Proteomic maps of human gastrointestinal stromal tumor subgroups

Yu Liu,1* Zhigui Li,1* Zhiqiang Xu,1 Xiuxiu Jin,1 Yanqiu Gong,1 Xuyang Xia,1 Yuqin Yao,1 Zhaofen Xu,2 Yong Zhou,1 Heng Xu,1 Shuangqing, Li,1 Yong Peng,1 Xiaoting Wu,† Lunzhi Dai,†∗

1Department of General Practice and Department of Gastrointestinal Surgery, State Key Laboratory of Biotherapy, West China Hospital, Sichuan University, and Collaborative Innovation Center of Biotherapy, Chengdu 610041, China.

2Department of Pathology, The Second People’s Hospital of Neijiang City, Sichuan province, Neijiang 641000, China.

*These authors contributed equally to this work.

†Correspondence: Email: lunzhi.dai@scu.edu.cn (Dr. Lunzhi Dai), wxt1@medmail.com.cn (Dr. Xiaoting Wu).

Running title: Proteome of GIST subgroups

Key words: Quantitative proteomics, immunohistochemical analysis, GIST, PPP2CB, PTPN1

Abbreviations:

GIST - Gastrointestinal stromal tumor

GIT - Gastrointestinal tract

MS – Mass spectrometry

DDA - Data dependent acquisition

FDR - False discovery rate
Proteome of GIST subgroups

NIH-L - National Institute of Health consensus criteria based very low/low risk

NIH-I - National Institute of Health consensus criteria based intermediate risk

NIH-H - National Institute of Health consensus criteria based high risk

KIT or CD117 - Mast/stem cell growth factor receptor Kit

DOG1 or ANO1 - Anoctamin-1

PTPN1 or PTP1B - Tyrosine-protein phosphatase non-receptor type 1

PPP2CB - Serine/threonine-protein phosphatase 2A catalytic subunit beta isoform

PLS-DA - Partial least squares discrimination analysis

TMT – Tandem mass tag

HPFs - high-power fields

IOD - Integrated option density

CV - Coefficient of variation

RSD - Relative standard deviation

GO - Gene Ontology

KEGG - Kyoto Encyclopedia of Genes and Genomes
Abstract

Gastrointestinal stromal tumor (GIST) is a common sarcoma of gastrointestinal tract (GIT) with high metastatic and recurrence rates, but the proteomic features are still less understood. Here we performed systematic quantitative proteome profiling of GIST from 13 patients classified into very low/low, intermediate and high risk subgroups. An extended cohort of GIST (n = 131) was used for immunohistochemical validation of proteins of interest. In total, 9177 proteins were quantified, covering 55.9% of the GIT transcriptome from The Human Protein Atlas. Out of the 9177 quantified proteins, 4930 proteins were observed in all 13 cases with 517 upregulated and 187 downregulated proteins in tumorous tissues independent of risk stage. Pathway analysis showed that the downregulated proteins were mostly enriched in metabolic pathway, while the upregulated proteins mainly belonged to spliceosome pathway. In addition, 131 proteins showed differentially expressed patterns among GIST subgroups with statistical significance. The 13 GIST cases were classified into 3 subgroups perfectly based on the expression of these proteins. The intensive comparison of molecular phenotypes and possible functions of quantified oncoproteins, tumor suppressors, phosphatases and kinases between GIST subgroups was carried out. Immunohistochemical analysis of the phosphatase PTPN1 (n = 117) revealed that the GIST patients with high PTPN1 expression had low chances of developing metastasis. Collectively, this work provides valuable information for understanding the inherent biology and evolution of GIST.
Proteome of GIST subgroups

Introduction

Gastrointestinal stromal tumor (GIST), a relatively newly defined pathologic entity, is the most common sarcoma of gastrointestinal tract (GIT), with an estimated annual incidence of 10-15 per million (1). It has been found that approximately 50-60% of GIST occurs in the stomach (2), 30-35% in small intestine, 5% in colorectum, and less than 1% in esophagus, which representing 0.1-3% of gastrointestinal malignancies (3). Approximately 10-20% of patients develop metastasis at the time of diagnosis (4). Previous studies have indicated that the 2-year and 5-year recurrence rates were 7.6% and 18.4%, respectively (5, 6). For operable GIST, recurrences mostly occurred within the first 5 years after surgery, and roughly 40% of patients with GIST developed metastasis during the 15 years follow-up period after surgery (7).

The mutations of mast/stem cell growth factor receptor Kit (KIT or CD117) (8, 9) and platelet-derived growth factor receptor A (PDGFRA) (10), which are considered as the major causes of GIST, mostly occurred in the muscle layer of the stomach or small intestine, thus it was postulated that GIST originated from the interstitial cells of Cajal or similar cells. Positive staining of KIT and anoctamin-1 (ANO1or DOG1) were observed in approximately 90% of GIST (11). GIST without KIT and PDGFRA mutation are considered to as wild type, which shelter mutations of other genes, such as succinate dehydrogenase (SDH ) and serine/threonine-protein kinase B-raf (BRAF) (12, 13).

Presently, several risk stratification criteria, such as the National Institute of Health (NIH) consensus criteria (14, 15), the modified NIH consensus criteria (16), and the Armed Forces Institute of Pathology (AFIP) criteria (17), have been used to evaluate the malignant potential of GIST. Generally, the large tumor size, high mitosis count, non-gastric site, presence of rupture and male sex were found to be associated with unfavorable outcomes (6, 7). NIH consensus classification criteria stratify GIST patients into four subgroups (very low, low, intermediate and high risk) based on the tumor size and mitotic rates (15). Except small GIST with a diameter no more than 1 cm, GIST with large tumor size and high mitosis count is more prone to recurrence.
Proteome of GIST subgroups

and metastasis typically in the liver and abdominal cavity (18). Large retrospective studies have also validated that patients with different risk stage stratification always have different disease-free survival ratio, increasing from very low-risk to high-risk GIST patients (7, 19, 20). Although remarkable progresses have been achieved to predict the potential malignancies of GIST, and a deep proteome coverage and unbiased proteomic studies of human GIST subgroups are still necessary (21-23).

In this study, we performed large-scale quantitative analysis of proteome between tumorous tissues (N) and paired adjacent non-tumorous tissues (T) of GIST, including 3 paired NIH consensus criteria based very low/low risk (NIH-L), 5 paired NIH consensus criteria based intermediate risk (NIH-I) and 5 paired NIH consensus criteria based high risk (NIH-H) GIST samples. Out of the 9177 quantified proteins in GIST proteome, 4930 proteins were quantified with good reproducibility (student-t test, \( p > 0.1 \)), and 517 proteins upregulated and 187 proteins downregulated in all 13 GIST tumorous tissues were found. Clustering and clustering enrichment analysis of 131 differentially expressed proteins clearly showed some distinctive biological processes and pathways enriched in GIST subgroups with significance, and partial least squares discrimination analysis (PLS-DA) confirmed the distinctive expression patterns of these 131 proteins in GIST subgroups. We further systematically assessed the expression patterns of oncoproteins, tumor suppressors (TSs), kinases and phosphatases, and discussed their potential functions in GIST. Immunohistochemical analysis of phosphatases tyrosine-protein phosphatase non-receptor type 1 (PTPN1 or PTP1B) and serine/threonine-protein phosphatase 2A catalytic subunit beta isoform (PPP2CB) in an extended cohort of GIST indicated that the GIST patients with high PTPN1 expression had low chances of developing metastasis. Collectively, this work is the first large-scale quantitative characterization of GIST proteome and will present valuable information for understanding the etiology of GIST progression.

Experimental procedures
Proteome of GIST subgroups

Experimental Design and Statistical Rationale

We performed in-depth proteomic analysis between tumorous and paired adjacent non-tumorous tissues of GIST, including 3 paired NIH-L, 5 paired NIH-I and 5 NIH-H GIST samples, using tandem mass tag (TMT) 10plex labeling, respectively. To reduce the influence of data-dependent acquisition, all samples were run in duplicates. To minimize the protein abundance difference among individuals, the protein intensity of tumor was normalized to that of its corresponding non-tumorous tissue. We used the \( p \)-value to assess the reproducibility of the technical duplicates based on student-t test. If the protein ratios quantified in duplicates had \( p \)-values (student-t test) more than 0.1, which meant that the quantified ratio difference between duplicates was small, the quantification of proteins was reproducible and these proteins were used for data mining.

Clinical samples collection

131 GIST patients were recruited from the department of gastrointestinal surgery, West China Hospital, with ethical approval from the Biomedical Ethics Committee of West China Hospital (Permission Number: 2017-254). All these cases were diagnosed as GIST by two independent pathologists according to Chinese consensus guidelines for diagnosis and management of GIST. To avoid protein degradation, all the tissues were frozen in liquid nitrogen in 30 min after surgery. There were the criteria for GIST samples collection here. Out of the 131 cases, 13 cases had tumors with paired adjacent non-tumorous tissues, and 118 cases had only tumorous tissues. All tumorous and adjacent non-tumorous specimens obtained from these patients were stored in West China Hospital Biobank of Sichuan University.

Protein extraction and isobaric labeling

The GIST tissues were shredded and lysed in RIPA buffer (1% NP-40, 0.5% (w/v) sodium deoxycholate, 150 mM NaCl, 50 mM Tris (pH = 7.5)) containing protease inhibitor, and then homogenized by Gentle-MACS (Miltenyi Biotec GmbH) under the procedure “protein 01. 01” for twice, followed by 5 min sonication under the condition of 0.3 s on and 1.7 s off with 195 watt of
Proteome of GIST subgroups

JY92-IIN (NingBoXinYi, China). The lysate was centrifuged at 20,000 rcf for 30 min, and the supernatant was transferred to a new tube. The protein concentration was measured by Bradford assay. Extracted proteins (50 µg) from each sample was reduced by 0.5 µmol Tris (2-carboxyethyl) phosphine (TCEP) with a final concentration of 10 mM at 56 °C for 1 h, alkylated with 1.0 µmol iodoacetamide with a final concentration of 20 mM in the dark at room temperature for additional 30 min, and then precipitated with methanol, chloroform and water (CH₃OH:CHCl₃:H₂O = 4:1:3). The precipitate was air-dried, and digested using sequence grade trypsin in triethylammonium bicarbonate (TEAB) buffer. The tryptic peptides of each sample were labeled with TMT (Thermo Fisher Scientific) reagents according to the manufacturer’s protocol. After quenching with 5% hydroxylamine, TMT-labeled peptides of each risk grade GIST were mixed and fractionated by reverse-phase C18 column.

Peptide fractionation

The TMT-labeled peptides were fractionated using reversed phase high performance liquid chromatography (RP-HPLC, agilent-1260) under basic pH. The mobile phase was composed of buffer A (98% H₂O with 2% ACN, 10 mM ammonium formate, pH = 10) and buffer B (90% ACN with 10% H₂O, 10mM ammonium formate, pH = 10). A standard 120 min LC gradient run was used and showed below: 0-10 min, 0%-8% Buffer B; 10-80 min, 8%-35% Buffer B; 80-95 min, 35%-60% Buffer B; 95-105 min, 60%-70% Buffer B; 105-120 min, 70%-100% Buffer B, and the flow rate was 1 mL/min. The peptides mixture was separated into 120 fractions and combined into 40 fractions. The combined fractions were dried in the speed vacuum, and used for mass spectrometry analysis.

LC-MS/MS analysis

The desalted peptides were lyophilized and resuspended in buffer A (2% ACN, 0.1% FA), LC-MS/MS analysis was performed using an EASY-nanoLC 1000 nanoflow LC instrument coupled to a high-resolution mass spectrometer (Q Exactive Plus, Thermo Fisher Scientific). A 100 µm
Proteome of GIST subgroups

(inner diameter) × 2 cm (length) of trap column and a 75 μm (inner diameter) × 12 cm (length) of analytical column were in-house pulled and packed with C18 particle (DIKMA). Data dependent acquisition (DDA) was performed in positive ion mode at the flow rate of 300 nL/min. MS spectra were acquired from 350 m/z to 1600 m/z with a resolution of 70,000 at m/z = 200. The automatic gain control (AGC) value was set at 3e6, with maximum injection time of 20 ms. For MS/MS scans, the top 15 most intense parent ions were selected with 1.6 m/z isolation window and fragmented with normalized collision energy (NCE) of 30%. The AGC value for MS/MS was set to a target value of 1e5, with a maximum injection time of 100 ms and a resolution of 35,000. Parent ions with a charge state of z = 1 or with unassigned charge states were excluded from fragmentation and the intensity threshold for selection was set to 2e5.

Data processing and analysis

All the raw files were searched against the Swiss-Prot human protein sequence database (20413 entries, 2017/01/14) in Maxquant (version 1.6). The precursor peptide mass tolerance was 10 ppm and a fragment ion mass tolerance was 0.02 Da. Two missed trypsin cleavages were allowed. Cysteine carbamidomethylation was set as a fixed modification. Oxidation of methionine and protein N-terminal acetylation were set as variable modifications. Peptides with <1% false discovery rate (FDR) were chosen for further data processing. Supplementary Table S1 contained 9177 quantified proteins of all GIST subgroups in duplicates, excluding the reverse and potential contaminant flagged, zero intensity proteins, as well as some proteins with single-peptide identification. Supplementary Table S2 contained quantified proteins of each GIST subgroup in duplicates. Supplementary Table S3 included 7230 proteins simultaneously quantified in all GIST cases. Supplementary Table S4 contained quantified proteins of each subgroup with good reproducibility between two technical repeats (student-t test, p >0.1). Supplementary Table S5 included 4930 quantified proteins in all GIST subgroups with good technical reproducibility (student-t test, p >0.1). Supplementary Table S6 contained 517 upregulated proteins and 187
Proteome of GIST subgroups

downregulated proteins in all subgroups. Supplementary Table S7 contains the GIT-specific and -unspecific genes in each GIST subgroup. Supplementary Table S8 contains the differentially expressed proteins between tumorous and paired adjacent non-tumorous tissues in each subgroup. Supplementary Table S9 showed Gene Ontology Cellular Component (GOCC) enriched results of differentially expressed proteins in each subgroup. Supplementary Table S10 included the differentially expressed proteins with statistical significance between GIST subgroups (student-t test, $p < 0.05$) and small difference within a GIST subgroup. Supplementary Table S11 included cluster-specific enrichment results of Gene Ontology Biological Processes (GOBP), GOCC and Kyoto Encyclopedia of Genes and Genomes (KEGG). Supplementary Table S12 included all oncoproteins and TSs simultaneously quantified in all GIST subgroups based on 138 well-annotated cancer driver genes (24). Supplementary Table S13 included all phosphatases and kinases simultaneously quantified in all GIST subgroups referring to Eukaryotic Kinase and Phosphatase Database (EKPD, 2019/1/1). Supplementary Table S14 included the immunohistochemical results of both PPP2CB and PTPN1. Supplementary Table S15 contained the statistics of PPP2CB and PTPN1 positive rates in GIST subgroups. Gene Annotation, including biological process, molecular function and cellular component, and KEGG pathway were performed by using DAVID 6.8 and gene set enrichment analysis (GSEA). Unsupervised-clustering and PLS-DA was applied to evaluate the difference of GIST subgroups. Pearson correlation coefficient analysis was used to confirm the difference and similarity among GIST subgroups, as well as the technical reproducibility in duplicates of each subgroup. The fold change statistics of kinases and phosphatases were performed in Excel. Volcano Plot was applied to show the significantly changed proteins in each subgroup.

Immunohistochemical staining

All tumor specimens diagnosed as GIST by two independent pathologists were made into the donor paraffin blocks. Tumor microarray and immunohistochemical staining were performed as
our previous study (25). The antibodies for immunohistochemistry were purchased from HuaBio (China, anti-PTPN1 antibody, Cat#RT1521; anti-PPP2CB antibody, Cat#ET1611-54). Image-Pro Plus 6.0 software was used to evaluate the intensity of protein expression.

Results

Proteome profiling of GIST subgroups

We collected 131 GIST cases deposited in the tumor biobank of West China Hospital. All 131 GIST cases were diagnosed by immunohistochemical staining of KIT, DOG1 and hematopoietic progenitor cell antigen CD34 (CD34), as well as histopathological review by two independent pathologists. Out of them, 13 tumors with their paired adjacent non-tumorous tissues were used for the following proteome profiling to study the molecular phenotypes during GIST progression (Table 1), while the remaining 118 cases with only tumorous tissues were used for immunohistochemical staining. In the 13 GIST cases, 3 patients were diagnosed with very low/low risk (tumor size < 2 cm, or mitotic count < 3/50 HPFs (high-power fields)) GIST, 5 patients with intermediate risk (tumor size 5-10 cm and mitotic count < 5/50 HPFs, or mitotic count 5-10/50 HPFs) GIST and 5 patients with high risk (mitotic count > 10/50 HPFs, or tumor size > 10 cm) GIST, according to NIH standard classification system (Table 1). It should be mentioned that one case belonging to NIH-I subgroup was KIT negative and included in the samples used for proteome profiling.

Quantitative proteomic analysis using isobaric isotope reagents such as TMT have been widely used to determine the relative protein ratios between different samples. In this study, to obtain the changing patterns of proteome between tumorous and non-tumorous tissues in GIST subgroups, we performed quantitative proteomics using TMT-10plex isobaric label reagent, with five biological repeats applied for both NIH-I and NIH-H GIST, as well as three biological repeats for NIH-L GIST (Figure 1A). To increase the throughput and minimize the co-isolated peptide contamination, the combined TMT labeled peptides were separated into 40 fractions by reversed
Proteome of GIST subgroups

phase high-performance liquid chromatography (RP-HPLC) under basic pH, and further analyzed by mass spectrometer with two technical replicates. The raw mass spectrometry (MS) data were searched using MaxQuant (version 1.6) against the Swiss-Prot human database (20413 entries, 2017/01/14), with protein false discovery rate (FDR) of 1% and MS2 tolerance of 10 ppm at peptide level (Figure 1A). After removing reverse, contaminant and zero intensity proteins, as well as proteins with single-peptide identification, a total number of 9177 proteins, representing about 55.9% of all GIT coding genes (26), were quantified in GIST samples (Figure 1B, and Supplementary Table S1). The median protein coverage was approximately 21.33% and the median number of unique peptides for each protein was 10.1. Specifically, 8650, 8287 and 7797 proteins were quantified in NIH-H, NIH-I and NIH-L GIST subgroups, respectively (Supplementary Figure S1A-D, and Supplementary Table S2). Good reproducibility was observed between technical repeats (Supplementary Figure S1E-G). Besides, 7230 quantified proteins were overlapped in all GIST samples (Figure 1C, and Supplementary Table S3). Then, we applied student-t test for the T/N ratios in duplicates and revealed that the p-values of 4930 proteins were greater than 0.1 between technical duplicates, indicating good reproducibility of protein quantification in between duplicates of each GIST subgroup, which were used for further data mining (Supplementary Figure S1H, and Supplementary Table S4-5).

To further demonstrate the reliability of our proteomic data, we first examined some well-known overexpressed proteins in GIST, such as KIT, protein PML (PML), tyrosine-protein kinase transmembrane receptor 2 (ROR2), ATP-dependent RNA helicase DDX39A/B (DDX39A/B ), apoptosis regulator Bcl-2 (BCL2), BTB/POZ domain-containing protein KCTD12 (KCTD12), BTB/POZ domain-containing protein KCTD10 (KCTD10), and ANO1 (Figure 1D). Notably, the changing patterns of all these proteins in our proteomic data were accurately consistent with previous reports (9, 10, 22, 23, 27, 28). Next, we calculated the Pearson correlation coefficient (r) of 4930 quantified proteins between GIST subgroups, and found that the coefficient between NIH-L and NIH-I subgroups was 0.77, between NIH-L and NIH-H subgroups was 0.77, and between
Proteome of GIST subgroups

NIH-I and NIH-H subgroups was 0.8 (Figure 1E), suggesting a closer similarity between NIH-I and NIH-H subgroups, consistent with their clinical output that recurrence-free survival of NIH-L GIST in significantly higher than that of NIH-I and NIH-H GIST (7). Collectively, we performed the deepest quantitative proteomics described to date in NIH-L, NIH-I and NIH-H GIST subgroups.

Common proteomic features of GIST subgroups

To pursue the general features of GIST proteome independent of risk grade, we processed the data to define unregulated and downregulated proteins in all GIST subgroups using the following filtration conditions, respectively. First, we selected the proteins whose T/N ratios in all 13 cases were simultaneously greater or less than 1.0, and 1011 proteins were obtained (Supplementary Table S6). Then, among the 1011 proteins, the proteins whose average T/N ratios in 13 cases were greater than 1.5 were defined as upregulated proteins, and the proteins whose average T/N ratios in 13 samples were less than 0.67 belonged to the downregulated proteins. As results, 517 upregulated proteins and 187 downregulated proteins in all subgroups were obtained (Figure 2A, and Supplementary Table S6). Gene Ontology (GO) analysis of these 704 common differentially expressed proteins revealed that the upregulated proteins significantly enriched in nucleus and cytoskeleton, while in the vesicles, mitochondrion and extracellular space only the downregulated proteins were enriched (Figure 2B). KEGG pathway analysis showed that the upregulated proteins mainly belonged to spliceosome, while the downregulated proteins were involved in metabolic pathways including carbon metabolism, glycolysis/gluconeogenesis, metabolism of xen by c-P450, TCA cycle, propanoate metabolism and drug metabolism, which related to the dysregulation of metabolism and detoxification in cancer (Figure 2C). In total, 28 proteins such as probable ATP-dependent RNA helicase DDX5 (DDX5), serine/arginine-rich splicing factor 1 (SRSF1) belonging to spliceosome were upregulated (Figure 2D, and Supplementary Figure S2A). DDX5, a RNA helicase regulating DNA replication and microRNA expression, was upregulated in majority of malignant breast cancer tissues (29, 30). SRSF1 was a member of the SR protein
family, which intensely associated with mRNA metabolism, including mRNA splicing, stability and translation (31), and found to be overexpressed in various cancer types as a potent proto-oncogene (29, 32, 33). RNA mis-splicing contributed to a great number of human diseases (34), and our data indicating a potential correlation between RNA mis-splicing and GIST risk stage. Some critical metabolic enzymes such as fructose-1,6-bisphosphatase 1/2 (FBP1/2), fructose-bisphosphate aldolase B (ALDOB), ribose-phosphate pyrophosphokinase 2 (PRPS2), phosphoenolpyruvate carboxykinase (PCK), alanine aminotransferase 1 (GPT), acyl-CoA synthetase short-chain family member 1/2 (ACSS1/2), succinyl-CoA ligase [ADP/GDP-forming] subunit alpha/beta (SUCLG1/2), 2-oxoglutarate dehydrogenase (OGDH), alcohol dehydrogenase [NADP(+)] (AKR1A1), isocitrate dehydrogenase [NADP] (IDH1, cytoplasmic), isocitrate dehydrogenase [NADP] (IDH2, mitochondrial) and aldehyde dehydrogenase (ALDH) were downregulated dramatically (Figure 2E-F). The dysregulation of enzymatic enzymes was closely associated with cancer. For example, the downregulation for IDH1/2 could decrease the effective level of a-ketoglutarate and inversely increase the stability of hypoxia-inducible factor 1-alpha (HIF-1A), a transcriptional factor facilitating the tumor growth under low oxygen environment (35), suggesting that downregulation of IDH1/2 might also involve in GIST progression. In addition, by statistically analyzing GIT specific proteins annotated in Human Protein Atlas (26), we found that these proteins were significantly downregulated compared with GIT unspecific proteins in all GIST subgroups (Supplementary Figure S2C, and Supplementary Table S7), indicating that loss of GIT tissue identity is a common feature of GIST.

The proteomic variances in GIST subgroups

Previous studies showed that the 5 years disease-free survival rates of very low/low, intermediate and high risks GIST patients were 96%, 54% and 20% separately (7), indicating the intrinsic molecular differences existing in GIST subgroups. To explore this, we first analyzed the proteomic features of GIST subgroups separately. For each subgroup, we processed the data using the
Proteome of GIST subgroups

following screening conditions: (1) The intensity difference of same protein in tumors and adjacent non-tumorous tissues should be significant (student-t test, $p < 0.05$), and the T/N ratio difference of same protein between technical duplicates should be less (student-t test, $p > 0.1$); (2) The average T/N ratio was greater than 1.5 and the ratio of each patient in this subgroup should be greater than 1.0, or the average T/N ratio was less than 0.67 and the ratio of each patient in this subgroup should be less than 1.0. Consequently, we obtained 896 upregulated and 93 downregulated proteins in NIH-L subgroup, 549 upregulated and 243 downregulated proteins in NIH-I subgroup, as well as 809 upregulated and 436 downregulated proteins in NIH-H subgroup (Supplementary Figure S3A, and Supplementary Table S8). Remarkably, kinesin-1 heavy chain (KIF5B), pre-B-cell leukemia transcription factor-interacting protein 1 (PBXIP1) and peptidyl-prolyl cis-trans isomerase FKBP10 (FKBP10) in NIH-L subgroup, protein disulfide-isomerase TMX3 (TMX3), tubulin alpha-1B chain (TUBA1B) and SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A member 5 (SMARCA5) in NIH-I subgroup, and myosin light chain 6B (MYL6B), sarcoplasmic/endoplasmic reticulum calcium ATPase 3 (ATP2A3), serine/arginine repetitive matrix protein 2 (SRRM2) and protein quaking (QKI) in NIH-H subgroup were dramatically upregulated, while histamine N-methyltransferase (HNMT) and a-kinase anchor protein 1 (AKAP1, mitochondrial) in NIH-L subgroup, glycerol-3-phosphate dehydrogenase [NAD(+)] (GPD1, cytoplasmic) and D-3-phosphoglycerate dehydrogenase (PHGDH) in NIH-I subgroup, and medium-chain specific acyl-CoA dehydrogenase (ACADM, mitochondrial), succinate-semialdehyde dehydrogenase (ALDH5A1, mitochondrial) and very long-chain specific acyl-CoA dehydrogenase (ACADVL, mitochondrial) in NIH-H subgroup were obviously downregulated (Supplementary Figure S3B, and Supplementary Table S4). Some of these proteins were known to be highly related to tumorigenesis. For example, overexpression of PBXIP1 could enhance the TGF-β induced epithelial–mesenchymal transition (EMT) (36), and activate the cell cycle check point G1/S and G2/M (37), indicating that the highly expressed PBXIP1 in tumor might play important roles in the early stage of GIST development.
Proteome of GIST subgroups

Next, we performed GOCC analysis of significantly expressed proteins in each GIST subgroup, respectively, using DAVID 6.8. We found that the extracellular exosome proteins were predominantly enriched in NIH-H subgroup, and some membrane proteins were mostly enriched in both NIH-L and NIH-I subgroups (Figure 3A, and Supplementary Table S9). GSEA analysis using tumor hallmark database (Version: h.all.v6.2.symbols.gmt [Hallmarks]) revealed that many signaling pathways showed similar regulatory patterns in three GIST subgroups (Figure 3B). For examples, G2M checkpoint and E2F target closely related to cell cycle were upregulated, while oxidative phosphorylation, xenobiotic metabolism, fatty acid metabolism and adipogenesis were downregulated in all subgroups. It should be mentioned that the cholesterol homeostasis pathway was almost unchanged in NIH-L subgroup, slightly downregulated in NIH-I subgroup, and dramatically downregulated in NIH-H subgroup (Figure 3B, and Supplementary Figure S3C). ACSS2, an important enzyme associated with homeostasis of cholesterol, was significantly downregulated in NIH-H GIST tumor. The low expression of ACSS2, exhibiting a tendency of elevated keratin, type II cytoskeletal 7 (KRT7) expression and decreased keratin, type II cytoskeletal 20 (KRT20) and homeobox protein CDX-2 (CDX2) expression, was an independent prognostic factor for poor 5-year progression-free survival in colorectal carcinoma (38).

Furthermore, to better known the proteomic variances, we screened the differentially expressed proteins, which have significant difference between GIST subgroups and small difference within a GIST subgroup, using two filtration conditions. First, the protein \( p \)-values (student-t test) were less than 0.05 between two GIST subgroups in at least two comparisons. Second, the relative standard deviation (RSD) of ratios of same protein in an identical GIST subgroup was less than 0.4, or the same protein T/N ratios of all cases in an identical GIST subgroup were greater or less than 1.0. Consequently, 131 proteins were identified as differentially expression proteins (Supplementary Table S10). Unbiased clustering and cluster-specific enrichment analysis using GOBP, GOCC, and KEGG (DAVID 6.8) revealed the biological process, cellular components and pathways enriched in different GIST subgroups (Figure 3C,
Proteome of GIST subgroups

and Supplementary Table S11). Pathways and biological functions enriched in NIH-L GIST included the endocytosis pathway and the protein transport processes. While in the NIH-H GIST, the ubiquitin mediated proteolysis processes was significantly enriched. PLS-DA analysis confirmed the distinctive features of these 131 differentially expressed proteins in GIST subgroups (39), and these 13 cases of GIST were classified into 3 subgroups perfectly based on the expression of these proteins (Figure 3D). Collectively, these analyses of specifically enriched proteins faithfully unveiled the unique proteomic features of GIST subgroups.

Oncoproteins and tumor suppressors in GIST

We next assessed the homeostasis of oncoproteins and TSs in GIST subgroups. The activation of oncogenes and inactivation of TS genes were the main causes of tumor progression. To disclose the oncoproteins and TSs associated with GIST risk stage, we selected them based on 138 well-annotated cancer driver genes (24) and summarized their fold-changes in all GIST subgroups (Figure 4A-B, and Supplementary Table S12). In total, 21 oncoproteins and 31 TSs were quantified in all 13 GIST cases (Supplementary Table S12). KIT, which belonged to PI3K pathway and was a hallmark of GIST, as well as some other oncoproteins including tyrosine-protein kinase JAK1 (JAK1), splicing factor U2AF 35 kDa subunit (U2AF1), splicing factor 3B subunit 1 (SF3B1) and serine/threonine-protein phosphatase 2A 65 kDa regulatory subunit A alpha isoform (PPP2R1A) was significantly upregulated in all GIST subgroups. Epidermal growth factor receptor (EGFR) is a well-demonstrated therapeutic target for tumor treatment, and its overexpression and activation is associated with the development of a wide variety of tumors (40). As showed in Figure 4A, the EGFR was highly expressed in NIH-H subgroup, but not NIH-L subgroup, indicating the EGFR pathway might involve in the late stage of GIST development. Moreover, we found some TSs including fizzy-related protein homolog (CDH1 or FZR1) and fizzy-related protein homolog (SMAD4) were downregulated in all GIST subgroups. Loss of function of
Proteome of GIST subgroups

CDH1 have a strongly increased incidence of gastric cancer (41). It was speculated that the downregulation of CDH1 were mostly likely associated with the progression of GIST.

**Kinases and phosphatases in GIST**

Kinases and phosphatases, two classes of essential cellular signaling regulators, have been widely used as targets for new drug development. These phosphorylation regulatory enzymes play very important roles in the evolution of cancers including GIST. For example, the overexpression of KIT, a tyrosine phosphorylation kinase, has been recognized as a significant feature in GIST. To wholly understand the roles of kinases and phosphatases during GIST progression, we selected kinases and phosphatases based on EKPD database, and 18 phosphatases (Figure 5A) and 32 kinases (Figure 5B) were quantified in all 13 GIST cases (Supplementary Table S13). Notably, most phosphatases and kinases were upregulated in GIST. Specifically, the kinases such as KIT, nik-related protein kinase (NRK), cyclin-dependent kinase 7 (CDK7), serine/threonine-protein kinase Nek6 (NEK6), serine-protein kinase ATM (ATM) and bromodomain-containing protein 3 (BRD3) were upregulated in tumors (Figure 5B). In contrast, we also found that the eukaryotic elongation factor 2 kinase (EEF2K) was downregulated in three GIST subgroups. EEF2K is a kinase to phosphorylate the substrate eEF2 and regulates the elongation stage of protein synthesis. It has a dual role in promoting or inhibiting tumorigenesis depending on the cancer types (42).

Compared with the kinases, the phosphatases were still understudied. Interestingly, the phosphatase PPP2CB showed distinctive expression patterns in GIST subgroups (Figure 5A). Previous studies revealed that PPP2CB was related to cell cycle arrest phenotype (43, 44) and established as a suppressor of NF-κB signaling (45). Reduced expression of PPP2CB were observed in prostate cancer (46). Our MS results disclosed that the PPP2CB was upregulated in tumor of low-risk subgroup and downregulated in tumor of high-risk subgroup (Supplementary Figure S4A, and Supplementary Table S10). And the T/N differences between subgroups were
Proteome of GIST subgroups

very significant (student-t test, \( p = 0.012 \) between NIH-H and NIH-I groups. \( p = 0.002 \) between NIH-H and NIH-L subgroups). These MS data were further confirmed by Western blots (Supplementary Figure S4B). Inspired by these results, we checked the absolute PPP2CB expression in tumors of an extended cohort of GIST cases (n =113) by immunohistochemical analysis using tissue microarray to explore the potential role of PPP2CB in different GIST risk stage (Supplementary Table S14). As results, we found that all the tumorous tissue spots belonging to NIH-L subgroups were positively stained, while only 86.1% and 90.4% of spots were positively stained in NIH-I and NIH-H subgroups, respectively (Supplementary Figure S4C, and Supplementary Table S15). The average integrated option density (IOD) of staining differences didn’t show statistical significance between GIST subgroups (Supplementary Figure S4D). We assumed that the protein abundance of PPP2CB differed greatly in normal states between individuals (47). Although the absolute PPP2CB expression in tumors of GIST subgroups didn’t have significant changes, the T/N ratios were very different. We assumed that high PPP2CB level was likely to be associated with low risk of GIST.

In addition, among all upregulated phosphatases (Figure 5A), PTPN1 gained our attention, because the function of PTPN1 in GIST was completely unknown and it was potentially associated with tumor metastasis (48, 49). To reveal the relationship between PTPN1 and GIST progression, we first verified the MS data of PTPN1 (Figure 5C) with Western blots, and found that the Western blotting results and MS data matched perfectly (Figure 5D). Then, we performed immunohistochemical analysis of 117 GIST tumorous tissues revealed high percentage of PTPN1 positive cases in NIH-L subgroup, and the positive rates of PTPN1 descending gradually from NIH-L to NIH-H subgroups (Figure 5E-G, and Supplementary Table S14-15), especially the average IOD of PTPN1 in NIH-L was significantly higher than that in NIH-H subgroup (student-t test, \( p = 0.037 \)). Further analysis revealed that the GIST patients with high PTPN1 had low chances of developing metastasis (Figure 5H). Taken together, these results indicated that
PTPN1 might be a potential suppressor of GIST metastasis, consistent with that PTPN1 stabilizes VE-cadherin–mediated cell–cell adhesions and controls cell motility and invasion (48, 49).

Discussion

In summary, we performed in-depth quantitative proteomics to reveal the proteomic features of GIST subgroups. In total, 9177 proteins were ambiguously quantified in GIST samples, covering almost 55.9% of the GIT transcriptome from The Human Protein Atlas. Among them, 704 with similar expression patterns and 131 proteins with distinctive expression patterns were identified in GIST subgroups. In addition, we performed immunohistochemical analysis in an extended GIST cohort to study the phosphatases PTPN1 and PPP2CB, which were found to be closely correlated with GIST metastasis and risk grade, respectively.

Previous studies showed that PTPN1 was a double-facet molecule in tumors. It acted as a tumor promoter in breast (50), non-small lung (51), ovarian (52) and prostate (53) cancers, but performed as a tumor suppressor in esophageal cancer (54) and lymphoma (55). Our immunohistochemical results suggested that the patients with very high expression of PTPN1 were metastasis-free, and the risk of metastasis increased as the PTPN1 abundance decreased (Figure 3F). Previous study indicated that PTPN1 acted as a tumor suppressor inhibiting the cell mobility and invasion, through negatively regulates VEGFR by binding to the VEGFR and stabilization of cell-cell adhesions via tyrosine dephosphorylation of VE-cadherin (48). Most likely, a similar mechanism might be associated with GIST tumor metastasis, and its function was largely dependent on its absolute abundance. The functions and mechanisms of PTPN1 during GIST progression required more intensive studies in the following researches.

Positive expression of KIT in GIST patients are very common (56). In all our GIST cases, only one patient was KIT negative. In this study we only compared the proteomic features of GIST subgroups without considering the influence of KIT in this work, actually, we found that the proteomic patterns between KIT negative and positive cases were very obvious in NIH-I subgroup.
Proteome of GIST subgroups

(Figure 3A). More KIT negative GIST cases may be required to profoundly study the impact of KIT on GIST proteome.

However, it should be mentioned that sample set used for proteomic discovery in this study was a little bit small, because the surgical principle in China for localized and potentially resectable GIST was R0 resection and it was very difficult to get paired adjacent non-tumorous tissue after surgery (57). The small sample size for discovery might affect the quantification accuracy of some proteins and make the global statistical analyses less robust. Some interesting and important discoveries in this study would be verified in a larger GIST cohort when the samples were available.

References

1. Mucciarini, C., Rossi, G., Bertolini, F., Valli, R., Cirilli, C., Rashid, I., Marcheselli, L., Luppi, G., and Federico, M. (2007) Incidence and clinicopathologic features of gastrointestinal stromal tumors. A population-based study. BMc cancer 7, 230
2. Pisters, P. W., Blanke, C. D., von Mehren, M., Picus, J., Sirulnik, A., Stealey, E., Trent, J. C., and re, G. S. C. (2011) A USA registry of gastrointestinal stromal tumor patients: changes in practice over time and differences between community and academic practices. Annals of oncology : official journal of the European Society for Medical Oncology 22, 2523-2529
3. Lewis, J. J., and Brennan, M. F. (1996) Soft tissue sarcomas. Current problems in surgery 33, 817-872
4. Woodall, C. E., 3rd, Brock, G. N., Fan, J., Byam, J. A., Scoggins, C. R., McMasters, K. M., and Martin, R. C., 2nd (2009) An evaluation of 2537 gastrointestinal stromal tumors for a proposed clinical staging system. Archives of surgery (Chicago, Ill. : 1960) 144, 670-678
5. Racz, J. M., Brar, S. S., Cleghorn, M. C., Jimenez, M. C., Azin, A., Atenafu, E. G., Jackson, T. D., Okrainec, A., and Quereshy, F. A. (2015) The accuracy of three predictive models in the evaluation of recurrence rates for gastrointestinal stromal tumors. Journal of surgical oncology 111, 371-376
6. Joensuu, H., Vehtari, A., Riihimäki, J., Nishida, T., Steigen, S. E., Brabec, P., Plank, L., Nilsson, B., Cirilli, C., Braconi, C., Bordoni, A., Magnusson, M. K., Linke, Z., Sufliarsky, J., Federico, M., Jonasson, J. G., Dei Tos, A. P., and Rutkowski, P. (2012) Risk of recurrence of gastrointestinal stromal tumour after surgery: an analysis of pooled population-based cohorts. The Lancet. Oncology 13, 265-274
7. Joensuu, H., Vehtari, A., Riihimäki, J., Nishida, T., Steigen, S. E., Brabec, P., Plank, L., Nilsson, B., Cirilli, C., Braconi, C., Bordoni, A., Magnusson, M. K., Linke, Z., Sufliarsky, J., Federico, M., Jonasson, J. G., Dei Tos, A. P., and Rutkowski, P. (2012) Risk of recurrence of gastrointestinal stromal tumour after surgery: an analysis of pooled population-based cohorts. The Lancet Oncology 13, 265-274
8. Hirota, S., Isozaki, K., Moriyama, Y., Hashimoto, K., Nishida, T., Ishiguro, S., Kawano, K., Hanada, M., Kurata, A., Takeda, M., Muhammad Tunio, G., Matsuzawa, Y., Kanakura, Y., Shinomura, Y., and Kitamura, Y. (1998) Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. Science 279, 577-580
9. Corless, C. L., Fletcher, J. A., and Heinrich, M. C. (2004) Biology of gastrointestinal stromal tumors. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 22, 3813-3825
10. Heinrich, M. C., Corless, C. L., Duensing, A., McGreevey, L., Chen, C. J., Joseph, N., Singer, S., Griffith, D. J., Haley, A., Town, A., Demetri, G. D., Fletcher, C. D., and Fletcher, J. A. (2003) PDGFRA activating mutations in gastrointestinal stromal tumors. *Science* 299, 708-710
11. Miettinen, M., Wang, Z. F., and Lasota, J. (2009) DOG1 antibody in the differential diagnosis of gastrointestinal stromal tumors: a study of 1840 cases. *The American journal of surgical pathology* 33, 1401-1408
12. Pasini, B., McWhinney, S. R., Bei, T., Matyakhina, L., Stergiopoulos, S., Muchow, M., Boikos, S. A., Ferrando, B., Pacak, K., Assie, G., Baudin, E., Chompret, A., Ellison, J. W., Briere, J. J., Rustin, P., Gimenez-Roqueplo, A. P., Eng, C., Carney, J. A., and Stratakis, C. A. (2008) Clinical and molecular genetics of patients with the Carney-Stratakis syndrome and germline mutations of the genes coding for the succinate dehydrogenase subunits SDHB, SDHC, and SDHD. *European journal of human genetics : EJHG* 16, 79-88
13. Agaram, N. P., Wong, G. C., Guo, T., Maki, R. G., Singer, S., Dematteo, R. P., Besmer, P., and Antonescu, C. R. (2008) Novel V600E BRAF mutations in imatinib-naive and imatinib-resistant gastrointestinal stromal tumors. *Genes, chromosomes & cancer* 47, 853-859
14. Fletcher, C. D. M., Berman, J. J., Corless, C., Gorstein, F., Lasota, J., Longley, B. J., Miettinen, M., O’Leary, T. J., Remotti, H., Rubin, B. P., Shmookler, B., Sobin, L. H., and Weiss, S. W. (2002) Diagnosis of gastrointestinal stromal tumors: A consensus approach. *Human Pathology* 33, 459-465
15. Joensuu, H. (2008) Risk stratification of patients diagnosed with gastrointestinal stromal tumor. *Human pathology* 39, 1411-1419
16. Rutkowski, P., Bylina, E., Wozniak, A., Nowecki, Z. I., Osuch, C., Matlok, M., Świtaj, T., Michej, W., Wróński, M., Głuszek, S., Kroc, J., Nasirowerska-Guttmejer, A., and Joensuu, H. (2011) Validation of the Joensuu risk criteria for primary resectable gastrointestinal stromal tumour – The impact of tumour rupture on patient outcomes. *European Journal of Surgical Oncology (EJSO)* 37, 890-896
17. Miettinen, M., and Lasota, J. (2006) Gastrointestinal stromal tumors: Pathology and prognosis at different sites. *Seminars in Diagnostic Pathology* 23, 70-83
18. Joensuu, H., Fletcher, C., Dimitrijevic, S., Silberman, S., Roberts, P., and Demetri, G. (2002) Management of malignant gastrointestinal stromal tumours. *The Lancet Oncology* 3, 655-664
19. Joensuu, H., Hohenberger, P., and Corless, C. L. (2013) Gastrointestinal stromal tumour. *The Lancet* 382, 973-983
20. Miettinen, M., El-Rifai, W. E., H, L., and Lasota, J. (2002) Evaluation of malignancy and prognosis of gastrointestinal stromal tumors: A review. *Human Pathology* 33, 478-483
21. Atay, S., Wilkey, D. W., Milhem, M., Merchant, M., and Godwin, A. K. (2018) Insights into the Proteome of Gastrointestinal Stromal Tumors-Derived Exosomes Reveals New Potential Diagnostic Biomarkers. *Molecular & cellular proteomics : MCP* 17, 495-515
22. Kikuta, K., Kubota, D., Saito, T., Orita, H., Yoshida, A., Tsuda, H., Suehara, Y., Katai, H., Shimada, Y., Toyama, Y., Sato, K., Yao, T., Kaneko, K., Beppu, Y., Murakami, Y., Kawai, A., and Kondo, T. (2012) Clinical proteomics identified ATP-dependent RNA helicase DDX39 as a novel biomarker to predict poor prognosis of patients with gastrointestinal stromal tumor. *Journal of proteomics* 75, 1089-1098
23. Suehara, Y., Kondo, T., Seki, K., Shibata, T., Fujii, K., Gotoh, M., Hasegawa, T., Shimada, Y., Sasaki, M., Shimoda, T., Kurosawa, H., Beppu, Y., Kawai, A., and Hirohashi, S. (2008) Pfiets in as a prognostic biomarker of gastrointestinal stromal tumors revealed by proteomics. *Clinical cancer research : an official journal of the American Association for Cancer Research* 14, 1707-1717
24. Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., and Kinzler, K. W. (2013) Cancer genome landscapes. *Science* 339, 1546-1558
Proteome of GIST subgroups

25. Li, Z., Xu, Z., Huang, Y., Zhao, R., Cui, Y., Zhou, Y., and Wu, X. (2018) The predictive value and the correlation of peripheral absolute monocyte count, tumor-associated macrophage and microvessel density in patients with colon cancer. *Medicine* 97, e10759

26. Uhlen, M., Fagerberg, L., Hallstrom, B. M., Lindskog, C., Oksvold, P., Mardinoglu, A., Sivertsson, A., Kampf, C., Stjostedt, E., Asplund, A., Olsson, I., Edlund, K., Lundberg, E., Navani, S., Szgyarto, C. A., Odeberg, J., Djureinovic, D., Taken, J. O., Hofer, S., Alm, T., Edqvist, P. H., Berling, H., Tegel, H., Mulder, J., Rockberg, J., Nilsson, P., Schwenk, J. M., Hamsten, M., von Feilitzen, K., Forsberg, M., Persson, L., Johansson, F., Zahlen, M., von Heijne, G., Nielsen, J., and Ponten, F. (2015) Proteomics. Tissue-based map of the human proteome. *Science* 347, 1260419

27. West, R. B., Corless, C. L., Chen, X., Rubin, B. P., Subramanian, S., Montgomery, K., Zhu, S., Ball, C. A., Nielsen, T. O., Patel, R., Goldblum, J. R., Brown, P. O., Heinrich, M. C., and van de Rijn, M. (2004) The Novel Marker, DOG1, Is Expressed Ubiquitously in Gastrointestinal Stromal Tumors Irrespective of KIT or PDGFRA Mutation Status. *The American Journal of Pathology* 165, 107-113

28. Edris, B., Espinosa, I., Muhlenberg, T., Mikels, A., Lee, C. H., Steigen, S. E., Zhu, S., Montgomery, K. D., Lazar, A. J., Lev, D., Fletcher, J. A., Beck, A. H., West, R. B., Nusse, R., and van de Rijn, M. (2012) ROR2 is a novel prognostic biomarker and a potential therapeutic target in leiomyosarcoma and gastrointestinal stromal tumour. *The Journal of pathology* 227, 223-233

29. Mazurek, A., Luo, W., Krasnitz, A., Hicks, J., Powers, R. S., and Stillman, B. (2012) DDX5 Regulates DNA Replication And Is Required For Cell Proliferation In A Subset Of Breast Cancer Cells. *Cancer Discovery*, CD-12-0116

30. Wang, D., Huang, J., and Hu, Z. (2012) RNA Helicase DDX5 Regulates MicroRNA Expression and Contributes to Cytoskeletal Reorganization in Basal Breast Cancer Cells. *Molecular & Cellular Proteomics* 11, M111.011932

31. Das, S., and Krainer, A. R. (2014) Emerging Functions of SRSF1, Splicing Factor and Oncoprotein, in RNA Metabolism and Cancer. *Molecular Cancer Research* 12, 1195-1204

32. Anczukow, O., Rosenberg, A. Z., Akerman, M., Das, S., Zhan, L., Karni, R., Muthuswamy, S. K., and Krainer, A. R. (2012) The splicing factor SRSF1 regulates apoptosis and proliferation to promote mammary epithelial cell transformation. *Nature structural & molecular biology* 19, 220-228

33. Karni, R., de Stanchina, E., Lowe, S. W., Sinha, R., Mu, D., and Krainer, A. R. (2007) The gene encoding the splicing factor SF2/ASF is a proto-oncogene. *Nature structural & molecular biology* 14, 185-193

34. Scotti, M. M., and Swanson, M. S. (2016) RNA mis-splicing in disease. *Nature reviews. Genetics* 17, 19-32

35. Zhao, S., Lin, Y., Xu, W., Jiang, W., Zha, Z., Wang, P., Yu, W., Li, Z., Gong, L., Peng, Y., Ding, J., Lei, Q., Guan, K. L., and Xiong, Y. (2009) Glioma-derived mutations in IDH1 dominantly inhibit IDH1 catalytic activity and induce HIF-1alpha. *Science* 324, 261-265

36. Shi, S., Zhao, J., Wang, J., Mi, D., and Ma, Z. (2017) HPIP silencing inhibits TGF-beta1-induced EMT in lung cancer cells. *International journal of molecular medicine* 39, 479-483

37. Feng, Y., Xu, X., Zhang, Y., Ding, J., Wang, Y., Zhang, X., Wu, Z., Kang, L., Liang, Y., Zhou, L., Song, S., Zhao, K., and Ye, Q. (2015) HPIP is upregulated in colorectal cancer and regulates colorectal cancer cell proliferation, apoptosis and invasion. *Scientific reports* 5, 9429

38. Bae, J. M., Kim, J. H., Oh, H. J., Park, H. E., Lee, T. H., Cho, N. Y., and Kang, G. H. (2017) Downregulation of acetyl-CoA synthetase 2 is a metabolic hallmark of tumor progression and aggressiveness in colorectal carcinoma. *Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc* 30, 267-277

39. Perez-Enciso, M., and Tenenhaus, M. (2003) Prediction of clinical outcome with microarray data: a partial least squares discriminant analysis (PLS-DA) approach. *Human genetics* 112, 581-592
Proteome of GIST subgroups

40. Seshacharyulu, P., Ponnusamy, M. P., Haridas, D., Jain, M., Ganti, A. K., and Batra, S. K. (2012) Targeting the EGFR signaling pathway in cancer therapy. *Expert opinion on therapeutic targets* 16, 15-31

41. Kluijt, I., Siemerink, E. J. M., Ausems, M. G. E. M., van Os, T. A. M., de Jong, D., Simões-Correia, J., van Krieken, J. H., Ligtenberg, M. J., Figueiredo, J., van Riel, E., Sijmons, R. H., Plukker, J. T. M., van Hillegersberg, R., Dekker, E., Oliveira, C., Cats, A., and Hoogerbrugge, N. (2012) CDH1-related hereditary diffuse gastric cancer syndrome: Clinical variations and implications for counseling. *International Journal of Cancer* 131, 367-376

42. Wang, X., Xie, J., and Proud, C. G. (2017) Eukaryotic Elongation Factor 2 Kinase (eEF2K) in Cancer. *Cancers* 9, 162

43. Su Y Q, Sugiura K, Sun F, et al. MARF1 regulates essential oogenic processes in mice[J]. *Science*, 2012, 335(6075): 1496-1499.

44. Bersani, C., Xu, L. D., Vilborg, A., Lui, W. O., and Wiman, K. G. (2014) Wig-1 regulates cell cycle arrest and cell death through the p53 targets FAS and 14-3-3sigma. *Oncogene* 33, 4407-4417

45. Li, S., Wang, L., Berman, M. A., Zhang, Y., and Dorf, M. E. (2006) RNAi screen in mouse astrocytes identifies phosphatases that regulate NF-kappaB signaling. *Molecular cell* 24, 497-509

46. Hornstein, M., Hoffmann, M. J., Alexa, A., Yamanaka, M., Muller, M., Jung, V., Rahnenfuhrer, J., and Schulz, W. A. (2008) Protein phosphatase and TRAIL receptor genes as new candidate tumor genes on chromosome 8p in prostate cancer. *Cancer genomics & proteomics* 5, 123-136

47. Wu, L., Candille, S. I., Choi, Y., Xie, D., Jiang, L., Li-Pook-Than, J., Tang, H., and Snyder, M. (2013) Variation and genetic control of protein abundance in humans. *Nature* 499, 79-82

48. Nakamura, Y., Patrushev, N., Inomata, H., Mehta, D., Urao, N., Kim, H. W., Razvi, M., Kini, V., Mahadev, K., Goldstein, B. J., McKinney, R., Fukai, T., and Ushio-Fukai, M. (2008) Role of Protein Tyrosine Phosphatase 1B in VEGF Signaling and Cell-Cell Adhesions in Endothelial Cells. *Circulation research* 102, 1182-1191

49. Lessard, L., Stuible, M., and Tremblay, M. L. (2010) The two faces of PTP1B in cancer. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics* 1804, 613-619

50. Wiener, J. R., Kerns, B. J., Harvey, E. L., Conaway, M. R., Iglehart, J. D., Berchuck, A., and Bast, R. C., Jr. (1994) Overexpression of the protein tyrosine phosphatase PTP1B in human breast cancer: association with p185c-erbB-2 protein expression. *Journal of the National Cancer Institute* 86, 372-378

51. Liu, H., Wu, Y., Zhu, S., Liang, W., Wang, Z., Wang, Y., Lv, T., Yao, Y., Yuan, D., and Song, Y. (2015) PTP1B promotes cell proliferation and metastasis through activating src and ERK1/2 in non-small cell lung cancer. *Cancer letters* 359, 218-225

52. Wiener, J. R., Hurteau, J. A., Kerns, B. J., Whitaker, R. S., Conaway, M. R., Berchuck, A., and Bast, R. C., Jr. (1994) Overexpression of the tyrosine phosphatase PTP1B is associated with human ovarian carcinomas. *American Journal of Obstetrics and Gynecology* 170, 1177-1183

53. Lessard, L., Labbé, D. P., Deblois, G., Bégin, L. R., Hardy, S., Mes-Masson, A.-M., Saad, F., Trotman, L. C., Giguère, V., and Tremblay, M. L. (2012) PTP1B Is an Androgen Receptor–Regulated Phosphatase That Promotes the Progression of Prostate Cancer. *Cancer research* 72, 1529-1537

54. Warabi, M., Nemoto, T., Ohashi, K., Kitagawa, M., and Hirokawa, K. (2000) Expression of Protein Tyrosine Phosphatases and Its Significance in Esophageal Cancer. *Experimental and Molecular Pathology* 68, 187-195

55. Dube, N., Bourdeau, A., Heinonen, K. M., Cheng, A., Loy, A. L., and Tremblay, M. L. (2005) Genetic ablation of protein tyrosine phosphatase 1B accelerates lymphomagenesis of p53-null mice through the regulation of B-cell development. *Cancer research* 65, 10088-10095

56. Medeiros, F., Corless, C. L., Duensing, A., Hornick, J. L., Oliveira, A. M., Heinrich, M. C., Fletcher, J. A., and Fletcher, C. D. (2004) KIT-negative gastrointestinal stromal tumors: proof of concept and therapeutic implications. *The American journal of surgical pathology* 28, 889-894
Proteome of GIST subgroups

57. Li, J., Ye, Y., Wang, J., Zhang, B., Qin, S., Shi, Y., He, Y., Liang, X., Liu, X., Zhou, Y., Wu, X., Zhang, X., Wang, M., Gao, Z., Lin, T., Cao, H., Shen, L., and Chinese Society Of Clinical Oncology Csco Expert Committee On Gastrointestinal Stromal, T. (2017) Chinese consensus guidelines for diagnosis and management of gastrointestinal stromal tumor. *Chinese journal of cancer research = Chung-kuo yen cheng yen chiu* 29, 281-293

**Acknowledgements**

This work was supported by the National Natural Science Foundation of China (No. 81570060 and 31870826), National Youth Thousand Talents Program, the Sichuan Thousand Talents Program, and National Key R&D Program of China (No. 2018YFC1312301, SQ2018YFC200004, and 2016YFC1201700). We would like to thank the Department of West China hospital biobank for its assistance in specimen collection, and the Department of Pathology of the Second People's Hospital of Neijiang City for providing technical assistance during the experiment.

**Data Availability**

All mass spectrometry raw data and the MaxQuant output tables have been deposited to iProX and are available using the iProX accession: IPX0001353000.

**Competing financial interests**

There is no competing financial interest.

**Figure Legends**

**Table 1. Clinicopathologic features of the collected GIST cases for proteome profiling.**

Thirteen GIST cases were selected for quantitative proteomic studies. All these tissues were frozen in liquid nitrogen in 30 min after surgery. Among them, three cases were diagnosed as very low/low risk, five cases as intermediate risk and five cases as high risk.

**Figure 1. Quantitative proteomic profiling of GIST.** (A) The workflow of quantitative proteomic analysis of human GIST subgroups. Three pairs of NIH-L GIST tissues, five pairs of NIH-L GIST tissues and five pairs of NIH-L GIST tissues were analyzed separately using TMT-10plex based quantitative method. (B) Venny diagram showed the percentage of GIT proteins identified in our
Proteome of GIST subgroups

GIST proteomics. (C) Venny diagram showed the number of quantified proteins in GIST subgroups. (D) The alteration of some well-known GIST related proteins. The change patterns were consistent with previous reports. (E) The Pearson correlation coefficient \((r)\) of 4930 quantified proteins used to assess the tissue similarity between GIST subgroups.

**Figure 2. Common proteomic features of GIST subgroups.** (A) The statistics showed the number of proteins upregulated or downregulated in all 13 GIST cases. (B) The cellular distribution of upregulated and downregulated proteins. The first number and the second number in the parentheses means the numbers of upregulated and downregulated proteins, respectively. (C) KEGG analysis of the upregulated and downregulated proteins, respectively. cyP-450 was short for cytochrome P450. (D) Proteins that were enriched in the spliceosome pathway. (E) The downregulated carbon metabolism, glycolysis/gluconeogenesis and TCA cycle pathways. (F) The quantified proteins downregulated in carbon metabolism, glycolysis/gluconeogenesis and TCA cycle pathways.

**Figure 3. Proteomic variances in GIST subgroups.** (A) GOCC analyses of significantly expressed proteins of each GIST subgroup. 989 proteins of NIH-L subgroup, 792 proteins of NIH-I subgroup and 1245 proteins of NIH-H subgroup were used for GOCC analyses. (B) GSEA enrichment for GIST subgroups. 6053 proteins of NIH-L subgroup, 6598 proteins of NIH-I subgroup and 6955 proteins of NIH-H subgroup were used for GSEA analyses. (C) Unbiased clustering and cluster-specific enrichment analysis of 131 differentially expressed proteins between GIST subgroups. The heatmap was drawn using “Heatmap” R package, and the rows were scaled. Z-score is a statistic normalization using 13 T/N ratios of GIST subgroups through an arithmetic included in “Heatmap” R package. (D) Partial least squares discrimination analysis (PLS-DA) of 131 differentially expressed proteins between GIST subgroups. PLS-DA analysis was performed using “mixOmics” R package.

**Figure 4. Oncoproteins and tumor suppressors in GIST.** (A) The changing patterns of quantified oncoproteins and TSs in all 13 GIST cases. And the size of circle positively correlated
Proteome of GIST subgroups

with the number of fold-change. Blue represented that the proteins were downregulated, and red mean that the proteins were upregulated. (B) The corresponding pathways that the quantified oncoproteins and TSs belonged to. The red color proteins were oncoproteins, and the black color proteins were TSs. In the parentheses, the first number represented the average ratio in NIH-L subgroup, the second number represented the average ratio in NIH-I subgroup, and the third number represented the average ratio in NIH-H subgroup.

Figure 5. Kinases and phosphatases in GIST. (A-B) The changing patterns of quantified phosphatases (A) and Kinases (B) in all 13 GIST cases. The heatmap was created in Excel. (C) The fold changes (T/N) of phosphatase PTPN1 in all GIST subgroups measured by MS. (D) Western blotting results of PTPN1 in paired tumorous and non-tumorous tissues of GIST subgroups. (E) Immunohistochemical analysis of PTPN1 using GIST tumorous tissues. The left picture represented PTPN1 with positive expression, and the right picture indicated PTPN1 was negatively expressed. (F) The average IOD of PTPN1 in GIST subgroups. P-values were calculated by student-t test. (G) The positive ratios of PTPN1 in GIST subgroups based on immunohistochemical analysis results of 117 GIST cases. 14/131 cases were excluded because the observable areas of these tissues in the microarray were too small. P-values were calculated by student-t test. (H) The relationship between PTPN1 abundance and tumor metastasis. High proportions of metastatic cases were observed in the groups with lowly expressed PTPN1.
Table 1. Clinicopathologic features of the proteomic set samples

| Patient No | Age | Gender | Site            | Tumor size in largest dimension (cm) | Mitosis count (per 50HFPs) | Risk classification * |
|------------|-----|--------|-----------------|--------------------------------------|---------------------------|-----------------------|
| 1          | 56  | M      | Stomach         | 7                                    | 12                        | High                  |
| 2          | 61  | F      | Stomach         | 15                                   | 10                        | High                  |
| 3          | 70  | F      | Stomach         | 5.5                                  | 7                         | High                  |
| 4          | 63  | F      | Small intestine | 10                                   | 16                        | High                  |
| 5          | 54  | F      | Stomach         | 6                                    | 9                         | High                  |
| 6          | 43  | F      | Stomach         | 5                                    | 8                         | Intermediate          |
| 7          | 74  | F      | Stomach         | 7                                    | 4                         | Intermediate          |
| 8          | 52  | F      | Stomach         | 6                                    | 3                         | Intermediate          |
| 9          | 59  | F      | Stomach         | 5                                    | 4                         | Intermediate          |
| 10         | 60  | M      | Stomach         | 3                                    | 7                         | Intermediate          |
| 11         | 55  | F      | Stomach         | 4                                    | 1                         | Low                   |
| 12         | 50  | M      | Stomach         | 2.8                                  | 3                         | Low                   |
| 13         | 58  | M      | Stomach         | 1.8                                  | 1                         | Low                   |

Note: M: male; F: female

*Prognostic classification based on tumor size and mitosis count.
A. Tissue collection
   Protein extraction
   Reduction & alkylation & digestion
   TMT labeling & quenching & pooling

   Data
      Fractionation
      LC-MS/MS
      Data analysis

B. GIST
   647 (4.2%)
   8530 (55.9%)
   6060 (39.9%)

   GIT

   NIH-L
   171 (1.9%)
   7230 (78.8%)
   474 (5.2%)

   NIH-I
   151 (1.6%)
   245 (2.7%)
   701 (7.6%)

   NIH-H
   205 (4%)
   7230 (78.8%)
   474 (5.2%)

D. Ratio (Tumor/Normal)
   NIH-L
   NIH-I
   NIH-H

E. Heatmap of expression levels for KIT, ANO1, KCTD10, KCTD12, DDX39B, ROR2, PML, BCL2, and DDX39A.
Figure 2

A) Scatter plot showing the distribution of 187 proteins and 517 proteins across different conditions.

B) Diagram illustrating the localization of proteins within cellular compartments:
- Extracellular matrix
- Extracellular space
- Lysosome
- Cytoskeleton
- Mitochondrion
- Nucleus
- Endoplasmic reticulum
- Golgi apparatus
- Vesicle
- Microfiber
- Cytosol

C) Heatmap representing the Spliceosome mRNA surveillance pathway and various metabolic pathways:
- Carbon metabolism
- Glycolysis/Gluconeogenesis
- Glutathione metabolism
- Drug metabolism
- Propanoate metabolism
- Citrate cycle (TCA cycle)

D) Heatmap showing the expression levels of various proteins in NIH-L, NIH-I, and NIH-H conditions.

E) Metabolic pathway diagram:
- Glucose
- Fructose-6-phosphate
- Fructose-1,6-bisphosphate
- Phosphoenolpyruvate
- Glyceraldehyde-3-phosphate
- Glyceraldehyde-3-phosphate dehydrogenase (GPDH)
- Phosphoenolpyruvate carboxykinase (PCK)
- Pyruvate kinase (PK)
- Acetyl-CoA
- Acetyl-CoA synthetase (ACSS)
- Aldehyde dehydrogenase (ALDH)
- Alcohol dehydrogenase (AKR1)
- Succinyl-CoA
- Succinyl-CoA synthetase (SUCLG1/2)

F) Bar graph showing the ratio (T/N) for various metabolites across NIH-L, NIH-I, and NIH-H conditions.
Figure 3

A

Catalytic step 2 spliceosome
Cell-cell adherens junction
Cytoplasm
Cytoplasmic stress granule
Cytoplasmic vesicle
Cytosol
Exon-exon junction complex
Extracellular exosome
Focal adhesion
Intracellular membrane-bounded
Organelle
Intracellular ribonucleoprotein complex
Lysosomal lumen
Membrane
Mitochondrial matrix
Nuclear envelope
Nuclear speck
Nucleolus
Nucleoplasm
Nucleus
Perinuclear region of cytoplasm
Protein complex
Spliceosomal complex
U4/U6 x U5 tri-snRNP complex
Viral nucleocapsid

B

G2M checkpoint
Interferon alpha response
E2F targets
Interferon gamma response
UV response DN
Hedgehog signaling
EMT
DNA repair
MYC targets V1
Mitotic spindle
Angiogenesis
Oxidative phosphorylation
Fatty acid metabolism
Adipogenesis
Xenobiotic metabolism
ROS pathway
Estrogen response late
Glycolysis
KRAS signaling DN
Androgen response
Cholesterol homeostasis
Hypoxia
Estrogen response early
Bile acid metabolism
P53 pathway
Peroxisome
Coagulation
Myogenesis

C

D

X-variante 1: 46% expl.var
X-variante 2: 20% expl.var

Sphingolipid biosynthetic process
Intracellular membrane-bounded organelle
Protein transport
Cytosol
Membrane budding
ESCRT III complex
Endocytosis
Regulation of proteasomal ubiquitin-dependent protein catabolic process
Cytoplasm
Ubiquitin mediated proteolysis

X-variante 1: 58% expl.var
X-variante 2: 9% expl.var
Figure 5

A) Heatmap showing gene expression levels for various genes in NIH-L, NIH-I, and NIH-H cells. The expression levels are color-coded with Log2(T/N) ratio.

B) Heatmap with additional gene expression data.

C) Bar graph showing protein ratio (T/N) for PTPN1 across NIH-L, NIH-I, and NIH-H.

D) Western blot images showing expression of PTPN1 in NIH-L, NIH-I, and NIH-H.

E) Images of positive and negative KIT staining.

F) Graph showing average IOD (*10^4) with ns significance.

G) Bar graph showing proportions of sample numbers.

H) Usage of IOD for classifying metastatic cases.

- IOD > 10^4: 0/25
- IOD: 10^3 ~ 10^4: 3/49
- IOD < 10^3: 4/43