RESEARCH ARTICLE

Global molecular dysfunctions in gastric cancer revealed by an integrated analysis of the phosphoproteome and transcriptome

**Tiannan Guo · Sze Sing Lee · Wai Har Ng ·
Yi Zhu · Chee Sian Gan · Jiang Zhu · Haixia Wang ·
Shiang Huang · Siu Kwan Sze · Oi Lian Kon**

Abstract We integrated LC-MS/MS-based and protein antibody array-based proteomics with genomics approaches to investigate the phosphoproteome and transcriptome of gastric cancer cell lines and endoscopic gastric biopsies from normal subjects and patients with benign gastritis or gastric cancer. More than 3,000 non-redundant phosphorylation sites in over 1,200 proteins were identified in gastric cancer cells. We correlated phosphoproteome data with transcriptome data sets and reported the expression of 41 protein kinases, 5 phosphatases and 65 phosphorylated mitochondrial proteins in gastric cancer cells. Transcriptional expression levels of 190 phosphorylated proteins were >2-fold higher in gastric cancer cells compared to normal stomach tissue. Pathway analysis demonstrated over-presentation of DNA damage response pathway and underscored critical roles of phosphorylated p53 in gastric cancer. This is the first study to comprehensively report the gastric cancer phosphoproteome. Integrative analysis of the phosphoproteome and transcriptome provided an expansive view of molecular signaling pathways in gastric cancer.

Keywords Gastric cancer · Phosphoproteome ·
Transcriptome · Protein antibody array · Protein kinase ·
Protein phosphatase · Mitochondria ·
DNA damage response

Abbreviations
RTK	Receptor tyrosine kinase
MS	Mass spectrometry
HPLC	High-performance liquid chromatography
FDR	False discovery rate
ERLIC	Electrostatic repulsion-hydrophilic interaction chromatography
SCX	Strong cation exchange
IMAC	Immobilized metal ion affinity chromatography
DDR	DNA damage response

T. Guo · S. S. Lee · W. H. Ng · O. L. Kon (✉)
Division of Medical Sciences, Humphrey Oei Institute of Cancer
Research, National Cancer Centre Singapore, 11 Hospital Drive,
Singapore 169610, Singapore
e-mail: dmskol@nccs.com.sg

T. Guo · Y. Zhu · C. S. Gan · S. K. Sze (✉)
School of Biological Sciences, Nanyang Technological
University, 60 Nanyang Drive, Singapore 637551, Singapore
e-mail: sksze@ntu.edu.sg

J. Zhu · H. Wang · S. Huang
Center for Stem Cell Research and Application, Union Hospital,
Huazhong University of Science and Technology,
430022 Wuhan, People's Republic of China

Introduction

Gastric cancer is the one of the most prevalent cancers and a major cause of cancer mortality worldwide [1]. Gastric cancer is generally refractory to curative radiotherapy and chemotherapy. Despite decades of steadily declining incidence, gastric cancer fatality rates remain paradoxically high in most countries. Multiple reasons contribute to gastric cancer's poor 5-year survival rate that remains low at about 33%. Surgical resection offers the only cure at present; therapy with conventional cytotoxic agents has only modest efficacy in extending survival [2].

Recent improvements in survival of some malignancies, including chronic myeloid leukemia, non-small cell lung

cancer and breast cancer, among others, owe much to advances in uncovering aberrantly active molecular pathways, from which molecule-targeting agents have been developed as new strategies to control cancers [3]. Experimentally and clinically validated agents include, but are not limited to, inhibitors of receptor and non-receptor tyrosine kinases (EGFR, HER2, HER3, insulin-like growth factor receptor, MET, fibroblast growth factor receptor and HSP 90 inhibitors), intracellular signaling pathways (PI3K, AKT, mTOR) and angiogenesis, and agents that interfere with DNA repair (PARP inhibitors) [4].

The efficacy of targeted agents appears to be cell context dependent. Deeper understanding is needed to identify molecular predictors of responses of cancer cells to such agents [5]. Cancer cells that are initially sensitive to suppression of a specific target commonly become resistant within 1 year of treatment [6]. Recent molecular mechanistic studies attribute acquired resistance to activation of alternative oncogenic signaling pathways that successfully bypass the point of inhibition. This is facilitated by extensive cross-talk known to exist among multiple receptor tyrosine kinase (RTK) signaling pathways [7]. Non-small cell lung cancer cells initially sensitive to EGFR inhibition acquire resistance by activating the MET-HER3-PI3K signaling pathway [8]. Acquired resistance of HER2-overexpressing breast cancers is dependent on activation of the insulin-like growth factor 1 receptor-PI3K/Akt signaling pathway [9]. There is evidence of improved efficacy when multiple targets are concomitantly suppressed [10]. Advancing these insights to durable clinical benefit will require in-depth understanding of oncogenic signaling networks in specific cancer types from which molecular predictors of response and strategies for avoiding or subverting acquired resistance may be devised.

Investigations into oncogenic signaling networks in gastric cancer have lagged relative to other common malignancies like lung cancer, breast cancer and leukemia. Therapeutics targeting HER2, EGFR, VEGFR, MET, IGFR and FGFR, which have proven efficacy in other cancers, are being tested in gastric cancer [11–15]. To date, a clinical trial directed at only one target, HER2, has reported significant but modest extension of survival of gastric cancer patients [11]. This remains to be confirmed in independent trials, especially as a related study found HER2 expression to be uncommon in gastric cancer and unrelated to prognosis [16]. A major challenge in developing targeted therapy is the current paucity of mechanistic understanding of gastric oncogenesis as mediated by signaling pathways [15].

Phosphoproteomic profiling sheds light on key components in oncogenic signaling networks [17, 18]. Although there are databases of the phosphoproteomes of various organisms and cell types, e.g., Phospho.ELM [19, 20],

PhosphoSitePlus [21], phosphoPep [22], PHOSIDA [23] and Uniprot, these have limited relevance for gastric cancer research because phosphorylation profiles are highly diverse and differ in a cell type-dependent manner. A global inventory of the kinome and phosphoproteome specific to gastric cancer has yet to be reported and is an essential first step to mapping disease-specific oncogenic pathways.

Protein kinases are the most frequently dysfunctional proteins in various cancers [24]. The human kinome intricately regulates phosphorylation of approximately one-third of the proteome. The gastric cancer kinome has not been systematically investigated yet. Protein phosphatases are the obverse of kinases and an equally important class of phosphorylation regulators [25]. More than 130 protein phosphatase genes have been identified in the human genome [26], several of which, e.g., CDC25, SHP2, phosphatase of regenerating liver (PRL) and CD45, have been proposed as critical oncoproteins and potential drug targets [27].

As an essential prerequisite to understanding oncogenic signaling networks in gastric cancer, we have performed a comprehensive investigation of the gastric cancer phosphoproteome and transcriptome. We report an expansive view of the gastric cancer oncokinome and phosphoproteome from an integrated bioinfomatics analysis of phosphoproteomics and transcriptome data sets.

Materials and methods

All chemicals were purchased from Sigma–Aldrich (St. Louis, MO) unless otherwise stated.

Cell culture and primary gastric tissues

Seventeen gastric cancer cell lines and endoscopic biopsies of stomach tissues were investigated in this study. AGS, Kato III, SNU1, SNU5, SNU16, NCIN87 and Hs746T were from the American Type Culture Collection (the ATCC, Manassas, VA). MKN7 and IM95 cells were from the Japan Health Science Research Resource Bank. All YCC cell lines were gifts from Dr. Sun Young Rha (Yonsei Cancer Center, Seoul, Korea). Normal stomach RNA samples were reference controls for transcriptome analysis. First Choice Human Stomach Total RNA was RNA from a single individual, whereas MVP Total RNA Human Stomach was pooled RNAs from two individuals. Fresh stomach biopsies were obtained from patients during gastroscopy performed for clinical indications and immediately frozen in liquid nitrogen before protein array analysis. After histopathological diagnosis, two histologically normal gastric biopsies, seven biopsies of benign gastritis and three pairs of gastric adenocarcinoma with their matched normal gastric tissues were analyzed (supplemental Table 1). Clinical specimens

were obtained in conformity with principles of the Declaration of Helsinki under a protocol approved by the SingHealth Centralised Institutional Review Board, Singapore.

Gene expression analysis

Transcriptomes of 17 gastric cancer cell lines and normal stomach RNA samples were analyzed using two microarray formats, i.e., Affymetrix HG-U133 and HG-U133 Plus 2.0 GeneChip®. Microarray data sets were averaged and normalized. Normal gastric tissue RNA served as reference controls to identify differentially expressed genes. Signal intensities of normal stomach tissue genes were averaged for each probe and used as divisors for cognate signal intensities of gastric cancer cell lines. The product values were regarded as the relative expression levels of the respective genes in gastric cancer. Values for probes belonging to the same gene were grouped and averaged.

Protein sample preparation

Gastric cancer cells were lysed in 50 mM HEPES (pH 7.5), 8 M urea, 75 mM NaCl, complete protease inhibitors cocktail (Roche Applied Science, Indianapolis, IN) and phosSTOP phosphatase inhibitors cocktail (Roche Applied Science). Proteins were reduced by adding dithiothreitol (final concentration 10 mM) to the sample solution at 33°C for 1 h, and then alkylated by adding iodoacetamide to a final concentration of 55 mM and incubating the samples at room temperature for 30 min, before diluting eight times with 50 mM HEPES (pH 7.5) and digestion with trypsin in a 1:100 (trypsin/protein) mass ratio. Protein concentrations were measured using bicinchoninic acid (BCA) assay. Peptide samples were desalted using SEP-PAK C18 cartridges (Waters Corp., Milford, MA) and vacuum-dried prior to phosphopeptide enrichment.

Phosphopeptide enrichment

Phosphopeptides were enriched using both ERLIC and SCX-IMAC as described [28]. Briefly, for ERLIC, approximately 2 mg of peptides was injected into a PolyLC PolyWAX LP column (4.6 × 200 mm, 5 μm particle size, 300Å pore size) mounted on a Shimadzu Prominence UFLC unit (Shimadzu Corporation, Kyoto, Japan). For SCX-IMAC, approximately 2 mg of peptides was fractionated using a PolySULFOETHYL A column (4.6 × 100 mm, 5 μm particle size, 200 Å pore size) on the UFLC unit. Each SCX fraction was dissolved in 100 μl of wash buffer (250 mM acetic acid with 30% acetonitrile, pH 2.6) and subsequently added to 20 μl of IMAC slurry (50% gel) (PHOS-Select, Sigma-Aldrich) for 1 h at room temperature with end-over-end rotation. Phosphopeptides were eluted with 100 μl of 200 mM Na_3PO_4 (pH 8.4) by incubating at room temperature for 5 min. Elution was repeated twice using 100 μl each of 50 mM Tris (pH 10) and 400 mM NH_4OH (pH 11). For each fraction, all the three eluents were combined immediately and pH adjusted to 2.6 using 10% formic acid. Peptides in salt solutions were desalted using SEP-PAK C18 cartridges and vacuum-dried.

LC-MS/MS analysis

Each dried peptide fraction was reconstituted in 0.1% formic acid and analyzed at least twice using an LTQ-FT ultra mass spectrometer (Thermo Fisher Scientific, Inc., Waltham, MA) coupled with a Prominence™ HPLC unit (Shimadzu), as described previously [29, 30] with some modifications. Briefly, the peptide samples were injected from an auto-sampler (Shimadzu) and concentrated in a Zorbax peptide trap (Agilent, Palo Alto, CA), and subsequently resolved in a capillary column (200 μm ID × 10 cm) packed with C18 AQ (5 μm particles, 100 Å pore size; Michrom BioResources, Auburn, CA). Mobile phase buffer A (0.1% formic acid in H_2O) and buffer B (0.1% formic acid in acetonitrile) were used to establish the 90-min gradient, which began with a ramp from 5 to 30% B over 66 min, followed by 10 min of 50% B and a ramp from 50 to 80% B in 4 min. The gradient was maintained at 80% B for 2 min before re-equilibrating the column at 5% B for 8 min. HPLC was operated at a constant flow rate of 20 μl/min, and a splitter was used to create a flow rate of approximately 300 nl/min at the electrospray emitter (Michrom BioResources). Samples were ionized in an ADVANCE™ CaptiveSpray™ Source (Michrom BioResources) with an electrospray potential of 1.5 kV. The gas flow was set at two, ion transfer tube temperature at 180°C and collision gas pressure at 0.85 mTorr. The LTQ-FT ultra was set to perform data acquisition in the positive ion mode. A full MS scan (range 350–2,000 m/z) was acquired in the FT-ICR cell at a resolution of 100,000 and a maximum ion accumulation time of 1,000 ms. The AGC target for FT was set at $1e^{+06}$, and precursor ion charge state screening was activated. The linear ion trap was used to collect peptides and measure peptide fragments generated by CID. The default AGC setting was used (full MS target at $3.0e^{+04}$, MS^n $1e^{+04}$) in the linear ion trap. The ten most intense ions above a 500 count threshold were selected for fragmentation in CID (MS2), which was performed concurrently with a one maximum ion accumulation time of 200 ms. Dynamic exclusion was activated for this process, with a repeat count of one, exclusion duration of 20 s and ±5 ppm mass tolerance. For CID, the activation Q was set at 0.25, isolation width (m/z) 2.0, activation time 30 ms and normalized collision energy at 35%.

Database search

The extract_msn (version 4.0) program in Bioworks Browser 3.3 (Thermo Electron, Bremen, Germany) was used to extract tandem MS spectra in the dta format from the raw data of LTQ-FT ultra. Dta files were then converted into MASCOT generic file format using an in-house program for each raw file. Intensity values and fragment ion m/z ratios followed the default setting. These data were used to obtain protein identities by searching against the IPI human protein database (version 3.70; 174,138 sequences) via multiple database search engines separately, including an in-house MASCOT server (version 2.2.03) (Matrix Science, Boston, MA), Sequest engine in Bioworks Browser, X!Tandem [31] (Tornado edition, version 2010.01.01.4) and OMSSA (command line version 2.1.7) [32]. All searches were limited to a maximum of two missed trypsin cleavages, mass tolerances of 10 ppm for peptide precursors (0.05 Da precursor tolerance for search in OMSSA) and 0.8-Da mass tolerances for fragment ions. The fixed modification was carbamidomethyl at Cys residues, whereas variable modifications were oxidation at Met residues and phosphorylation at Ser, Thr and Tyr residues. A combination of target and reverse sequence version decoy databases were used in Mascot, Sequest and OMSSA for estimation of false discovery rates (FDR). Here, $FDR = 2 \times Md/(Md + Mt)$, where Md represents the number of decoy matches, and Mt is the number of target matches. In X!Tandem, FDR was estimated by a default algorithm. FDR was adjusted to <1% for all searches by regulating cutoff values for peptide scores or expectation values. Output results from these engines were analyzed using in-house scripts.

Motif analysis

Phosphorylated sites for serines, threonines and tyrosines were submitted to Motif-X algorithm (http://motif-x.med.harvard.edu) for motif extraction, using the IPI human database as background. Extendible peptide sequences were centered on each phosphorylation site and extended to 13 amino acids (±6 residues). The minimum reported number of occurrences for a given motif was set at 2% of the total number of phosphorylation sites found for a given residue. Significance was set at 0.000001. Scansite [33] was also employed to predict the most likely kinases responsible for the phosphorylation sites in gastric cancer phosphoproteome.

Protein antibody array experiments

Protein lysates from stomach biopsies were probed for phosphorylated signaling proteins using Proteome Profiler™ antibody arrays (R&D Systems, Minneapolis, MN).

Manufactured in sets of two, the arrays interrogate 46 kinases and kinase substrates, with specific anti-phospho-amino acid antibodies spotted in duplicate. Experiments were performed according to the supplier's instructions. Briefly, 100 μg protein lysate was diluted with blocking buffer in 5:1 ratio and incubated overnight with pre-blocked nitrocellulose membranes. After three washes, the membranes were incubated with a mixture of biotinylated detection antibodies for 2 h at room temperature. Phosphorylated proteins were detected on washed membranes using streptavidin-horseradish peroxidase provided with the arrays and a chemiluminescent substrate reagent (Amersham ECL™ Western Blotting System, GE Healthcare, UK) on Amersham Hyperfilm™ ECL (GE Healthcare, UK). Developed x-ray films were scanned on a GS-800 Calibrated Densitometer (Bio-Rad Laboratories, UK). Pixel intensities for each spotted antibody were analyzed using Axon GenePix Pro 6.0 (Molecular Devices, USA).

Pathway analysis

Canonical pathway mapping was performed using ingenuity pathway analysis (IPA) application (www.ingenuity.com) against ingenuity pathway knowledge base.

Results

LC–MS/MS-based phosphoproteomic analysis of gastric cancer cell lines

Owing to the substoichiometric nature of protein phosphorylation, it is essential to enrich phosphopeptides in shotgun LC-MS/MS analysis [18]. Multiple enrichment methods are recommended for comprehensive shotgun phosphoproteome analysis [28, 34]. We employed two methods, electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) and SCX-IMAC, to enrich phosphopeptides [28, 35]. The benefits of using two different enrichment methods are shown in supplemental Figure 1. SCX-IMAC and ERLIC increased the coverage of SNU5 phosphoproteome by 122% and 58%, respectively. Only 8% of non-redundant phosphopeptides were identified by both methods. Five cell lines, i.e., SNU5, SNU1, AGS, YCC1 and KatoIII, were included in the phosphoproteomic analysis. Different gastric cancer cells are heterogeneous, and their phosphoproteomes exhibit different characteristics. However, due to the qualitative nature of this study, we did not compare phosphoproteomes between cell lines; instead, we combined spectral data from all cell lines to achieve a comprehensive picture of the gastric cancer phosphoproteome from diverse gastric cancer cells.

The complete translation of MS spectra obtained in LC-MS/MS experiments into peptide and protein assignments remains a major computational challenge in proteomics. Multiple protein sequence database search algorithms are available to interpret MS spectra, including Mascot, Sequest, X!Tandem [31] and OMSSA [32], among others. The sensitivity and specificity of database search engines are subject to substantial variations. While most studies are dependent on a single database search engine, the use of multiple database search engines has been shown to enhance the sensitivity of shotgun proteomics considerably [36, 37]. In this study, we analyzed MS spectra using four different database search engines. False discovery rates (FDRs) were set at <1% for all searches (supplemental Table 2). The benefit of using multiple engines is illustrated in Fig. 1, showing that spectra not identified in one engine, e.g., Mascot, could be characterized by another, e.g., Sequest.

Mascot and Sequest identified 718 and 441 phosphorylated proteins, respectively. A total of 210 and 227 phosphoproteins were identified by X!Tandem and OMSSA, respectively. The advantage of combining multiple database search engines in phosphoproteomics is further shown in

Fig. 2. A single engine identified only 17–59% of all phosphoproteins. By combining results from four search engines, the numbers of phosphoproteins and unique phosphorylation sites were substantially increased to a total of 3,021 unique phosphorylated peptides in 1,211 phosphorylated proteins from gastric cancer cells (supplemental Table 3). Among these, 547 (18%) phosphorylation sites and 295 (24%) phosphoproteins were identified by at least two search engines. Non-redundant phosphorylation sites comprised 2,144 phosphorylated serines, 673 phosphorylated threonines and 204 phosphorylated tyrosines. The distribution of pS, pT and pY was 71, 22 and 7%, respectively. These results are consistent with other findings that in some proteins phosphorylation sites with high occupancy are likely associated with serine, whereas those of low occupancy involve threonine and tyrosine [38]. Compared to previous reports of global phosphoproteome profiling [28], higher percentages of low-abundance phosphorylated threonines and tyrosines were identified in this study, reflecting increased sensitivity and greater phosphoproteome coverage of our workflow that combined different phosphopeptide enrichment methods (ERLIC and SCX-IMAC) and used multiple MS spectra interpretation approaches.
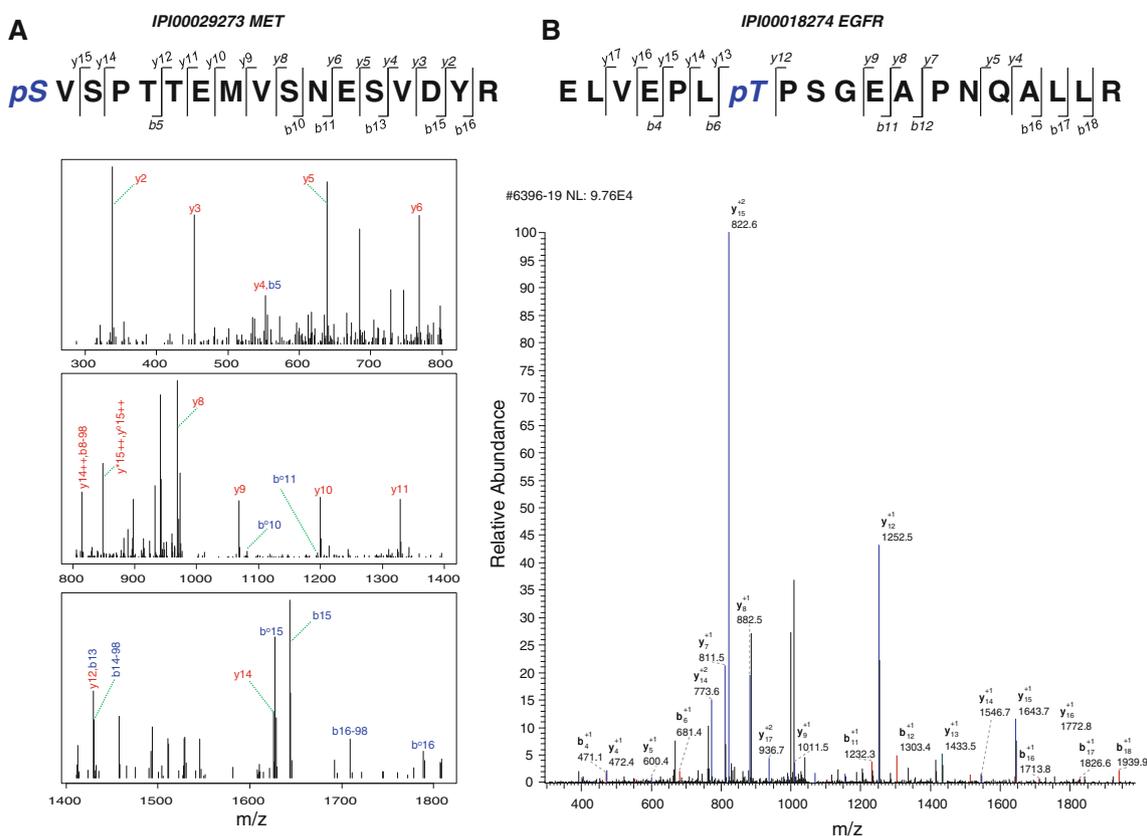

Fig. 1 MS spectra interpreted by Mascot and Sequest. Annotated MS/MS spectra for peptides from **a** MET, identified via Mascot and **b** EGFR, identified via Sequest are shown. Spectrum annotated by Mascot is split into three parts according to mass range. Detected b ions and y ions are annotated
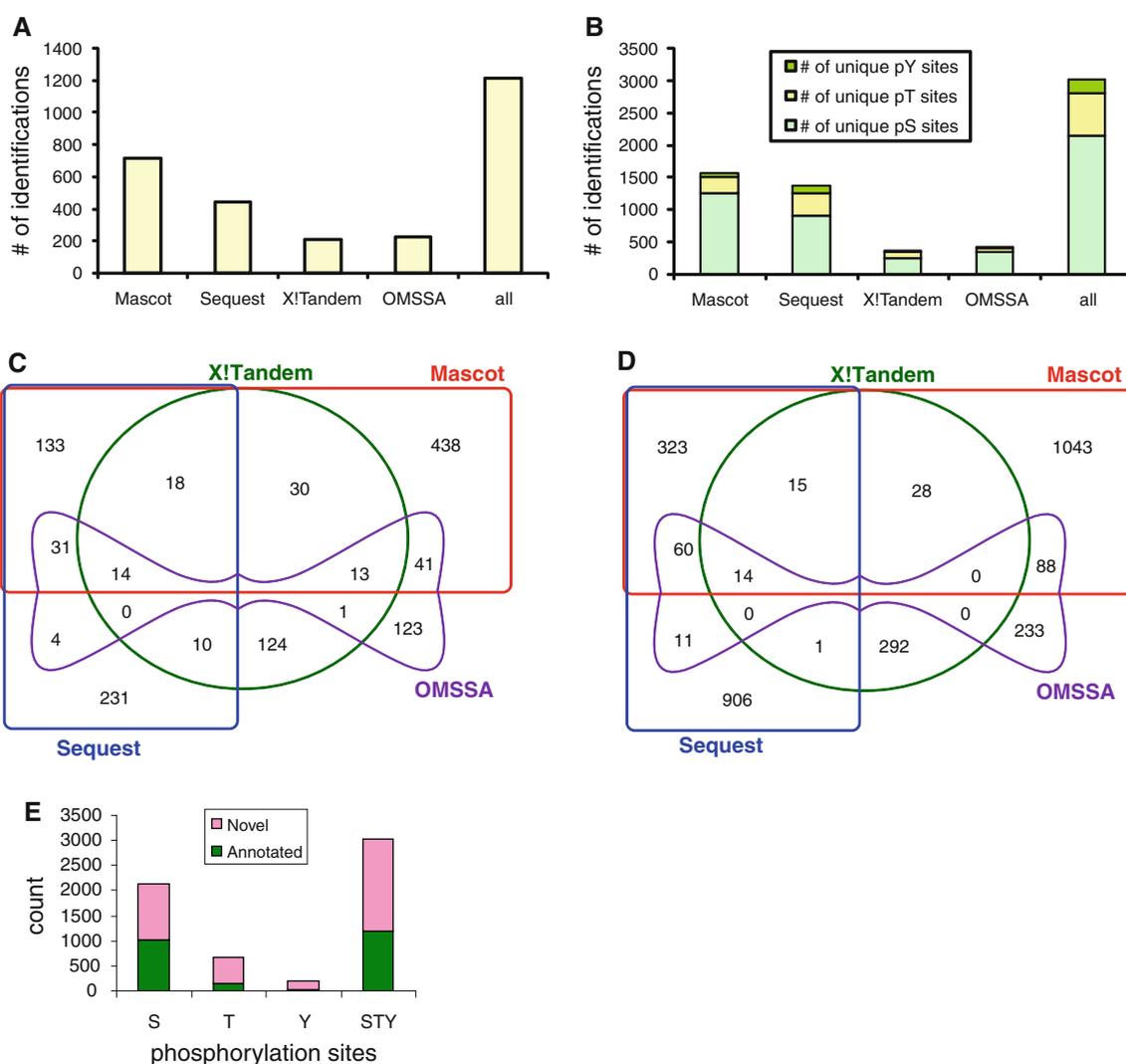
Fig. 2 Identification and characterization of gastric cancer phospho-proteome. **a** Number of phosphorylated proteins identified from database search engines Mascot, Sequest, X!Tandem and OMSSA. Unique phosphoproteins identified by overlapping results from the four engines are shown. **b** Number of non-redundant phosphorylation sites (pS, pT and pY, respectively) for each engine and the total number from all engines are shown. **c** *Venn diagram* of 1,211 non- redundant phosphoproteins identified by four database search engines. **d** *Venn diagram* of 3,021 non-redundant phosphorylation sites identified by four database search engines. **e** Summary of gastric cancer phosphoproteome by known annotated and novel phosphorylation sites. Non-redundant phosphorylation sites in gastric cancer cells were compared with Uniprot human database

Confidence measures for correct localization of phosphorylation sites

In shotgun proteomics, it is often difficult to pinpoint the correct position of phosphorylation sites with single amino acid resolution, especially for multiple phosphorylated peptides. To localize phosphorylation sites accurately, we first undertook a computational assessment of the phosphorylation site assignment using the Ascore algorithm [39]. As shown in supplemental Figure 2, 64% of the localizations were assigned with >90% confidence ($P < 0.05$) and 56% with >95% confidence ($P < 0.01$). Near certainty (>99% confidence, $P < 0.001$) of localization was achieved

for 44% of the data set. It should be noted that the Ascore algorithm did not take into account 207 phosphopeptides with unambiguous localization, i.e., those for which the number of potential phosphorylation sites was equal to the number of phosphorylation sites. After including these 207 unequivocal phosphopeptides, the number of localizations with 99, 95 and 90% confidence increased from 726, 930 and 1,056 to 915, 1,117 and 1,242, respectively. This indicated that the majority of phosphorylation assignments were of high confidence.

The quality of identification was further supported by the fact that many phosphorylation sites were found multiple times and in peptides that contained different numbers

and forms of phosphorylation sites. For instance, a phosphorylation site could be identified from fully or partially trypsin-digested peptides, with/without oxidized methionine, peptides with different charges and peptides with different numbers of phosphorylation modifications. As shown in supplemental Figure 3, 51% of phosphopeptides were singly phosphorylated, 27% were doubly phosphorylated, and 11% were triply phosphorylated. Only 1% of phosphopeptides carried four or more phosphates. This distribution was similar to phosphopeptides characterized in an earlier report [28]. Phosphopeptides detected in MS were ionized with different charges, as shown in supplemental Figure 3.

To further confirm phosphorylation site localization of the whole data set, we performed a final manual inspection of MS/MS spectra of phosphopeptides. All the identified MS/MS spectra with their database search identification information are listed in supplemental Table 3. In most cases, multiple spectra were interpreted as a single phosphopeptide sequence; only the spectrum with the highest identification score was manually inspected and is supplied in our website (http://proteomics.sbs.ntu.edu.sg/).

Characterizing the gastric cancer phosphoproteome

To characterize the gastric cancer phosphoproteome, we first checked whether the phosphoproteins we identified in this study were also present in other human phosphoproteome data sets. Of the 3,021 phosphorylation sites we identified, 1,194 (40%) were annotated in the Uniprot database. Thus, our data revealed 1,827 novel phosphorylation sites in gastric cancer (Fig. 2e).

Subcellular localizations of gastric cancer phosphoproteins were annotated using Gene Ontology (Fig. 3a). The majority were localized to the nucleus (38%), cytoplasm (34%) and plasma membrane (11%). It is noteworthy that we uncovered 141 non-redundant phosphorylation sites in 65 mitochondrial proteins, 108 phosphorylation sites (77%) of which have not been documented previously (supplemental Table 4).

Biological process classification showed that transcription, RNA splicing, signal transduction, mRNA processing, cell cycle and DNA damage responses were dominant processes represented in the gastric cancer phosphoproteome. Proteins involved in protein transport, apoptosis, anti-apoptosis, protein phosphorylation, differentiation, adhesion and proliferation were also phosphorylated in gastric cancer cells (Fig. 3b).

Motif analysis of gastric cancer phosphoproteins

Protein kinases phosphorylate their substrates at specific motifs. Motif analysis thus helps to shed light on the presence of activated protein kinases. To infer the identities of protein kinases that are active in gastric cancer, we analyzed our phosphoproteome data using Motif-X [40]. By limiting the significance to no more than 0.000001, 11 pS motifs and 2 pT motifs were identified, each occurring in a minimum of 41 pS and 14 pT peptide sequences. The enriched motifs were further annotated according to the Human Protein Reference Database [41]. Logo-like representations of the motifs are shown in Fig. 4. Five acidic motifs associated with casein kinase 2 (CK2) and G protein-coupled receptor kinase (GPCR kinase) were identified, and one basic motif
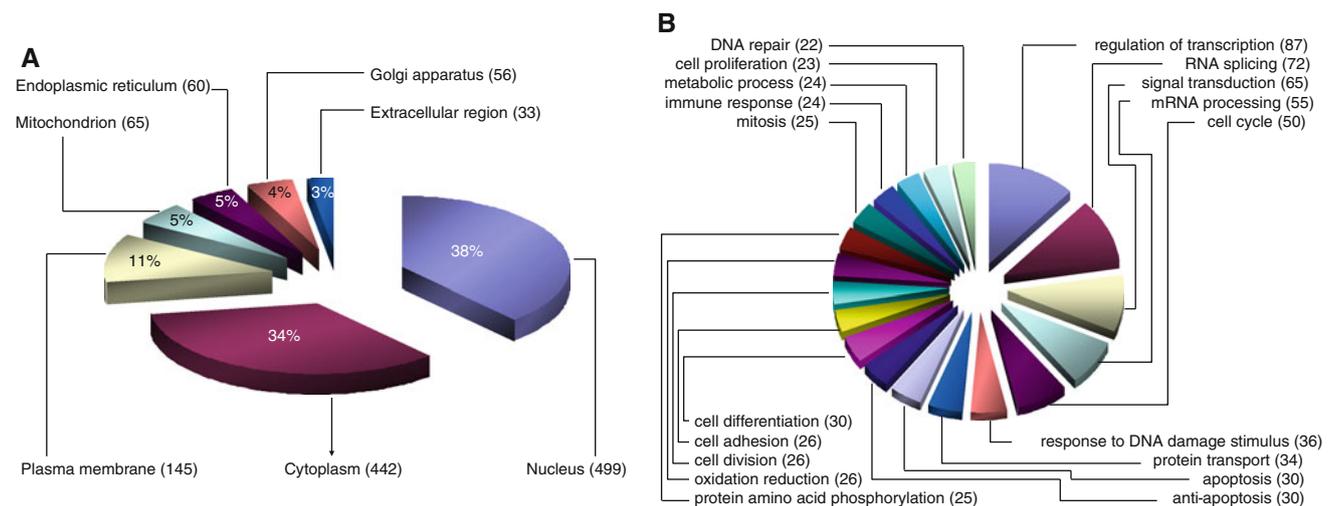

Fig. 3 Classification of phosphoproteins based on Gene Ontology. Phosphorylated proteins in gastric cancer cells were classified according to subcellular localization (**a**) and biological process (**b**). Only biological processes with a hit number over 20 are shown

identified was predicted to be specific to protein kinase A (PKA), PKC and AKT. Four proline-directed motifs were also identified. These were predicted to reflect activation of MAP kinase (MAPK), extracellular signal-regulated kinase 1/2 (ERK1/2), PKA, AKT, PKC, glycogen synthase kinase 3 alpha/beta (GSK3A/B) and CDK5. Motif-X analysis

failed to identify any pY motif from a total of 193 non-redundant pY peptides, probably due to the low-abundance of tyrosine-phosphorylated peptides. To evaluate the kinase specificity of the tyrosine phosphopeptides, we individually checked pY peptides based on known motifs retrieved from the literature [41] using in-house programs. This revealed
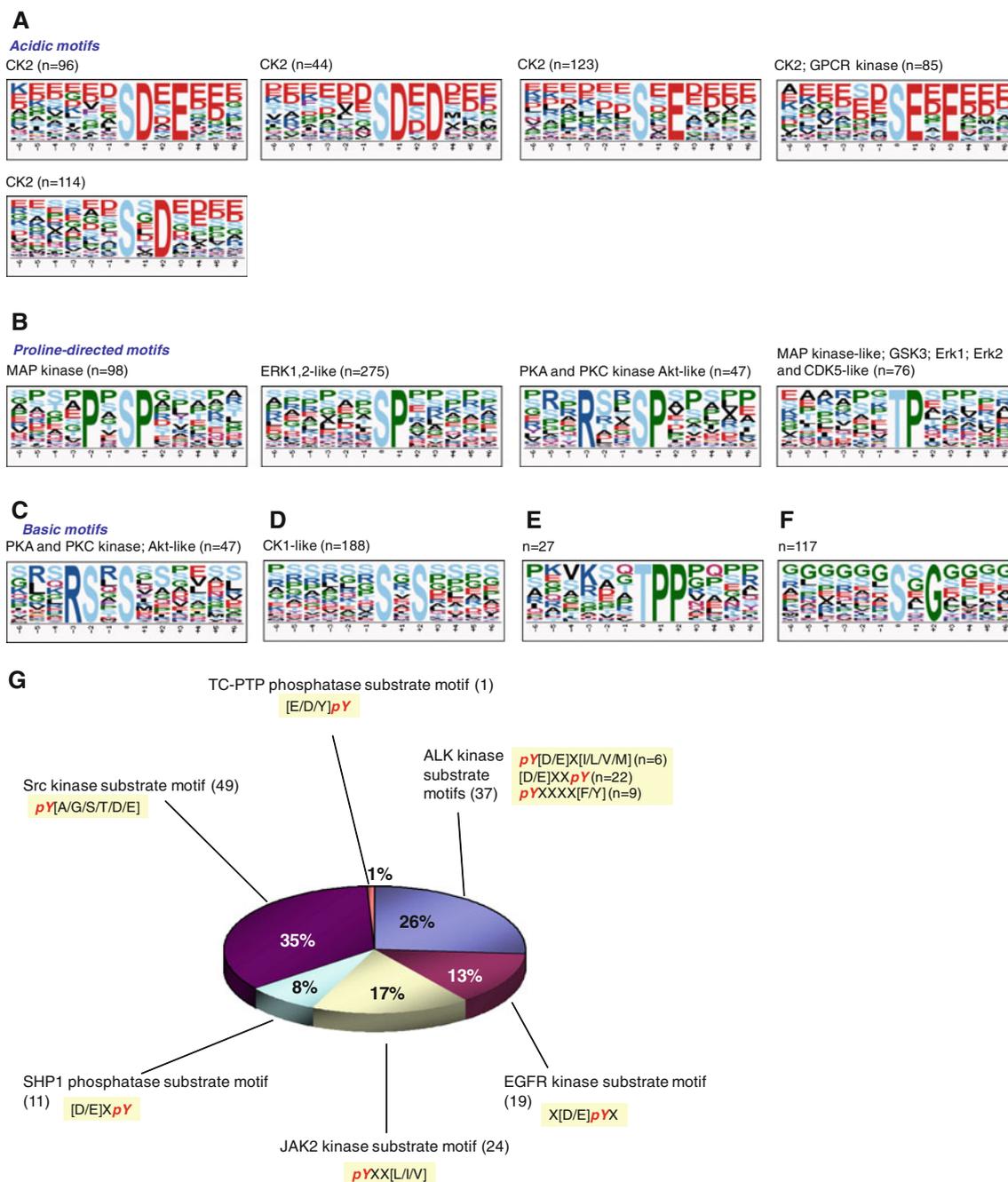
Fig. 4 Motif analysis of gastric cancer phosphoproteome. Non-redundant 13-mer phosphorylated peptide sequences were analyzed in Motif-X. Motifs were classified according to annotations in Human Protein Reference Database. Logo-like representations of motifs are classified into acidic (**a**), proline-directed (**b**), basic (**c**) and others (**d**–**f**). Tyrosine-phosphopeptide motif was not identified by Motif-X, and thus manual evaluation of a total of 193 non-redundant pY peptides was carried out for known motifs. Tyrosine phosphorylation motifs and their counts in gastric cancer cells are shown in (**g**)

six types of motifs, i.e., anaplastic lymphoma receptor tyrosine kinase (ALK), EGFR, JAK2, SHP1, Src kinase substrate motifs and TC-PTP phosphatase substrate motif in 193 non-redundant tyrosine phosphopeptides (Fig. 4g).

Motif-X analysis is based on phosphopeptide sequences that were detected in LC-MS/MS experiments. However, as our analysis may still have missed some low-abundance phosphopeptides, we employed Scansite to analyze kinase motifs in the full protein sequence database of the gastric cancer phosphoproteome. This identified motifs for ABL, AKT, AMPHI, ATM, CAM, CASN, CDC2, CDK5, CLK2, CORT, CRK, DNA-PK, EGFR, ERK1, FGR, FYN, GRB2, GSK3, INSR, ITK, ITSN, LCK, NCK, p38, p85, PDGFR, PDK1, PDZ, PIP3, PKA, PKC, PLCg, SHC, SHIP, SRC and 14-3-3 (supplemental Table 5).

Protein kinases and phosphatases in gastric cancer

In LC-MS/MS-based phosphoproteomics experiments, we were able to identify 15 phosphorylated protein kinases, i.e., adaptor-associated kinase 1 (AAK1), calcium/calmodulin-dependent serine protein kinase (CASK), CDK3, DYRK1B, EGFR, GSK3B, insulin receptor (INSR), mitogen-activated protein kinase kinase 2 (MAP2K2), MET, polycystic kidney disease 2 (PKD2), protein kinase N2 (PKN2), PI-3-kinase-related kinase SMG1, serine/arginine-rich protein-specific kinase 2 (SRPK2), NCK interacting kinase (TNIK) and tau tubulin kinase 2 (TTBK2) (Table 1).

We also found ten phosphorylation sites in five protein phosphatases (phosphorylation sites shown in parentheses), i.e., PTPN12 (S435), PTPN14 (S312, S314), PTPRA (S171, S172, T161), PTPRF (T1801, T1811, T1825) and MTMR7 (S213) in gastric cancer cells (Table 1). Four of these, i.e., PTPN12, PTPN14, PTPRA and PTPRF, belong to classical transmembrane protein tyrosine phosphatases. Nine of the ten phosphorylation sites have never been reported in the literature. As most of these protein phosphatases tended to be overexpressed in the 17 gastric cancer cells, this class of enzymes may participate in modulating the phosphoproteome in gastric cancer.

We next evaluated the expression of these kinase and phosphatase genes in our transcriptome data sets of 17 gastric cancer cell lines that quantified the expression of >12,000 genes relative to pooled normal stomach tissues. Relative expression of 221 protein kinase and 80 protein phosphatase genes were quantified in 17 gastric cancer cell lines (supplemental Tables 6 and 7). These data showed overexpression of subsets of protein kinase and phosphatase genes. Taking the geometric mean of 17 cell lines, PLK1, NEK2, CDC2, FGFR4, TRRAP, MELK, MET, PBK, PLK2 and TTK were the top ten overexpressed protein kinase genes, while the top ten over-expressed

protein phosphatase genes were DUSP9, CDC25B, PTPRU, DUSP14, CDKN3, PTPN14, PTPRF, TPTE, PTPN12 and MTMR10.

The relative expression of protein kinases and phosphatases that were phosphorylated in gastric cancer (Table 1) confirmed EGFR, MET and CDKs as overexpressed and activated kinases, and also revealed many novel kinases whose involvement in gastric cancer was hitherto unknown. These novel gastric cancer protein kinases include fibroblast growth factor receptor 4 (FGFR4), nemo-like kinase (NLK) and NIMA (never in mitosis gene a)-related kinase 2 (NEK2), among others. Although protein kinase N2 (PKN2) has not yet been linked to any cancer type, it had unusually high transcriptional expression and was phosphorylated in gastric cancer.

Phosphoproteomics of primary gastric tissues using antibody arrays

To extend our study of gastric cancer phosphoproteome from cell lines to in vivo clinical samples for the detection of low-abundance phosphoproteins that are beyond the sensitivity of LC-MS/MS-based phosphoproteomics, we utilized antibody arrays that interrogated 46 phosphorylated signaling molecules to investigate the kinome in flash frozen gastric tissues obtained by endoscopic biopsies. These tissues comprised two histologically normal antral biopsies, seven cases of benign gastritis, and three pairs of gastric adenocarcinoma (two intestinal histotype and one diffuse histotype) with their cognate matched normal tissues. All tissues were frozen within seconds after biopsy. Compared to absent signals in the phosphate-buffered saline-spotted negative controls, the antibody array results revealed the expression of 40 phosphoproteins in gastric tissues (Figs. 5, 6, supplemental Figure 4). Of these, 27 were phosphorylated protein kinases (Table 1).

Several phosphorylated proteins displayed substantially stronger signals in tumor tissues than in normal antral tissues and benign gastritis samples, implying potentially critical roles in gastric cancer. They were TP53 (S15, S392, S46), SRC (Y419), YES (Y426), STAT5b (Y699), nitric oxide synthase 3 (eNOS) (S1177), STAT2 (Y689), STAT6 (Y641), MEK1/2 (S218/S222, S222/S226), AKT (S308), ribosomal S6 kinase 1 (RSK1) (S221, S380), RSK2 (S221, S380), RSK3 (S380) and ribosomal protein S6 kinase I (p70S6K) (T229, T389, T421/S424).

Our results also showed that inflammation in gastric tissues induced substantial changes in phosphoproteins. Benign gastritis samples had clearly different patterns of phosphorylated signaling molecules compared to normal stomach tissues. Tyrosine 412 of Src family tyrosine kinase FGR/SRC2 was highly phosphorylated in gastritis, but not in normal or cancerous stomach tissues. Other

Table 1 Phosphorylated protein kinases and phosphatases in gastric cancer

Protein/gene	Protein kinase/ phosphatase family	Description	Phosphorylation site	Identified by	Relative expression level in gastric cancer transcriptome	If annotated[a]	If documented in gastric cancer[b]	If documented in other cancers[c]
Protein kinase								
AAK1/AP2	Other	AP2 associated kinase 1; adaptor-associated kinase 1	S623	MS	0.97	Yes	No	Yes
AAK1/AP2	Other	AP2 associated kinase 1; adaptor-associated kinase 1	S624	MS	0.97	Yes	No	Yes
AAK1/AP2	Other	AP2 associated kinase 1; adaptor-associated kinase 1	T620	MS	0.97	Yes	No	Yes
AKT/AKT1	AGC	AKT1 kinase	S473	Antibody	2.30	Yes	Yes	Yes
AKT/AKT1	AGC	AKT1 kinase	T308	Antibody	2.30	Yes	Yes	Yes
AMPKa1/PRKAA1	CAMK	Protein kinase, AMP-activated, alpha 1 catalytic subunit	T174	Antibody	1.03	Yes	No	Yes
AMPKa2/PRKAA2	CAMK	Protein kinase, AMP-activated, alpha 2 catalytic subunit	T172	Antibody	1.39	Yes	No	Yes
CASK	CAMK	Calcium/calmodulin-dependent serine protein kinase (MAGUK family)	S192	MS	1.82	No	No	Yes
CDK3	CMGC	Cyclin-dependent kinase 3	T42	MS	1.65	No	Yes	Yes
CDK3	CMGC	Cyclin-dependent kinase 3	Y43	MS	1.65	No	Yes	Yes
DYRK1B	CMGC	Dual-specificity tyrosine-(Y)-phosphorylation regulated kinase 1B	Y273	MS	2.32	Yes	No	Yes
EGFR	TK	Epidermal growth factor receptor	T648	MS	2.20	No	No	No
EGFR	TK	Epidermal growth factor receptor	S695	MS	2.20	Yes	No	No
EGFR	TK	Epidermal growth factor receptor	T693	MS	2.20	Yes	No	No
ERK1/MAPK3	CMGC	Mitogen-activated protein kinase 3	T185/Y187[d]	Antibody	0.79	Yes	Yes	Yes
ERK1/MAPK3	CMGC	Mitogen-activated protein kinase 3	T202/Y204[d]	Antibody	0.79	Yes	Yes	Yes
ERK2/MAPK1	CMGC	Mitogen-activated protein kinase 1	T185/Y187[d]	Antibody	1.46	Yes	Yes	Yes
ERK2/MAPK1	CMGC	Mitogen-activated protein kinase 1	T202/Y204[d]	Antibody	1.46	Yes	Yes	Yes
FAK/PTK2	TK	Protein tyrosine kinase 2	Y397	Antibody	1.39	Yes	Yes	Yes
FGR	TK	Proto-oncogene tyrosine-protein kinase FGR	Y412	Antibody	0.18	Yes	No	Yes
FYN	TK	Proto-oncogene tyrosine-protein kinase fyn	Y420	Antibody	0.34	Yes	Yes	Yes
GSK3A	CMGC	Glycogen synthase kinase 3 alpha	S21/S9[d]	Antibody	1.52	Yes	No	Yes
GSK3B	CMGC	Glycogen synthase kinase 3 beta	S21/S9[d]	Antibody	1.48	Yes	No	Yes

Table 1 continued

Protein/gene	Protein kinase/ phosphatase family	Description	Phosphorylation site	Identified by	Relative expression level in gastric cancer transcriptome	If annotated[a]	If documented in gastric cancer[b]	If documented in other cancers[c]
GSK3B	CMGC	Glycogen synthase kinase 3 beta	Y216	MS	1.48	Yes	No	Yes
HCK	TK	Tyrosine-protein kinase HCK	Y411	Antibody	0.26	Yes	Yes	Yes
INSR	TK	Insulin receptor; CD220	S720	MS	0.62	No	No	Yes
JNK1/MAPK8	CMGC	Mitogen-activated protein kinase 8	T183/Y185[d]	Antibody	1.28	Yes	Yes	Yes
JNK1/MAPK8	CMGC	Mitogen-activated protein kinase 8	T221/Y223[d]	Antibody	1.28	Yes	Yes	Yes
LCK	TK	Lymphocyte-specific protein tyrosine kinase	Y394	Antibody	0.23	Yes	Yes	Yes
LYN	TK	Yamaguchi sarcoma viral (v-yes-1) related oncogene homolog	Y397	Antibody	0.79	Yes	Yes	Yes
MARK2	CAMK	MAP/microtubule affinity-regulating kinase 2	S423	MS	1.28	Yes	Yes	Yes
MEK1/MAP2K1	STE	Mitogen-activated protein kinase 1	S218/S222[d]	Antibody	1.64	Yes	Yes	Yes
MEK1/MAP2K1	STE	Mitogen-activated protein kinase 1	S222/S226[d]	Antibody	1.64	Yes	Yes	Yes
MEK2/MAP2K2	STE	Mitogen-activated protein kinase 2	S218/S222[d]	Antibody	1.64	Yes	Yes	Yes
MEK2/MAP2K2	STE	Mitogen-activated protein kinase 2	S222/S226[d]	Antibody	1.64	Yes	Yes	Yes
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	S1006	MS	4.33	No	No	No
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	S1008	MS	4.33	No	No	No
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	T1011	MS	4.33	No	No	No
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	T678	MS	4.33	No	No	No
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	T992	MS	4.33	No	No	No
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	T993	MS	4.33	No	No	No
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	Y666	MS	4.33	No	No	No
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	S1000	MS	4.33	Yes	No	No
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	S988	MS	4.33	Yes	No	No

Table 1 continued

Protein/gene	Protein kinase/ phosphatase family	Description	Phosphorylation site	Identified by	Relative expression level in gastric cancer transcriptome	If annotated[a]	If documented in gastric cancer[b]	If documented in other cancers[c]
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	S990	MS	4.33	Yes	No	No
MET	TK	Met proto-oncogene (hepatocyte growth factor receptor)	S997	MS	4.33	Yes	No	No
MSK1/RPS6KA5	AGC	Ribosomal protein S6 kinase, 90 kDa, polypeptide 5	S376/S360[d]	Antibody	0.29	Yes	No	Yes
MSK2/RPS6KA4	AGC	Ribosomal protein S6 kinase, 90 kDa, polypeptide 4	S376/S360[d]	Antibody	2.27	Yes	No	Yes
MTOR/FRAP1	Atypical	Mechanistic target of rapamycin (serine/threonine kinase)	S2448	Antibody	1.25	Yes	Yes	Yes
P38a/MAPK14	STE	Mitogen-activated protein kinase 14	T180/Y192[d]	Antibody	2.37	Yes	No	Yes
P70S6K/RPS6KB1	AGC	Ribosomal protein S6 kinase, 70 kDa, polypeptide 1	T229	Antibody	1.02	Yes	Yes	Yes
p70S6K/RPS6KB1	AGC	Ribosomal protein S6 kinase, 70 kDa, polypeptide 1	T389	Antibody	1.02	Yes	Yes	Yes
p70S6K/RPS6KB1	AGC	Ribosomal protein S6 kinase, 70 kDa, polypeptide 1	T421/S424[d]	Antibody	1.02	Yes	Yes	Yes
PKD2	CAMK	Polycystic kidney disease 2 (autosomal dominant)	S812	MS	0.55	Yes	Yes	Yes
PKN2	AGC	Polycystic kidney disease 2 (autosomal dominant)	S582	MS	1.70	Yes	No	No
PKN2	AGC	Polycystic kidney disease 2 (autosomal dominant)	S583	MS	1.70	Yes	No	No
PYK2/PTK2B	TK	Protein tyrosine kinase 2 beta	Y402	Antibody	1.13	Yes	Yes	Yes
RSK1/RPS6KA1	AGC	Ribosomal protein S6 kinase, 90 kDa, polypeptide 1	S221	Antibody	1.06	Yes	No	Yes
RSK1/RPS6KA1	AGC	Ribosomal protein S6 kinase, 90 kDa, polypeptide 1	S380	Antibody	1.06	Yes	No	Yes
RSK2/RPS6KA3	AGC	Ribosomal protein S6 kinase, 90 kDa, polypeptide 3	S221	Antibody	2.46	Yes	No	Yes
RSK2/RPS6KA3	AGC	Ribosomal protein S6 kinase, 90 kDa, polypeptide 3	S380	Antibody	2.46	Yes	No	Yes
RSK3/RPS6KA2	AGC	Ribosomal protein S6 kinase, 90 kDa, polypeptide 2	S380	Antibody	0.39	Yes	No	Yes
SMG1	Atypical	Phosphatidylinositol 3-kinase-related protein kinase	S2940	MS	0.93	No	No	Yes

Table 1 continued

Protein/gene	Protein kinase/ phosphatase family	Description	Phosphorylation site	Identified by	Relative expression level in gastric cancer transcriptome	If annotated[a]	If documented in gastric cancer[b]	If documented in other cancers[c]
SMG1	Atypical	Phosphatidylinositol 3-kinase-related protein kinase	S2946	MS	0.93	No	No	Yes
SMG1	Atypical	Phosphatidylinositol 3-kinase-related protein kinase	T2947	MS	0.93	No	No	Yes
SRC	TK	Proto-oncogene tyrosine-protein kinase	Y419	Antibody	1.11	Yes	Yes	Yes
SRPK2	CMGC	Serine/arginine-rich protein-specific kinase 2	S496	MS	1.53	No	No	Yes
SRPK2	CMGC	Serine/arginine-rich protein-specific kinase 2	T492	MS	1.53	No	No	Yes
SRPK2	CMGC	Serine/arginine-rich protein-specific kinase 2	S494	MS	1.53	Yes	No	Yes
SRPK2	CMGC	Serine/arginine-rich protein-specific kinase 2	S497	MS	1.53	Yes	No	Yes
SRPK2	CMGC	Serine/arginine-rich protein-specific kinase 2	T498	MS	1.53	Yes	No	Yes
TNIK	STE	TRAF2 and NCK interacting kinase	S680	MS	1.25	Yes	No	Yes
TTBK2	CK1	Tau tubulin kinase 2	T1070	MS	0.75	No	No	No
YES/YES1	TK	Yamaguchi sarcoma viral oncogene homolog 1	Y426	Antibody	0.86	Yes	Yes	Yes
Protein phosphatase								
PTPN14	PTP	Tyrosine-protein phosphatase non-receptor-type 14	S312	MS	3.58	No	No	Yes
PTPN14	PTP	Tyrosine-protein phosphatase non-receptor-type 14	S314	MS	3.58	No	No	Yes
PTPRF	PTP	Isoform 1 of receptor-type tyrosine-protein phosphatase F	T1801	MS	3.26	No	No	Yes
PTPRF	PTP	Isoform 1 of receptor-type tyrosine-protein phosphatase F	T1811	MS	3.26	No	No	Yes
PTPRF	PTP	Isoform 1 of receptor-type tyrosine-protein phosphatase F	T1825	MS	3.26	No	No	Yes
PTPN12	PTP	Tyrosine-protein phosphatase non-receptor-type 12	S435	MS	2.89	Yes	No	Yes
PTPRA	PTP	cDNA FLJ56484, highly similar to receptor-type tyrosine-protein phosphatase alpha	S171	MS	2.89	No	Yes	Yes

Table 1 continued

Protein/gene	Protein kinase/phosphatase family	Description	Phosphorylation site	Identified by	If annotated[a]	Relative expression level in gastric cancer transcriptome	If documented in gastric cancer[b]	If documented in other cancers[c]
PTPRA	PTP	cDNA FLJ56484, highly similar to receptor-type tyrosine-protein phosphatase alpha	S172	MS	No	2.89	Yes	Yes
PTPRA	PTP	cDNA FLJ56484, highly similar to receptor-type tyrosine-protein phosphatase alpha	T161	MS	No	2.89	Yes	Yes
MTMR7	DSP	Isoform 1 of myotubularin-related protein 7	S213	MS	No	0.86	No	No

List of protein kinases and phosphatases identified by LC-MS/MS-based and protein antibody array-based phosphoproteomics analysis of gastric cancer cell lines and primary gastric tissues. The expression of each protein kinase and phosphatase gene in the transcriptomes of 17 gastric cancer cell lines relative to normal stomach tissues is shown. The values are the average of 17 gastric cancer cell lines

[a] The phosphorylation site was annotated previously in the literature

[b] The protein's role in gastric cancer was documented previously in the literature

[c] The protein's role in other cancers was documented previously in the literature

[d] The antibody detects both phosphorylation sites in the protein kinase

phosphorylation sites specifically associated with stomach inflammation included Y402 in PYK2, S78/S82 in HSP27, T202/Y204 and T185/Y187 in ERK1/2, T180/Y192 in p38a, Y118 in paxillin, S63 in c-Jun and Y701 in STAT1. Several tyrosine kinases appeared deactivated in gastritis compared to normal stomach, including FAK, YES, FYN, HCK, JUN, CHK2, LCK, GSK3A/B, AMOKa1 and p70S6K. Compared to gastritis, cancerous tissues exhibited higher levels of nuclear phosphoproteins including TP53, STATs, CREB and CHK2 as well as tyrosine kinases such as GSK3A/B, FAK, FYN, LCK, AMPKA1, JNK, HCK and p70S6K. It is noteworthy that matched cancerous and non-cancerous tissues from the same patient had very similar phospho-proteome patterns, consistent with field cancerization in this disease [42, 43] (supplemental Figure 4 and Figure 6).

LC-MS/MS-based and antibody array-based phospho-proteomics analysis jointly identified 74 phosphorylation sites in 41 protein kinases in gastric cancer cell lines and primary stomach tissues (Table 1). Eighteen of these phosphorylation sites (24%) are novel. Literature mining revealed that 37 of the 41 identified protein kinases (90%) have been implicated in a range of different non-gastric cancers, whereas only 19 (46%) have been associated with gastric cancer.

Discussion

In this study, we have integrated LC-MS/MS-based phos-phoproteomic, protein antibody array and transcriptomic techniques, undergirded by bioinformatic analysis, to gen-erate an expansive view of phosphoproteome and molecular signaling pathways in gastric cancer. This is the first com-prehensive view of the gastric cancer phosphoproteome.

Phosphoproteins are the most pervasive signaling mol-ecules, whereas many overexpressed proteins are likely to be critical in carcinogenesis. We have investigated the phosphoproteome of both gastric cancer cell lines and clinical samples. Protein antibody array-based phospho-proteomics was employed to detect low-abundance phosphorylated proteins in clinical tissues. Since the main focus of this study is not to compare phosphoproteomes between cancer and normal samples, only three pairs of gastric adenocarcinoma with their cognate matched normal tissues, in addition to nine normal and benign samples, were included in this study. However, to characterize the differential expression of the phosphoproteins as identified in gastric cancer, more comprehensive clinical investiga-tions are required.

It is noteworthy that the overlap between the phospho-proteome from LC-MS/MS and the phosphoproteome from protein antibody array is negligible. One reason is that the commercially available protein antibody array for probing
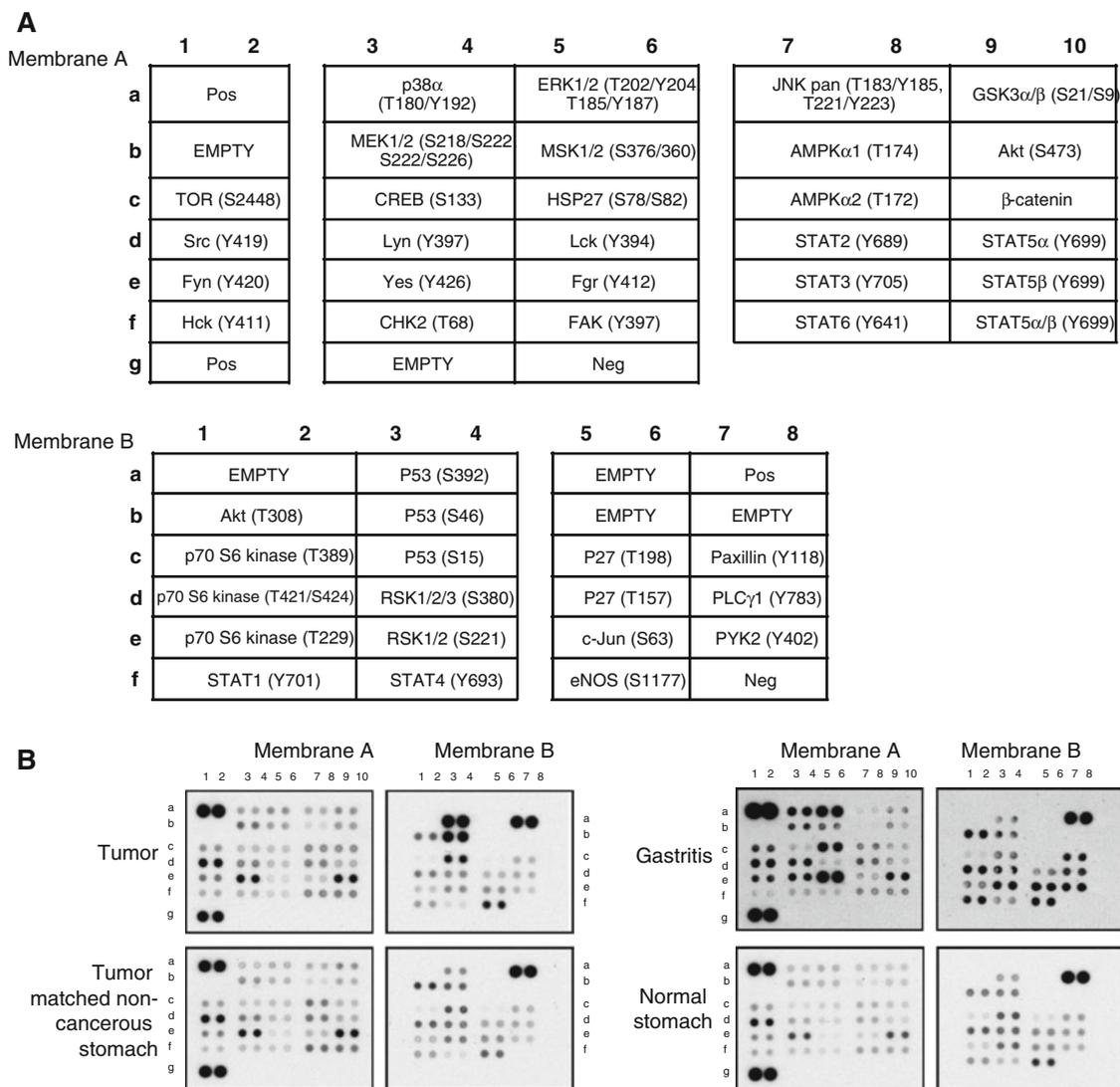
A

Membrane A

	1	2	3	4	5	6	7	8	9	10
a	Pos		p38α (T180/Y192)		ERK1/2 (T202/Y204 T185/Y187)		JNK pan (T183/Y185, T221/Y223)		GSK3α/β (S21/S9)	
b	EMPTY		MEK1/2 (S218/S222 S222/S226)		MSK1/2 (S376/360)		AMPKα1 (T174)		Akt (S473)	
c	TOR (S2448)		CREB (S133)		HSP27 (S78/S82)		AMPKα2 (T172)		β-catenin	
d	Src (Y419)		Lyn (Y397)		Lck (Y394)		STAT2 (Y689)		STAT5α (Y699)	
e	Fyn (Y420)		Yes (Y426)		Fgr (Y412)		STAT3 (Y705)		STAT5β (Y699)	
f	Hck (Y411)		CHK2 (T68)		FAK (Y397)		STAT6 (Y641)		STAT5α/β (Y699)	
g	Pos		EMPTY		Neg					

Membrane B

	1	2	3	4	5	6	7	8
a	EMPTY		P53 (S392)		EMPTY		Pos	
b	Akt (T308)		P53 (S46)		EMPTY		EMPTY	
c	p70 S6 kinase (T389)		P53 (S15)		P27 (T198)		Paxillin (Y118)	
d	p70 S6 kinase (T421/S424)		RSK1/2/3 (S380)		P27 (T157)		PLCγ1 (Y783)	
e	p70 S6 kinase (T229)		RSK1/2 (S221)		c-Jun (S63)		PYK2 (Y402)	
f	STAT1 (Y701)		STAT4 (Y693)		eNOS (S1177)		Neg	

B

Fig. 5 Representative antibody array images of primary gastric tissues. Proteome Profiler™ Human Phospho-Kinase Array Kit (R&D Systems) was used to simultaneously detect phosphorylation sites in a panel of protein kinases and key signaling proteins in fresh frozen primary endoscopic gastric tissues, i.e., normal, gastritis and gastric cancer tissues. **a** Layout of protein antibody array composed of membrane A and membrane B. **b** Representative images of protein arrays of two cases of primary gastric cancers, each with its matched non-cancerous tissue, one case each of normal stomach and benign gastritis biopsies. Refer to supplemental Figure 4 for all images

phosphoproteins contains only 46 phosphoproteins. In addition, most of these proteins are low-abundance signaling proteins that are rarely identified by LC-MS/MS approaches due to the dynamic range.

Integrating phosphoproteome and transcriptome data sets is a powerful strategy for understanding cancer biology and mining potential gastric cancer biomarkers. Moreover, cancer therapeutics is being transformed by highly efficacious agents targeted at abnormally activated oncogenic tyrosine kinases. Focusing on phosphorylated proteins that were >2-fold transcriptionally overexpressed, we identified 190 dysregulated phosphoproteins (supplemental Table 9). Our study confirmed previous reports that *MET* transcriptional overexpression (>40-fold higher than normal stomach tissues) is a prominent feature of some gastric cancer cells [44, 45], while our phosphoproteomics data set identified the presence of MET in its phosphorylated and active state. Selective inhibition of MET is known to kill *MET*-overexpressing gastric cancer cells effectively [14, 46] and is the rationale for ongoing clinical trials of MET inhibitors for gastric cancer therapy. Our data also showed overexpression of several genes whose protein products were phosphorylated and have been proposed as useful prognostic markers and/or therapeutic targets for gastric cancer, including EGFR [47], TOP2A [48], minichromosome maintenance 2 (MCM2) [49], erythropoietin-producing hepatocellular (Eph) A2 receptor [50], CTNNB1 [51] and hepatoma-derived
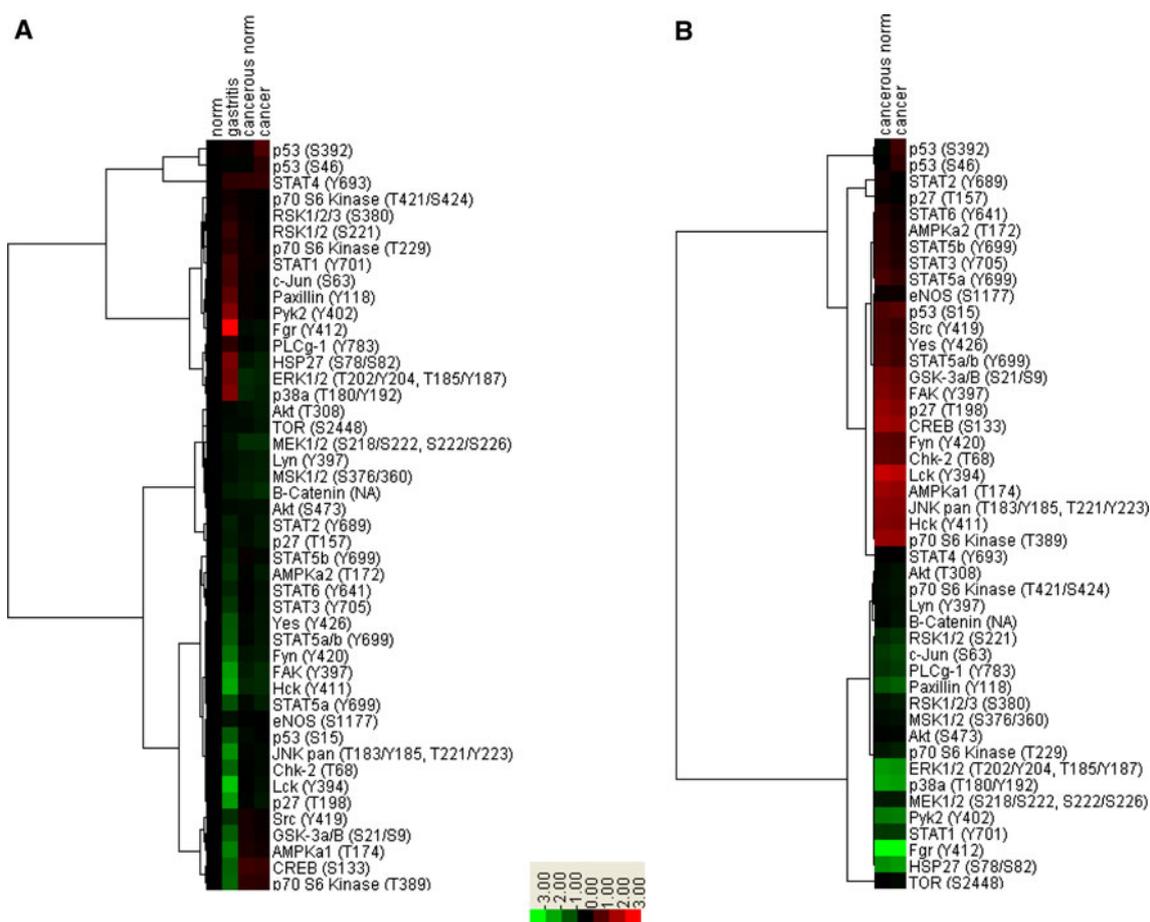
Fig. 6 Phosphorylated signaling molecules determined by antibody array analysis of primary gastric tissues. One pooled sample of histologically normal stomach tissue from two individuals, seven cases of histologically benign gastritis, as well as three pairs of gastric adenocarcinoma and matched normal tissues were analyzed on protein antibody arrays. Each phosphorylation site for every sample was detected in duplicate. **a** Normalized intensities of gastritis tissues, cancer and matched normal tissues relative to normal stomach tissues are shown. **b** Normalized intensities of a case of gastric cancer and its matched normal tissue relative to benign gastritis tissues are shown

growth factor (HDGF) [52]. The data sets also reveal novel overexpressed and phosphorylated proteins whose roles in gastric cancer have yet to be defined, such as EIF2S3, LMNB2, KIF23, SLC7A5/CD98 and MCM3 (supplemental Table 9), although some have been associated with other types of cancers. For instance, SLC7A5/CD98 is a proposed prognostic indicator of adult acute leukemia [53], breast cancer [54], lung cancer [55] and renal cancer [56]. Our integrated analyses suggest that such molecules could provide helpful insights into processes underlying gastric oncogenesis.

The DNA damage response (DDR) pathway appears overrepresented in the pathway analysis of the 190 overexpressed phosphoproteins. DNA damage in the absence of physiological repair responses is the origin of many diseases, including cancers [57]. DDR comprises a variety of signaling pathways, which are activated by DNA damage and replication stress, and are transduced by kinase cascades, mainly through a pair of protein kinases, ataxia telangiectasia mutated (ATM) and ATR (ATM and Rad3-related). Both ATM and ATR in turn phosphorylate a number of substrates, including checkpoint kinase 1 (CHK1) and CHK2, and influence cell cycle, DNA repair, DNA replication and many other biological processes involving nucleic acids, as well as diverse signaling pathways like the insulin/IGF-1-PI3K-AKT pathway [58]. As shown in Fig. 7, our data sets identified overexpression of mRNA levels of many components in this pathway. Moreover, phosphorylation of some critical player in this pathway was identified. Specifically, our data suggest that hyperphosphorylated TP53 might be one characteristic of gastric cancer. While normal stomach tissues consistently displayed basal levels of phosphorylated TP53, cancerous tissues from both intestinal-type gastric adenocarcinomas had markedly elevated levels of TP53 phosphorylated at S392, S46 and S15. In contrast, matched non-cancerous gastric tissue from the same patients displayed only basal phosphorylation (supplemental Figure 4).
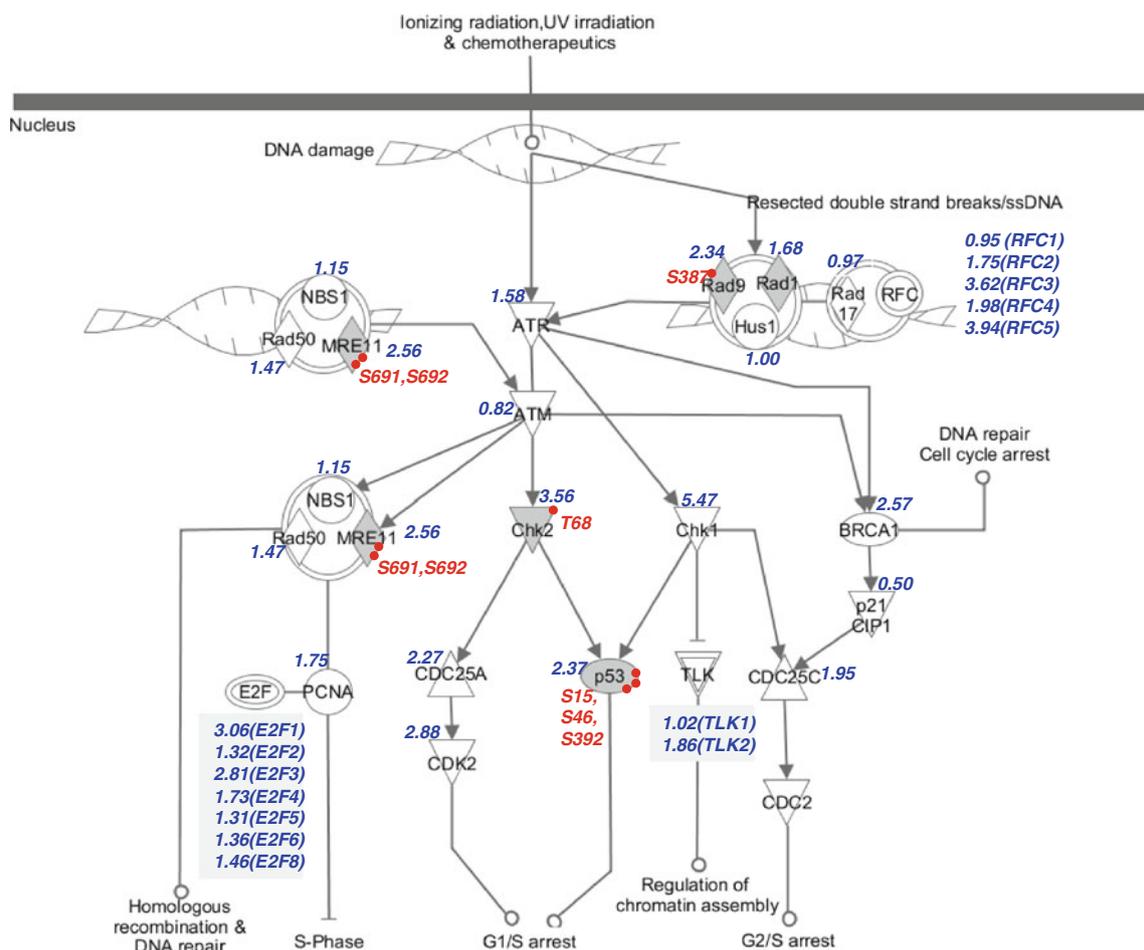
Fig. 7 DNA damage response pathway in gastric cancer. The pathway is modified based on the cell cycle checkpoint control pathway from ingenuity pathway analysis (IPA). Overexpressed phosphoproteins are shaded in grey. The relative mRNA expression level of proteins is shown in *blue*. Phosphorylation sites are shown in *red*

The integrated approach we adopted generated an unbiased view of the gastric oncokinome. The human kinome contains 518 protein kinases classified into 10 groups based on catalytic domain sequence similarities, i.e., AGC, CAMK, CK1, CMGC, STE, TK, TKL, RGC, Atypical and Other [59]. Protein phosphatases play equally critical roles in setting the levels of protein phosphorylation in cells and in regulating many physiological processes [26]. However, proportionately much less research has focused on protein phosphatases in cancer cells. Protein phosphatases are classified according to their substrate specificities into protein tyrosine phosphatase (PTP), serine/threonine phosphatase (STP), protein histidine phosphatase (PHP) and dual-specific phosphatases (DSPs) [26, 27]. Like protein kinases, phosphorylation of protein phosphatases is an important regulatory mechanism [26]. Relative expression levels of protein kinases and phosphatases, as well as their phosphorylation status, are functionally crucial to cancer phenotypes. By integrating transcriptional expression levels of 221 protein kinase and 80 protein phosphatase genes in

17 gastric cancer cell lines with phosphoproteomic data, our data help to define the dynamic molecular terrain of protein kinases and protein phosphatases (Table 1) from which key pathways in gastric oncogenesis may be discerned.

It is also worth noting that 30 overexpressed phosphoproteins (16%) were associated with mitochondria, implying critical roles for this organelle in gastric oncogenesis (supplemental Table 4). Mitochondria are pivotal in cell metabolism, survival and apoptosis. Several protein kinases and protein tyrosine phosphatases are known to reside in mitochondria, whereas other mitochondrial proteins are themselves kinase substrates. As well as being the target of all major kinase signaling pathways, intramitochondrial signaling also occurs [60, 61]. Mitochondrial phosphoproteomes of mammalian cardiomyocytes [62], hepatocytes [63], pancreatic beta-cells[64], yeast[65] and *Arabidopsis thaliana* [66] have been reported. However, there is as yet no systematic documentation of mitochondrial phosphoproteins in cancer cells. Our data demonstrated that TOMM20 (translocase of outer mitochondrial membrane 20) was

overexpressed and phosphorylated in some gastric cancer cells. This protein is a central receptor component of the TOM complex (translocase of the outer membrane of mitochondria) that recognizes and translocates cytosolically synthesized mitochondrial preproteins. In addition to TOMM20, several mitochondrial proteins were also dysregulated in gastric cancer. Mitochondrial ribosomal proteins (MRPS16, MRPL11 and DAP3) were all phosphorylated and highly expressed, reflecting active synthesis of mitochondrial proteins. Among other phosphorylated mitochondrial proteins we identified were proteins of the electron transfer chain, mitochondrial permeability transition pore, mitochondrial ribosomal proteins, as well as various enzymes involved in apoptosis and metabolism. These data not only support the role of phosphorylation in regulating mitochondrial proteins, but also point to key roles of mitochondrial functions in oncogenic processes.

Conclusion

In conclusion, this is the most comprehensive report to date of the phosphoproteome of gastric cancer cells. We also provide the first documentation of gastric cancer kinome and phosphatome at both transcriptional and post-translational levels. Moreover, we documented phosphorylated mitochondrial proteins. Nonetheless, this study marks an early phase of unraveling global oncogenic signaling networks in gastric cancer as many of the phosphoproteins identified here are completely novel. Hence, elucidation of their functions and roles in gastric cancer require further investigations.

References

1. Peek RM Jr, Blaser MJ (2002) Helicobacter pylori and gastrointestinal tract adenocarcinomas. Nat Rev Cancer 2:28–37
2. Wagner AD, Moehler M (2009) Development of targeted therapies in advanced gastric cancer: promising exploratory steps in a new era. Curr Opin Oncol 21:381–385
3. Sawyers C (2004) Targeted cancer therapy. Nature 432:294–297
4. Di Cosimo S, Baselga J (2010) Management of breast cancer with targeted agents: importance of heterogeneity. [corrected]. Nat Rev Clin Oncol 7:139–147
5. Esteva FJ, Yu D, Hung MC, Hortobagyi GN (2010) Molecular predictors of response to trastuzumab and lapatinib in breast cancer. Nat Rev Clin Oncol 7:98–107
6. Daub H, Specht K, Ullrich A (2004) Strategies to overcome resistance to targeted protein kinase inhibitors. Nat Rev Drug Discov 3:1001–1010
7. Xu AM, Huang PH (2010) Receptor tyrosine kinase coactivation networks in cancer. Cancer Res 70:3857–3860
8. Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, Lindeman N, Gale CM, Zhao X, Christensen J, Kosaka T, Holmes AJ, Rogers AM, Cappuzzo F, Mok T, Lee C, Johnson BE, Cantley LC, Janne PA (2007) MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling. Science 316:1039–1043
9. Nahta R, Yu D, Hung MC, Hortobagyi GN, Esteva FJ (2006) Mechanisms of disease: understanding resistance to HER2-targeted therapy in human breast cancer. Nat Clin Pract Oncol 3:269–280
10. Stommel JM, Kimmelman AC, Ying H, Nabioullin R, Ponugoti AH, Wiedemeyer R, Stegh AH, Bradner JE, Ligon KL, Brennan C, Chin L, DePinho RA (2007) Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies. Science 318:287–290
11. Cutsem EV, Kang Y, Chung H, Shen L, Sawaki A, Lordick F, Hill J, Lehle M, Feyereislova A, Bang Y (2009) Efficacy results from the ToGA trial: a phase III study of trastuzumab added to standard chemotherapy (CT) in first-line human epidermal growth factor receptor 2 (HER2)-positive advanced gastric cancer (GC). J Clin Oncol 27:18s Suppl. abstr LBA4509
12. Dragovich T, Campen C (2009) Anti-EGFR-targeted therapy for esophageal and gastric cancers: an evolving concept. J Oncol 2009:804108
13. Iwasaki J, Nihira S (2009) Anti-angiogenic therapy against gastrointestinal tract cancers. Jpn J Clin Oncol 39:543–551
14. Comoglio PM, Giordano S, Trusolino L (2008) Drug development of MET inhibitors: targeting oncogene addiction and expedience. Nat Rev Drug Discov 7:504–516
15. Arkenau HT (2009) Gastric cancer in the era of molecularly targeted agents: current drug development strategies. J Cancer Res Clin Oncol 135:855–866
16. Grabsch H, Sivakumar S, Gray S, Gabbert HE, Muller W (2010) HER2 expression in gastric cancer: rare, heterogeneous and of no prognostic value—conclusions from 924 cases of two independent series. Cell Oncol 32:57–65
17. Huang PH, White FM (2008) Phosphoproteomics: unraveling the signaling web. Mol Cell 31:777–781
18. Macek B, Mann M, Olsen JV (2009) Global and site-specific quantitative phosphoproteomics: principles and applications. Annu Rev Pharmacol Toxicol 49:199–221
19. Diella F, Gould CM, Chica C, Via A, Gibson TJ (2008) Phospho.ELM: a database of phosphorylation sites—update 2008. Nucleic Acids Res 36:D240–D244
20. Diella F, Cameron S, Gemund C, Linding R, Via A, Kuster B, Sicheritz-Ponten T, Blom N, Gibson TJ (2004) Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins. BMC Bioinformatics 5:79
21. http://www.phosphosite.org
22. Bodenmiller B, Campbell D, Gerrits B, Lam H, Jovanovic M, Picotti P, Schlapbach R, Aebersold R (2008) PhosphoPep—a database of protein phosphorylation sites in model organisms. Nat Biotechnol 26:1339–1340
23. Gnad F, Ren S, Cox J, Olsen JV, Macek B, Oroshi M, Mann M (2007) PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites. Genome Biol 8:R250
24. Knight ZA, Lin H, Shokat KM (2010) Targeting the cancer kinome through polypharmacology. Nat Rev Cancer 10:130–137
25. Tonks NK (2006) Protein tyrosine phosphatases: from genes, to function, to disease. Nat Rev Mol Cell Biol 7:833–846
26. Alonso A, Sasin J, Bottini N, Friedberg I, Osterman A, Godzik A, Hunter T, Dixon J, Mustelin T (2004) Protein tyrosine phosphatases in the human genome. Cell 117:699–711

27. Vintonyak VV, Antonchick AP, Rauh D, Waldmann H (2009) The therapeutic potential of phosphatase inhibitors. Curr Opin Chem Biol 13:272–283

28. Gan CS, Guo T, Zhang H, Lim SK, Sze SK (2008) A comparative study of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) versus SCX-IMAC-based methods for phosphopeptide isolation/enrichment. J Proteome Res 7:4869–4877

29. Guo T, Gan CS, Zhang H, Zhu Y, Kon OL, Sze SK (2008) Hybridization of pulsed-Q dissociation and collision-activated dissociation in linear ion trap mass spectrometer for iTRAQ quantitation. J Proteome Res 7:4831–4840

30. Zhu Y, Guo T, Park JE, Li X, Meng W, Datta A, Bern M, Lim SK, Sze SK (2009) Elucidating in vivo structural dynamics in integral membrane protein by hydroxyl radical footprinting. Mol Cell Proteomics 8:1999–2010

31. Fenyo D, Beavis RC (2003) A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes. Anal Chem 75:768–774

32. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH (2004) Open mass spectrometry search algorithm. J Proteome Res 3:958–964

33. Obenauer JC, Cantley LC, Yaffe MB (2003) Scansite 2.0: proteome-wide prediction of cell signaling interactions using short sequence motifs. Nucleic Acids Res 31:3635–3641

34. Bodenmiller B, Mueller LN, Mueller M, Domon B, Aebersold R (2007) Reproducible isolation of distinct, overlapping segments of the phosphoproteome. Nat Methods 4:231–237

35. Alpert AJ (2008) Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides. Anal Chem 80:62–76

36. Alves G, Wu WW, Wang G, Shen RF, Yu YK (2008) Enhancing peptide identification confidence by combining search methods. J Proteome Res 7:3102–3113

37. Zhang H, Guo T, Li X, Datta A, Park JE, Yang J, Lim SK, Tam JP, Sze SK (2010) Simultaneous characterization of glyco- and phospho-proteomes of mouse brain membrane proteome with electrostatic repulsion hydrophilic interaction chromatography (ERLIC). Mol Cell Proteomics 9:635–647

38. Sarg B, Helliger W, Talasz H, Forg B, Lindner HH (2006) Histone H1 phosphorylation occurs site-specifically during interphase and mitosis: identification of a novel phosphorylation site on histone H1. J Biol Chem 281:6573–6580

39. Beausoleil SA, Villen J, Gerber SA, Rush J, Gygi SP (2006) A probability-based approach for high-throughput protein phosphorylation analysis and site localization. Nat Biotechnol 24:1285–1292

40. Schwartz D, Gygi SP (2005) An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets. Nat Biotechnol 23:1391–1398

41. Amanchy R, Periaswamy B, Mathivanan S, Reddy R, Tattikota SG, Pandey A (2007) A curated compendium of phosphorylation motifs. Nat Biotechnol 25:285–286

42. Slaughter DP, Southwick HW, Smejkal W (1953) Field cancerization in oral stratified squamous epithelium; clinical implications of multicentric origin. Cancer 6:963–968

43. Subramanian H, Roy HK, Pradhan P, Goldberg MJ, Muldoon J, Brand RE, Sturgis C, Hensing T, Ray D, Bogojevic A, Mohammed J, Chang JS, Backman V (2009) Nanoscale cellular changes in field carcinogenesis detected by partial wave spectroscopy. Cancer Res 69:5357–5363

44. Kuniyasu H, Yasui W, Yokozaki H, Kitadai Y, Tahara E (1993) Aberrant expression of c-met mRNA in human gastric carcinomas. Int J Cancer 55:72–75

45. Heideman DA, Snijders PJ, Bloemena E, Meijer CJ, Offerhaus GJ, Meuwissen SG, Gerritsen WR, Craanen ME (2001) Absence of tpr-met and expression of c-met in human gastric mucosa and carcinoma. J Pathol 194:428–435

46. Smolen GA, Sordella R, Muir B, Mohapatra G, Barmettler A, Archibald H, Kim WJ, Okimoto RA, Bell DW, Sgroi DC, Christensen JG, Settleman J, Haber DA (2006) Amplification of MET may identify a subset of cancers with extreme sensitivity to the selective tyrosine kinase inhibitor PHA-665752. Proc Natl Acad Sci USA 103:2316–2321

47. Nicholson RI, Gee JM, Harper ME (2001) EGFR and cancer prognosis. Eur J Cancer 37 (Suppl 4):S9–S15

48. Kanta SY, Yamane T, Dobashi Y, Mitsui F, Kono K, Ooi A (2006) Topoisomerase II alpha gene amplification in gastric carcinomas: correlation with the HER2 gene. An immunohistochemical, immunoblotting, and multicolor fluorescence in situ hybridization study. Hum Pathol 37:1333–1343

49. Tokuyasu N, Shomori K, Nishihara K, Kawaguchi H, Fujioka S, Yamaga K, Ikeguchi M, Ito H (2008) Minichromosome maintenance 2 (MCM2) immunoreactivity in stage III human gastric carcinoma: clinicopathological significance. Gastric Cancer 11:37–46

50. Yuan W, Chen Z, Wu S, Ge J, Chang S, Wang X, Chen J (2009) Expression of EphA2 and E-cadherin in gastric cancer: correlated with tumor progression and lymphogenous metastasis. Pathol Oncol Res 15:473–478

51. Zhou YN, Xu CP, Han B, Li M, Qiao L, Fang DC, Yang JM (2002) Expression of E-cadherin and beta-catenin in gastric carcinoma and its correlation with the clinicopathological features and patient survival. World J Gastroenterol 8:987–993

52. Yamamoto S, Tomita Y, Hoshida Y, Takiguchi S, Fujiwara Y, Yasuda T, Doki Y, Yoshida K, Aozasa K, Nakamura H, Monden M (2006) Expression of hepatoma-derived growth factor is correlated with lymph node metastasis and prognosis of gastric carcinoma. Clin Cancer Res 12:117–122

53. Nikolova M, Guenova M, Taskov H, Dimitrova E, Staneva M (1998) Levels of expression of CAF7 (CD98) have prognostic significance in adult acute leukemia. Leuk Res 22:39–47

54. Esseghir S, Reis-Filho JS, Kennedy A, James M, O'Hare MJ, Jeffery R, Poulsom R, Isacke CM (2006) Identification of transmembrane proteins as potential prognostic markers and therapeutic targets in breast cancer by a screen for signal sequence encoding transcripts. J Pathol 210:420–430

55. Kaira K, Oriuchi N, Imai H, Shimizu K, Yanagitani N, Sunaga N, Hisada T, Ishizuka T, Kanai Y, Endou H, Nakajima T, Mori M (2009) Prognostic significance of L-type amino acid transporter 1 (LAT1) and 4F2 heavy chain (CD98) expression in early stage squamous cell carcinoma of the lung. Cancer Sci 100:249–254

56. Prager GW, Poettler M, Schmidinger M, Mazal PR, Susani M, Zielinski CC, Haitel A (2009) CD98hc (SLC3A2), a novel marker in renal cell cancer. Eur J Clin Invest 39:304–310

57. Jackson SP, Bartek J (2009) The DNA-damage response in human biology and disease. Nature 461:1071–1078

58. Harper JW, Elledge SJ (2007) The DNA damage response: 10 years after. Mol Cell 28:739–745

59. Manning G, Whyte DB, Martinez R, Hunter T, Sudarsanam S (2002) The protein kinase complement of the human genome. Science 298:1912–1934

60. Thomson M (2002) Evidence of undiscovered cell regulatory mechanisms: phosphoproteins and protein kinases in mitochondria. Cell Mol Life Sci 59:213–219

61. Horbinski C, Chu CT (2005) Kinase signaling cascades in the mitochondrion: a matter of life or death. Free Radic Biol Med 38:2–11

62. Deng N, Zhang J, Zong C, Wang Y, Lu H, Yang P, Wang W, Young GW, Korge P, Lotz C, Doran P, Liem DA, Apweiler R, Weiss JN, Duan H, Ping P (2010) Phosphoproteome analysis reveals regulatory sites in major pathways of cardiac

mitochondria. Mol Cell Proteomics (In press) (doi:10.1074/mcp. M110.000117)

63. Deng WJ, Nie S, Dai J, Wu JR, Zeng R (2010) Proteome, phosphoproteome, and hydroxyproteome of liver mitochondria in diabetic rats at early pathogenic stages. Mol Cell Proteomics 9:100–116

64. Cui Z, Hou J, Chen X, Li J, Xie Z, Xue P, Cai T, Wu P, Xu T, Yang F (2010) The profile of mitochondrial proteins and their phosphorylation signaling network in INS-1 beta cells. J Proteome Res 9:2898–2908

65. Reinders J, Wagner K, Zahedi RP, Stojanovski D, Eyrich B, van der Laan M, Rehling P, Sickmann A, Pfanner N, Meisinger C (2007) Profiling phosphoproteins of yeast mitochondria reveals a role of phosphorylation in assembly of the ATP synthase. Mol Cell Proteomics 6:1896–1906

66. Ito J, Taylor NL, Castleden I, Weckwerth W, Millar AH, Heazlewood JL (2009) A survey of the *Arabidopsis thaliana* mitochondrial phosphoproteome. Proteomics 9:4229–4240