and the three molecular subtypes, PX1, PX2, and PX3, classified by Ge *et al.* As listed in [supplemental Table S15](#), the upregulated pathways of the three subtypes in our work were mainly ECM related, and the downregulated pathways were mainly energy metabolism related. In contrast, the PX1 subtype didn't clearly imply its downregulated pathway; meanwhile, the PX2 and PX3 showed the downregulated pathways related with energy metabolism and translation, respectively. As for the upregulated pathways, the PX1 and PX2 were concentrated in transcription and cell-cycle-related functions, while the PX3 exhibited enrichment of immune-system-related pathways. The interstudy differences in DEPs and their enriched pathways could be attributed to the following two factors: 1) different schemes of subtype classifications adopted by the two works may highlight different functional aspects for each subtype and 2) the LCM used our work reduced interfering signals from other types of cells present in tumors, while Ge's work made use of bulk tissues for the analysis. Nevertheless, it should be recognized that the results of both studies reflected only parts of the gastric cancers, and further investigations featuring an advanced scale and depth are needed to fully characterize the gastric cancer.

With an emphasis on SRCC's unique characteristics in comparison to ACs, ten proteins were revealed to have distinct expression patterns between SRCC and ACs in this work. To examine whether these patterns were supported at a transcription level, a transcriptomic gastric tumor data set including 407 samples (32 normal tissues, 363 AC tumors, and 12 SRCC tumors) generated in a TCGA project (TCGA-STAD, <https://portal.gdc.cancer.gov/projects/TCGA-STAD>) was retrieved. The mRNA abundances in FPKM were normalized, and values for the ten relevant genes were extracted, as shown in [supplemental Fig. S16](#). Among them, CEACAM5, MUC2, MUC5B, and MRPS11 had similar SRCC/AC differences in their transcripts and proteins, which made them solid SRCC-specific indicators. Carcinoembryonic antigen-related cell adhesion molecules, encoded by the CEACAM5 gene, had long been recognized as a tumor-associated transmembrane protein. Its overexpression was observed in gastric and colon cancers (49, 50). Besides its intercellular adhesive role played in various types of tissues (51), CEACAM5 also possess a series of tumor-promoting functions such as the disruption of cell polarization, inhibition of cellular differentiation, and anoikis (52–54). A biomarker study carried out by Zhou *et al.* (55) associated CEACAM5 expression with a worse prognosis of gastric cancer. When it comes to SRCC, the presence of CEACAM5 was not consistent. Immune staining results demonstrated in Terada's study suggested that CEACAM5 had a higher level of expression in gastric and colorectal SRCC (56), while Warner *et al.* (57) reviewed 20 prostate SRCC cases only to find four CEACAM5 positive cases. Nevertheless, as this study and independent TCGA data set

revealed that CEACAM5 was specifically highly expressed in gastric SRCC compared with AC, there is a potential opportunity to develop a unique therapy for SRCC by targeting CEACAM5, whose protein product is located at the tumor cell surface. In fact, such a strategy was already conceptualized and experimented for colorectal cancer (58). Mucin 2 and mucin 5B are two mucus-comprising proteins widely produced and secreted by epithelial goblet cells under physiological conditions. One of the functions of mucin 2 is that suppression of inflammation occurs at mucous epithelia, a deficit of which was postulated to be a promoting factor of colon cancer (59, 60). However, in the case of gastric SRCC, the overexpression of the secreted mucins doesn't necessarily contribute to a positive effect since a significant amount of mucins are stored in the intracellular droplets of signet ring cells, which potentially indicates a disruption of the physiological secretion of mucins. Further examination of the expression levels of the mucin secretion-related proteins, including rab3 GTPase-activating protein, protein unc-13 homologs, protein unc-18 homologs, syntaxins, synaptotagmins, synaptosomal-associated proteins, and vesicle-associated membrane proteins (61) in our proteomics data, didn't support this postulation. The unique morphology of SRCC complicates the function of overexpressed mucins, and one can hope future investigations can harness the complication in treating SRCC. The mitochondrial ribosomal small subunit 11, encoded by MRPS11, was shown to be expressed at a specifically lower level in SRCC compared with AC. The expression of MRPS11 was correlated with a favored outcome in colorectal cancer (62), but its functional association with cancer is yet to be discovered.

As emphasized in the results, the complement cascade and its regulation were found to be characteristically upregulated pathways (Fig. 5). Concerning the cancer-related complement cascade deregulation, as reviewed by Afshar-Kharghan (63), many previous studies have been carried out, covering glioblastoma and melanoma as well as cervical, ovarian, lung, colorectal, breast, thyroid, and bladder cancer. The complement cascade carried out double-sided functions in the development of various tumors. On the one hand, it promotes the elimination of tumor cells by activating adaptive immune systems and forming MAC in the microenvironment, which directly induces apoptosis in tumor cells; on the other hand, the complement cascade promotes the proliferation of tumor cells *via* anaphylatoxin signaling. For the complement cascade in gastric cancers, very limited findings were available. Chen *et al.* (64) revealed that the expressions of complement proteins C5b, C6, C7, C8, and C9 were tumor-related and differentiating stage-dependent in gastric adenocarcinoma, while Inoue *et al.* (65) reported that a complement regulator, CD55 was constantly more highly expressed in gastric cancer cells than in the normal gastric tissues. Other complement cascade-related proteins, including C1r, C1s, C3, and the most central C4b, as well as multiple complement regulators, lacked documentation until the present study. With regard complement-related proteins in

gastric SRCC, only C1q was reported to be associated with the tumor development (66). For the first time, our study discovered a bulk of the proteins in complement cascade pathways highly sensitive in SRCC tissues. Although the implications of the complement cascade are incomplete and naive, its importance in the host immune system has attracted studies related to a wide range of diseases. As pointed out by Kleczko *et al.* (67), therapies targeting the complement cascade have already been experimented against immune-system-related diseases such as rheumatoid arthritis and age-related macular degeneration. Judging by the quantitative proteomes profiled by the present study, it was assumed that the evasion of complement-induced cell death by up-tuning the complement negative regulators was a significant characteristic of SRCC, and targeting the CRPs might be an effective approach to inhibit SRCC.

Although we discovered the association of the complement cascade activation with gastric SRCC in a subtype-constrained manner, it should be noted that the proteins involved in the complement cascade were largely missing in any of the 14 gastric cancer cells. In fact, approximately 80% (61) of proteins in the complement cascade in gastric tumor tissues were not reflected by any gastric cancer cell lines. This caveat needs to be noted when cell lines are used to model tumors, where molecular events occurring in the tumor microenvironments, such as the complement cascade, are lost in cell lines.

ADDITIONAL INFORMATION

Ethics Approval and Consent to Participate

The collection, storage, and usage of human tissues were covered by consents to participate and approved by the Institutional Review Board on Bioethics and Biosafety of BGI (NO. FT 15168). The present study was performed in accordance with the Declaration of Helsinki.

DATA AVAILABILITY

All of the mass spectrometry proteomics data have been deposited to the PRIDE Archive (<http://www.ebi.ac.uk/pride/archive/>) via the PRIDE partner repository with the data set identifier PXD022405, PXD022627, and PXD022952.

Supplemental data—This article contains [supplemental data](#) (68–72).

Funding and additional information—This work was supported by the funding from the National Key R&D Program of China (2017YFC0908403) and the National Key Basic Research Program of China (973 program) (No. 2014CBA02002 and No. 2014CBA02005).

Author contributions—Y. F.: Study design, lab research, data analysis, and article writing. B. B.: Study design and lab research. Y. L.: Lab research, data analysis, and article writing. Y. R.: Lab management and supervision. Y. L.: Lab research.

F. Z.: Lab research. X. L.: Study design and funding acquisition. J. Z.: Lab management and funding acquisition. G. H.: Data analysis and funding acquisition. F. C.: Study design and supervision. Q. Z.: Lab management and supervision. S. L.: Study design, lab management, supervision, and article writing.

Conflict of interest—The authors declare no competing interests.

Abbreviations—The abbreviations used are: AC, adenocarcinoma; CRP, complemental regulation protein; DEP, differentially expressed protein; DIA, data-independent acquisition; GSEA, gene set enrichment analysis; H&E, hematoxylin and eosin; LCM, laser capture microdissection; PDAC, poorly differentiated adenocarcinoma; SRCC, signet ring cell carcinoma; TME, tumor microenvironment; WMDAC, well-moderately differentiated adenocarcinoma.

Received September 16, 2020, and in revised form, December 25, 2020. Published, MCPRO Papers in Press, March 3, 2021, <https://doi.org/10.1016/j.mcpro.2021.100068>

REFERENCES

- Bosman, F. T., Carneiro, F., Hruban, R. H., and Theise, A. N. (2010) *WHO Classification of Tumours of the Digestive System*, 4th Ed, IRAC, France
- Bamboato, Z. M., Tang, L. H., Vinuela, E., Kuk, D., Gonen, M., Shah, M. A., Brennan, M. F., Coit, D. G., and Strong, V. E. (2014) Stage-stratified prognosis of signet ring cell histology in patients undergoing curative resection for gastric adenocarcinoma. *Ann. Surg. Oncol.* **21**, 1678–1685
- Kim, D. Y., Park, Y. K., Joo, J. K., Ryu, S. Y., Kim, Y. J., Kim, S. K., and Lee, J. H. (2004) Clinicopathological characteristics of signet ring cell carcinoma of the stomach. *ANZ J. Surg.* **74**, 1060–1064
- Chon, H. J., Hyung, W. J., Kim, C., Park, S., Kim, J. H., Park, C. H., Ahn, J. B., Kim, H., Chung, H. C., Rha, S. Y., Noh, S. H., and Jeung, H. C. (2017) Differential prognostic implications of gastric signet ring cell carcinoma: Stage adjusted analysis from a single high-volume center in Asia. *Ann. Surg.* **265**, 946–953
- Zhang, F., Ren, G., Lu, Y., Jin, B., Wang, J., Chen, X., Liu, Z., Li, K., Nie, Y., Wang, X., and Fan, D. (2009) Identification of TRAK1 (trafficking protein, kinesin-binding 1) as MGB2-Ag: A novel cancer biomarker. *Cancer Lett.* **274**, 250–258
- Yoshii, M., Tanaka, H., Ohira, M., Muguruma, K., Iwauchi, T., Lee, T., Sakurai, K., Kubo, N., Yashiro, M., Sawada, T., and Hirakawa, K. (2012) Expression of Forkhead box P3 in tumour cells causes immunoregulatory function of signet ring cell carcinoma of the stomach. *Br. J. Cancer* **106**, 1668–1674
- Ryu, J. W., Kim, H. J., Lee, Y. S., Myong, N. H., Hwang, C. H., Lee, G. S., and Yom, H. C. (2003) The proteomics approach to find biomarkers in gastric cancer. *J. Korean Med. Sci.* **18**, 505–509
- Jang, J. S., Cho, H. Y., Lee, Y. J., Ha, W. S., and Kim, H. W. (2004) The differential proteome profile of stomach cancer: Identification of the biomarker candidates. *Oncol. Res.* **14**, 491–499
- Li, W., Li, J. F., Qu, Y., Chen, X. H., Qin, J. M., Gu, Q. L., Yan, M., Zhu, Z. G., and Liu, B. Y. (2008) Comparative proteomics analysis of human gastric cancer. *World J. Gastroenterol.* **14**, 5657–5664
- Bagnell, C. R., Jr. (2006) *Laser Capture Microdissection. Molecular Diagnostics*. Springer, Berlin/Heidelberg, Germany: 219–224
- Gillet, L. C., Navarro, P., Tate, S., Röst, H., Selevsek, N., Reiter, L., Bonner, R., and Aebersold, R. (2012) Targeted data extraction of the MS/MS spectra generated by data-independent acquisition: A new concept for consistent and accurate proteome analysis. *Mol. Cell. Proteomics* **11**, O111.016717
- Bauer, D. F. (1972) Constructing confidence sets using rank statistics. *J. Am. Stat. Assoc.* **67**, 687–690

13. Hope, A. C. A. (1968) A simplified Monte Carlo significance test procedure. *J. R. Stat. Soc. Series B Stat. Methodol.* **30**, 582–598
14. Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., and Vilo, J. (2016) g:Profiler—a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res.* **44**, W83–W89
15. Croft, D., Mundo, A. F., Haw, R., Milacic, M., Weiser, J., Wu, G., Caudy, M., Garapati, P., Gillespie, M., Kamdar, M. R., Jassal, B., Jupe, S., Matthews, L., May, B., Palatnik, S., et al. (2014) The reactome pathway knowledgebase. *Nucleic Acids Res.* **42**, D472–D477
16. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545
17. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., et al. (2011) Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830
18. Rosenberger, G., Koh, C. C., Guo, T., Röst, H. L., Kouvonen, P., Collins, B. C., Heusel, M., Liu, Y., Caron, E., Vichalkovski, A., Faini, M., Schubert, O. T., Faridi, P., Ebhardt, H. A., Matondo, M., et al. (2014) A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci. Data* **1**, 140031
19. Yang, S., Shin, J., Park, K. H., Jeung, H. C., Rha, S. Y., Noh, S. H., Yang, W. I., and Chung, H. C. (2007) Molecular basis of the differences between normal and tumor tissues of gastric cancer. *Biochim. Biophys. Acta* **1772**, 1033–1040
20. Lapis, K., and Timár, J. (2002) Role of elastin-matrix interactions in tumor progression. *Semin. Cancer Biol.* **12**, 209–217
21. Liang, L., Zhao, K., Zhu, J. H., Chen, G., Qin, X. G., and Chen, J. Q. (2019) Comprehensive evaluation of FKBP10 expression and its prognostic potential in gastric cancer. *Oncol. Rep.* **42**, 615–628
22. Chen, J., Wang, X., Hu, B., He, Y., Qian, X., and Wang, W. (2018) Candidate genes in gastric cancer identified by constructing a weighted gene co-expression network. *PeerJ* **6**, e4692
23. Zhang, Q., Jin, X. S., Yang, Z. Y., Wei, M., Zhu, X. C., Wang, P., Liu, B. Y., and Gu, Q. L. (2013) Upregulated expression of LOX is a novel independent prognostic marker of worse outcome in gastric cancer patients after curative surgery. *Oncol. Lett.* **5**, 896–902
24. Peng, L., Ran, Y. L., Hu, H., Yu, L., Liu, Q., Zhou, Z., Sun, Y. M., Sun, L. C., Pan, J., Sun, L. X., Zhao, P., and Yang, Z. H. (2009) Secreted LOXL2 is a novel therapeutic target that promotes gastric cancer metastasis via the Src/FAK pathway. *Carcinogenesis* **30**, 1660–1669
25. Yang, W., Lai, Z., Li, Y., Mu, J., Yang, M., Xie, J., and Xu, J. (2019) Immune signature profiling identified prognostic factors for gastric cancer. *Chin. J. Cancer Res.* **31**, 463–470
26. Guo, L., Guo, X. B., Jiang, J. L., Zhang, J. N., Ji, J., Liu, B. Y., Zhu, Z. G., and Yu, Y. Y. (2010) Discovery and verification of matrix gla protein, a TNM staging and prognosis-related biomarker for gastric cancer. *J. Pathol.* **39**, 436–441
27. Cheong, J.-H., Yang, H.-K., Kim, H., Kim, W. H., Kim, Y.-W., Kook, M.-C., Park, Y.-K., Kim, H.-H., Lee, H. S., Lee, K. H., Gu, M. J., Kim, H. Y., Lee, J., Choi, S. H., Hong, S., et al. (2018) Predictive test for chemotherapy response in resectable gastric cancer: A multi-cohort, retrospective analysis. *Lancet Oncol.* **19**, 629–638
28. Li, H., Yu, B., Li, J., Su, L., Yan, M., Zhang, J., Li, C., Zhu, Z., and Liu, B. (2015) Characterization of differentially expressed genes involved in pathways associated with gastric cancer. *PLoS One* **10**, e0125013
29. Quan, Y., Zhang, Y., Lin, W., Shen, Z., Wu, S., Zhu, C., and Wang, X. (2018) Knockdown of long non-coding RNA MAP3K20 antisense RNA 1 inhibits gastric cancer growth through epigenetically regulating miR-375. *Biochem. Biophys. Res. Commun.* **497**, 527–534
30. Leung, S. Y., Chen, X., Chu, K. M., Yuen, S. T., Mathy, J., Ji, J., Chan, A. S., Li, R., Law, S., Troyanskaya, O. G., Tu, I. P., Wong, J., So, S., Botstein, D., and Brown, P. O. (2002) Phospholipase A2 group IIA expression in gastric adenocarcinoma is associated with prolonged survival and less frequent metastasis. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 16203–16208
31. Xing, X. F., Li, H., Zhong, X. Y., Zhang, L. H., Wang, X. H., Liu, Y. Q., Jia, S. Q., Shi, T., Niu, Z. J., Peng, Y., Du, H., Zhang, G. G., Hu, Y., Lu, A. P., Li, J. Y., et al. (2011) Phospholipase A2 group IIA expression correlates with prolonged survival in gastric cancer. *Histopathology* **59**, 198–206
32. Wang, X., Huang, C. J., Yu, G. Z., Wang, J. J., Wang, R., Li, Y. M., and Wu, Q. (2013) Expression of group IIA phospholipase A2 is an independent predictor of favorable outcome for patients with gastric cancer. *Hum. Pathol.* **44**, 2020–2027
33. Guan, Z., Zhang, J., Song, S., and Dai, D. (2013) Promoter methylation and expression of TIMP3 gene in gastric cancer. *Diagn. Pathol.* **8**, 110
34. Warburg, O. (1956) On the origin of cancer cells. *Science* **123**, 309–314
35. Hanahan, D., and Weinberg, R. A. (2011) Hallmarks of cancer: The next generation. *Cell* **144**, 646–674
36. Nirmal, A. J., Regan, T., Shih, B. B., Hume, D. A., Sims, A. H., and Freeman, T. C. (2018) Immune cell gene signatures for profiling the microenvironment of solid tumors. *Cancer Immunol. Res.* **6**, 1388–1400
37. Varela, J. C., Atkinson, C., Woolson, R., Keane, T. E., and Tomlinson, S. (2008) Upregulated expression of complement inhibitory proteins on bladder cancer cells and anti-MUC1 antibody immune selection. *Int. J. Cancer* **123**, 1357–1363
38. Bansil, R., and Turner, B. S. (2006) Mucin structure, aggregation, physiological functions and biomedical applications. *Curr. Opin. Colloid Interf. Sci.* **11**, 164–170
39. Matsushita, M., Endo, Y., Taira, S., Sato, Y., Fujita, T., Ichikawa, N., Nakata, M., and Mizuochi, T. (1996) A novel human serum lectin with collagen- and fibrinogen-like domains that functions as an opsonin. *J. Biol. Chem.* **271**, 2448–2454
40. Turner, M. W. (1996) Mannose-binding lectin: The pluripotent molecule of the innate immune system. *Immunol. Today* **17**, 532–540
41. Sim, R. B., Day, A. J., Moffatt, B. E., and Fontaine, M. (1993) Complement factor I and cofactors in control of complement system convertase enzymes. *Methods Enzymol.* **223**, 13–35
42. Ozen, A., Comrie, W. A., Ardy, R. C., Domínguez Conde, C., Dalgic, B., Beser, Ö. F., Morawski, A. R., Karakoc-Aydiner, E., Tutar, E., Baris, S., Ozca, F., Serwas, N. K., Zhang, Y., Matthews, H. F., Pittaluga, S., et al. (2017) CD55 deficiency, early-onset protein-losing enteropathy, and thrombosis. *N. Engl. J. Med.* **377**, 52–61
43. Soames, C. J., and Sim, R. B. (1997) Interactions between human complement components factor H, factor I and C3b. *Biochem. J.* **326**(Pt 2), 553–561
44. Campbell, W. D., Lazoura, E., Okada, N., and Okada, H. (2002) Inactivation of C3a and C5a octapeptides by carboxypeptidase R and carboxypeptidase N. *Microbiol. Immunol.* **46**, 131–134
45. Tschopp, J., Chonn, A., Hertig, S., and French, L. E. (1993) Clusterin, the human apolipoprotein and complement inhibitor, binds to complement C7, C8 beta, and the b domain of C9. *J. Immunol.* **151**, 2159–2165
46. Jaccard, P. (1912) The distribution of the flora in the alpine zone.1. *New Phytol.* **11**, 37–50
47. Ho, T. K. (1995) Random decision forest. Proceedings of 3rd international conference on document analysis and recognition. *IEEE* **1**, 278–282
48. Ge, S., Xia, X., Ding, C., Zhen, B., Zhou, Q., Feng, J., Yuan, J., Chen, R., Li, Y., Ge, Z., Ji, J., Zhang, L., Wang, J., Li, Z., Lai, Y., et al. (2018) A proteomic landscape of diffuse-type gastric cancer. *Nat. Commun.* **9**, 1012
49. Jothy, S., Yuan, S. Y., and Shirota, K. (1993) Transcription of carcinoembryonic antigen in normal colon and colon carcinoma. *In situ hybridization study and implication for a new in vivo functional model.* *Am. J. Pathol.* **143**, 250–257
50. Kodera, Y., Isobe, K., Yamauchi, M., Satta, T., Hasegawa, T., Oikawa, S., Kondoh, K., Akiyama, S., Itoh, K., and Nakashima, I. (1993) Expression of carcinoembryonic antigen (CEA) and nonspecific crossreacting antigen (NCA) in gastrointestinal cancer; the correlation with degree of differentiation. *Br. J. Cancer* **68**, 130–136
51. Benchimol, S., Fuks, A., Jothy, S., Beauchemin, N., Shirota, K., and Stanners, C. P. (1989) Carcinoembryonic antigen, a human tumor marker, functions as an intercellular adhesion molecule. *Cell* **57**, 327–334
52. Eidelman, F. J., Fuks, A., DeMarte, L., Taheri, M., and Stanners, C. P. (1993) Human carcinoembryonic antigen, an intercellular adhesion molecule, blocks fusion and differentiation of rat myoblasts. *J. Cell Biol.* **123**, 467–475
53. Ordoñez, C., Screaton, R. A., Ilantzis, C., and Stanners, C. P. (2000) Human carcinoembryonic antigen functions as a general inhibitor of anoikis. *Cancer Res.* **60**, 3419–3424

54. Ilantzis, C., DeMarte, L., Screatton, R. A., and Stanners, C. P. (2002) Deregulated expression of the human tumor marker CEA and CEA family member CEACAM6 disrupts tissue architecture and blocks colonocyte differentiation. *Neoplasia* **4**, 151–163
55. Zhou, J., Fan, X., Chen, N., Zhou, F., Dong, J., Nie, Y., and Fan, D. (2015) Identification of CEACAM5 as a biomarker for prewarning and prognosis in gastric cancer. *J. Histochem. Cytochem.* **63**, 922–930
56. Terada, T. (2013) An immunohistochemical study of primary signet-ring cell carcinoma of the stomach and colorectum: III. Expressions of EMA, CEA, CA19-9, CDX-2, p53, Ki-67 antigen, TTF-1, vimentin, and p63 in normal mucosa and in 42 cases. *Int. J. Clin. Exp. Pathol.* **6**, 630–638
57. Warner, J. N., Nakamura, L. Y., Pacelli, A., Humphreys, M. R., and Castle, E. P. (2010) Primary signet ring cell carcinoma of the prostate. *Mayo Clin. Proc.* **85**, 1130–1136
58. Conaghan, P., Ashraf, S., Tytherleigh, M., Wilding, J., Tchilian, E., Bicknell, D., Mortensen, N. J., and Bodmer, W. (2008) Targeted killing of colorectal cancer cell lines by a humanised IgG1 monoclonal antibody that binds to membrane-bound carcinoembryonic antigen. *Br. J. Cancer* **98**, 1217–1225
59. Van der Sluis, M., De Koning, B. A., De Bruijn, A. C., Velcich, A., Meijerink, J. P., Van Goudoever, J. B., Büller, H. A., Dekker, J., Van Seuning, I., Renes, I. B., and Einerhand, A. W. (2006) Muc2-deficient mice spontaneously develop colitis, indicating that MUC2 is critical for colonic protection. *Gastroenterology* **131**, 117–129
60. Velcich, A., Yang, W., Heyer, J., Fragale, A., Nicholas, C., Viani, S., Kuchelapati, R., Lipkin, M., Yang, K., and Augenlicht, L. (2002) Colorectal cancer in mice genetically deficient in the mucin Muc2. *Science* **295**, 1726–1729
61. Davis, C. W., and Dickey, B. F. (2008) Regulated airway goblet cell mucin secretion. *Annu. Rev. Physiol.* **70**, 487–512
62. Cavalieri, D., Dolara, P., Mini, E., Luceri, C., Castagnini, C., Toti, S., Maciag, K., De Filippo, C., Nobili, S., Morganti, M., Napoli, C., Tonini, G., Baccini, M., Biggeri, A., Tonelli, F., et al. (2007) Analysis of gene expression profiles reveals novel correlations with the clinical course of colorectal cancer. *Oncol. Res.* **16**, 535–548
63. Afshar-Kharghan, V. (2017) The role of the complement system in cancer. *J. Clin. Invest.* **127**, 780–789
64. Chen, J., Yang, W. J., Sun, H. J., Yang, X., and Wu, Y. Z. (2016) C5b-9 staining correlates with clinical and tumor stage in gastric adenocarcinoma. *Appl. Immunohistochem. Mol. Morphol.* **24**, 470–475
65. Inoue, T., Yamakawa, M., and Takahashi, T. (2002) Expression of complement regulating factors in gastric cancer cells. *Mol. Pathol.* **55**, 193–199
66. Wasserfallen, J. B., Spaeth, P., Guillou, L., and Pécoud, A. R. (1995) Acquired deficiency in C1-inhibitor associated with signet ring cell gastric adenocarcinoma: A probable connection of antitumor-associated antibodies, hemolytic anemia, and complement turnover. *J. Allergy Clin. Immunol.* **95**(1 Pt 1), 124–131
67. Kleczko, E. K., Kwak, J. W., Schenk, E. L., and Nemenoff, R. A. (2019) Targeting the complement pathway as a therapeutic strategy in lung cancer. *Front. Immunol.* **10**, 954
68. Reiter, L., Rinner, O., Picotti, P., Hüttenhain, R., Beck, M., Brusniak, M.-Y., Hengartner, M. O., and Aebersold, R. (2011) mProphet: automated data processing and statistical validation for large-scale SRM experiments. *Nat. Methods* **8**, 430–435
69. McKinney, W. (2010) Data Structures for Statistical Computing in Python. *Proc 9th Python Sci Conf.* 51–56
70. Oliphant, T., and Millma, J. K. (2006) *A guide to NumPy*. Trelgol Publishing, USA
71. Van Der Maaten, L., and Hinton, G. (2008) Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605
72. Rauniyar, N. (2015) Parallel reaction monitoring: a targeted experiment performed using high resolution and high mass accuracy mass spectrometry. *Int. J. Mol. Sci.* **16**, 28566–28581