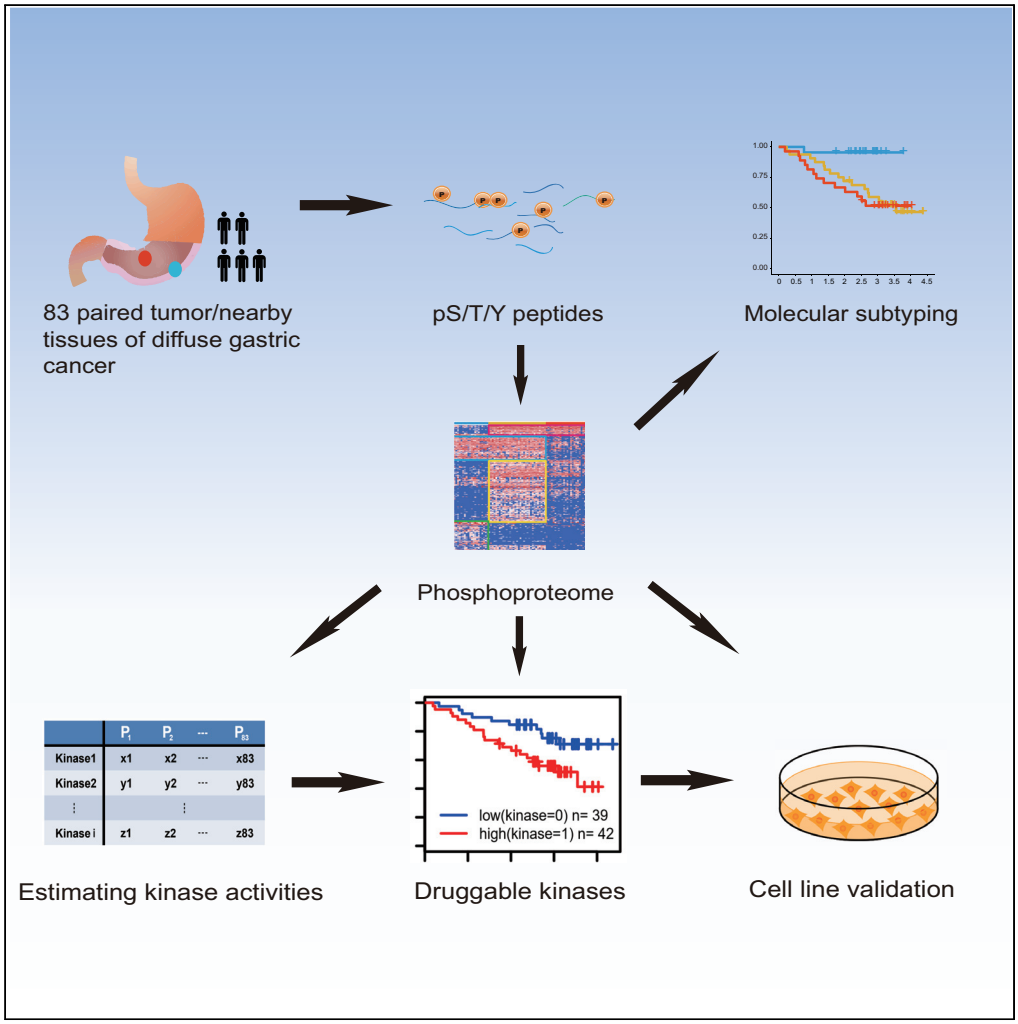
iScience

CellPress

Article

Phosphoproteomics Enables Molecular Subtyping and Nomination of Kinase Candidates for Individual Patients of Diffuse-Type Gastric Cancer



Mengsha Tong, Chunyu Yu, Jinwen Shi, ..., Lin Shen, Tingting Li, Jun Qin

hefc@nic.bmi.ac.cn (F.H.)
linshenpku@163.com (L.S.)
litt@hsc.pku.edu.cn (T.L.)
jqin1965@126.com (J.Q.)

HIGHLIGHTS

A phosphoproteomic landscape of diffuse-type gastric cancer (DGC) was depicted

DGC could be classified into three subtypes based on phosphorylation data

A bioinformatics workflow was used to identify 16 kinases as potential drug targets

A patient-specific strategy for nomination of kinases was proposed

Tong et al., iScience 22, 44–57
December 20, 2019
https://doi.org/10.1016/j.isci.2019.11.003

iScience

CellPress

Article

Phosphoproteomics Enables Molecular Subtyping and Nomination of Kinase Candidates for Individual Patients of Diffuse-Type Gastric Cancer

Mengsha Tong,[1,5,8] Chunyu Yu,[2,7,8] Jinwen Shi,[1,8] Wenwen Huang,[3,8] Sai Ge,[3,8] Mingwei Liu,[1] Lei Song,[1] Dongdong Zhan,[6] Xia Xia,[1] Wanlin Liu,[1] Jinwen Feng,[6] Wenhao Shi,[1] Jiafu Ji,[3] Jing Gao,[3] Tieliu Shi,[6] Weimin Zhu,[1] Chen Ding,[1,4] Yi Wang,[1] Fuchu He,[1,4,*] Lin Shen,[3,*] Tingting Li,[2,7,*] and Jun Qin[1,4,9,*]

SUMMARY

The diffuse-type gastric cancer (DGC) constitutes a subgroup of gastric cancer with poor prognosis and no effective molecular therapies. Here, we report a phosphoproteomic landscape of DGC derived from 83 tumors together with their nearby tissues. Based on phosphorylation, DGC could be classified into three molecular subtypes with distinct overall survival (OS) and chemosensitivity. We identified 16 kinases whose activities were associated with poor OS. These activated kinases covered several cancer hallmark pathways, with the MTOR signaling network being the most frequently activated. We proposed a patient-specific strategy based on the hierarchy of clinically actionable kinases for prioritization of kinases for further clinical evaluation. Our global data analysis indicates that in addition to finding activated kinase pathways in DGC, large-scale phosphoproteomics could be used to classify DGCs into subtypes that are associated with distinct clinical outcomes as well as nomination of kinase targets that may be inhibited for cancer treatments.

INTRODUCTION

Gastric cancer (GC), one of the most common and fatal diseases in East Asian, is a heterogeneous disease with diverse histological and molecular characteristics (Wang et al., 2011; Cancer Genome Atlas Research Network, 2014). The widely used Lauren classification stratifies GC into diffuse-type, intestinal-type, and mixed types (Wu et al., 2019). Diffuse-type gastric cancer (DGC) has the worst prognosis and lacks treatment options, particularly for targeted therapy.

Over the past decades, genomic landscapes of GC has been mapped (Cancer Genome Atlas Research Network, 2014; Tan et al., 2011; Cristescu et al., 2015). The Cancer Genome Atlas (TCGA) project uncovered four molecular subtypes of gastric cancer. The Asian Cancer Research Group (ACRG) also described four molecular subtypes based on gene expression data of GC (Cancer Genome Atlas Research Network, 2014). Previously, we mapped the proteomic landscape of DGC of 84 paired tumors and their nearby tissues and showed that based on the altered protein expression alone, DGC could be subtyped into three major classes (PX1-3) that are associated with clinical outcomes (Ge et al., 2018). Our study allowed the nomination of more than 20 proteins that function in cancer growth, ROS and metabolism, cell-cell adhesion and adjunction, as well as immune-response pathways as potential drug targets.

The rapid development of kinase inhibitors raised the hope for targeted therapy and even truly individualized therapy (Ferguson and Gray, 2018). The concurrent development of phosphoproteomics that focused on identification and quantification of phosphorylated amino acid residues in proteins in biological specimens provided a needed support for the utilization of kinase inhibition as a therapy (Casado et al., 2017; Wu et al., 2019). The state-of-the-art phosphoproteomics now can measure tens and thousands of phosphorylation sites, from which kinase activities can be inferred and targeted kinase therapies may be developed (Casado et al., 2017; Wu et al., 2019). Large-scale mappings of phosphorylation landscapes were carried out in several cancers including breast cancer (Mertins et al., 2016), ovarian cancer (Mertins et al., 2016), prostate cancer (Drake et al., 2016), lung cancer (Rikova et al., 2007), blastoma brain cancer (Liu et al., 2018), gastric cancer (Mun et al., 2019), and hepatocellular carcinoma (Jiang et al., 2019). But the utilization of the phosphoproteomics data was poor, and there were few reports of using

Continued

[1]State Key Laboratory of Proteomics, Joint Laboratory of Gastrointestinal Oncology, Beijing Proteome Research Center, National Center for Protein Sciences (The PHOENIX Center, Beijing), Beijing Institute of Lifeomics, Beijing 102206, China

[2]Department of Biomedical Informatics, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China

[3]Department of Gastrointestinal Oncology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Peking University Cancer Hospital and Institute, Beijing, China

[4]State Key Laboratory of Genetic Engineering and Collaborative Innovation Center for Genetics and Development, School of Life Sciences, Institute of Biomedical Sciences, Fudan University, Shanghai 200433, China

[5]State Key Laboratory of Cellular Stress Biology, Innovation Center for Cell Signaling Network, School of Life Sciences, Xiamen University, Xiamen, Fujian 361102, China

[6]Center for Bioinformatics, East China Normal University, Shanghai 200241, China

[7]Institute of Systems Biomedicine, School of Basic Medical Sciences, Peking University Health Science Center, Beijing 100191, China

[8]These authors contributed equally

[9]Lead Contact

phosphoproteomics to subtype cancers to predict patient prognosis or nominate kinase targets for treatments in a systematic manner.

Here, we report the phosphorylation landscape of DGC as a part of a concerted effort of the Chinese Human Proteome Project (CNHPP). We carried out a quantitative measurement of phosphoproteomics of 83 DGC tumors and their matching nearby tissues. Applying an unbiased bioinformatics analysis workflow that we developed recently (Tong et al., 2018), we showed that using phosphoproteomics data alone, DGC could be classified into three subtypes that associated with distinct clinical outcomes and we nominated druggable kinase candidates for each individual patient, allowing for drug prioritization on an individualized basis for DGC patients.

RESULTS

General Features of the DGC Phosphoproteome

We enriched phosphopeptides with TiO2-coupled beads from tryptic digests of tumor samples and their nearby tissues of 83 DGC patients obtained from the tumor tissue bank of Beijing Cancer Hospital (Ge et al., 2018) (Table S1) and measured enriched phosphopeptides with LC-MS/MS. Phosphorylation was identified through database using a data analysis platform Firmiana (Feng et al., 2017) (Figure 1A). After quality control including the false discovery rate (FDR) of peptides and Mascot delta score (Wu et al., 2011) for phosphorylation sites (Figure 1B), 28,016 phosphorylation sites were used for subsequent analysis (Figure S1, Table S2), which include 22,744 (81.18%) phosphoserine sites, 4,889 (17.45%) phosphothreonine sites, and 383 (1.37%) phosphotyrosine sites (Figure 1C).

Among the 28,016 phosphorylation sites in the DGC phosphoproteomics, 21,282 sites were found in both tumors and nearby tissues; 4,447 sites and 2,287 sites were detected only in tumors and nearby tissues, respectively (Figure S1B). Principle component analysis (PCA) revealed that tumors could be separated from the nearby tissues based on phosphorylation profiles (Figure 1D). An SAM (significance analysis of microarray) analysis (Tusher et al., 2001) identified 445 upregulated and 819 downregulated phosphorylation sites in tumors compared with nearby tissues (FDR <0.01 by SAM and differential expression percentage >0.5 or <−0.5, Figure S2A, Table S3). Proteins with upregulated phosphorylation in tumors were enriched in cell-cycle-related pathways (DNA replication and cell division), cell-cell adhesion, DNA repair, and mRNA splicing pathways (Figure 1E, Table S4), whereas proteins with upregulated phosphorylation in the nearby tissues were enriched in cell-cell adhesion, gastric acid secretion, and regulation of Rho protein signal transduction.

Subtypes of DGC and Their Associations with OS and Chemosensitivity

Based on the intensity of 28,016 phosphorylation sites in the tumors (Figure 1B, Table S2), we employed consensus clustering (Wilkerson and Hayes, 2010) to identify DGC subtypes. Three clusters (Ph1-3) were apparent (Figures S2B, S2E, and S3A). We further identified 302 differentially phosphorylated sites in tumors among the subtypes (Anova, FDR<0.001, Table S4). As shown in Figure 2A, hierarchical clustering revealed four groups of phosphorylated sites (Tables S5 and S6). To investigate subtype-specific pathway alterations, we further identified significantly altered phosphorylation sites between tumors and nearby tissues within the same subtype (FDR <0.01 by SAM and differential expression percentage >0.5 or <–0.5, Table S7). Ph1 to Ph3 contained 28, 133, and 5 uniquely upregulated phosphorylation sites in tumors, respectively, and many phosphorylation sites were downregulated in tumor (Figure 2B). These dysregulated phosphoproteins suggested that the Ph1 subtype was more or less normal in the basic function of the stomach while upregulated rRNA processing and RNA polymerase II promoter activity (Figures 2C and 2D, Table S7). The Ph2 subtype mainly upregulated DNA metabolic process and DNA repair while losing the basic function of the stomach including gastric acid secretion (Figures 2C and 2D, Table S7). The Ph3 subtype upregulated chromosome segregation and mainly lost cell-cell interaction and communications (Figures 2C and 2D, Table S7)

We further investigated the OS of these patients. The Ph1 subtype had the best OS, whereas both Ph2 and Ph3 had worse survival (logrank p=0.012, Figure 2E). Moreover, we found that the Ph2 patients tend to be more sensitive to chemotherapy (logrank p = 0.044, Figure 2F), but the Ph1 and Ph3 groups exhibited no statistically significant prognosis improvement by chemotherapy (logrank p> 0.1, Figure 2F). In addition, a multivariable Cox analysis showed that the phosphoproteomics subtyping remained significantly associated with patients' OS after adjusting for age, gender, adjuvant chemotherapy, tumor site, stage, and

*Correspondence:
hefc@nic.bmi.ac.cn (F.H.),
linshenpku@163.com (L.S.),
litt@hsc.pku.edu.cn (T.L.),
jqin1965@126.com (J.Q.)

https://doi.org/10.1016/j.isci.2019.11.003

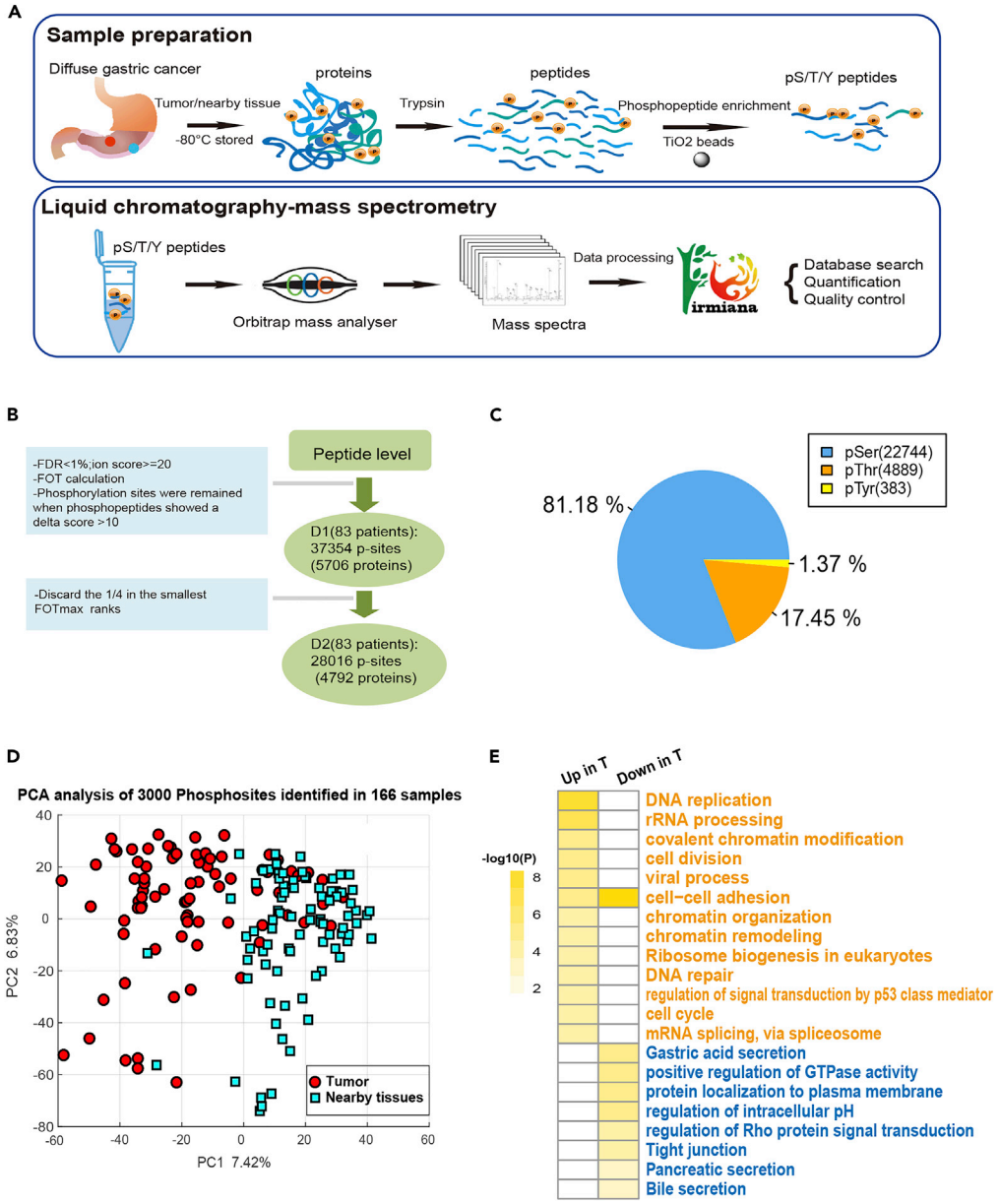


Figure 1. A Summary of Phosphoproteome Analysis of Diffuse-Type Gastric Cancer

(A) General workflow of the phosphopeptide enrichment and quantitative mass spectrometry protocol.

(B)Phosphoproteomic datasets filtered at different levels for various statistical analyses.

(C) Distribution of modification types.

(D) Principal component analysis (PCA) to visualize tumor and nearby tissue samples.

(E) Top ranked pathways that are significantly altered in tumors compared with nearby tissues (FDR<0.1, the minimum proteins in one pathway is 5). Yellow words represent the pathways enriched by the proteins with upregulated phosphorylation sites in tumors compared with nearby tissues; blue words represent the pathways enriched by the proteins with downregulated phosphorylation sites in tumors compared with nearby tissues.

Also see Figure S1.

TP53 (using Ph1 as the reference, HR = 8.67, p = 0.038 for Ph2; HR = 9.87, p = 0.029 for Ph3, Table 1). Other clinical characteristics were also associated with the Ph1–3 subtypes. Firstly, age (≥50 years vs <50 years), stage (I/II/III/IV), and MSI status (microsatellite unstable vs microsatellite stable) in each subtype were significantly different (Table S8, Figure S3B, Fisher-test, p<0.05). The Ph1 subtype contained younger

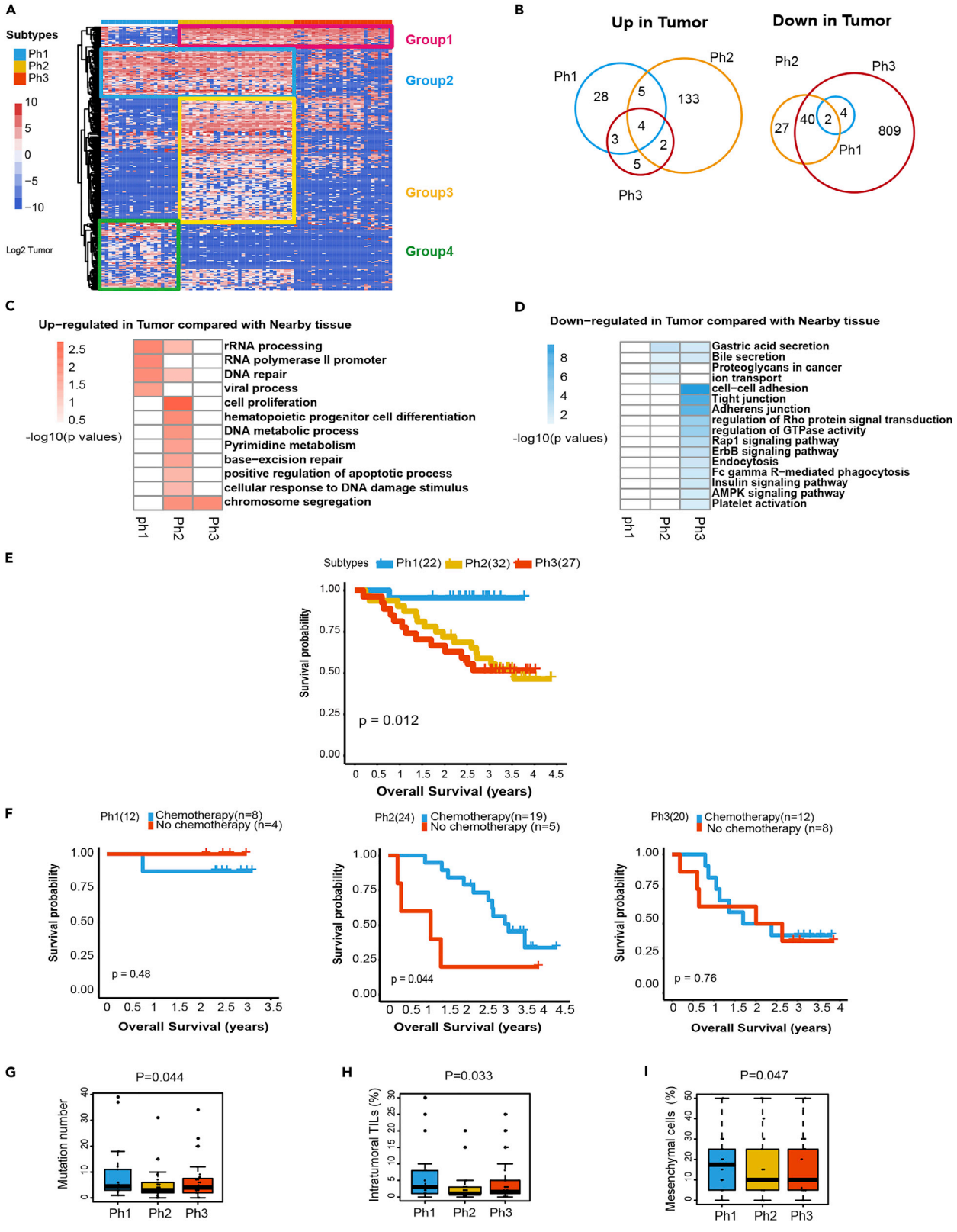

A
Subtypes
Ph1
Ph2
Ph3
10
5
0
−5
−10
Log2 Tumor
Group1
Group2
Group3
Group4
B
Up in Tumor
Ph1
Ph2
Ph3
28
5
133
4
3
2
5
Down in Tumor
Ph2
Ph3
Ph1
27
40
2
4
809
C
Up−regulated in Tumor compared with Nearby tissue
2.5
2
1.5
1
0.5
-log10(p values)
rRNA processing
RNA polymerase II promoter
DNA repair
viral process
cell proliferation
hematopoietic progenitor cell differentiation
DNA metabolic process
Pyrimidine metabolism
base−excision repair
positive regulation of apoptotic process
cellular response to DNA damage stimulus
chromosome segregation
ph1
Ph2
Ph3
D
Down−regulated in Tumor compared with Nearby tissue
8
6
4
2
-log10(p values)
Gastric acid secretion
Bile secretion
Proteoglycans in cancer
ion transport
cell−cell adhesion
Tight junction
Adherens junction
regulation of Rho protein signal transduction
regulation of GTPase activity
Rap1 signaling pathway
ErbB signaling pathway
Endocytosis
Fc gamma R−mediated phagocytosis
Insulin signaling pathway
AMPK signaling pathway
Platelet activation
ph1
Ph2
Ph3
E
Subtypes
Ph1(22)
Ph2(32)
Ph3(27)
Survival probability
1.00
0.75
0.50
0.25
0.00
p = 0.012
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5
Overall Survival (years)
F
Ph1(12)
Chemotherapy(n=8)
No chemotherapy (n=4)
Survival probability
p = 0.48
0 0.5 1 1.5 2 2.5 3 3.5
Overall Survival (years)
Ph2(24)
Chemotherapy(n=19)
No chemotherapy (n=5)
Survival probability
p = 0.044
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5
Overall Survival (years)
Ph3(20)
Chemotherapy(n=12)
No chemotherapy (n=8)
Survival probability
p = 0.76
0 0.5 1 1.5 2 2.5 3 3.5 4
Overall Survival (years)
G
P=0.044
Mutation number
Ph1
Ph2
Ph3
H
P=0.033
Intratumoral TILs (%)
Ph1
Ph2
Ph3
I
P=0.047
Mesenchymal cells (%)
Ph1
Ph2
Ph3

Figure 2. Phosphoproteome Subtyping of DGC with Different Overall Survival and Chemosensitivity

(A) 301 sites differentially phosphorylated among the subtypes (Anova, FDR<0.001).

(B) The number of differentially expressed phosphorylation sites in tumors compared with nearby tissues for each type.

(C and D) Heatmaps of selected phosphoproteins representing major altered signaling pathways in each type.

(E) The association of molecular subtypes with overall survival of patients; Kaplan-Meier analysis, p value from logrank test.

(F) The association of adjuvant chemotherapy with overall survival in each subtype. Only the patients who have the chemotherapy information were shown.

(G–I) Other clinical parameters across 83 patients in Ph1–3. Intratumoral TILs, intratumoral tumor-infiltrating lymphocytes.

Also see Figures S2 and S3.

patients (<50 years), as well as early stage (stage II) patients, whereas Ph2 and Ph3 contained older and more advanced stage (stage III–IV) patients. Cancer driver gene mutation rate that we measured previously (Ge et al., 2018) seemed to be higher in Ph1 than in Ph2 and Ph3 (Figure 2G, p = 0.044, Wilcox test), for example, the Ph1 subtype contained three out of five patients with MSI. Secondly, Ph1 was enriched with higher intratumoral TILs (tumor-infiltrating lymphocytes) and mesenchymal cells than Ph2 and Ph3 (p<0.05, Wilcox test Figures 2H and 2I). It has been reported that increasing intratumoral TILs implies a better prognosis in GC (Kang et al., 2017; Yu et al., 2016; Grogg et al., 2003). These correlations were consistent with and may explain the OS difference between the three molecular subtypes. Previously, we subtyped DGC of the same samples into three subtypes (PX1-3) based on protein profiling alone (Ge et al., 2018). As showed in Figure S3B, most of the patients with worse survival in the Ph2 and Ph3 (85.25%) were also classified to PX2-3, demonstrating that the two methods reached good agreement. Moreover, Ph1 appeared to include more patients with better survival than PX1 did, suggesting that phosphoproteomics-based subtyping may be superior to that from protein profiling for correlation with OS.

Nomination of Kinases as Potential Therapeutic Targets

As the activity of a kinase can be inferred by the intensity of its substrates (Casado et al., 2013), we built a kinase-substrate dataset that contained 3,321 substrates and 250 kinases to find activated kinases from the measured substrates (Table S9, Figure 3A). Our DGC phosphoproteomes contained 1,896 protein substrates in the kinase-substrate database (see Methods, Table S9). For each kinase, the average number of substrate sites is 13 and the median number is 4 (Figure S4, Table S9). We then calculated the normalized value of p-site-FOT^{TiO2}/protein-FOT^{profiling} to correct for altered protein expression (Wu et al., 2011) and obtained a normalized ratio of p-site between tumor and nearby tissue for each patient. We used the average fold difference for all detected substrate sites for the same kinase as a measurement of the kinase activation/inhibition (Table S10). The highly activated kinases (with top ranked values of average fold difference and top frequencies detected in the patients) included PRKACA, CSNK2A1, CDK1/2, MAPK1/3, GSK3B, PRKCA, AKT1, CDK4, and CDK6 (Figure S5A). We also adapted three other methods including Z-test, kinase substrate enrichment analysis (KSEA), and the multiple linear regression (MLR) model to identify more activated kinases (Hernandez-Armenta et al., 2017) (Figure 3A, see Methods). We computed "kinase activity" using each of these four methods for every patient and generated a kinase-patient matrix, respectively (Figure 3A).

In order to evaluate the accuracy of the four prediction methods, we calculated correlation of the predicted kinases activities with phosphorylation intensities of the kinase activation loops. Phosphorylation of the activation loop is often critical in regulating kinase activity in many cases (Nolen et al., 2004). We found that the kinase activities predicted by the four methods were positively correlated with the intensities of the 43 activation loop phosphorylation sites that we could measure (Figure 3A, panel 2; Figure S6). The Mean value, KSEA, and regression methods performed significantly better than the Z-test method (Wilcox test p<0.05; Figure 3A, panel 2) and were retained for subsequent analysis.

We then postulated that activated kinases might be therapeutic targets if their activities are associated with poor OS. We stratified patients into two groups according to high and low kinase activity. The cut-off value was individually determined by the lowest p value of OS according to the logrank test (Figure 3A, panel 3; Table S11). We identified 19 kinases whose activities were significantly associated with poor OS (10 from the mean value method, 7 from MLR, and 8 from KSEA) (Figure 3A, panel 4). At least two substrates were identified for each kinase with a fold change of >1.5 between tumor and the near-by tissues. Finally, 16 non-overlap kinases were found after controlling the logrank p value <0.05 at least in three continuous cut-offs, which was to reduce the false-positive rate of the survival analysis (Figures S6–S8). The higher activities of these kinases were significantly associated with poor OS (logrank p<0.05, HR>2, Figure 3B, Table 2). We thus nominated these 16 kinases as potential therapeutic kinase targets (PTKT). Using the

Characteristics (n)	Univariate Analysis		Multivariate Analysis	
	HR (95% CI)	p Value	HR (95% CI)	p Value
Age[a]	1.033 (1.002–1.065)	**0.036**	1.039 (0.99–1.08)	0.054
Gender				
Male (50)	1		0.68 (0.30–1.55)	0.36
Female (31)	0.662 (0.323–1.36)	0.26		
Adjuvant chemotherapy			0.60 (0.23–1.58)	0.3
Without (17)	1			
With[b] (64)	0.439 (0.201–0.961)	**0.04**		
Tumor site				
Cardia, GEJ (19)	1			
Body (32)	1.484 (0.569–3.870)	0.42	1.62 (0.53–4.98)	0.4
Antrum (26)	1.342 (0.487–3.696)	0.57	1.27 (0.38–4.27)	0.7
Clinical stage[a]				
(Ib to IV)	1.82 (1.01–3.29)	**0.048**	2.09 (1.12–3.88)	**0.02**
TP53 mutation				
Wild-type (45)	1			
Mutant (36)	1.23 (0.60–2.51)	0.58	1.10 (0.53–2.31)	0.79
Phosphoproteome cluster				
Ph1 (22)	1			
Ph2 (32)	10.53 (1.39–79.64)	**0.023**	8.67 (1.12–67.01)	**0.038**
Ph3 (27)	11.64 (1.52–89.14)	**0.018**	9.86 (1.26–76.95)	**0.029**

Table 1. Univariate and Multivariate Analysis of Overall Survival in 81 Patients

GEJ, gastroesophageal junction; HR, hazard ratio; CI, confidence interval.

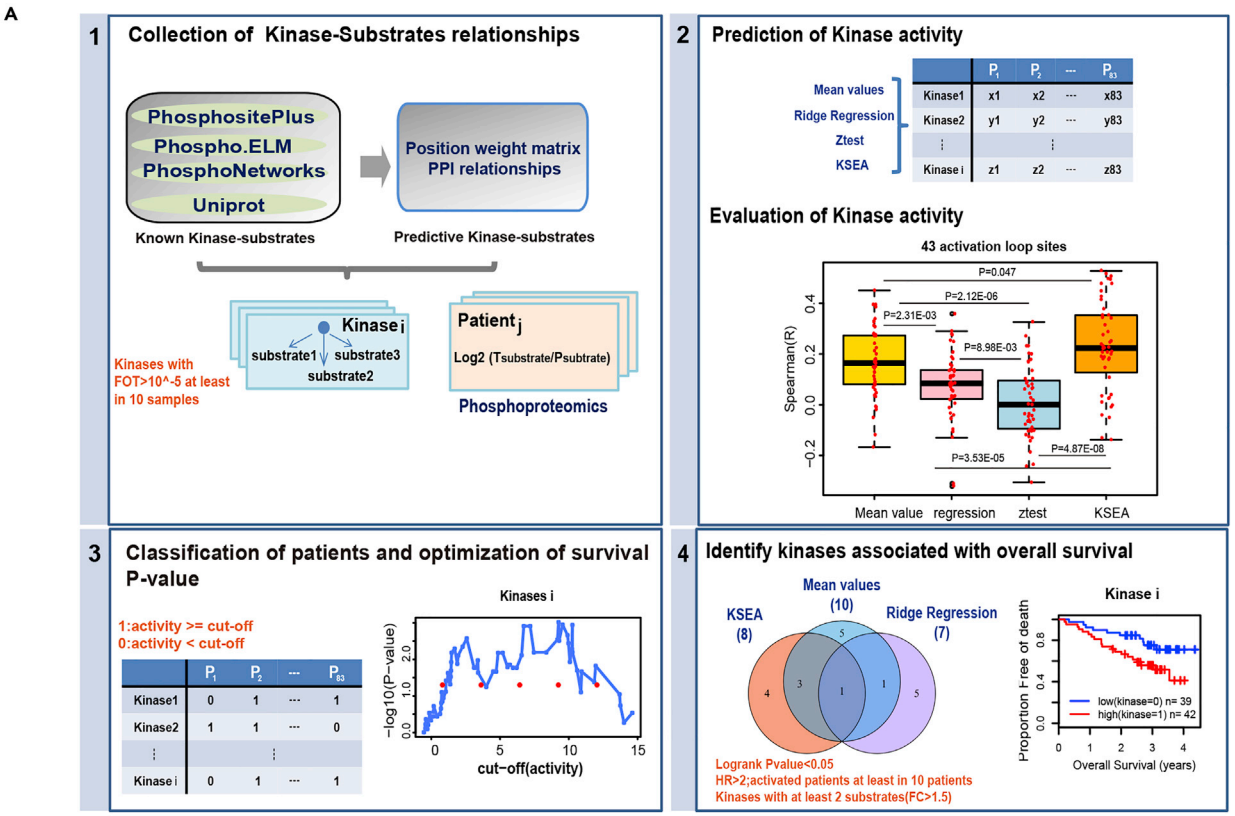
[a]Continuous variable.

[b]Patients proceed at least one cycle of adjuvant chemotherapy. Significant data are emphasized in bold.

same strategy, we could identify only two kinases, namely SRPK2 and MAPKAPK2, whose increase in protein abundance correlated with poor OS (Table S11, Figure S9). SRPK2 was the only kinase that was identified as PTKT based on activity or abundance. Therefore, using kinase activity could identify more PTKTs.

Major Dysregulated Kinase Pathways in DGC

The above 16 PTKTs could be classified into seven pathways: (1) MTOR signaling, (2) cell cycle and apoptosis, (3) MAPK signaling, (4) cell adhesion and actin organization, (5) microtubule affinity, (6) RNA splicing, and (7) G protein-coupled receptor (Figure 4A). The MTOR signaling pathway contained six PTKTs, including PDPK1, MTOR, RPS6KB1, RPS6KA3, PKN1, and STK11, representing the most frequently dysregulated pathway in DGC (Figure 4A). It was reported that overexpression of PDPK1 was associated with poor OS in gastric carcinoma (Gagliardi et al., 2018). The cell cycle pathway ranked the second highly dysregulated pathway including PLK1, CDK7, CDK4, and CDK6. Upregulation of PLK1 and CDK7 in gastric cancer correlated with poor OS (Otsu et al., 2016; Wang et al., 2016). Moreover, CDK4/6 inhibitors demonstrated significant activity against several solid tumors (Goel et al., 2017). Recently, it was reported that MTORC1-S6K1 and CK1 phosphorylate SRPK2 to induce its nuclear translocation and turn on a splicing program to activate lipid metabolism to fuel cancer growth (Lee et al., 2017). Our findings that the activities of MTOR as well as its downstream kinases RPS6KB1/RPS6KA3 and SRPK2 were associated with poor OS (Figure 3B) strongly support the notion that the MTOR-S6K1-SRPK2 signaling is at work in DGC


A
1
Collection of Kinase-Substrates relationships
PhosphositePlus
Phospho.ELM
PhosphoNetworks
Uniprot
Known Kinase-substrates
Position weight matrix
PPI relationships
Predictive Kinase-substrates
Kinase i
substrate1
substrate3
substrate2
Kinases with
FOT>10^-5 at least
in 10 samples
Patient j
Log2 (Tsubstrate/Psubstrate)
Phosphoproteomics
2
Prediction of Kinase activity
Mean values
Ridge Regression
Ztest
KSEA
Kinase1
Kinase2
Kinase i
Evaluation of Kinase activity
43 activation loop sites
Spearman(R)
P=0.047
P=2.12E-06
P=2.31E-03
P=8.98E-03
P=4.87E-08
P=3.53E-05
Mean value
regression
ztest
KSEA
3
Classification of patients and optimization of survival P-value
1:activity >= cut-off
0:activity < cut-off
Kinases i
−log10(P−value)
cut−off(activity)
4
Identify kinases associated with overall survival
Mean values
(10)
KSEA
(8)
Ridge Regression
(7)
Logrank Pvalue<0.05
HR>2;activated patients at least in 10 patients
Kinases with at least 2 substrates(FC>1.5)
Kinase i
Proportion Free of death
low(kinase=0) n= 39
high(kinase=1) n= 42
Overall Survival (years)

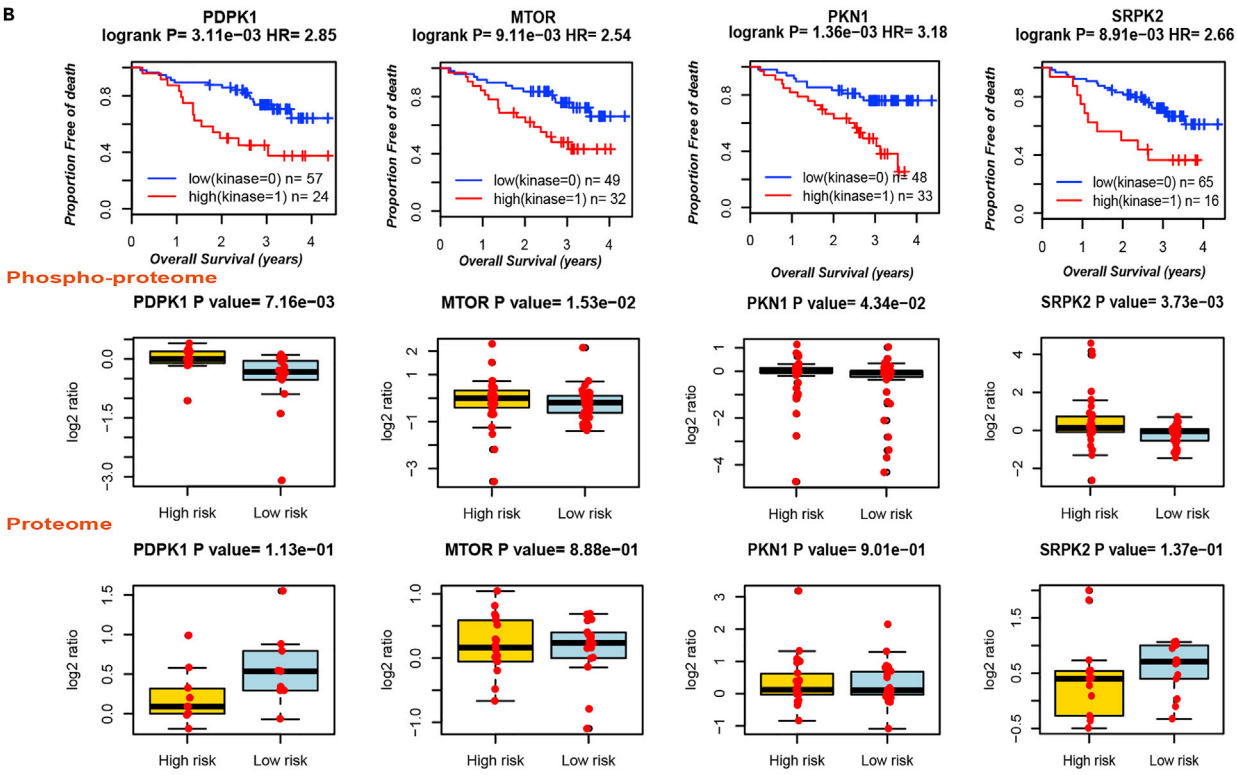

B
PDPK1
logrank P= 3.11e−03 HR= 2.85
MTOR
logrank P= 9.11e−03 HR= 2.54
PKN1
logrank P= 1.36e−03 HR= 3.18
SRPK2
logrank P= 8.91e−03 HR= 2.66
Proportion Free of death
low(kinase=0) n= 57
high(kinase=1) n= 24
low(kinase=0) n= 49
high(kinase=1) n= 32
low(kinase=0) n= 48
high(kinase=1) n= 33
low(kinase=0) n= 65
high(kinase=1) n= 16
Overall Survival (years)
Phospho-proteome
PDPK1 P value= 7.16e−03
MTOR P value= 1.53e−02
PKN1 P value= 4.34e−02
SRPK2 P value= 3.73e−03
log2 ratio
High risk
Low risk
Proteome
PDPK1 P value= 1.13e−01
MTOR P value= 8.88e−01
PKN1 P value= 9.01e−01
SRPK2 P value= 1.37e−01

Figure 3. Nominating Potential Druggable Kinases for DGC

(A) The workflow of the identification of druggable kinases.

(B) Survival curves of the four kinases and boxplots of the kinase substrates in the phosphoproteome and proteome data. Low (Kinase = 0)/high (Kinase = 1): patients with kinase activity lower/above than cut-off. For kinases were both identified by more than one kinase activity prediction method, only the smallest p value of the kinase was shown.

Also see Figures S4–S10.

tumorigenesis. SRPK2 was reported to be overexpressed in several cancer types (Lee and Abdel-Wahab, 2016) including leukemia, lung, colon, and head and neck squamous cell carcinoma; its overexpression in gastric cancer was not reported yet. Summaries of other kinases are provided in Table 2. Table S12 displayed the support validation evidences for kinases in published paper and the immunohistochemistry results of kinases in 11 stomach cancer tissues from human protein atlas (https://www.proteinatlas.org). Among the 16 kinases we identified, 10 kinases have been validated as potential targets in different cancers.

In addition, we identified 287 unfavorable prognostic phosphorylation sites, for which higher expression of a given phosphorylation site was correlated with a poor patient survival outcome (FDR<0.1, Table S13). The best ratio cut-offs were determined using the same strategy as the kinase identification. Interestingly, we found these unfavorable prognostic phosphorylation sites had three co-expression modules (Figure 4B). A functional gene ontology (GO) analysis and kinase enrichment (see Methods) were performed for the modules (Table S14). The Module1 was related to Rho protein signal transduction and cell-cell adhesion. Module 2 contained many phosphorylation sites associated with RNA splicing. These phosphorylation sites included S312-THOC5, S597-SRRM1, S549-SRRM1, S965-SCF1, and S796-SCAF11 (Figure 4B). The Module 3 enriched in cell growth and regulation of small GTPase-mediated signal transduction pathways. Three representative phosphorylation sites, S1507-AKAP13, S1261-MTOR, and T1353- MYO9B, were displayed in Figure 4B. Many kinases such as AKT1, RPS6KB1, RPS6KA1, RPS6KA3, PRKCA, PRKCD, RAF1, and PAK4 were enriched. These co-expression modules results confirmed that Rho protein signal transduction, cell-cell adhesion, MTOR signaling pathway, and RNA splicing were major altered pathways in DGC.

Nomination of PTKTs and Inhibitors for Individual Patients

To help nominate PTKTs for DGC patients, we calculated kinase activity for each individual patient. We then selected the top three PTKTs according to their activated kinase activities for the 83 patients (Figure 4B). It was evident that each patient had a unique pattern of activation for the top three PTKTs. The most frequently activated PTKT, CSNK2A1, was identified in 42/83 (50%) of the patients, whereas the second most frequently activated kinases MTOR, MAPK3, and GRK6 were each identified in 25/83 (30%) of the patients and the least frequently activated PTKT, PDPK1, was only identified in 4/83 (5%) of the patients.

Silmitasertib, the CSNK2A1 inhibitor, was in several clinical trials for treating cancers as a single agent or in combination with chemotherapy (Chon et al., 2015). Thus, we nominated Silmitasertib as an actionable reagent. The second best available actionable kinase inhibitors were the MTOR inhibitor, everolimus, and the MAPK3 inhibitor Ulixertinib. The oral MTOR inhibitor everolimus was evaluated in the phase III GRANITE-1 (First Gastric Antitumor Trial With Everolimus; Clinical Trial No. NCT00879333) (Ohtsu et al., 2013) and did not significantly improve OS for advanced gastric cancer. We speculate that inhibition of MTOR might be more effective if patients with hyperactive MTOR were selected in the trial. Ulixertinib is a potent and selective small molecule inhibitor of ERK1 and ERK2. It holds promise as a treatment for ERK-dependent cancers including colon cancer and melanoma with BRAF mutation (Germann et al., 2017). Although the frequency of BRAF mutation is rare in DGC, its downstream effector MAPK3 kinase activity is elevated, rendering them amenable to Ulixertinib treatment. Given the diverse PTKT activation profile of the 83 DGC patients, we propose that kinase inhibitions as a means of cancer treatment needs to be done on an individual basis for this cohort of patients.

Validation of CSNK2A1 Inhibition with Silmitasertib in GC Cell Lines

To test the hypothesis that the computed kinase activity from phosphoproteomics data could predict sensitivity to kinase inhibition, we repeated the above approach in gastric cancer cell lines so that we could carry out experimental validation. We measured ten GC cell lines and obtained a quantitate proteome of 8,973 proteins and a phosphoproteome of 10,332 phosphorylation sites (see Methods, Tables S15 and S16, Figures S11 and S12). We then measured IC_{50} values of Silmitasertib in the ten GC cell lines. The activity of

Kinases	Method	HR	p Value[a]	#Patients[b]	Summary
PI3K/AkT/mTOR pathway					
PDPK1	KSEA	2.85	0.0031	24	AKT1 activation
MTOR	Regression	2.54	0.0091	32	Cell growth and metabolism
RPS6KB1	Regression	2.29	0.021	27	Cell growth and proliferation
RPS6KA3	Mean values	2.69	0.0051	29	Cell growth and proliferation
PKN1	Mean values	3.18	0.00136	33	Phosphorylated by PDPK1
STK11	KSEA	2.58	0.0084	34	A tumor suppressor
Cell cycle; apoptosis					
CDK7	Mean values	2.66	0.012	15	Cell cycle
CSNK2A1	KSEA; regression	2.25; 2.45	0.028; 0.014	22; 21	Cell cycle; apoptosis
CSNK2A2	Mean values	2.39	0.016	22	Cell cycle; apoptosis
ATM	Mean values	3.3	0.0007	29	DNA damage
MAPK pathway					
MAPK3	Mean values	2.67	0.0066	35	Proliferation, differentiation, and survival
MAP3K7	Regression	3.35	0.0012	15	Proliferation, differentiation, and survival
Cell adhesion and actin organization					
PAK4	Regression	2.49	0.013	17	Actin organization and cell adhesion
Microtubule affinity					
MARK2	Mean values; regression	3.22; 2.39	0.00096; 0.014	20; 27	Controls the stability of microtubules
RNA splicing					
SRPK2	Regression	2.66	0.0090	16	Pre-mRNA splicing; cell cycle regulation and cell apoptosis
G protein-coupled receptor					
GRK6	Mean values	2.75	0.011	47	G protein-coupled receptor kinase

Table 2. Kinases Associated with Overall Survival of DGC Patients

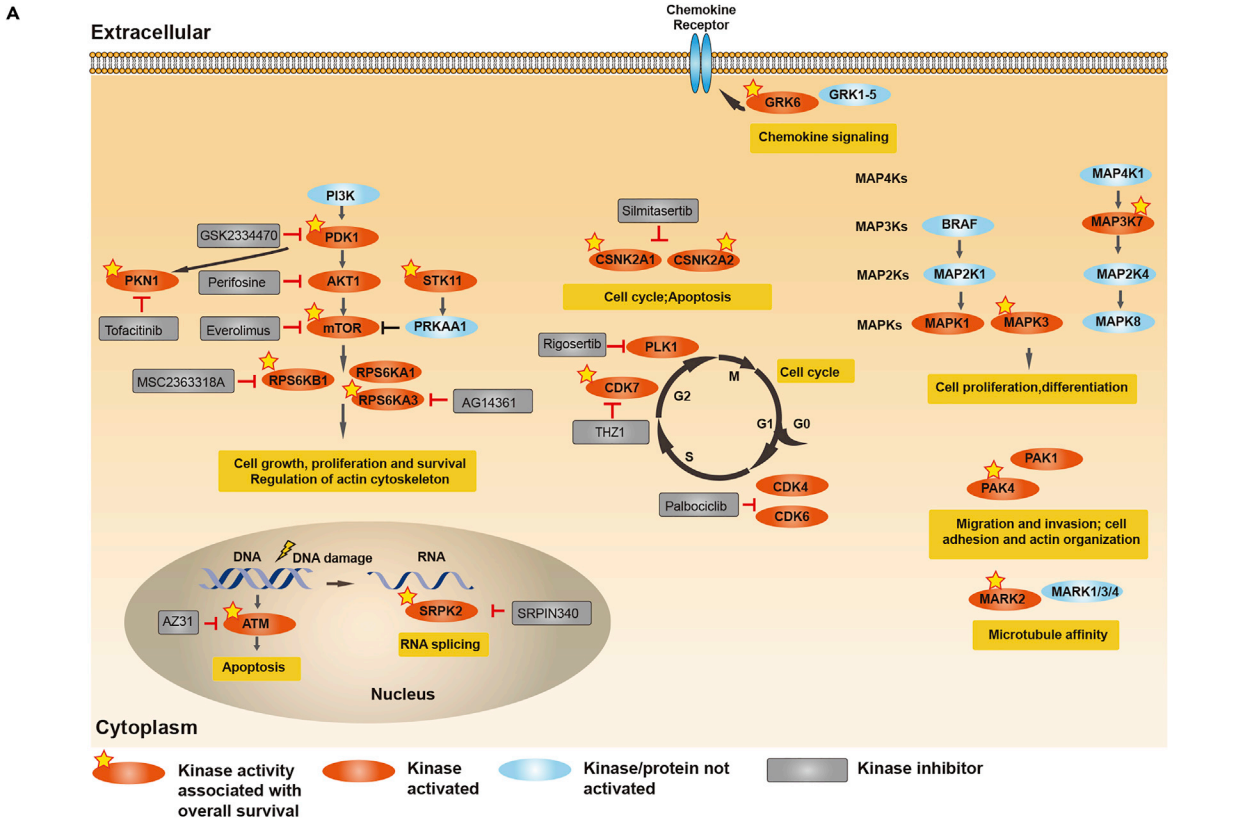
[a]Logrank p values.

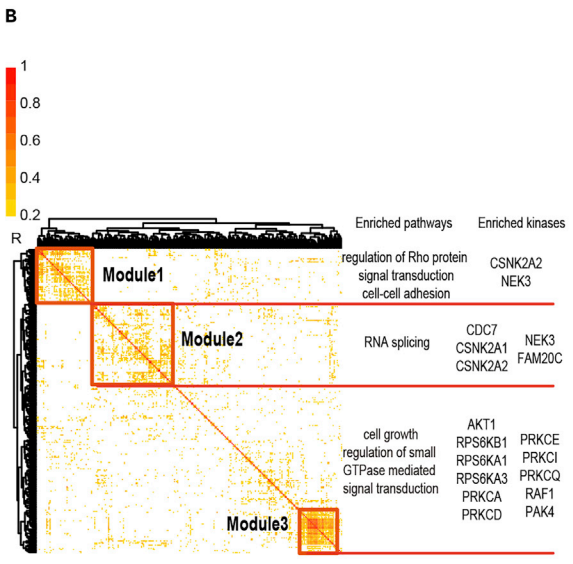
[b]Number of patients with a kinase activated.

CSNK2A1 for each cell line was calculated by the mean value method based on the phosphoproteomics data. It was evident that the computed CSNK2A1 activity in cancer cell lines was negatively correlated with the IC_{50} value of Silmitasertib (Spearman, R = −0.47, p = 0.04, Figure 5). We also found that the abundance of CSNK2A1 in the ten gastric cancer cell lines was also negatively correlated with IC_{50} value of the Silmitasertib (Spearman, R = 0.69, p = 0.013, Figure 5). It seems that both the activity and abundance of CSNK2A1 could predict Silmitasertib sensitivity of GC cell lines.

DISCUSSION

We measured phosphoproteomes of 83 DGC patients to paint the phosphoproteomic landscape of DGC. Using a global phosphoproteomics data analysis workflow that we developed recently (Tong et al., 2018), we demonstrated that large-scale phosphoproteomics alone could be used to classify DGCs into subtypes that are associated with distinct clinical outcomes as well as nomination of kinase targets for further clinical evaluation.


A
Extracellular
Chemokine Receptor
GRK6
GRK1-5
Chemokine signaling
PI3K
GSK2334470
PDK1
PKN1
Perifosine
AKT1
STK11
Tofacitinib
Everolimus
mTOR
PRKAA1
MSC2363318A
RPS6KB1
RPS6KA1
RPS6KA3
AG14361
Cell growth, proliferation and survival
Regulation of actin cytoskeleton
Silmitasertib
CSNK2A1
CSNK2A2
Cell cycle;Apoptosis
Rigosertib
PLK1
CDK7
THZ1
M
G2
G1
G0
S
Cell cycle
CDK4
Palbociclib
CDK6
MAP4Ks
MAP4K1
MAP3Ks
BRAF
MAP3K7
MAP2Ks
MAP2K1
MAP2K4
MAPKs
MAPK1
MAPK3
MAPK8
Cell proliferation,differentiation
PAK1
PAK4
Migration and invasion; cell adhesion and actin organization
MARK2
MARK1/3/4
Microtubule affinity
DNA
DNA damage
RNA
AZ31
ATM
SRPK2
SRPIN340
RNA splicing
Apoptosis
Nucleus
Cytoplasm
Kinase activity associated with overall survival
Kinase activated
Kinase/protein not activated
Kinase inhibitor


B
1
0.8
0.6
0.4
0.2
R
Enriched pathways
Enriched kinases
Module1
regulation of Rho protein signal transduction cell-cell adhesion
CSNK2A2
NEK3
Module2
RNA splicing
CDC7
CSNK2A1
CSNK2A2
NEK3
FAM20C
Module3
cell growth regulation of small GTPase mediated signal transduction
AKT1
RPS6KB1
RPS6KA1
RPS6KA3
PRKCA
PRKCD
PRKCE
PRKCI
PRKCQ
RAF1
PAK4

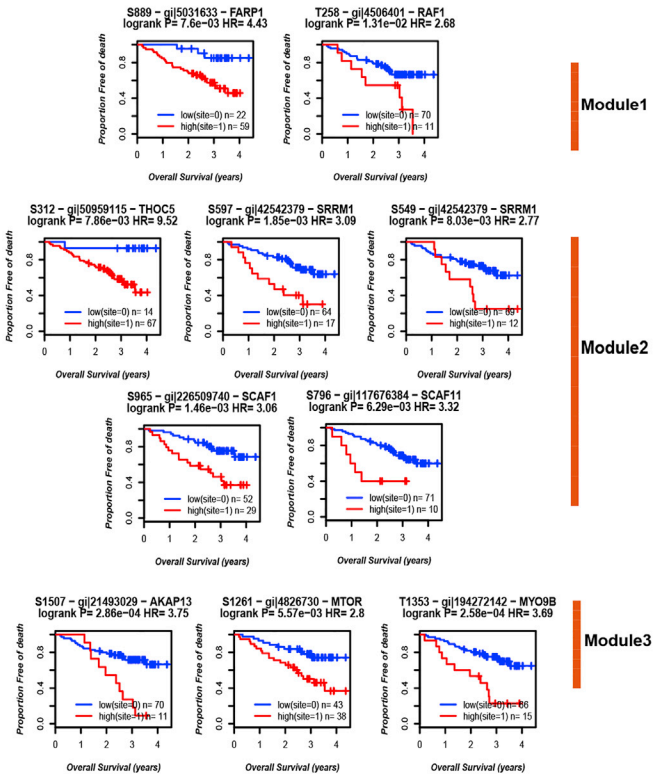

C
S889 – gi|5031633 – FARP1
logrank P= 7.6e−03 HR= 4.43
T258 – gi|4506401 – RAF1
logrank P= 1.31e−02 HR= 2.68
Module1
S312 – gi|50959115 – THOC5
logrank P= 7.86e−03 HR= 9.52
S597 – gi|42542379 – SRRM1
logrank P= 1.85e−03 HR= 3.09
S549 – gi|42542379 – SRRM1
logrank P= 8.03e−03 HR= 2.77
Module2
S965 – gi|226509740 – SCAF1
logrank P= 1.46e−03 HR= 3.06
S796 – gi|117676384 – SCAF11
logrank P= 6.29e−03 HR= 3.32
S1507 – gi|21493029 – AKAP13
logrank P= 2.86e−04 HR= 3.75
S1261 – gi|4826730 – MTOR
logrank P= 5.57e−03 HR= 2.8
T1353 – gi|194272142 – MYO9B
logrank P= 2.58e−04 HR= 3.69
Module3
Proportion Free of death
Overall Survival (years)

Figure 4. Major Pathways Mapped by the Kinases Activated in DGC

(A) Potential kinase targets in DGC.

(B) Coexpression modules of unfavorable phosphorylation sites prognostic markers.

(C) Survival curves of the phosphorylation sites within the three modules.

We previously subtyped the same DGC patients with protein profiling data (PX1–3) (Ge et al., 2018). Our current study suggests that subtyping with phosphoproteomics data may be more accurate, as the Ph1 group that is associated with best OS retrieved some patients assigned to the PX2 and PX3 groups but with better OS. It remains to be tested whether the accuracy of prediction for OS and chemosensitivity can be further improved by using both whole proteomics and phosphoproteomics data. It will be a significant challenge to translate the current findings into clinical practice.

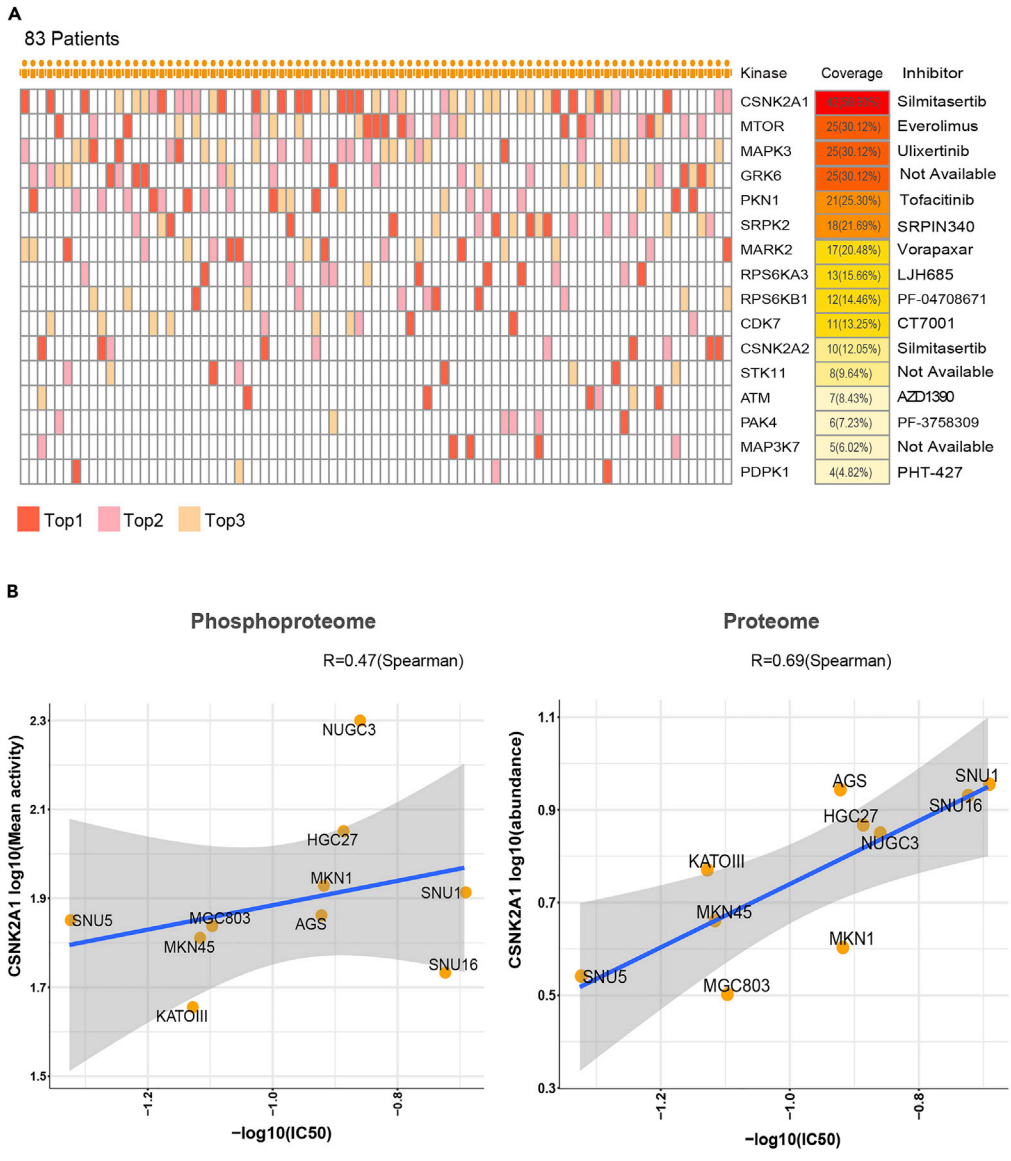


Figure 5. A Patient-Specific Strategy for Nomination of Kinases

(A) Patient-specific kinase inhibitors for DGC.

(B) The relationship of CSNK2A1 activity with the IC_{50} of Silmitasertib.

Also see Figures S11 and S12.

Protein kinases have become a major class of drug targets, and kinase inhibitors have demonstrated their efficacy in the treatment of many cancers (Wu et al., 2015, 2016; Fabbro et al., 2015). In this study, we investigated kinase activation in a systematic manner to handle large-scale phosphoproteomics data. Our analyses successfully dysregulated kinase pathways in DGC that conformed to the general conceptual framework of cancer hallmarks. In addition to the finding that the MTOR signaling network was indeed the key for DGC signaling as expected, the finding that SRPK2 kinase was activated in DGC on both protein abundance and kinase activity suggests that RNA splicing may play a significant role in DGC. It is necessary to delineate the exact signaling pathways to pinpoint the substrates of SRPK2 and what downstream biochemical pathways they regulate.

Noticeably, clinical trials of kinase inhibitors of HER2, EGFR, AKT1, and MTOR have been successfully used in other cancer treatment but failed in DGC (Choi et al., 2016). We speculate that one of the major causes might be the selection of patients in the clinical trials. Take MTOR as an example: among the 25 DGC patients whose MTOR kinase activity was among the top three most activated in our cohort, only eight were the number one most activated potential therapeutic kinases. Stratifying patients with activated MTOR activity might increase the success rate for MTOR inhibition to work in DGC patients.

It is exciting to find about 50% of the DGC patients with hyperactivation of CSNK2A1 and an actionable inhibitor, Silmitasertib, is available. This suggests the possibility of translating this finding into clinical actions. Moreover, we validated such a finding in 10 gastric cancer cell lines that calculated CSNK2A1 kinase activity from our bioinformatics workflow could predict the sensitivity to CSNK2A1 inhibition with Silmitasertib. Recently, a more potent CSNK2A1 inhibitor was reported to have anticancer activity in cell line and mouse models (Oshima et al., 2019). Given that DGC patients lack any targeted therapy options, it is worthy trying in a preclinical setting to test whether CSNK2A1 inhibition would demonstrate efficacy.

The goal of individualized treatment in cancer is paradoxical in practice. It is more accurate and elegant to perform analysis of each individual tumor to find the druggable targets, but it is almost cost prohibitive to do so in clinical practice. The best compromise would be to find activated kinases with high frequency in a cohort of cancer, demonstrate their preclinical efficacy, and then develop a practical companion test for each high-frequency kinase so that a clinical trial can be carried out selecting the right patients. This work is the first step in this line of approach, as we have identified and nominated a collection of kinases that activated with high frequency and are correlated with OS in DGC.

Limitations of the Study

Our current bioinformatics workflow only used less than 20% of the data due to the limitation of the kinase-substrate database, in which we could only analyze phosphorylation data included in the database. How to make use of the rest of the 80% of data would further strengthen the utility of the method as well as enhance our understanding of DGC and hold the promise to find novel pathways and drug targets. We note that it is necessary to further develop the current method to use in a clinical setting. Biopsy samples instead of the resected samples, for example, from stomach endoscopy may be used to acquire the phosphoproteomics data. An improvement of the current methodology is to make use of tiny amount of the biopsy samples to acquire data and be able to analyze the data for cancer molecular subtyping and nomination of actionable kinase targets.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTALINFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2019.11.003.

ACKNOWLEDGMENTS

This work was supported by the National Program on Key Basic Research Project (973 Program) (973 Program, 2014CBA02000), National International Cooperation Grant (2014DFB30010 and 2014DFA33160), Beijing Municipal Science and Technology "Frontier Project" (Z131100005213003), National Key Research and Development Program of China (2017YFC0908404 and 2018YFA0507504), National Natural Science

Foundation of China (61773025), and Shanghai Municipal Science and Technology Major Project (2017SHZDZX01). This work is dedicated to the memory of Dr. Bei Zhen.

AUTHOR CONTRIBUTIONS

J.Q.,T.L., L. S., and F.H. directed and designed research; M.T.,C.Y.,J.S, L.S.,D.Z.,X.X.,W.L.,J.F., and W.S. performed analyses of mass spectrometry data and adapted algorithms and software for data analysis; S.G.,M.L. and W.H. coordinated acquisition, distribution, and quality evaluation of tumor and nearby tissue samples; J. J.,J.G.,T.S.,W.Z., and C.D contributed new reagent/analytic tools and provided professional suggestions; J.Q.,Y.W.,T.L, M.T., and C.Y wrote the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: March 25, 2019
Revised: September 25, 2019
Accepted: November 1, 2019
Published: December 20, 2019

REFERENCES

Cancer Genome Atlas Research Network (2014). Comprehensive molecular characterization of gastric adenocarcinoma. Nature *513*, 202–209.

Casado, P., Hijazi, M., Britton, D., and Cutillas, P.R. (2017). Impact of phosphoproteomics in the translation of kinase-targeted therapies. Proteomics *17*, 1600235.

Casado, P., Rodriguez-Prados, J.C., Cosulich, S.C., Guichard, S., Vanhaesebroeck, B., Joel, S., and Cutillas, P.R. (2013). Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells. Sci. Signal. *6*, rs6.

Choi, Y.Y., Noh, S.H., and Cheong, J.H. (2016). Molecular dimensions of gastric cancer: translational and clinical perspectives. J.Pathol. Transl. Med. *50*, 1–9.

Chon, H.J., Bae, K.J., Lee, Y., and Kim, J. (2015). The casein kinase 2 inhibitor, CX-4945, as an anti-cancer drug in treatment of human hematological malignancies. Front. Pharmacol. *6*, 70.

Cristescu, R., Lee, J., Nebozhyn, M., Kim, K.M., Ting, J.C., Wong, S.S., Liu, J., Yue, Y.G., Wang, J., Yu, K., et al. (2015). Molecular analysis of gastric cancer identifies subtypes associated with distinct clinical outcomes. Nat. Med. *21*, 449–456.

Drake, J.M., Paull, E.O., Graham, N.A., Lee, J.K., Smith, B.A., Titz, B., Stoyanova, T., Faltermeier, C.M., Uzunangelov, V., Carlin, D.E., et al. (2016). Phosphoproteome integration reveals patient-specific networks in prostate cancer. Cell *166*, 1041–1054.

Fabbro, D., Cowan-Jacob, S.W., and Moebitz, H. (2015). Ten things you should know about protein kinases: IUPHAR Review 14. Br. J. Pharmacol. *172*, 2675–2700.

Feng, J., Ding, C., Qiu, N., Ni, X., Zhan, D., Liu, W., Xia, X., Li, P., Lu, B., Zhao, Q., et al. (2017). Firmiana: towards a one-stop proteomic cloud platform for data processing and analysis. Nat. Biotechnol. *35*, 409–412.

Ferguson, F.M., and Gray, N.S. (2018). Kinase inhibitors: the road ahead. Nat. Rev. Drug Discov. *17*, 353–377.

Gagliardi, P.A., Puliafito, A., and Primo, L. (2018). PDK1: at the crossroad of cancer signaling pathways. Semin. Cancer Biol. *48*, 27–35.

Ge, S., Xia, X., Ding, C., Zhen, B., Zhou, Q., Feng, J., Yuan, J., Chen, R., Li, Y., Ge, Z., et al. (2018). A proteomic landscape of diffuse-type gastric cancer. Nat. Commun. *9*, 1012.

Germann, U.A., Furey, B.F., Markland, W., Hoover, R.R., Aronov, A.M., Roix, J.J., Hale, M., Boucher, D.M., Sorrell, D.A., Martinez-Botella, G., et al. (2017). Targeting the MAPK signaling pathway in cancer: promising preclinical activity with the novel selective ERK1/2 inhibitor BVD-523 (Ulixertinib). Mol. Cancer Ther. *16*, 2351–2363.

Goel, S., Decristo, M.J., Watt, A.C., Brinjones, H., Sceneay, J., Li, B.B., Khan, N., Ubellacker, J.M., Xie, S., Metzger-Filho, O., et al. (2017). CDK4/6 inhibition triggers anti-tumour immunity. Nature *548*, 471–475.

Grogg, K.L., Lohse, C.M., Pankratz, V.S., Halling, K.C., and Smyrk, T.C. (2003). Lymphocyte-rich gastric cancer: associations with Epstein-Barr virus, microsatellite instability, histology, and survival. Mod. Pathol. *16*, 641–651.

Hernandez-Armenta, C., Ochoa, D., Goncalves, E., Saez-Rodriguez, J., and Beltrao, P. (2017). Benchmarking substrate-based kinase activity inference using phosphoproteomic data. Bioinformatics *33*, 1845–1851.

Jiang, Y., Sun, A., Zhao, Y., Ying, W., Sun, H., Yang, X., Xing, B., Sun, W., Ren, L., Hu, B., et al.; Chinese Human Proteome Project Consortium (2019). Proteomics identifies new therapeutic targets of early-stage hepatocellular carcinoma. Nature *567*, 257–261.

Kang, B.W., Kim, J.G., Lee, I.H., Bae, H.I., and Seo, A.N. (2017). Clinical significance of tumor-infiltrating lymphocytes for gastric cancer in the era of immunology. World J. Gastrointest. Oncol. *9*, 293–299.

Lee, G., Zheng, Y., Cho, S., Jang, C., England, C., Dempsey, J.M., Yu, Y., Liu, X., He, L., Cavaliere, P.M., et al. (2017). Post-transcriptional regulation of de Novo Lipogenesis by mTORC1-S6K1-SRPK2 signaling. Cell *171*, 1545–1558.e18.

Lee, S.C., and Abdel-Wahab, O. (2016). Therapeutic targeting of splicing in cancer. Nat. Med. *22*, 976–986.

Liu, J.J., Sharma, K., Zangrandi, L., Chen, C., Humphrey, S.J., Chiu, Y.T., Spetea, M., Liu-Chen, L.Y., Schwarzer, C., and Mann, M. (2018). In vivo brain GPCR signaling elucidated by phosphoproteomics. Science *360*.

Mertins, P., Mani, D.R., Ruggles, K.V., Gillette, M.A., Clauser, K.R., Wang, P., Wang, X., Qiao, J.W., Cao, S., Petralia, F., et al. (2016). Proteogenomics connects somatic mutations to signalling in breast cancer. Nature *534*, 55–62.

Mun, D.G., Bhin, J., Kim, S., Kim, H., Jung, J.H., Jung, Y., Jang, Y.E., Park, J.M., Kim, H., Jung, Y., et al. (2019). Proteogenomic characterization of human early-onset gastric cancer. Cancer Cell *35*, 111–124.e10.

Nolen, B., Taylor, S., and Ghosh, G. (2004). Regulation of protein kinases; controlling activity through activation segment conformation. Mol. Cell *15*, 661–675.

Ohtsu, A., Ajani, J.A., Bai, Y.X., Bang, Y.J., Chung, H.C., Pan, H.M., Sahmoud, T., Shen, L., Yeh, K.H., Chin, K., et al. (2013). Everolimus for previously treated advanced gastric cancer: results of the randomized, double-blind, phase III GRANITE-1 study. J. Clin. Oncol. *31*, 3935–3943.

Oshima, T., Niwa, Y., Kuwata, K., Srivastava, A., Hyoda, T., Tsuchiya, Y., Kumagai, M., Tsuyuguchi, M., Tamaru, T., Sugiyama, A., et al. (2019). Cell-based screen identifies a new potent and highly selective CK2 inhibitor for modulation of circadian rhythms and cancer cell growth. Sci. Adv. *5*, eaau9060.

Otsu, H., Iimori, M., Ando, K., Saeki, H., Aishima, S., Oda, Y., Morita, M., Matsuo, K., Kitao, H., Oki, E., and Maehara, Y. (2016). Gastric cancer patients with high PLK1 expression and DNA Aneuploidy correlate with poor prognosis. Oncology *91*, 31–40.

Rikova, K., Guo, A., Zeng, Q., Possemato, A., Yu, J., Haack, H., Nardone, J., Lee, K., Reeves, C., Li, Y., et al. (2007). Global survey of phosphotyrosine signaling identifies oncogenic kinases in lung cancer. Cell *131*, 1190–1203.

Tan, I.B., Ivanova, T., Lim, K.H., Ong, C.W., Deng, N., Lee, J., Tan, S.H., Wu, J., Lee, M.H., Ooi, C.H., et al. (2011). Intrinsic subtypes of gastric cancer, based on gene expression pattern, predict survival and respond differently to chemotherapy. Gastroenterology *141*, 476–485, 485.e1-11.

Tong, M., Yu, C., Zhan, D., Zhang, M., Zhen, B., Zhu, W., Wang, Y., Wu, C., He, F., Qin, J., and Li, T. (2018). Molecular subtyping of cancer and nomination of kinase candidates for inhibition with phosphoproteomics: reanalysis of CPTAC ovarian cancer. EBioMedicine *40*, 305–317.

Tusher, V.G., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc. Natl. Acad. Sci. U S A *98*, 5116–5121.

Wang, K., Kan, J., Yuen, S.T., Shi, S.T., Chu, K.M., Law, S., Chan, T.L., Kan, Z., Chan, A.S., Tsui, W.Y., et al. (2011). Exome sequencing identifies frequent mutation of ARID1A in molecular subtypes of gastric cancer. Nat. Genet. *43*, 1219–1223.

Wang, Q., Li, M., Zhang, X., Huang, H., Huang, J., Ke, J., Ding, H., Xiao, J., Shan, X., Liu, Q., et al. (2016). Upregulation of CDK7 in gastric cancer cell promotes tumor cell proliferation and predicts poor prognosis. Exp. Mol. Pathol. *100*, 514–521.

Wilkerson, M.D., and Hayes, D.N. (2010). ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. Bioinformatics *26*, 1572–1573.

Wu, P., Nielsen, T.E., and Clausen, M.H. (2015). FDA-approved small-molecule kinase inhibitors. Trends Pharmacol. Sci. *36*, 422–439.

Wu, P., Nielsen, T.E., and Clausen, M.H. (2016). Small-molecule kinase inhibitors: an analysis of FDA-approved drugs. Drug Discov. Today *21*, 5–10.

Wu, R., Dephoure, N., Haas, W., Huttlin, E.L., Zhai, B., Sowa, M.E., and Gygi, S.P. (2011). Correct interpretation of comprehensive phosphorylation dynamics requires normalization by protein expression changes. Mol. Cell. Proteomics *10*, M111 009654.

Wu, X., Xing, X., Dowlut, D., Zeng, Y., Liu, J., and Liu, X. (2019). Integrating phosphoproteomics into kinase-targeted cancer therapies in precision medicine. J. Proteomics *191*, 68–79.

Yu, X., Zhang, Z., Wang, Z., Wu, P., Qiu, F., and Huang, J. (2016). Prognostic and predictive value of tumor-infiltrating lymphocytes in breast cancer: a systematic review and meta-analysis. Clin. Transl. Oncol. *18*, 497–506.

ISCI, Volume *22*

Supplemental Information

Phosphoproteomics Enables Molecular Subtyping and Nomination of Kinase Candidates for Individual Patients of Diffuse-Type Gastric Cancer

Mengsha Tong, Chunyu Yu, Jinwen Shi, Wenwen Huang, Sai Ge, Mingwei Liu, Lei Song, Dongdong Zhan, Xia Xia, Wanlin Liu, Jinwen Feng, Wenhao Shi, Jiafu Ji, Jing Gao, Tieliu Shi, Weimin Zhu, Chen Ding, Yi Wang, Fuchu He, Lin Shen, Tingting Li, and Jun Qin

Supplementary Information

Transparent Methods

Patient samples

Tumor samples and their nearby tissues of 83 DGC patients were collected from the tumor tissue bank of Beijing Cancer Hospital after being evaluated by pathologists. The details of the biospecimen collection and clinical data annotation were described in the previous study (Ge et al., 2018).

Cell lines and reagents

Ten GC cell lines were purchased from Cobioer Biosciences Co., Ltd (Nanjing, Jiangsu, China). NUGC3, MKN45, SNU16, AGS, MGC803, SNU1, MKN1 and HGC27 were cultured in RPMI 1640 medium (Gibco, USA) supplemented with 10% FBS, while KATOIII and SNU5 were cultured in IMDM medium (Gibco, USA) supplemented with 20% FBS. All cell lines were maintained at 37◦C with 5% CO2. To perform MTT assay, cells were plated in tissue culture treated 96-well plates. After overnight adhesion to the plate, cells were incubated with the indicated concentrations of Silmitasertib (MedChem Express, USA) for 48 h. Cell Counting Kit-8 reagent (MedChem Express, USA) was added to 10% and incubated for 1.5 h at 37◦C before reading the absorbance at 450 nm on a plate reader. MTT assay was performed to determine the half-maximal inhibitory concentration (IC_{50}) of Silmatisertib in the GC cells.

Phosphopeptide enrichment and analysis by mass spectrometry.

Tissue samples were minced and lysed in lysis buffer (8M Urea, 100mM Tris Hydrochloride, pH8.0) containing protease and phosphatase Inhibitors (Thermo Scientific) followed by 1 min sonication (3s on and 3s off, amplitude 25%). The lysate

was centrifuged at 16,000g for 10 min at 4 °C and the supernatant was collected as whole tissue extract. Protein concentration was determined by Bradford protein assay. Extracts from each sample (2 mg protein) was reduced with 10 mM dithiothreitol (DTT) at 56°C for 30 min and alkylated with 20 mM iodoacetamide (IAA) at room temperature in the dark for additional 30 min. Samples were then digested using the FASP method(Wisniewski et al., 2009) with trypsin; tryptic peptides were separated in a home-made reverse-phase C18 column in a pipet tip. Then, 15mg TiO2-coupled beads were incubated with 500ul Binding Buffer(BB) at RT for 10min. Separate TiO2 into three 1.5ml EP tubes equally, 5mg for each and centrifuge 2,000g for 2min. The peptides were re-solved digested peptides with 6ul BB and combine with 5mg incubated TiO2 at RT for 30 min and centrifuge 1,000g for 2min. Repeat the solved procedures for the dried phospho-/peptides twice with the 5 mg TiO2 and then discard the supernatant. Peptides were eluted and separated into 6 fractions using a stepwise gradient of increasing acetonitrile (0%, 2%, 5%, 8%, 10%, 40%) at pH 10. The 6 fractions were combined to 3 fractions, dried in a vacuum concentrator (Thermo Scientific).

Then the tumor tissue samples were analyzed by Orbitrap Fusion mass spectrometers (Thermo Fisher Scientific, Rockford, IL, USA) coupled with an Easy-nLC 1000 nanoflow LC system (Thermo Fisher Scientific). The injected peptides were separated on a reverse-phase nano-HPLC C18 column (precolumn, 3 μ m, 120 Å, 2 cm ×100 μ m i.d.; analytical column, 1.9 μ m, 120 Å, 12 cm ×150 μ m i.d.) at a flow rate of 500 nL/min with a 75 min gradient of 5−30% solvent B (0.1% formic acid in

acetonitrile). For peptide ionization, 2000 V was applied and a 320 ° C capillary temperature was used. For detection with Fusion mass spectrometry, a precursor scan was carried out in the Orbitrap by scanning m/z 300 -1400 with a resolution of 120,000 at 200 m/z. The most intense ions selected under top-speed mode were isolated in Quadrupole with a 1.6 m/z window and fragmented by HCD with normalized collision energy of 35%, then measured in the linear ion trap using the rapid ion trap scan rate. Automatic gain control targets were 5×10^5 ions with a max injection time of 50 ms for full scans and 5×10^3 with 35 ms for MS/MS scans. Dynamic exclusion time was set as 18s.

For cell line samples, a total of 0.5mg cell protein lysates were digested with trypsin for phosphorylated peptides enrichment. 0.15ml binding buffer (80% ACN, 5% trifluoroacetic acid (TFA (Sigma-Aldrich)), and 1M lactic acid (Sigma-Aldrich)) was added to the digested peptides. Vortex vigorously at room temperature for 1min and centrifugated 10min at 14000g. Supernate was transferred to a new tube and 3mg TiO2 beads were added. Incubation was processed at room temperature for 30min on rotor with middle speed. Then beads were collected by centrifugation for 2min at 2000 × g, and the supernatant was used to repeat the enrichment for twice with new TiO2. Beads were combined and wash in a clear tube with 500uL wash buffer (50% ACN and 5% TFA) for 5times. After that, beads were transferred into a 0.2ml StageTip with two pieces of C18. Another twice wash was proceed in tip with 100uL wash buffer per time. Phosphorylated peptides were eluted five times with 100uL elution buffer (40% ACN, and 18% aqua ammonia) per time. Peptide samples were concentrated in a SpeedVac.

The cell lines samples were analyzed by the same Orbitrap Fusion mass spectrometers. The injected peptides were separated on a reverse-phase nano-HPLC C18 column (precolumn, 3 μ m, 120 Å, 2 cm ×100 μ m i.d.; analytical column, 1.9 μ m, 120 Å, 30 cm ×150 μ m i.d.) at a flow rate of 600 nL/min with a 150 min gradient of 6−40% solvent B. Dynamic exclusion time was set as 25s.

MS data processing

All the MS data were processed in the Firmiana database(Feng et al., 2017). Raw files were searched against the human National Center for Biotechnology Information (NCBI) Refseq protein database (updated on 07-04-2013, 32015 entries) by Mascot 2.3 (Matrix Science Inc). The mass tolerances were 20 ppm for precursor and 0.5 Da for product ions for Fusion. Up to two missed cleavages were allowed. The search engine set cysteine carbamidomethylation as a fixed modification and Phospho(ST), Phospho(Y), Acetyl (Protein N-term) and Oxidation(M), as variable modifications. The data were also searched against a decoy database so that peptide identifications were accepted at a false discovery rate (FDR) of 1%. Peptides with 1% FDR and Mascot ion score greater than 20 were selected for further analysis. Proteins with at least 2 unique peptides with 1% FDR at the peptide level and Mascot ion score greater than 20 were remained. Phosphorylation sites were reported when phosphopeptides showed a delta score >10, otherwise the precise modification site was deemed ambiguous(Savitski et al., 2011). For the peptide's quantification, we used the area under the curve of a peptide feature. Label-free protein quantifications were calculated using a label-free, intensity-based absolute quantification (iBAQ) approach(Schwanhausser et al., 2011). The

fraction of total (FOT) was used to represent the normalized abundance of a particular protein/peptide across samples. FOT of peptide was defined as a peptide's peak area value divided by the total area values of all identified peptides within one sample. FOT of protein was defined as a protein's iBAQ divided by the total iBAQ of all identified proteins within one sample. The FOT was multiplied by 10^6 for the ease of presentation (Supplementary Table S2).

The following filter criterias were applied for each statistical analysis shown in Fig. 1B: 1) Dataset 1 (D1) included all 37,354 identified phosphorylation sites on 1% of global FDR and Mascot ion score greater than 20. 2) For dataset 2 (D2), we excluded phosphorylation sites whose maximum FOT in all experiments less than the 3rd quartile value. The missing data were imputed with the minimum values. After missing value imputation, phosphorylation sites were quantile normalized. For the Proteomics data, the same filter criteria was used.

Statistical analysis

Principle Component Analysis (PCA) was performed to visualize separation of tumors and nearby tissues. Significance Analysis of Microarray (SAM) (samr R package) (Tusher et al., 2001) was performed to find differentially expressed phosphorylation sites between tumors and paired nearby tissues of all 83 patients and within each clusters. Data type was set as two class paired. Delta value was set respectively to meet FDR<0.01. The differentially expressed phosphorylation sites defined here must meet the following criteria: 1) q value less than 0.01, 2) differentially expressed percentage larger than 50%, which was calculated using the following formula:

$$\text{Differentially expressed percentage} = \frac{\left| N_{\text{tumor up}} - N_{\text{tumor down}} \right|}{N_{\text{total detected}}} \quad (1)$$

where $N_{\text{tumor up}}$ means the number of patients with Tumor-Nearby tissue ratio lager than 3, $N_{\text{tumor down}}$ means the number of patients with Tumor-Nearby tissue ratio less than 1/3, $N_{\text{total detected}}$ means the number of patients with the phosphorylation site detected. To generate an appropriate list for the function enrichment, phosphorylation sites with FDR <0.01 by SAM and differential expression percentage >0.5 or < -0.5 were selected. The corresponding proteins with the phosphorylation sites were used to perform pathways enrichment on the David website(Huang da et al., 2009).

Consensus clustering was performed using the R package ConsensusClusterPlus (Wilkerson and Hayes, 2010). Fisher's exact test and Wilcoxon rank-sum test were used to assess different frequencies of clinical characteristics between groups. Kaplan-Meier method was used to perform the survival analysis and the difference was tested using the log-rank test. Coefficient value, which equals to ln (Hazard Ratio), was calculated from Cox proportional hazards regression analysis. P values less than 0.05 were considered as significantly different. Overall survival was used as primary endpoint. Clinical variables analyzed with P value less than 0.05 using single variant analysis were chosen to enter Cox regression multivariate analysis. The R package 'survival' was used for survival statistical tests. The multivariate Cox proportional-hazards regression model was used to evaluate the independent prognostic value of the signature after adjusting for clinical factors including tumor stage, adjuvant chemotherapy age, gender, tumor sites, and TP53 mutation.

All statistical analyses were performed using the R 3.31 and Python 3.6.2 (with Anaconda 5.1.0)

Construction of kinase–substrate interaction dataset

The comprehensive kinase–substrate interaction dataset was constructed through integration of the data from the four public human protein databases; PhosphoSitePlus, Phospho.ELM, PhosphoNetworks and UniprotKB, leading to generation of 15,530 non-redundant interactions (Supplementary Table S9), which were validated in Swiss-Prot database. We then expanded the kinase-substrate relationships for other sites in the DGC dataset based on the Position weight matrixes (PWMs) score and protein-protein interactions collected from eight public databases: BioGrid, PioPlex, CCSB, DIP, HPRD, IntAct, MINT, and PINA. Briefly, if a phosphorylation site meets both of the following criteria, it will be predicted as the substrate site of a specific kinase: 1) similar sequence pattern with known substrate sites of the kinase; 2) the substrate protein interacted with the kinase.

Nominate kinases as potential therapeutic targets

Based on the assumption that the activation of a kinase is reflected by the phosphorylation state of its substrates, four different methods including Mean values, multiple linear regression (MLR) model, KSEA algorithm and Z-test (Hernandez-Armenta et al., 2017) were used to estimate the kinase activity after normalization of phosphopeptide abundance by protein abundance.

If phosphosite *j* was a substrate of kinase *i*, otherwise 0. *i* ranged from 1 to n (n = 250 for all 250 kinases), y_j was the relative value of phosphosite *j*, and was computed as below:

$$y_j = \log_2 \frac{(T_j + 1) / (TP_j + 1)}{(N_j + 1) / (NP_j + 1)} \quad (2)$$

T_j and N_j were quantitative values of phosphosite *j* in tumor and nearby tissue; TP_j and NP_j were the quantitative values of the protein corresponding to phosphosite *j* in tumor and nearby tissue. To smooth categorical data, $T_j + 1$ instead of T_j was used to calculate y_j.

For each kinase, the predicted kinase-patient matrix of activity was used to classify the patients into two groups: patients with higher kinase activity and patients with lower kinase activity. The activity cut-off value of each kinase was determined by the lowest log-rank P value of overall survival (Fig.3A). Then, the significance of the overall survival between the two groups was tested by log-rank test.

Enrichment of kinases in co-expression modules

For each kinase with more than 30 substrates, distribution of PWM scores of phosphosites in different modules was calculated, and distributions of PWM scores of all detected phosphosites were regarded as the background. Kinases with significantly higher PWM scores compared with background distribution (One-tail Mann-Whitney U test ,P value <0.01) in each module were selected.

Data and Software Availability

All raw data and the Mascot output tables have been deposited to iProX (http://www.iprox.org/) and can be accessed with the accession IPX0001234000.

Supplementary figures and legends

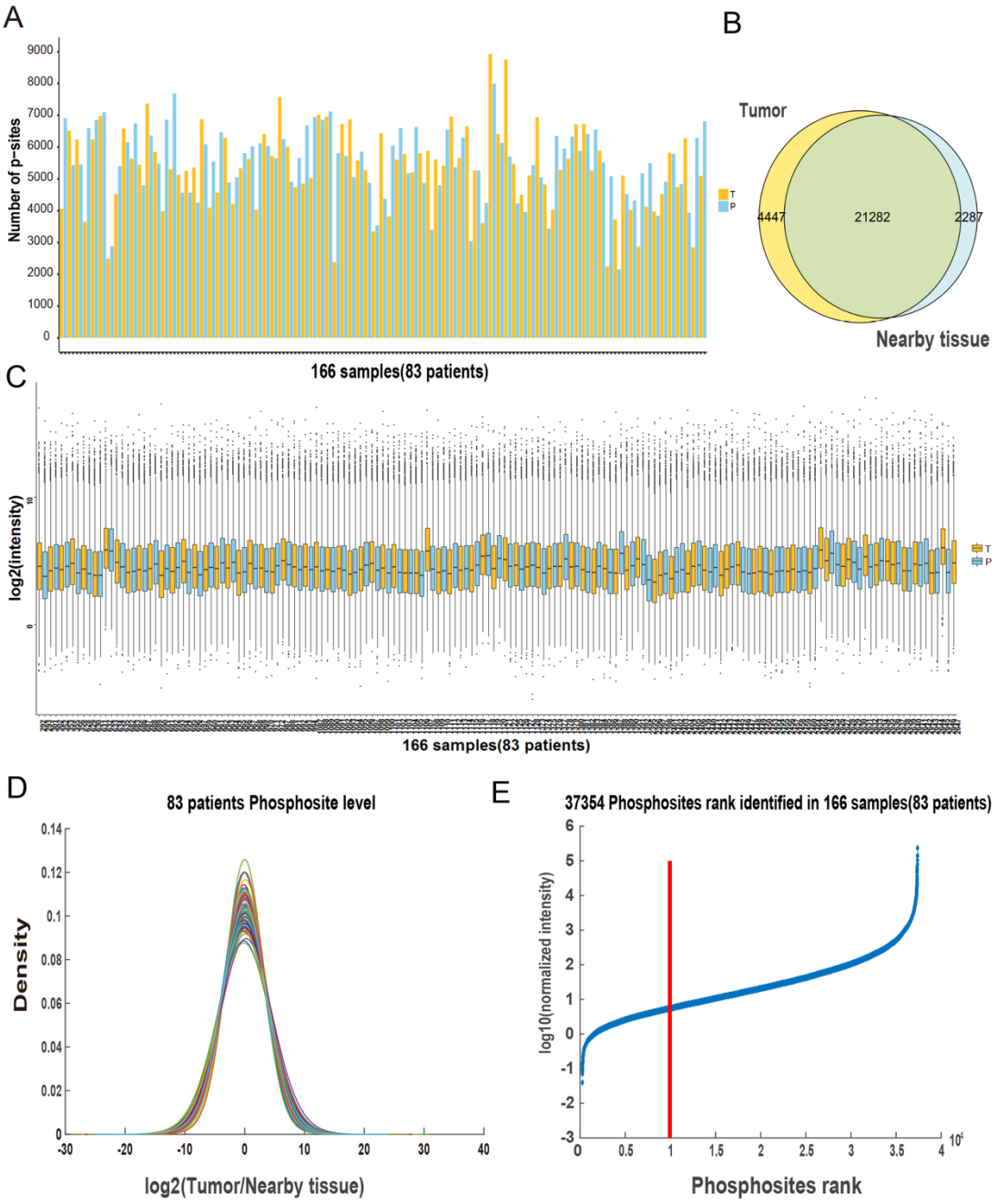


Fig.S1

Fig. S1. Global phosphorylation sites identifications in DGC, Related to Figure 1.

A. The number of phosphorylation sites detected in 83 patients; **B.** Numbers of phosphorylation sites detected in tumors and nearby tissues; **C.** Distribution of $\log_2$ transformed FOT of the phosphorylation sites in 166 samples; **D.** Global tumor to nearby tissue ratios distributions for 83 patients; **E.** Ranks of the phosphorylation sites based on the FOT.

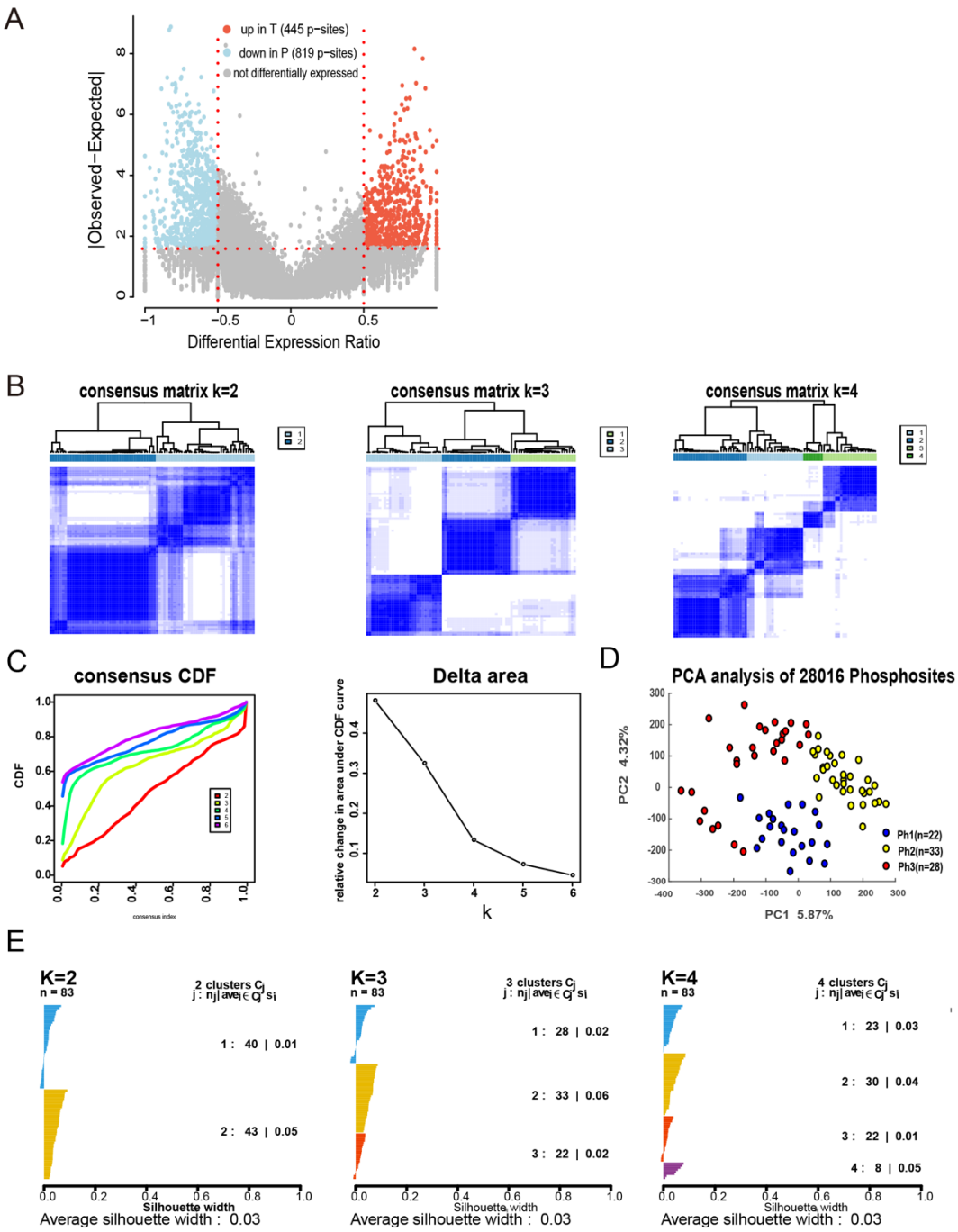

A
up in T (445 p−sites)
down in P (819 p−sites)
not differentially expressed
|Observed−Expected|
Differential Expression Ratio
B
consensus matrix k=2
consensus matrix k=3
consensus matrix k=4
C
consensus CDF
CDF
consensus index
Delta area
relative change in area under CDF curve
k
D
PCA analysis of 28016 Phosphosites
PC2 4.32%
PC1 5.87%
Ph1(n=22)
Ph2(n=33)
Ph3(n=28)
E
K=2
n = 83
2 clusters Cj
1 : 40 | 0.01
2 : 43 | 0.05
Silhouette width
Average silhouette width : 0.03
K=3
n = 83
3 clusters Cj
1 : 28 | 0.02
2 : 33 | 0.06
3 : 22 | 0.02
Silhouette width
Average silhouette width : 0.03
K=4
n = 83
4 clusters Cj
1 : 23 | 0.03
2 : 30 | 0.04
3 : 22 | 0.01
4 : 8 | 0.05
Silhouette width
Average silhouette width : 0.03

Fig.S2

Fig. S2.Consensus clustering based on the phosphorylation sites, Related to Figure2.

A. A volcano plot for differentially expressed phosphorylation sites between tumors and nearby tissues; **B.** Consensus matrices of the 83 patients from k=2 to k=4; **C.** The cumulative distribution function (CDF) plots; **D.** Principal component analysis (PCA) to visualize the Ph1-Ph3 subtypes; **E.** The silhouette plots of the three subtypes.

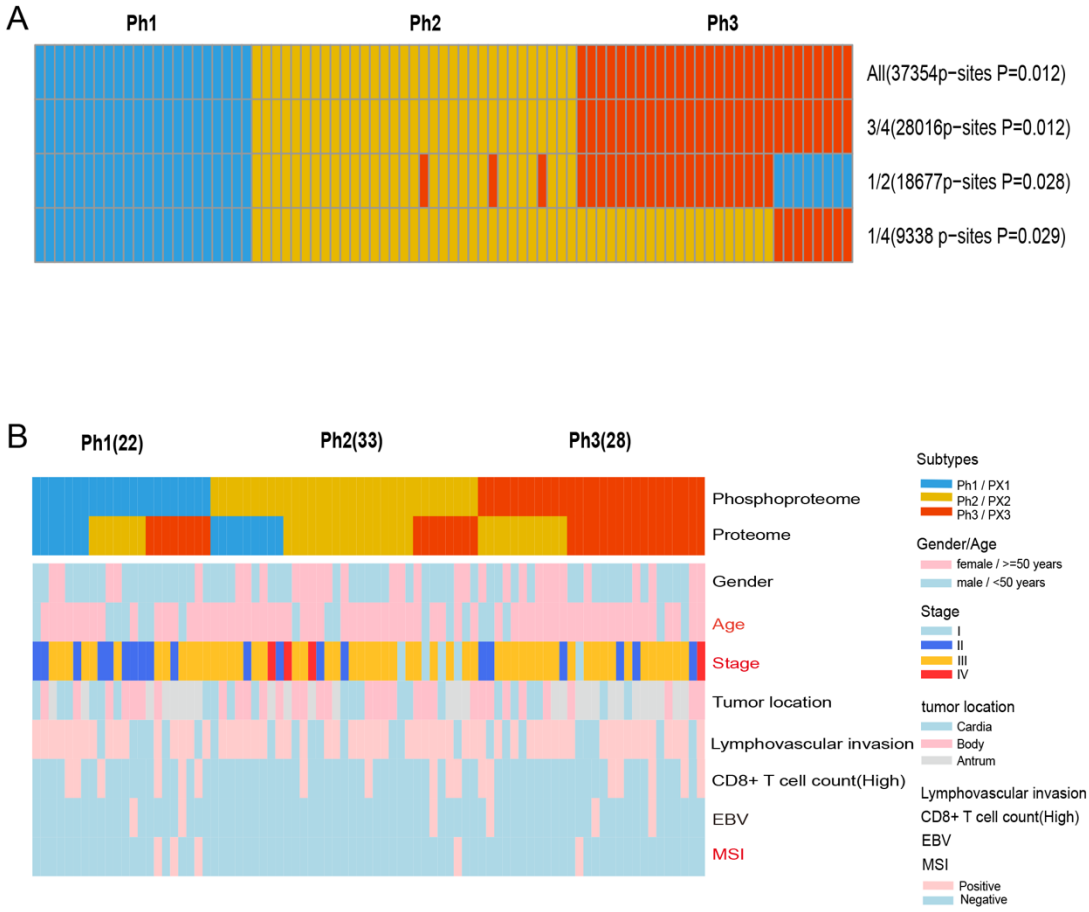

A
Ph1
Ph2
Ph3
All(37354p−sites P=0.012)
3/4(28016p−sites P=0.012)
1/2(18677p−sites P=0.028)
1/4(9338 p−sites P=0.029)
B
Ph1(22)
Ph2(33)
Ph3(28)
Phosphoproteome
Proteome
Gender
Age
Stage
Tumor location
Lymphovascular invasion
CD8+ T cell count(High)
EBV
MSI
Subtypes
Ph1 / PX1
Ph2 / PX2
Ph3 / PX3
Gender/Age
female / >=50 years
male / <50 years
Stage
I
II
III
IV
tumor location
Cardia
Body
Antrum
Lymphovascular invasion
CD8+ T cell count(High)
EBV
MSI
Positive
Negative
C
Mutation genes(Fisher−P<0.1)
Ph1
Ph2
Ph3
SPTA1
KRAS
PLEKHA6
NF1
RNF43
DNMT3A
CASP8
MAP2K4
FLI1
ERG
SMARCA4
ALPK2

Fig.S3

Fig. S3.Other clinical characteristics among the three subtypes, Related to Figure2.

A. Clusters assigned by consensus clustering with different number of phosphorylation sites with different thresholds (all, top3/4, top1/2 and top1/4); **B.** Other clinical parameters across 83 patients in Ph1-3. CD8+ T-high, CD8 positive T cells number ≥ 298 per μm2; MSI-H, microsatellite instability high; EBV, Epstein-Barr virus status; **C.** Mutation genes in the three subtypes;

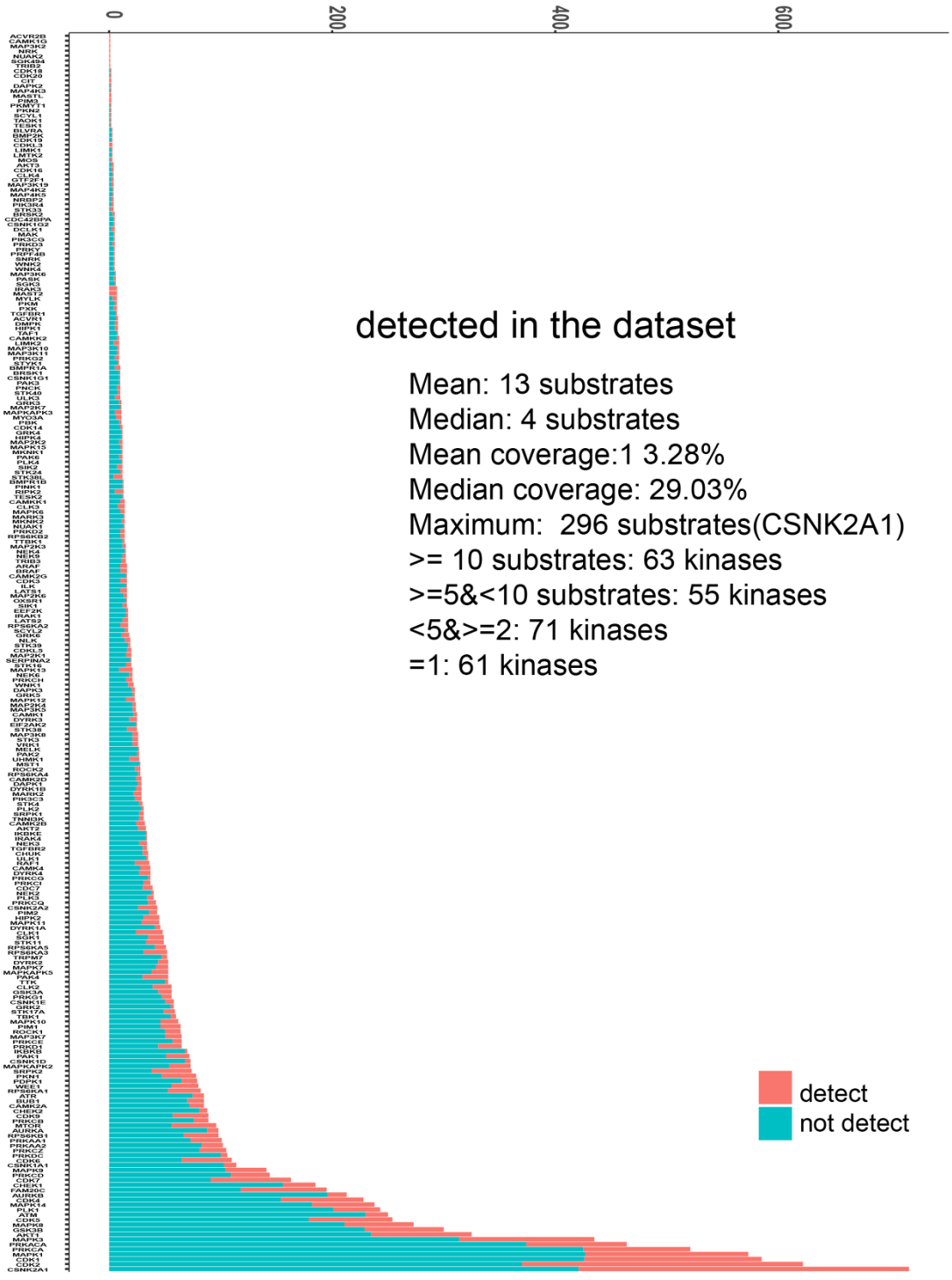


Fig.S4

Fig. S4. Numbers of substrates detected for each kinase, Related to Figure 3.

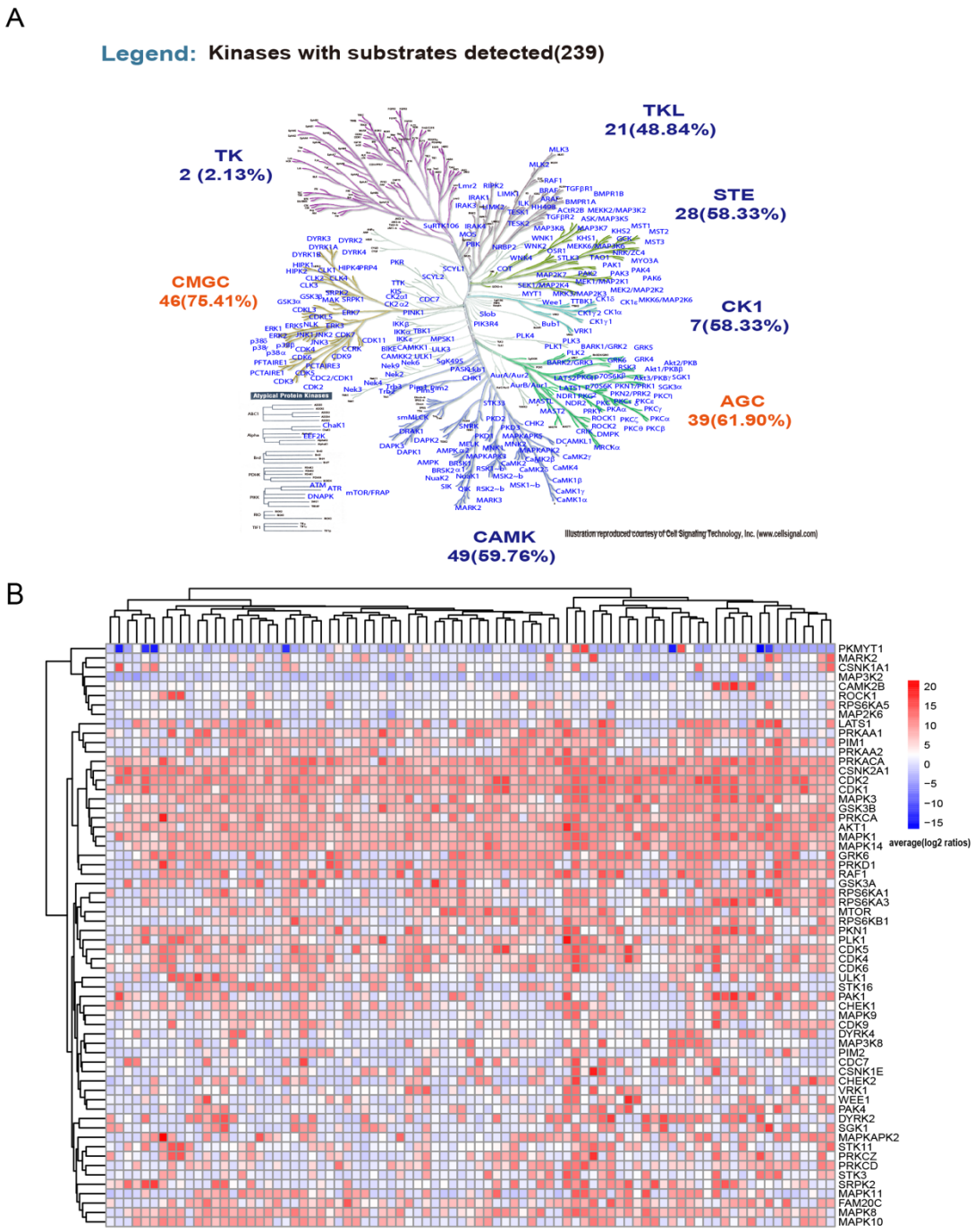


Fig.S5

Fig. S5. Kinome tree and kinase activity pattern in DGC, Related to Figure3.

A. Kinome tree annotated using Kinome Render from Cell Signalling Technology, Inc.(www.cellsignal.com); **B.** The average fold difference for all detected substrate sites for the same kinase.

Fig.S6

Fig. S6. Linear regression results of the activation loop phosphorylation sites, Related to Figure 3.

Examples for the correlations between the kinases activities predicted by the **A.** Mean values **B.** KSEA **C.** regression methods with the intensities of the activation loop phosphorylation sites. R^2: r square in the linear regression analysis.

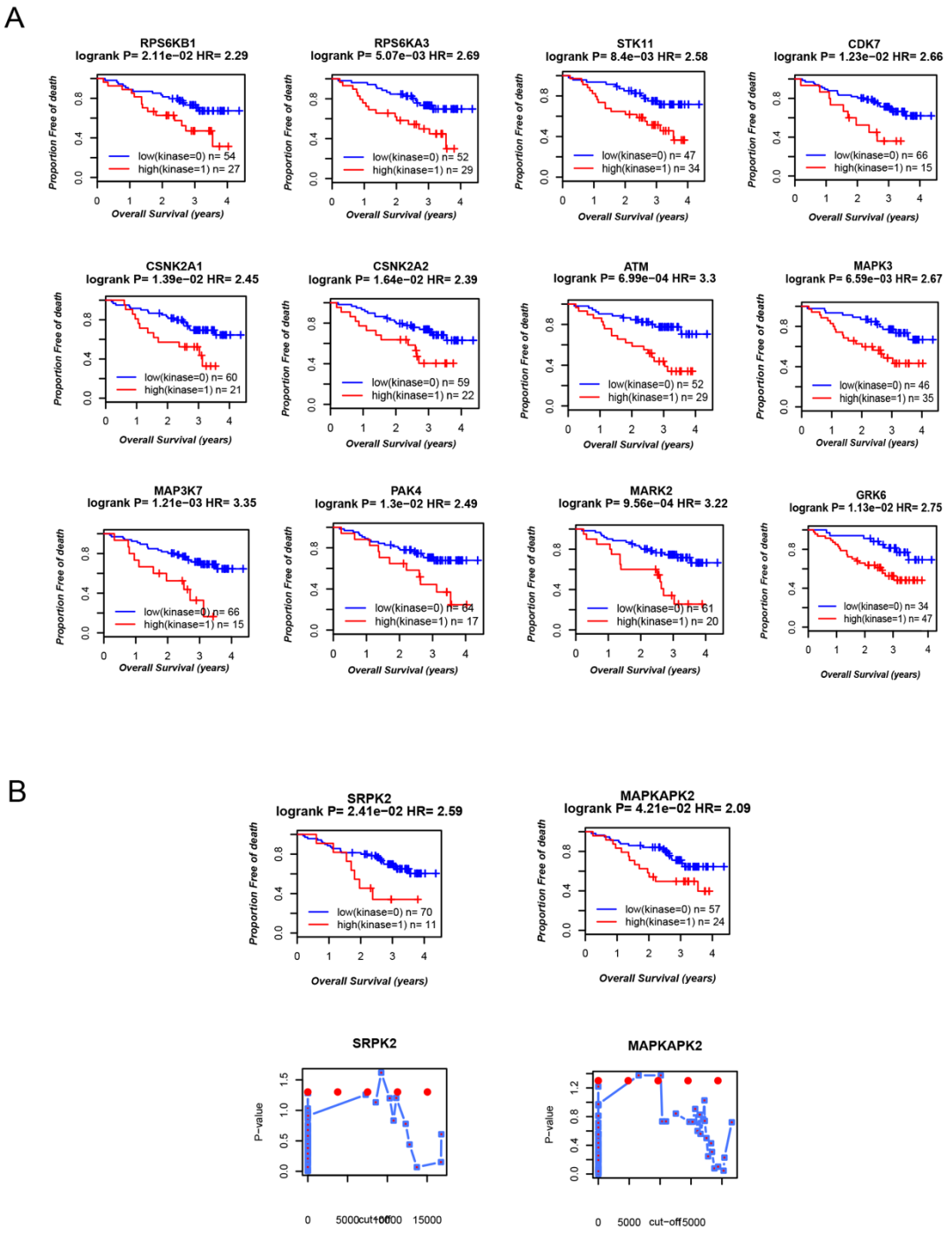


Fig.S7

Fig.S7. Cutoff curves of kinase activity predicted by the mean value (A) and KSEA methods (B), Related to Figure 3.

Each P-value was obtained by the log-rank test between the patients with kinase activity above the cut-off and the patients with kinase activity below the cut-off.

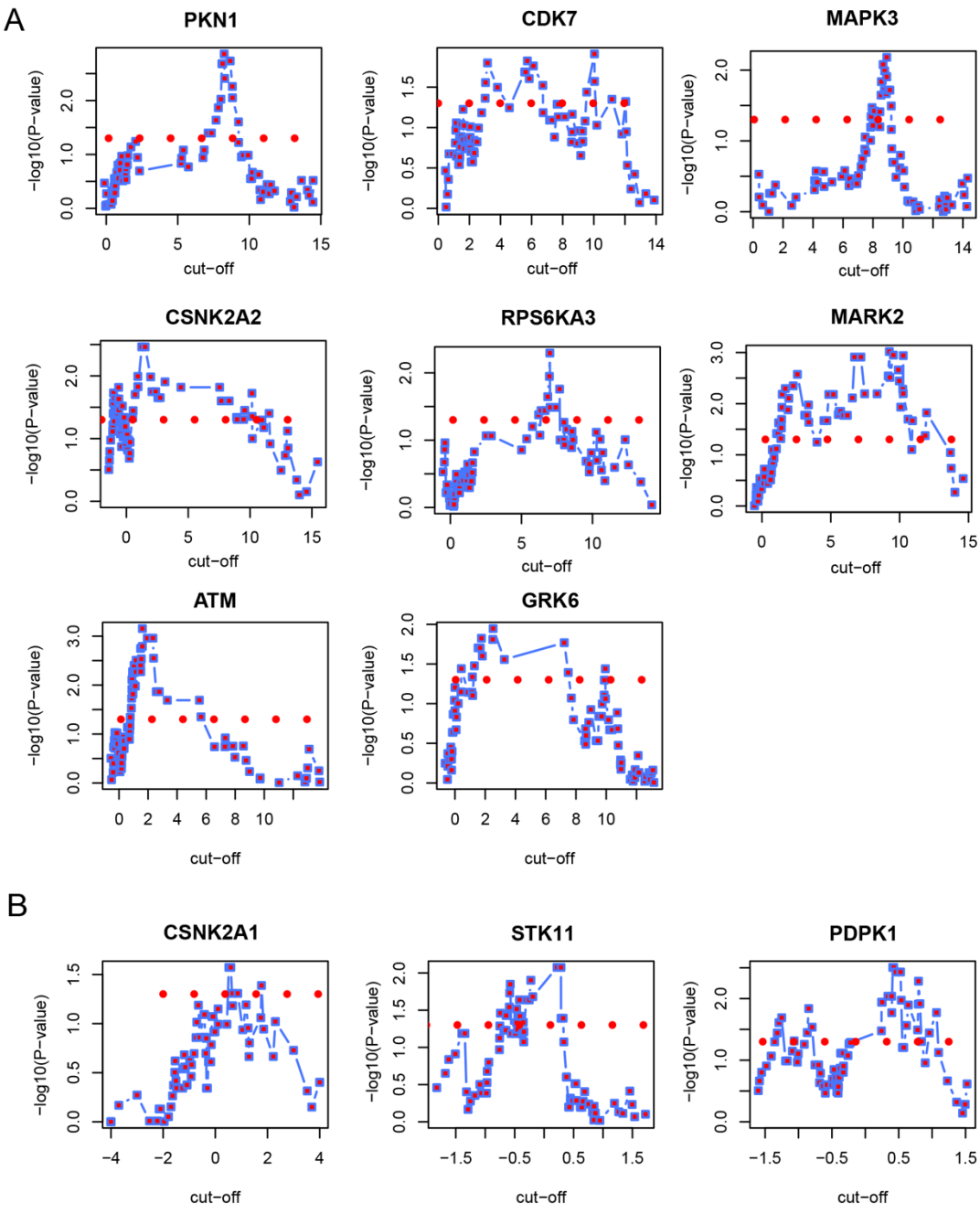


Fig.S8

Fig.S8. Cutoff curves of kinases activity predicted by the Ridge (A) and Z-test methods(B), Related to Figure 3.

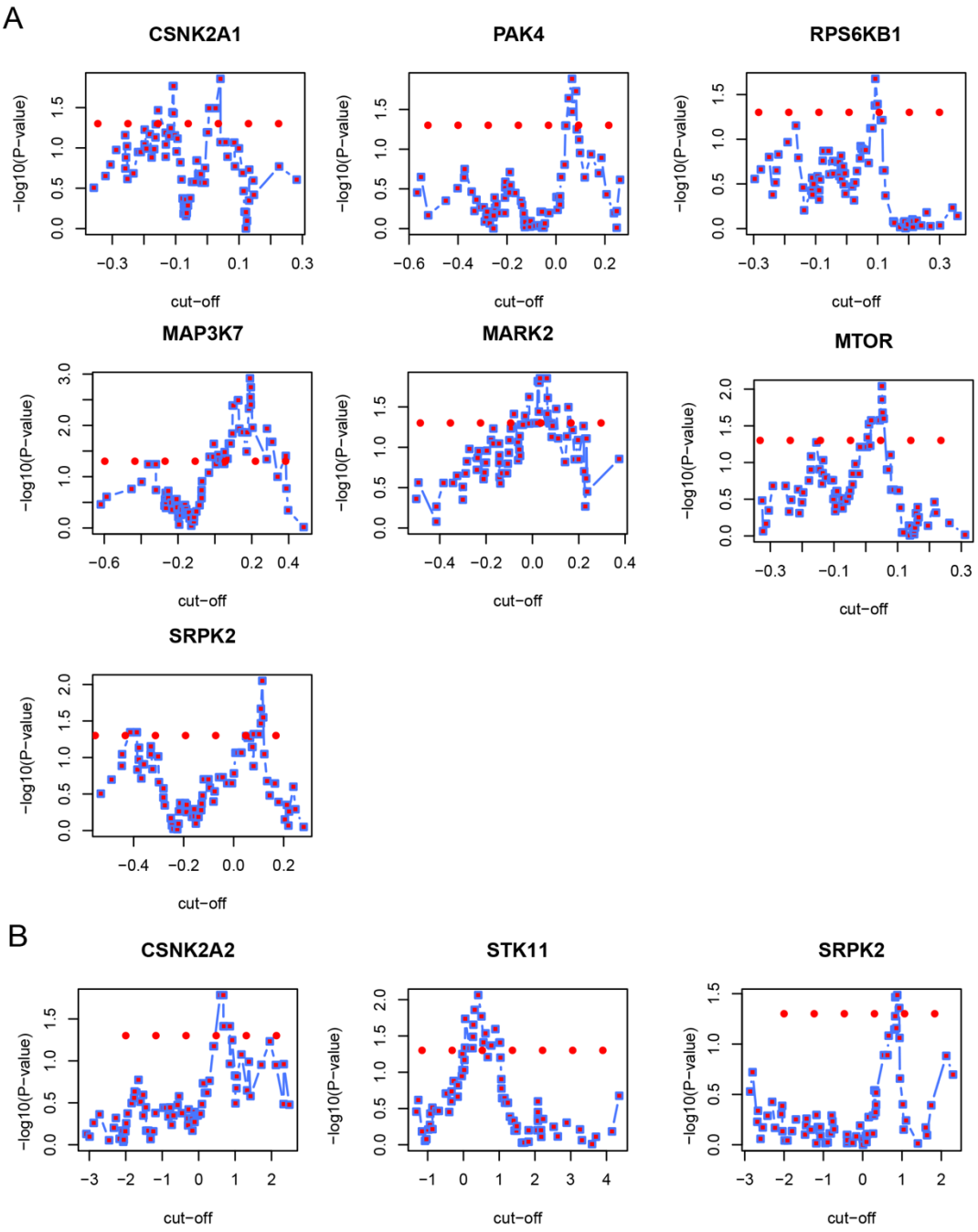


Fig.S9

Fig.S9. Heatmaps of substrates phosphorylation difference in two prognostic group, Related to Figure 3.

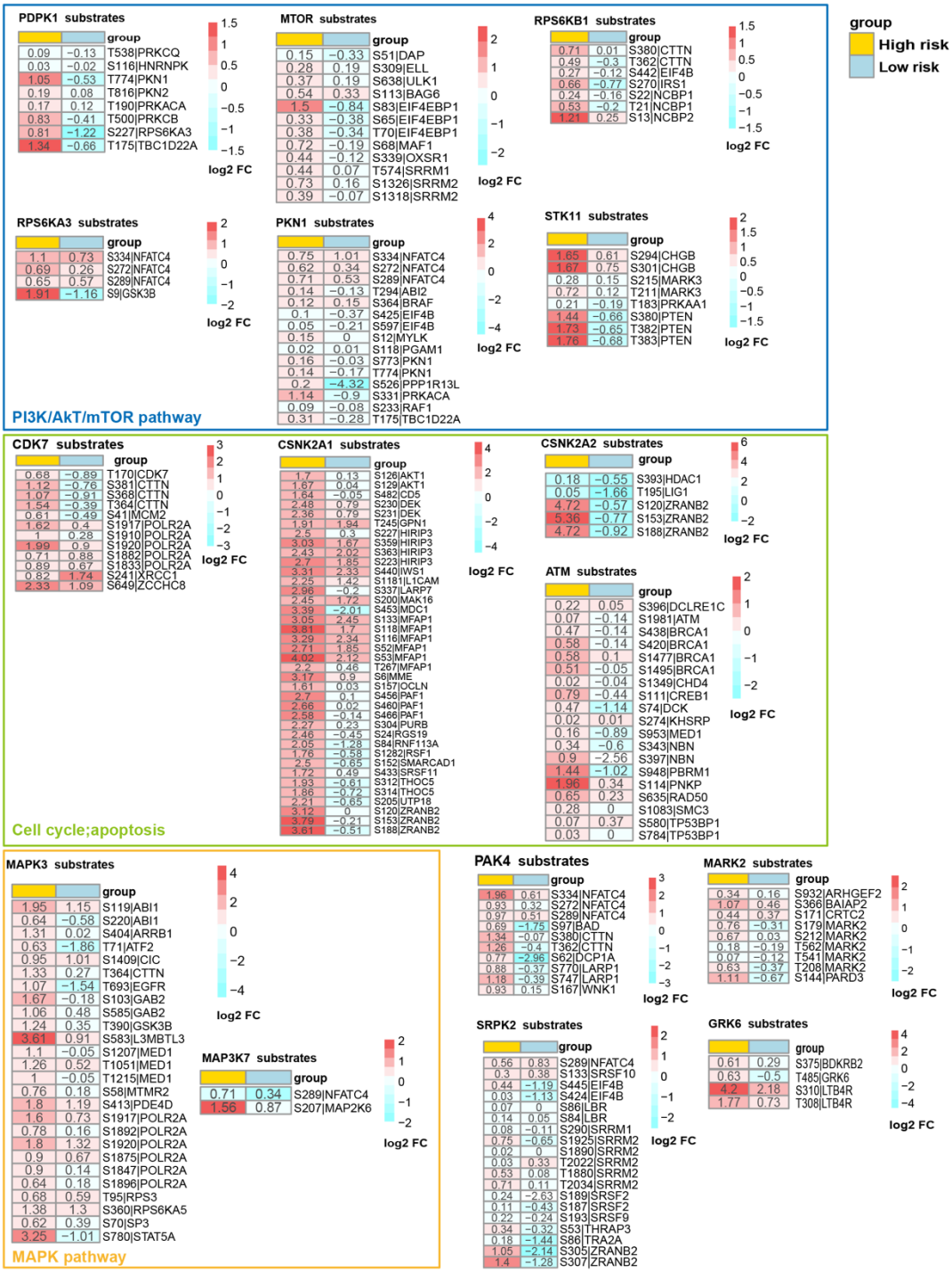

group
High risk
Low risk
PDPK1 substrates
group
0.09 −0.13 T538|PRKCQ
0.03 −0.02 S116|HNRNPK
1.05 −0.53 T774|PKN1
0.19 0.08 T816|PKN2
0.17 0.12 T190|PRKACA
0.83 −0.41 T500|PRKCB
0.81 −1.22 S227|RPS6KA3
1.34 −0.66 T175|TBC1D22A
log2 FC
MTOR substrates
group
0.15 −0.33 S51|DAP
0.28 0.19 S309|ELL
0.37 0.19 S638|ULK1
0.54 0.33 S113|BAG6
1.5 −0.84 S83|EIF4EBP1
0.33 −0.38 S65|EIF4EBP1
0.38 −0.34 T70|EIF4EBP1
0.72 −0.19 S68|MAF1
0.44 −0.12 S339|OXSR1
0.44 0.07 T574|SRRM1
0.73 0.16 S1326|SRRM2
0.39 −0.07 S1318|SRRM2
log2 FC
RPS6KB1 substrates
group
0.71 0.01 S380|CTTN
0.49 −0.3 T362|CTTN
0.27 −0.12 S442|EIF4B
0.66 −0.77 S270|IRS1
0.24 −0.16 S22|NCBP1
0.53 −0.2 T21|NCBP1
1.21 0.25 S13|NCBP2
log2 FC
RPS6KA3 substrates
group
1.1 0.73 S334|NFATC4
0.69 0.26 S272|NFATC4
0.65 0.57 S289|NFATC4
1.91 −1.16 S9|GSK3B
log2 FC
PKN1 substrates
group
0.75 1.01 S334|NFATC4
0.62 0.34 S272|NFATC4
0.71 0.53 S289|NFATC4
0.14 −0.13 T294|ABI2
0.12 0.15 S364|BRAF
0.1 −0.37 S425|EIF4B
0.05 −0.21 S597|EIF4B
0.15 0 S12|MYLK
0.02 0.01 S118|PGAM1
0.16 −0.03 S773|PKN1
0.14 −0.17 T774|PKN1
0.2 −4.32 S526|PPP1R13L
1.14 −0.9 S331|PRKACA
0.09 −0.08 S233|RAF1
0.31 −0.28 T175|TBC1D22A
log2 FC
STK11 substrates
group
1.65 0.61 S294|CHGB
1.67 0.75 S301|CHGB
0.28 0.15 S215|MARK3
0.72 0.12 T211|MARK3
0.21 −0.19 T183|PRKAA1
1.44 −0.66 S380|PTEN
1.73 −0.65 T382|PTEN
1.76 −0.68 T383|PTEN
log2 FC
PI3K/AkT/mTOR pathway
CDK7 substrates
group
0.68 −0.89 T170|CDK7
1.12 −0.76 S381|CTTN
1.07 −0.91 S368|CTTN
1.54 −0.39 T364|CTTN
0.61 −0.49 S41|MCM2
1.62 0.4 S1917|POLR2A
1 0.28 S1910|POLR2A
1.99 0.9 S1920|POLR2A
0.71 0.88 S1882|POLR2A
0.89 0.67 S1833|POLR2A
0.82 1.74 S241|XRCC1
2.33 1.09 S649|ZCCHC8
log2 FC
CSNK2A1 substrates
group
1.7 0.13 S126|AKT1
1.67 0.04 S129|AKT1
1.64 −0.05 S482|CD5
2.48 0.79 S230|DEK
2.36 0.79 S231|DEK
1.91 1.94 T245|GPN1
2.5 0.3 S227|HIRIP3
3.03 1.67 S359|HIRIP3
2.43 2.02 S363|HIRIP3
2.7 1.85 S223|HIRIP3
3.31 2.33 S440|IWS1
2.25 1.42 S1181|L1CAM
2.96 −0.2 S337|LARP7
2.45 1.72 S200|MAK16
3.39 −2.01 S453|MDC1
3.05 2.45 S133|MFAP1
3.81 1.7 S118|MFAP1
3.29 2.34 S116|MFAP1
2.71 1.85 S52|MFAP1
4.02 2.12 S53|MFAP1
2.2 0.46 T267|MFAP1
3.17 0.9 S6|MME
1.61 0.03 S157|OCLN
2.7 0.1 S456|PAF1
2.66 0.02 S460|PAF1
2.58 −0.14 S466|PAF1
2.27 0.23 S304|PURB
2.46 −0.45 S24|RGS19
2.05 −1.28 S84|RNF113A
1.76 −0.58 S1282|RSF1
2.5 −0.65 S152|SMARCAD1
1.72 0.49 S433|SRSF11
1.93 −0.61 S312|THOC5
1.86 −0.72 S314|THOC5
2.21 −0.65 S205|UTP18
3.12 0 S120|ZRANB2
3.79 −0.21 S153|ZRANB2
3.61 −0.51 S188|ZRANB2
log2 FC
CSNK2A2 substrates
group
0.18 −0.55 S393|HDAC1
0.05 −1.66 T195|LIG1
4.72 −0.57 S120|ZRANB2
5.36 −0.77 S153|ZRANB2
4.72 −0.92 S188|ZRANB2
log2 FC
ATM substrates
group
0.22 0.05 S396|DCLRE1C
0.07 −0.14 S1981|ATM
0.47 −0.14 S438|BRCA1
0.58 −0.14 S420|BRCA1
0.58 0.1 S1477|BRCA1
0.51 −0.05 S1495|BRCA1
0.02 −0.04 S1349|CHD4
0.79 −0.44 S111|CREB1
0.47 −1.14 S74|DCK
0.02 0.01 S274|KHSRP
0.16 −0.89 S953|MED1
0.34 −0.6 S343|NBN
0.9 −2.56 S397|NBN
1.44 −1.02 S948|PBRM1
1.96 0.34 S114|PNKP
0.65 0.23 S635|RAD50
0.28 0 S1083|SMC3
0.07 0.37 S580|TP53BP1
0.03 0 S784|TP53BP1
log2 FC
Cell cycle;apoptosis
MAPK3 substrates
group
1.95 1.15 S119|ABI1
0.64 −0.58 S220|ABI1
1.31 0.02 S404|ARRB1
0.63 −1.86 T71|ATF2
0.95 1.01 S1409|CIC
1.33 0.27 T364|CTTN
1.07 −1.54 T693|EGFR
1.67 −0.18 S103|GAB2
1.06 0.48 S585|GAB2
1.24 0.35 T390|GSK3B
3.61 0.91 S583|L3MBTL3
1.1 −0.05 S1207|MED1
1.26 0.52 T1051|MED1
1 −0.05 T1215|MED1
0.76 0.18 S58|MTMR2
1.8 1.19 S413|PDE4D
1.6 0.73 S1917|POLR2A
0.78 0.16 S1892|POLR2A
1.8 1.32 S1920|POLR2A
0.9 0.67 S1875|POLR2A
0.9 0.14 S1847|POLR2A
0.64 0.18 S1896|POLR2A
0.68 0.59 T95|RPS3
1.38 1.3 S360|RPS6KA5
0.62 0.39 S70|SP3
3.25 −1.01 S780|STAT5A
log2 FC
MAP3K7 substrates
group
0.71 0.34 S289|NFATC4
1.56 0.87 S207|MAP2K6
log2 FC
MAPK pathway
PAK4 substrates
group
1.96 0.61 S334|NFATC4
0.93 0.32 S272|NFATC4
0.97 0.51 S289|NFATC4
0.69 −1.75 S97|BAD
1.34 −0.07 S380|CTTN
1.26 −0.4 T362|CTTN
0.77 −2.96 S62|DCP1A
0.88 −0.37 S770|LARP1
1.18 −0.39 S747|LARP1
0.93 0.15 S167|WNK1
log2 FC
MARK2 substrates
group
0.34 0.16 S932|ARHGEF2
1.07 0.46 S366|BAIAP2
0.44 0.37 S171|CRTC2
0.76 −0.31 S179|MARK2
0.67 0.03 S212|MARK2
0.18 −0.19 T562|MARK2
0.07 −0.12 T541|MARK2
0.63 −0.37 T208|MARK2
1.11 −0.67 S144|PARD3
log2 FC
SRPK2 substrates
group
0.56 0.83 S289|NFATC4
0.3 0.38 S133|SRSF10
0.44 −1.19 S445|EIF4B
0.03 −1.13 S424|EIF4B
0.07 0 S86|LBR
0.14 0.05 S84|LBR
0.08 −0.11 S290|SRRM1
0.75 −0.65 S1925|SRRM2
0.02 0 S1890|SRRM2
0.03 0.33 T2022|SRRM2
0.53 0.08 T1880|SRRM2
0.71 0.11 T2034|SRRM2
0.24 −2.63 S189|SRSF2
0.11 −0.43 S187|SRSF2
0.22 −0.24 S193|SRSF9
0.34 −0.32 S53|THRAP3
0.18 −1.44 S86|TRA2A
1.05 −2.14 S305|ZRANB2
1.4 −1.28 S307|ZRANB2
log2 FC
GRK6 substrates
group
0.61 0.29 S375|BDKRB2
0.63 −0.5 T485|GRK6
4.2 2.18 S310|LTB4R
1.77 0.73 T308|LTB4R
log2 FC

Fig.S10

Fig.S10. Two kinases whose overexpressed protein abundance as measured in the profiling experiments associated with overall survival, Related to Figure 3.

A.Survival and **B.** Cutoff curves of the 2 kinases; Low(Kinase=0)/high(Kinase=1): patients with kinase abundance lower/above than cut-off; Each P-value was obtained by the log-rank test between the patients with kinase abundance above the cut-off and the patients with kinase abundance below the cut-off.

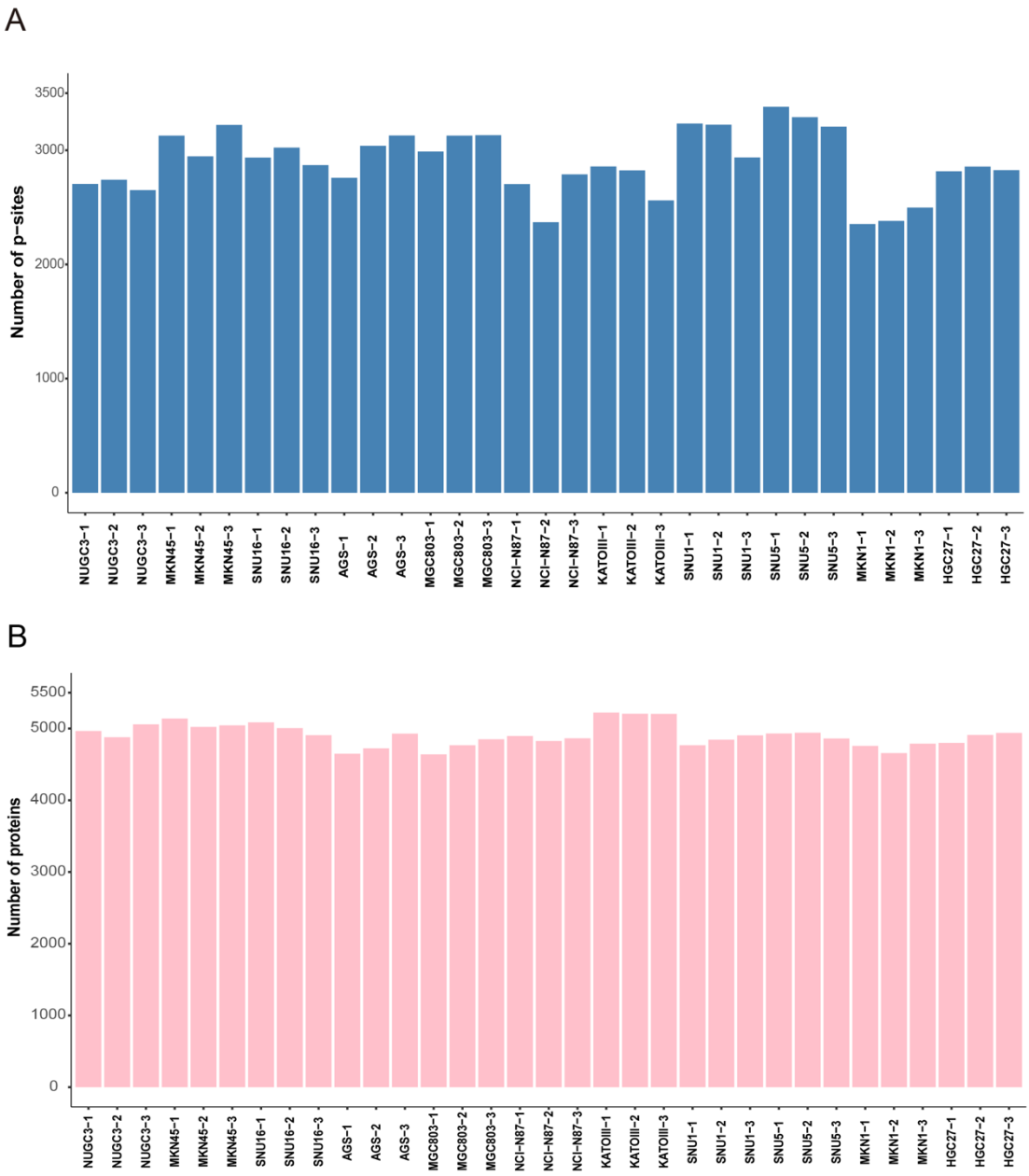


Fig.S11

Fig.S11.Global phosphorylation sites and proteins identifications in gastric cancer cell lines, Related to Figure 5.

A. The number of phosphorylation sites detected in gastric cancer cells; **B.** The number of proteins detected in gastric cancer cells.

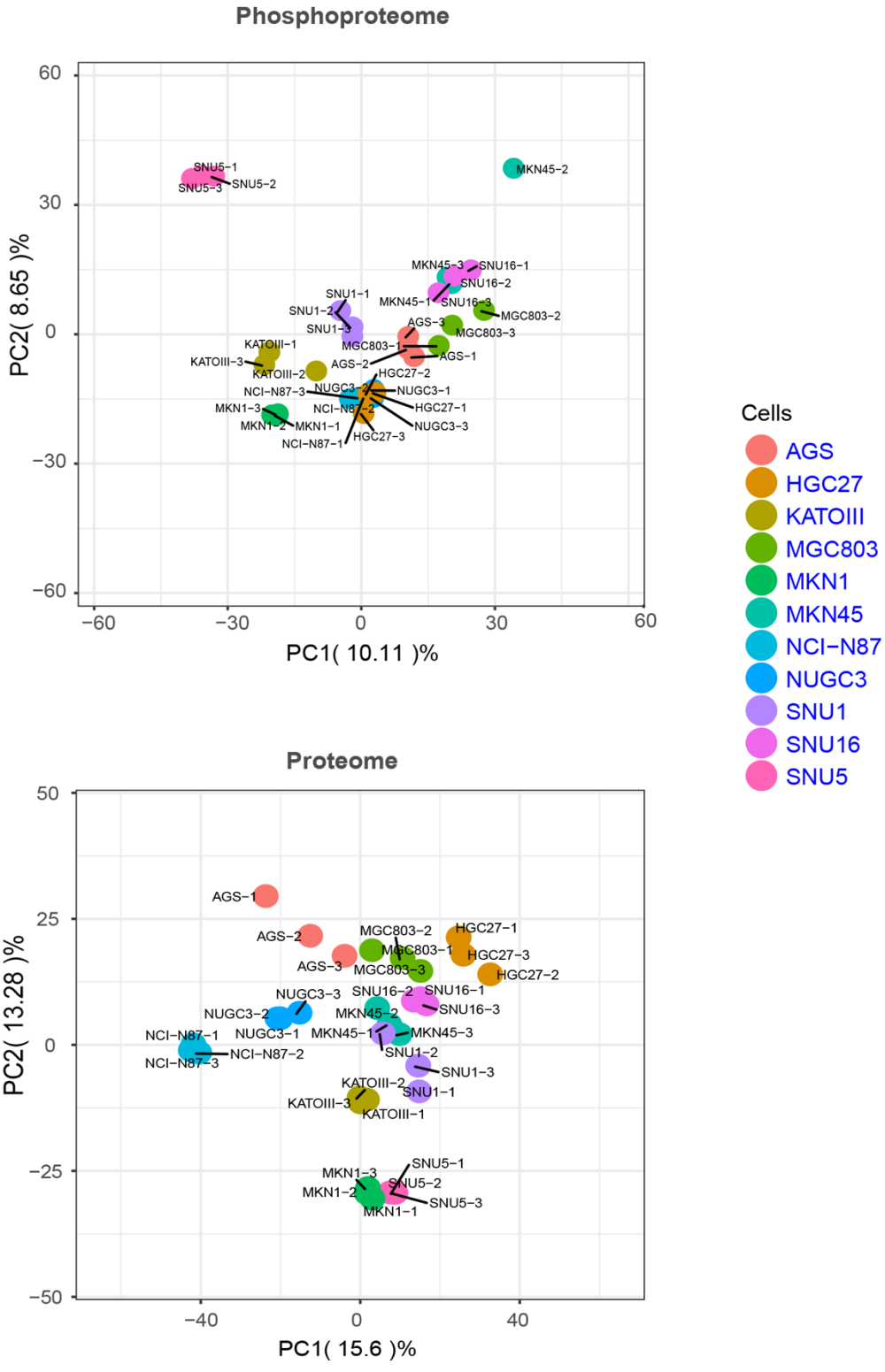

Phosphoproteome
PC2(8.65)%
PC1(10.11)%
60
30
0
−30
−60
−60
−30
0
30
60
SNU5−1
SNU5−3
SNU5−2
MKN45−2
MKN45−3
SNU16−1
SNU16−2
SNU1−1
SNU1−2
MKN45−1
SNU16−3
MGC803−2
SNU1−3
AGS−3
MGC803−3
KATOIII−1
MGC803−1
KATOIII−3
AGS−1
KATOIII−2
AGS−2
HGC27−2
NCI−N87−3
NUGC3−2
NUGC3−1
MKN1−3
NCI−N87−2
HGC27−1
MKN1−2
MKN1−1
NUGC3−3
NCI−N87−1
HGC27−3
Proteome
PC2(13.28)%
PC1(15.6)%
50
25
0
−25
−50
−40
0
40
AGS−1
AGS−2
MGC803−2
HGC27−1
MGC803−1
HGC27−3
AGS−3
MGC803−3
HGC27−2
NUGC3−3
SNU16−2
SNU16−1
NUGC3−2
MKN45−2
SNU16−3
NCI−N87−1
NUGC3−1
MKN45−1
MKN45−3
NCI−N87−2
SNU1−2
NCI−N87−3
SNU1−3
KATOIII−2
SNU1−1
KATOIII−3
KATOIII−1
SNU5−1
MKN1−3
SNU5−2
MKN1−2
SNU5−3
MKN1−1
Cells
AGS
HGC27
KATOIII
MGC803
MKN1
MKN45
NCI−N87
NUGC3
SNU1
SNU16
SNU5

Fig.S12

Fig.S12. PCA analysis of gastric cancer cells in both the proteomes and phosphoproteomes data, Related to Figure 5.

The phosphorylation sites or the proteins with higher abundance, whose maximum intensity in all experiments ranked in the top 75% were used for the PCA analysis.

Supplementary Tables

Table S1: Clinical characteristics of 83 patients, Related to Table 1.

Table S2: iFOT of 28,016 phosphorylation sites in the DGC phosphoproteome after quality control, Related to Figure 1.

Table S3: Anotation of phosphorylation sites and the corresponding proteins for all samples/Ph1/Ph2/Ph3, Related to Figure 1 and Figure 2.

Table S4: Pathways that are significantly altered in tumors compared with nearby tissues (FDR<1%), Related to Figure 2.

Table S5: The log2iFOT in tumors after quantile normalization of 302 phosphorylation site identified by anova analysis of three subtypes (FDR<0.001) , Related to Figure 2.

Table S6: Pathways and kinases that are enriched with phosphorylation sites clustrered by the anova analysis(P<0.05) , Related to Figure 2.

Table S7: Pathways that are significantly altered in tumors compared with nearby tissues for each subtype (P<0.05) , Related to Figure 2.

Table S8: Baseline Characteristics of the Ph1,Ph2 and Ph3 subtypes(n=83) , Related to Figure 2.

Table S9: Kinase-substrate relationships from public databases, Related to Figure3.

Table S10: Kinases with average log2ratios of sustrates from 83 DGC patients, Figure 3.

Table S11: Nomination of kinases as potential therapeutic targets, Related to Figure 3.

Table S12: The validation results of candidate kinase targets from published paper and the immunohistochemistry results from Human protein atlas database, Related to Figure 3.

Table S13: 287 unfavorable phosphorylation sites (FDR<0.1) , Related to Figure4.

Table S14: Pathways and kinases that are enriched with coexpression phosphorylation sites, Related to Figure 4.

Table S15: The characteristics of gastric cancer cells, Figure 5.

Table S16: The proteomes and phosphoproteomes data of gastric cancer cells, Related to Figure 5.

Reference

FENG, J., DING, C., QIU, N., NI, X., ZHAN, D., LIU, W., XIA, X., LI, P., LU, B., ZHAO, Q., NIE, P., SONG, L., ZHOU, Q., LAI, M., GUO, G., ZHU, W., REN, J., SHI, T. & QIN, J. 2017. Firmiana: towards a one-stop proteomic cloud platform for data processing and analysis. *Nat Biotechnol,* 35**,** 409-412.

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., HU, Y., LI, Y., LI, Y., GAO, J., CHEN, L., XU, J., ZHANG, C., JUNG, S. Y., CHOI, J. M., JAIN, A., LIU, M., SONG, L., LIU, W., GUO, G., GONG, T., HUANG, Y., QIU, Y., HUANG, W., SHI, T., ZHU, W., WANG, Y., HE, F., SHEN, L. & QIN, J. 2018. A proteomic landscape of diffuse-type gastric cancer. *Nat Commun,* 9**,** 1012.

HERNANDEZ-ARMENTA, C., OCHOA, D., GONCALVES, E., SAEZ-RODRIGUEZ, J. & BELTRAO, P. 2017. Benchmarking substrate-based kinase activity inference using phosphoproteomic data. *Bioinformatics,* 33**,** 1845-1851.

HUANG DA, W., SHERMAN, B. T. & LEMPICKI, R. A. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc,* 4**,** 44-57.

SAVITSKI, M. M., LEMEER, S., BOESCHE, M., LANG, M., MATHIESON, T., BANTSCHEFF, M. & KUSTER, B. 2011. Confident phosphorylation site localization using the Mascot Delta Score. *Mol Cell Proteomics,* 10**,** M110 003830.

SCHWANHAUSSER, B., BUSSE, D., LI, N., DITTMAR, G., SCHUCHHARDT, J., WOLF, J., CHEN, W. & SELBACH, M. 2011. Global quantification of mammalian gene expression control. *Nature,* 473**,** 337-42.

TUSHER, V. G., TIBSHIRANI, R. & CHU, G. 2001. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A,* 98**,** 5116-21.

WILKERSON, M. D. & HAYES, D. N. 2010. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics,* 26**,** 1572-3.

WISNIEWSKI, J. R., ZOUGMAN, A., NAGARAJ, N. & MANN, M. 2009. Universal sample preparation method for proteome analysis. *Nat Methods,* 6**,** 359-62.