Proteomic studies on gastric cancer to understand the underlying molecular mechanisms of oncogenesis

Guo, Tiannan

2012

Guo, T. N. (2012). Proteomic studies on gastric cancer to understand the underlying molecular mechanisms of oncogenesis. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/48095

https://doi.org/10.32657/10356/48095

NANYANG TECHNOLOGICAL UNIVERSITY

PROTEOMIC STUDIES ON GASTRIC CANCER TO UNDERSTAND THE UNDERLYING MOLECULAR MECHANISMS OF ONCOGENESIS

TIANNAN GUO

SCHOOL OF BIOLOGICAL SCIENCES

2012

Proteomic Studies on Gastric Cancer to Understand the Underlying Molecular Mechanisms of Oncogenesis

Tiannan Guo

School of Biological Sciences

A thesis submitted to Nanyang Technological University in fulfilment of the requirement for the degree of Doctor of Philosophy

2012

Acknowledgements

It would not have been possible to write this doctoral thesis without the help and support of the kind people around me, to only some of whom it is possible to give particular mention here.

First of all, I owe my deepest gratitude to my supervisor Professor Sze Siu Kwan (Newman) from Nanyang Technological University and my co-supervisor Professor Kon Oi Lian from National Cancer Centre Singapore. This thesis would not have been possible without their generous support, warm and insightful guidance, detailed and constructive comments.

I am grateful to my thesis advisory committee members, Professor Lin Chunling (Valerie) and Professor Su I-Hsin, for their invaluable discussions and suggestions.

I am grateful to have learned from my labmates. I learned phosphopeptide enrichment and iTRAQ experiment from Dr Gan Chee Sian, western blotting from Ms Zhu Yi, membrane protein analysis from Dr Zhang Huoming, some statistical analysis from Ms Ong Whee Sze. Ms Zhu Jiang and Ms Wang Haixia helped with the mitochondrial functional analysis using flow cytometry.

Some data I have analyzed in this thesis were acquired by my colleagues. Microarray data sets obtained from 17 gastric cancer cell lines and normal stomach tissues, which have been referred to in Chapter 2, 4 and 5, were from Ms Sze Sing Lee (Louise) in the Kon lab. Protein antibody array experiments of primary stomach tissues (Chapter

2) were performed by Ms Sze Sing Lee (Louise) and Ms Ng Wai Har. Flow cytometric analysis of 14 gastric cancer cell lines was done by Ms Fan Lingling from the Huang lab in Wuhan, China. The tissue microarray of 49 pairs of primary gastric adenocarcinomas and their matched adjacent non-cancer tissues (Chapter 4) was performed by Ms Ng Wai Har, Dr Mengfatt Ho (Mac), Dr Wei Keat Wan and Dr Kiat Hon Lim.

All YCC cell lines were a gift from Dr. Sun Young Rha, Korea. The potent MET inhibitor PHA-665752 was a gift from Dr James Christensen, Pfizer, USA.

Our research was supported by Ministry of Education (ARC: T206B3211 to SKS), the Agency for Science, Technology and Research Biomedical Research Council (BMRC: 07/1/22/19/531 to SKS) of Singapore, National Cancer Centre Singapore Research Foundation, and a PhD scholarship from Nanyang Technological University.

And last but not least, I would like to thank my family for constant support and encouragement.

Table of Contents

Summary

Gastric cancer ranks as the fourth most common cancer and the second leading cause of cancer mortality globally. One of the main reasons for gastric cancers' poor outlook is the limited knowledge of the underlying molecular mechanisms of gastric oncogenesis. Proteomics has evolved rapidly in recent years and is now recognised as a powerful suite of tools to systematically dissect molecular abnormalities in cancer cells. However, proteomics has yet to be intensively applied to gastric cancer; consequently few global insights into gastric oncogenesis to date have emerged from this approach. In this thesis, we employed advanced proteomic techniques to study phosphorylated proteins, methylated proteins, and cell surface proteins of multiple gastric cancer cells on a large scale. Proteomic data were correlated with transcriptome data sets to gain deeper insights into aberrantly expressed proteins in gastric cancer cells. We found MET, a receptor tyrosine kinase, to be a dominant aberrant protein that was overexpressed in some gastric cancers. A quantitative proteomic approach was utilized to investigate molecular events associated with MET-directed therapy in gastric cancer cells. Remarkably MET was found to be present in the mitochondria of gastric cancer cells, as it was on the plasma membrane. Moreover, mitochondrial MET was identified as a direct target of the MET kinase inhibitor, PHA-665752. Taken together, data presented in this thesis offer a systematic and unbiased profile of multiple molecular abnormalities in gastric cancer cells, and has uncovered a novel mechanism of action of molecularly-directed cancer therapy.

Abbreviations

2-DE	Two Dimensional gel Electrophoresis
AAK1	Adaptor-Associated Kinase 1
ACO	Acyl-CoA Oxidase
AdoMet	S-adenosylmethionine
ALDO	Aldolase
ALK	Anaplastic Lymphoma receptor tyrosine Kinase
ANT	Adenine Nucleotide Translocator
ATCC	American Type Culture Collection
ATM	Ataxia Telangiectasia Mutated
BCA	BicinChoninic Acid
CAD	Collisionally Activated Dissociation
CAM	Cellular Adhesion Molecules
CASK	Calcium/calmodulin-dependent Serine protein Kinase
CD	Cluster of Differentiation
CDC2	Cell Division Cycle 2
CDK	Cyclin-Dependent Kinase
CHK1	Checkpoint Kinase 1
CHM	Cell Homogenization Medium
CK2	Casein Kinase 2
CLK2	Cell division cycle-Like Kinase 2
CORT	CORTistatin
COX	Cytochrome c Oxidase

CS	Citric Synthase
CYC	Cytochrome c
DAVID	Database for Annotation, Visualization and Integrated Discovery
DDR	DNA Damage Response
DSP	Dual-Specific Phosphatase
DTT	Dithiothreitol
DYRK	Dual specificity tyrosine(Y)-phosphorylation-Regulated Kinase
EGFR	Epidermal Growth Factor Receptor
ENO	Enolase
EPHA2	Erythropoietin-Producing Hepatocellular A2 Receptor
ERK	Extracellular signal-Regulated Kinase
ERLIC	Electrostatic Repulsion-Hydrophilic Interaction Chromatography
ETC	Electron Transfer Chain
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
FGFR	Fibroblast Growth Factor Receptor
FH	Fumarate Hydratase
GAPDH	Glyceraldehyde-3-Phosphate DeHydrogenase
GAR	Glycine-Arginine-Rich
GC	Gastric Cancer
GPCR	G Protein-Coupled Receptor
GPI	Glucose-6-Phosphate Isomerase
GRB2	Growth factor Receptor-Bound protein 2
GSK	Glycogen Synthase Kinase
HBSS	Hanks' Balanced Salt Solution

HDGF	Hepatoma-Derived Growth Factor
HER	Human Epidermal growth factor Receptor
HGF	Hepatocyte Growth Factor
HK	hexokinase
HPLC	High Performance Liquid Chromatography
IDH	Isocitrate DeHydrogenase
IGFR	Insulin-Like Growth Factor
IMAC	Immobilized Metal Affinity Chromatography
IMM	Inner Mitochondrial Membrane
INSR	Insulin Receptor
IPA	Ingenuity Pathway Analysis
IPI	International Protein Index
ITK	IL2-inducible T-cell Kinase
iTRAQ	isobaric Tags for Relative and Absolute Quantitation
ITSN	InTerSectiN
JAK2	Janus Kinase 2
LCK	LymphoCyte-specific protein tyrosine Kinase
LDH	Lactate DeHydrogenase
MALDI	Matrix-Assisted Laser Desorption/Ionization
MAP2K2	Mitogen-Activated Protein Kinase Kinase 2
MCM2	Minichromosome Maintenance 2
MDH	Malate Dehydrogenase
MMP	Mitochondrial Membrane Potential
mPTP	mitochondrial Permeability Transition Pore
MRPS16	Mitochondrial Ribosomal Proteins 16

MS	Mass Spectrometry
MTMR	MyoTubularin-related protein
mTOR	mammalian Target Of Rapamycin
NCK	Non-Catalytic region of tyrosine Kinase adaptor protein
NDUF	NADH dehydrogenase (ubiquinone)
NEK2	NIMA (never in mitosis gene a)-related Kinase 2
NLK	Nemo-Like Kinase
OGDH	OxoGlutarate DeHydrogenase
OMM	Outer Mitochondrial Membrane
OMSSA	Open Mass Spectrometry Search Algorithm
PBS	Phosphate-Buffered Saline
PC	Pyruvate Carboxylase
PDGFR	Platelet-Derived Growth Factor Receptor
PDH	Pyruvate DeHydrogenase
PDK	Pyruvate Dehydrogenase Kinase
PFKP	6-phosphofructokinase
PGAM	Phosphoglycerate Autase
PGK	Phosphoglycerate Kinase
PHP	Protein Histidine Phosphatase
PI3K	PhosphatidylInositol 3-Kinase
PIP3	PhosphoInositide-binding Protein 3
PK	Pyruvate Kinase
PKA	Protein Kinase A
PKC	Protein Kinase C
PKD2	Polycystic Kidney Disease 2

PKN2	Protein Kinase N2
PLC	PhosphoLipase C
PQD	Pulsed-Q Dissociation
PRL	Phosphatase of Regenerating Liver
PRMT	Arginine Methyltransferase
PTM	Post-Translational Modification
PTP	Protein Tyrosine Phosphatase
PTPN	Protein Tyrosine Phosphatase, Non-receptor type
PTPR	Protein Tyrosine Phosphatase, Receptor type
Q-TOF	Quadrupole Time-Of-Flight
RTK	Receptor Tyrosine Kinase
SCX	Strong Cation eXchange
SDH	Succinate DeHydrogenase
SDS-PAGE	Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis
SHC	Src Homology 2 domain Containing transforming protein
SHIP	SH2 domain-containing Inositol Phosphatase
SHP1	Src homology region 2 domain-containing phosphatase-1
SILAC	Stable Isotope Labeling by Amino Acids in Cell culture
SRPK2	Serine/arginine-rich Protein-specific Kinase 2
STP	Serine/Threonine Phosphatase
SUCL	Succinate-CoA Ligase
TC-PTP	T-cell Protein Tyrosine Phosphatase
TiO_2	Titanium dioxide
TNIK	NCK Interacting Kinase
TOF	Time-Of-Flight

TOMM20 Translocase of Outer Mitochondrial Membrane 20

TPI1 triosephosphate isomerase 1

TPP Trans-Proteome Pipeline

TTBK2 Tau Tubulin Kinase 2

VDAC Voltage-Dependent Anion Channel

VEGFR Vascular Endothelial Growth Factor Receptor

Chapter 1. Introduction

Gastric cancer

Gastric cancer (GC) or stomach cancer is malignant tumor arising from gastric epithelium. Gastric cancer dates from antiquity, having been recorded by the Greek physician, Hippocrates (*ca.* 460-370 BCE) [1]. Gastric cancer has a distinctive geographic distribution of high incidence in East Asia, Eastern and Southern Europe, Central and South America.

Gastric cancer is twice as common among men as in women. It is uncommon below the age of 45 years, most patients being 60–80 years old at diagnosis [2]. In various countries, at least 50% of all newly diagnosed cases already have metastatic disease for which there is no effective curative treatment. Moreover, among patients without metastases, only about half are suitable for gastrectomy with curative intent [3]. Relapse rates after curative gastrectomy are high [3], possibly because primary carcinomas develop in a wide field of cancerization. Thus, the fatality rate for gastric cancer remains high in most countries, except Japan, because surgery is currently the only treatment with curative potential.

The lethality of gastric cancer is not, *a priori*, intractable or inevitable. Measures that prevent or halt gastric oncogenesis, sensitive and specific diagnostic methods for early stage disease, and the development of novel chemo- and/or biological therapy can all be expected to improve an otherwise bleak situation. Better food sanitation and increased intake of fresh vegetables and fruits are credited with lowering the

incidence of gastric cancer. Higher standards of personal hygiene decrease the risk of infection with *Helicobacter pylori*, a Gram negative rod that is a class I gastric carcinogen in humans [3].

A positive trend over several decades has been the steadily declining incidence in most countries attributed to improved living standards, such that gastric cancer now ranks lower, at fourth place, among the most prevalent cancers worldwide. Regrettably, decreased gastric cancer incidence has not been matched by a proportionate decline in mortality [4]. Overall 5-year survival from GC, except in Japan, is only about 20% [5]. Indeed, the fact that GC causes more deaths annually than all other cancers, except lung cancer, is an urgent call to action to enhance the effectiveness of current methods of prevention, diagnosis and treatment [6].

There are recognizably distinct histotypes that have been classified in several ways. The most commonly used histological classification is that of Lauren who identified two main histotypes *i.e.* intestinal and diffuse types [7]. Intestinal-type tumours predominate in countries of high GC prevalence. The diffuse histotype tends to occur in younger patients without male predominance, is typical of hereditary GC arising from germline mutations, is more invasive and has poorer treatment outcomes. However histotype is not a clinically useful prognostic indicator.

During the past several decades, numerous molecular abnormalities have been reported in various cancers, including gastric cancer. It is now evident that individual cancers can vary markedly from each other. It is estimated that, for most cancers, 1000~10000 somatic mutations are present in a single tumor [8]. Thus, no single gene

or protein could ever completely explain the biology of cancer. For this reason alone, cancer cells need to be investigated and understood at the systems level.

A turning point in cancer research was the completion of sequencing of the human genome, with its attendant technological advances [9]. Next-generation genomic approaches enable unbiased sequencing and analysis of tumor cells [6]. Compared to conventional hypothesis-based studies, high-throughput techniques that interrogate a global class of macromolecules unfettered by prior assumptions provide invaluable clues to understanding oncogenic mechanisms. Systematic investigation of cancer genomes has demonstrated the power to develop molecular tumor taxonomies, discover diagnostic, prognostic and predictive biomarkers. These insights are increasingly enabling targeted therapeutics and biologically informed clinical management of cancer susceptibility [10]. While much of recent cancer research has been performed on tumor DNA and RNA, it is clear that nucleic acids are not the direct executors of life activities. Cellular functions occur as a network which is mainly constructed by protein-protein interactions. Moreover, the majority of clinically validated biomarkers are proteins [11]. Effective molecular targets in targeted cancer therapeutics are also mainly proteins [12]. Therefore, these considerations motivate global investigation of the protein inventory of GC as an urgent need.

Proteomics

Proteomics refers to a suite of techniques for system-wide investigations of proteins for discovery-based research [13]. Instead of studying individual proteins, proteomics

permits analysis of thousands of proteins in a single experiment for uncovering novel biological functions of the system. The capacity to achieve this was accelerated after completion of the human genome project which provided the sequence of genes encoding the human proteome. During my PhD training, this field has experienced phenomenal technical advancements.

Only several years ago, two-dimensional gel electrophoresis, abbreviated as 2-DE or 2-D electrophoresis, was the most widely used technique for proteomic studies. The technique separates proteins in the first dimension on the basis of the isoelectric point or electric charge of protein molecules. This is followed by electrophoretic resolution in the second dimension by a different property, usually molecular size by SDS-PAGE (sodium dodecyl sulfate polyacrylamide gel electrophoresis). Proteins are visualized as "spots" by various staining methods, which can be excised and digested individually for analysis by mass spectrometry (MS). MALDI (matrix-assisted laser desorption/ionization)-TOF (time-of-flight) or MALDI-TOF/TOF are the most common instruments that are coupled with 2-DE-based proteomics for identification of proteins. This method has inherent limitations. First, proteins must first be resolved in SDS-PAGE, thereby introducing a distinct bias to aqueous-soluble proteins. Hydrophobic proteins, such as membrane proteins, are difficult to analyze in 2-DE experiments. The capacity of electrophoresis also limits the amount of proteins that can be analyzed. Second, only visible proteins can be excised for digestion and MS analysis, thus introducing an additional level of bias to high-abundance proteins. Proteins of low-abundance are mostly invisible and will be missed in this workflow. Moreover, MALDI-MS has limited sensitivity in identifying low-abundance proteins, compared to linear ion trap mass spectrometers and Q-TOF mass spectrometers. Only

a few hundred of proteins of high abundance are typically identified after tedious experimental procedures. Nevertheless, 2-DE remains the method of choice for many laboratories engaged in proteomic studies due to its low cost and ease of maintenance. A survey of the current literature shows that the majority of gastric cancer proteomic studies were performed using 2-DE.

The development and introduction of shotgun proteomics was a tremendous advance in the field. Here proteins from any complex sample are digested into short peptides by a proteolytic enzyme. The resulting peptides are subsequently separated by multiple-dimensional liquid chromatography (LC), typically 2 dimensional LC with the first dimensional separation property orthogonal to the second dimensional separtion. Each fraction from the first dimensional LC, after desalting and cleaning, is analyzed with a high performance liquid chromatography (HPLC) coupled to tandem mass spectrometery (LC-MS/MS) via a nanospray ionization. Linear ion trap and Q-TOF are usually the mass spectrometers of choice for shotgun proteomics. Coupled with protein sequence database searching, shotgun proteomics is capable of identifying thousands of proteins in a proteome in a single experiment. By combining shotgun proteomics with subproteome enrichment, it is feasible to gain unprecedented depth of coverage and comprehensive data of post-translationally modified peptides and proteins that reside in specific subcellular organelles including membranes. Moreover, expression levels of thousands of proteins can also be measured simultaneously using isobaric labeling.

Oncogenic signaling pathways and phosphoproteomics of gastric cancers

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 molecularly-targeted agents have been developed as new strategies to control cancers [12]. 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), HSP90, intracellular signaling pathways (PI3K, AKT, mTOR) and angiogenesis, and agents that interfere with DNA repair (PARP inhibitors) [14].

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 [15]. Cancer cells that are initially sensitive to suppression of a specific target commonly become resistant within one year of treatment [16]. 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 [17]. Non-small cell lung cancer cells initially sensitive to EGFR inhibition acquire resistance by activating MET-HER3-PI3K signaling pathway [18]. Acquired resistance of HER2-overexpressing breast cancers is dependent on activation of insulin-like growth factor 1 receptor-PI3K/Akt signaling pathway [19]. There is evidence of improved efficacy when multiple targets are concomitantly suppressed [20]. 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, that have proven efficacy in other cancers, are being tested in gastric cancer [21-25]. To date, a clinical trial directed at only one target, HER2, has reported significant but modest extension of survival of gastric cancer patients [21]. 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 [26]. A major challenge in developing targeted therapy is the current paucity of mechanistic understanding of gastric oncogenesis as mediated by signaling pathways [27].

Dynamic changes in the phosphorylation state of proteins are key functions of protein kinases and phosphoprotein phosphatases that mediate and modulate a broad variety of cell functions. In contrast to conventional biochemical assays which typically investigate individual phosphorylation sites, MS-based proteomics permits profiling of proteins phosphorylated at multiple sites on a large scale.

Phosphoproteomic profiling of cancer cells sheds light on key components in oncogenic signaling networks, which are enriched of biomarkers of potential diagnostic and prognostic importance, and may also be candidate therapeutic targets [28-30]. The phosphoproteome of cancer cells has been investigated using antibody arrays, but this approach is limited by the high cost and low throughput [31].

Phosphoproteins can also be purified by immunoaffinity binding followed by LC-MS/MS analysis [32]; however, this is also a low-throughput method. Recent developments in shotgun proteomics of enriched phosphopeptides permit in-depth profiling of phosphoproteomes, and have boosted new insights into specific oncogenic signaling pathways, such as TGF-beta in colon cancer [33], HER2 in breast cancer [34], IL2 in chronic lymphocytic leukemia [35], Ephrin B3 in lung cancer [36], and DNA damage response in melanoma cells [37]. However, a comprehensive and unbaised view of the cancer phosphoproteome of human cancers has not been reported yet, despite several phosphoproteomic reports on various organisms and cell types *e.g.* Phospho.ELM [38], PhosphoSitePlus [39], phosphoPep [40], PHOSIDA [41] and Uniprot. These phosphoproteomic studies on non-human or non-cancer cells have limited relevance for gastric cancer research because phosphorylation profiles are highly diverse and differ in a cell type-dependent manner. In the current literature, only a few publications have investigated phosphoproteomes of human cancers but with relatively obsolete techniques with limited detection capacity. Thus, only about 20 phosphoproteins were identified from a set of five prostate cancer specimens [42]. Nearly 50 phosphorylated proteins were reported to be secreted by a gastric cancer cell line [43]. One hundred and eighty-one phosphoproteins were identified in a lung cancer cell line [44] and 260 in nine primary multiple myeloma bone marrow samples [45]. A total of 385 [46] and 296 [47] phosphoproteins were characterized in prostate cancer cell lines. No large-scale phosphoproteomic study of gastric cancer was reported prior to our report documented in Chapter 2.

Owing to the substoichiometric nature of phosphorylated proteins, it is essential to enrich phosphorylated proteins and peptides prior to MS analysis. Several methods

are available for phosphopeptide enrichment in shotgun proteomics. These include phosphoramidate chemistry [48], strong cation exchange (SCX) [49], SCX-IMAC (immobilized metal affinity chromatography) [50], titanium dioxide (TiO_2) [51] and electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) [52]. We have performed a comparative study of several phosphopeptides enrichment techniques including SCX, SCX-IMAC and ERLIC, and found that no single method is sufficient to enrich all phosphopeptides [53]. Hence, a combination of several different phosphopeptide enrichment methods increases the depth of coverage of the phosphoproteome (Figure 1.1), consistent with an earlier comparative study which investigated the performance of phosphoramidate chemistry, IMAC and TiO_2 [54].

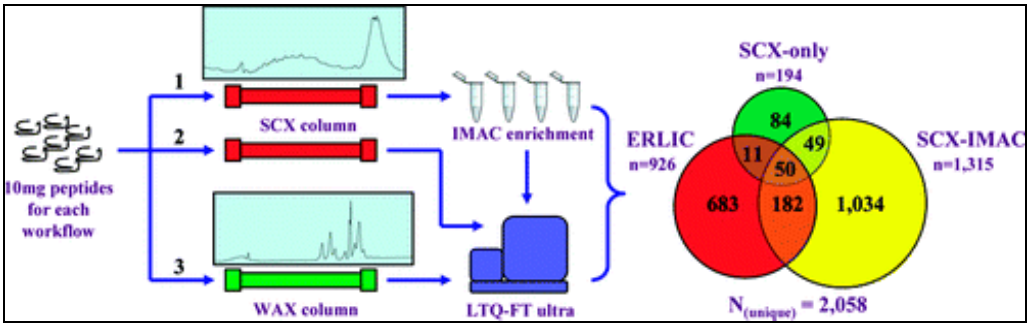


Figure 1.1. A comparative study of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) *versus* SCX-IMAC-based methods for phosphopeptide isolation/enrichment. We compared ERLIC with the well-established SCX-IMAC for identifying phosphopeptides in EGF-treated A431 cells. The ERLIC approach detected a higher number of phosphopeptides (17311) than SCX-IMAC (4850), but it detected fewer unique phosphopeptides (926) than SCX-IMAC (1315). Only 12% of unique phosphopeptides were common to both approaches, suggesting that more comprehensive phosphoproteome coverage is generated by complementing SCX-IMAC with ERLIC [53].

An obvious benefit of utilizing multiple complementary phosphopeptide enrichment techniques is the broader view of phosphorylation-mediated signaling pathways. We have shown this by generating a more detailed map of EGFR signaling in A431

epidermoid carcinoma cells by combining three different phosphopeptide enrichment methods (Figure 1.2) [53].

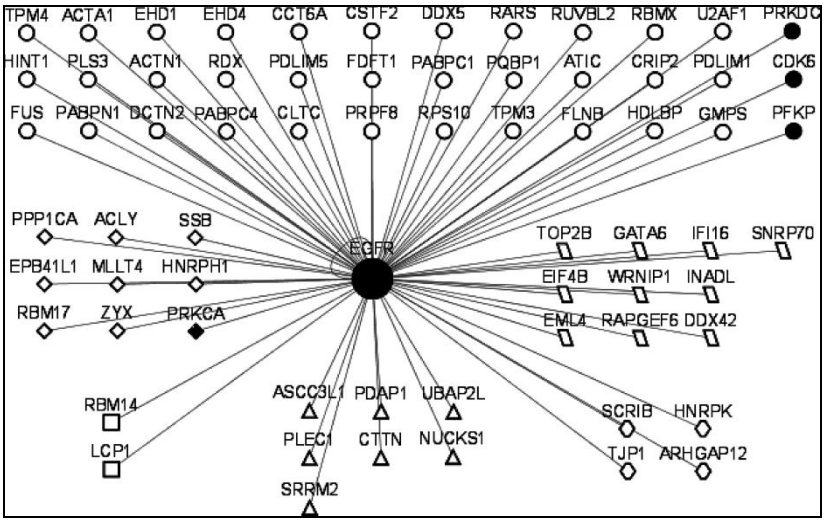


Figure 1.2. EGFR protein substrate interaction network. Non-phosphorylated proteins appear as circles, while phosphorylated proteins identified in different workflows appear in shapes as follows. Parallelogram, ERLIC; diamond, SCX-IMAC; hexagon, ERLIC and SCX-IMAC; triangle, ERLIC, SCX-IMAC, and SCX-only; square, SCX-IMAC and SCX-only. Kinases are shown as filled black symbols. Information about protein kinases was retrieved from NetworKIN [55].

In Chapter 2, we have performed a comprehensive investigation of the gastric cancer phosphoproteome and transcriptome. LC-MS/MS-based phosphoproteomics profiling identified >3,000 phosphorylation sites in >1,200 proteins in five gastric cancer cell lines, whereas protein antibody-array-based phosphoproteomics further identified 46 phosphorylated signaling proteins in clinically procured stomach tissues. The phosphoproteomics data set was correlated with large-scale gastric cancer transcriptional data sets of quantitative expression of >12,000 genes in a panel of 17 gastric cancer cell lines relative to normal stomach. We report an expansive view of the gastric cancer oncokinome and phosphoproteome from an integrated bioinfomatics analysis of phosphoproteomics and transcriptome data sets.

Methyl proteomics of cancers

While phosphorylation negatively charges a protein and thereby modulates its structure and biological activities, another type of post-translational modification, *i.e.* protein methylation, also has active roles in regulating protein functions and signal transduction by modulating the hydrophobicity and charge of proteins [56-59]. Methylation of proteins at lysines and arginines has been known for decades [60-62]. Methylation was later found to modify various residues including lysine, arginine, histidine, proline, and the N- and C–terminal residues [56, 58, 63, 64]. Of these, methylation at lysine and arginine have been most extensively studied [62]. Methylation of proteins alters not only their hydrophobicity, but also their charges. The charge state of lysine and arginine residues is slightly altered when their amino groups are methylated [65]. Although methylation is more stable than phosphorylation in proteins, recent findings have nonetheless established protein methylation to be a reversible process [56]. Remarkably, the same amino acid residue can be methylated to different degrees. Thus, lysine residues may be mono-, di- or trimethylated, while arginine residues are mono- and dimethylated [62]. Different methylation states of the same residue are known to have different impacts on the structure and/or function of the modified substrate proteins [65].

Histones are the best studied methylated proteins [57, 63]. Lysine methylation also occurs in non-histone proteins, including p53, cytochrome *c*, VEGFR1 and ribosomal proteins [65]. Lysine methylation enhanced VEGFR kinase activity and altered the transcriptional activity of p53 [122, 123]. Interestingly, monomethylation and dimethylation of the same lysine residue had different effects on p53 activity [66, 67]. Protein arginine methylation is relatively well-studied. Currently eleven mammalian

protein arginine methyltransferases (PRMTs) have been identified [61]. Six of them catalyze the transfer of a methyl group from S-adenosylmethionine (AdoMet) to arginine residues[68]. Arginines flanked by one or more glycines are frequently methylated. Proteins that harbor glycine-arginine-rich (GAR) motifs are common targets for PRMTs [69]. Dozens of proteins have been reported to be methylated at arginine residues [57, 63]. It is worth noting that methylation sometimes occurs at the same residue that could also be subject to other covalent modifications. For example, the same lysine residue in p53 may be modified by acetylation, ubiquitination, sumoylation and methylation, each with distinct functional effects [65].

Methylated proteins regulate a variety of cellular functions including transcription, mRNA splicing, signal transduction and DNA repair, among other [56-58, 70]. Given its multifaceted effects, it is not surprising that protein methylation has been implicated in the pathogenesis of multiple diseases, including cancers in which anti-tumor effects have been ascribed to protein methylation [56, 68]. Other investigators have suggested that PRMTs are promising targets for targeted therapies in prostate and breast cancer [68]. However, our knowledge of protein methylation is still fragmentary, and the breadth and depth of protein methylation remains to be fully documented [62, 68].

Mass spectrometry-based proteomics offers a powerful approach for large scale investigation of methylated proteins. Instead of investigating methylation sites individually using conventional methods, over 200 novel putatively methylated proteins were reported from HeLa cells by MS-based proteomics [71]. Eighteen methylated proteins were uncovered from a proteomic analysis of rat liver Golgi

complex [72]. About 60 methylation sites were identified in immunoprecipitated methylated peptides from HeLa cells cultured in heavy methyl SILAC [73]. In this study, cells were grown in either normal or isotope-labeled methionine. The cell lysates were mixed prior to LC-MS/MS analysis, so that methylated peptides with methyl group donated by AdoMet displayed characteristic isotopic peak pairs in mass spectra, which added confidence in identifying methylation sites. Another study reported that not all methylated peptides exhibited isotopic peak pairs in heavy methyl SILAC experiments, probably due to presence of methyl donors other than AdoMet. In addition, methylation also takes place *in vitro* during sample preparation [74]. Given over-expressed target proteins and careful inspection of spectra, this study demonstrated the power of MS-based proteomics in identifying novel methylation sites. Another comprehensive proteomic study uncovered 155 methylated proteins from *Leptospira interrogans* [75].

One major hurdle in profiling methylated peptides on a large scale is that methylated proteins are usually present in substoichiometric amount. Efficient enrichment of methylated proteins or peptides is challenging because, unlike phosphorylation, methylation only partially alters the hydrophobicity and charge of the modified residues. In some studies, antibodies were raised to specifically bind methylated peptides. However, the specificity of these antibodies is unsatisfactory [73]. An alternative method is to map methylation sites in a particular protein after having been enriched by immunopurification; however, this approach can hardly be considered high-throughput [74].

In Chapter 3, we deployed advanced MS-based proteomics to globally profile methylated cellular proteins. We reasoned that exhaustive fractionation of the proteome by various chromatographic methods should enable methylated peptides of low abundance to be identified with high confidence. We demonstrated the utility of this approach by effectively identifying methylated proteins in a gastric cancer cell line, SNU5. Heavy methyl SILAC was further used to validate methylated peptides utilizing AdoMet and methionine as methyl donors.

Membrane proteomics of cancers

Membrane proteins have critical functions in both normal and malignant cells [76]. Some membrane proteins, especially glycosylated plasma membrane proteins, have proved to be major cancer biomarkers [77, 78]. Moreover, as plasma membrane proteins are more easily accessible, 60% of current drug targets are on the cell surface although they account for about 22% of all human protein-coding genes [79]. Plasma membrane proteins mediate several fundamental cellular functions such as regulation of cell growth, apoptosis and signal transduction and are widely regarded as the most important class of membrane proteins. Characterization of plasma membrane proteins in cancer and stem cells is an active area of research [80-82]. Even now, detection of selected cell surface proteins using immunohistochemistry and flow cytometry has become standard clinical practice in cancer diagnosis and prognosis [83, 84] .

The cluster of differentiation (CD) molecules are a group of cell surface proteins that were identified initially in hematopoietic cells, and subsequently found in multiple cell types. A total of 389 human CD proteins were curated by the UniProt Swiss-Prot

Protein Knowledgebase as of January 2011 [85]. CD molecules are best known from studies of cell surface proteins as they encompass a plethora of vital functions. They include protein kinases, receptors, ligands and enzymes. Expression patterns of CD antigens on cancer cells have been proposed as rational, cancer-specific diagnostic and prognostic biomarkers, as well as potential molecular targets for therapeutic intervention [86].

Receptor tyrosine kinases (RTKs), the most extensively studied cell surface membrane molecules [87], comprise 58 members grouped into 20 subfamilies based on kinase domain sequences [88]. RTKs currently dominate as diagnostic and therapeutic biomarkers of several types of cancer whose cells express mutationally activated, overexpressed or amplified RTKs that initiate and/or sustain malignant transformation [89, 90]. The development of inhibitors against HER2, KIT, EGFR, ALK and VEGFR are exemplars of disease-specific RTK biomarkers that are both diagnostic and therapeutic. Compared to other common cancers, there is a dearth of molecularly-directed therapeutic agents for GC. However, recent evidence for a modest survival benefit in advanced GC patients who received trastuzumab with chemotherapy suggests that targeted therapy for GC has relatively untapped potential [91]. To this end, there is an urgent need to comprehensively document RTKs that are dysregulated in GC using large-scale characterization of plasma membrane proteins (17). Genomic approaches that interrogate tens of thousands of genes inherently fail to provide direct evidence of subcellular localization and are thus not optimally suited for investigating cell surface proteins. MS-based proteomics is a proven method for developing systematic inventories of membrane proteins [76, 92-94]. Few publications have documented the surface proteome of cancer cells. Conventional 2D-

gel-based proteomics permits only limited throughput and sensitivity. 2D-gel electrophoresis coupled with MALDI-MS analysis identified only 33 plasma membrane proteins from cell lines derived from several cancers including neuroblastoma, lung adenocarcinoma, colon adenocarcinoma, acute lymphoblastic leukemia, and ovarian cancer [95]. Another study applying 2D-gel electrophoresis coupled with MALDI-MS to six human colorectal carcinoma biopsies characterized around 20 proteins of potential pathobiological importance [96]. Recent advances of LC-MS/MS-based proteomics allowed higher throughput. A study that compared the plasma membrane proteomes of proliferating and differentiated human colorectal carcinoma cells identified a total of 1125 proteins [97]. More than 100 cell surface proteins of acute myeloid leukemia cell lines were analyzed recently [94]. However, a comprehensive and unbiased inventory of the GC membrane proteome has yet to be reported.

In Chapter 4, we demonstrate a strategy to systematically investigate membrane proteins from multiple cell sources by exploiting high-throughput technologies and computational methods. Analysis of transcriptome data sets integrated with LC-MS/MS-based membrane proteome data sets revealed insights into membrane proteins at both mRNA and protein levels. This has generated the first global view of the membrane proteome in GC. Our data showed CD molecules to be extensively but variably expressed on the surface of GC cell lines derived and established from different patients. We found similarly variable expression of RTKs. The existence of distinct signatures of the plasma membrane proteome potentially stratifies GC into different molecular subtypes. Integrated analysis showed that cell surface proteins, including CD molecules and RTKs, are functionally organized in complex

interactomes. We validated our strategy by characterizing cell surface proteins by flow cytometry, immunostaining primary GC tumor tissues and by antibody array analysis of activated RTKs in cancerous and benign gastric epithelium.

Quantitative proteomics

Simultaneous quantitative proteomic profiling of multiple biological states in a high-throughput manner holds significant potential for biological and biomedical discovery. This has encouraged rapid development in biological mass spectrometry methods for quantitative proteomics [98]. As MS is not inherently quantitative, protein or peptide samples are usually labeled with stable isotopes for relative quantitation. Quantitative information can be acquired either from MS spectra, such as stable isotope labeling by amino acids in cell culture (SILAC) [99], or MS/MS spectra *e.g.* using isobaric tags for relative and absolute quantitation (iTRAQ) [100].

SILAC has gained global popularity in quantitative proteomics in recent years [99]. In this method, cells are grown in a customized medium which replaces a specific amino acid, usually lysine and/or arginine, with their respective heavy stable isotope form. After culturing for a few generations, all the proteins in the cells are labeled with the heavy amino acid(s). Cell lysate from heavy cells is mixed with that from normal cells, and analyzed in a typical shotgun proteomic approach. Peptides from different samples display a characteristic peak pair in the resulting mass spectra. The abundance of peptides/proteins from each cell culture is inferred from the peak area/intensities. SILAC experiments are well suited for quantification of post-translational modifications. However, the technique is limited by the requirement for

culture *in vitro*. Cells and tissues that have not been cultured in SILAC medium are not suitable for this workflow.

iTRAQ is currently the most widely used approach for high-throughput protein quantitation by MS/MS. It enables simultaneous quantitation of up to 4 (4-plex iTRAQ) or 8 (8-plex iTRAQ) different biological samples. Furthermore, iTRAQ stable labelling reagents are incorporated "post-harvest" to protein samples at the peptide level via chemical reactions [100], allowing accurate protein quantitation from specimens of diverse origins, including cell lines, tissue samples, biological fluids, and so forth.

Linear ion trap MS is one of the most extensively employed MS instruments in current proteomic research, mainly due to its unsurpassed sensitivity as well as high ion capacity, fast scan rate, ease of use and relatively low cost. Application of iTRAQ in linear ion trap mass spectrometers has been a great challenge because most of the low *m*/*z* iTRAQ reporter ions are not detected/captured in the ion trap after collisionally activated dissociation (CAD) fragmentation. iTRAQ reporter ions can be detected in many mass spectrometers in MS/MS mode, such as MALDI-TOF-TOF and Q-TOF [100], with the exception of the ion trap. In theory, iTRAQ reporter ions can only be detected with linear ion trap MS instruments with CAD fragmentation method if the precursor ion is a small peptide with m/z less than 3 times of iTRAQ reporter ions m/z.

The recent introduction of a novel fragmentation method, pulsed-Q dissociation (PQD) is touted as a more robust method for iTRAQ reporter ions detection in the

linear ion trap MS [101]. We have developed a novel method which intergrated both CAD and PQD to perform iTRAQ experiments in linear ion trap MS (Figure 1.3) [102]. With the PQD−CAD hybrid method, we identified and quantified 1610 proteins from a gastric cancer cell SNU5 in a single proteomic experiment with high accuracy with respect to protein quantification [87, 102].

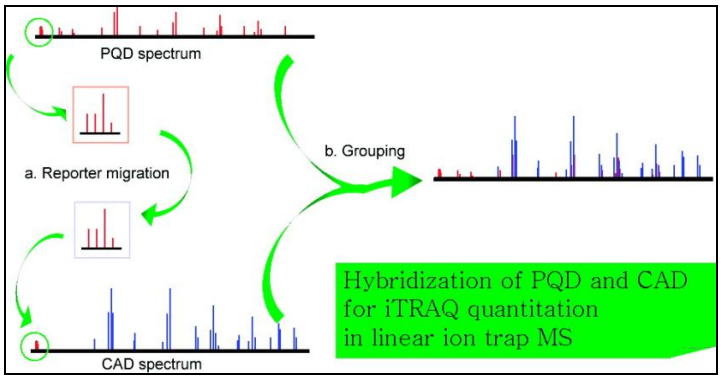


Figure 1.3. Combined pulsed-Q dissociation (PQD) and collisionally activated dissociation (CAD) in linear ion trap mass spectrometer for iTRAQ quantitation. Integrating the advantages of PQD and CAD fragmentation methods in a PQD−CAD hybrid mode, together with PQD optimization and data manipulation with a bioinformatics algorithm, resulted in a robust, sensitive and accurate iTRAQ quantitative proteomic workflow. The workflow was superior to the default PQD setting when profiling the proteome of a gastric cancer cell line, SNU5. Taken together, we established an optimized PQD-CAD hybrid workflow in LTQ-MS for iTRAQ quantitative proteomic profiling that may have wide applications in biological and biomedical research [102].

We further optimized the iTRAQ workflow, and established another protocol for iTRAQ quantification using multiple dimensional chromatography coupled with a Q-TOF mass spectrometer. The protocol resulted in enhanced proteome coverage and high quality of protein quantification, and has been successfully employed in this thesis for investigation of MET-directed targeted therapy in gastric cancer cells (Chapter 5 and reference [102, 103]).

MET "addiction" in gastric cancer cells

Recent improvements in survival of some malignancies owe much to advances in uncovering aberrantly active molecular pathways, against which molecularly-targeted agents have been developed as new strategies to control cancers [12, 104]. However, molecular mechanisms underlying the curious dependence of some cancer cells, which contain multiple genomic, genetic and epigenetic abnormalities, on a single oncogenic molecule (the phenomenon of "oncogene addiction") are incompletely understood [105-107].

Receptor tyrosine kinases (RTKs) are the most extensively studied oncogenic targets and RTK inhibitors have proven anti-cancer therapeutic efficacy. A receptor tyrosine kinase, MET, whose ligand is hepatocyte growth factor (HGF), is frequently amplified and overexpressed [108, 109] in gastric cancer, the second highest cause of cancer mortality globally [110, 111]. Human gastric cancer cell lines harboring *MET* amplicons and overexpressing MET are readily induced to apoptosis by selective inhibitors of MET [112, 113], several of which are under active development for clinical use [114]. One of the selective small molecular inhibitors, PHA-665752, (molecule weight of 641.61, Figure 1.4), specifically suppresses tyrosine phosphorylation of MET. PHA-665752 has >50-fold higher selectivity for MET than for other tyrosine and serine/threonine kinases [115]. The inhibition of MET-kinase function by PHA-665752 on cancer cells had been confirmed with siRNA knockdown of MET, and a number of downstream effectors of MET signaling pathways were confirmed to be effectively abrogated by this compound [112, 115]. PHA-665752 has been widely used as a potent and selective tool for the evaluation of MET-dependent cellular functions and signal transduction [112, 116-125].

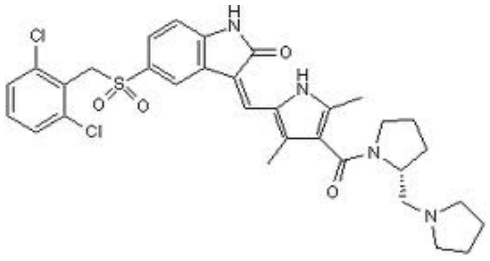


Figure 1.4. Structure of MET inhibitor PHA-665752.

The fact that only a subset of cancers is sensitive to killing by MET-directed therapeutics (hereafter referred to as sensitive cells) [114], raises an unexplained paradox. MET-overexpressing cancer cells could reasonably be expected to be more tolerant of MET kinase inhibition compared to cancer cells that do not overexpress MET. In reality, the opposite occurs. The underlying molecular mechanisms are incompletely understood.

In Chapter 5, to investigate this paradox we undertook a systematic exploration of responses of a MET-overexpressing gastric cancer cell line, SNU5, to sublethal MET inhibition using iTRAQ-based quantitative proteomics approach. Our results unexpectedly showed a predominant perturbation of mitochondrial proteins in response to MET inhibition. Next, we found that MET inhibition was rapidly associated with altered mitochondrial functions. These observations raised the possibility that mitochondria might be a direct target of MET inhibition. Both protein immunoblotting and confocal microscopy showed the presence of highly activated MET in the mitochondria of sensitive cancer cells. Furthermore, we observed that activating phosphorylation of tyrosine residues of mitochondrial MET was critically modulated by sublethal PHA-665752 treatment.

Chapter 2. An Integrated Analysis of the Phosphoproteome and Transcriptome Provides an Expansive View of Molecular Signaling Pathways in Gastric Cancer

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 phosphoproteome and transcriptome provided an expansive view of molecular signaling pathways in gastric cancer.

Material and Methods

All chemicals were purchased from Sigma-Aldrich (St Louis, MO, USA) 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 American Type Culture Collection (the ATCC, Manassas, Virginia, USA). MKN7 and IM95 cells were from Japan Health Science Research Resource Bank. All YCC cell lines were a gift from Dr. Sun Young Rha (Yonsei Cancer Center, Seoul, Korea). Normal stomach RNA samples were reference controls for transcriptome analysis. FirstChoice Human Stomach Total RNA was RNA from a single individual, while 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, 2 histologically normal gastric biopsies, 7 biopsies of benign gastritis, and 3 pairs of gastric adenocarcinoma with their matched normal gastric tissues were analyzed. 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 GC 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, protease inhibitor cocktail (COMPLETE, Roche Applied Science, Indianapolis, IN, USA) and phosphatase inhibitor cocktail (PHOSTOP, Roche Applied Science). Proteins were reduced in 10 mM DTT for 1 hr at 33 ℃, and then alkylated with 55 mM iodoacetamide for 30 min at room temperature, before diluting 8 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.

Phosphopeptide enrichment

Phosphopeptides were enriched using both ERLIC and SCX-IMAC as described [53]. Briefly, for ERLIC, approximately 2 mg of peptides were injected into a PolyLC

PolyWAX LP column (4.6 × 200 mm, 5 μm particle size, 300 Å pore size) mounted on a Shimadzu Prominence UFLC unit (Shimadzu Corp., Kyoto, Japan). For SCX-IMAC, approximately 2 mg of peptides were 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 hr 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). The eluate was immediately pH-adjusted to pH 2.60 using 10 % formic acid. Peptides in salt solutions were vacuum-dried and desalted using SEP-PAK C18 cartridges (Waters Corp., Milford, MA, USA).

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, USA) coupled with a Prominence™ HPLC unit (Shimadzu), as described previously [102] 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, USA), and subsequently resolved in a capillary column (200 μm ID x 10 cm) packed with C18 AQ (5 μm particles, 100 Å pore size; Michrom BioResources, Auburn, CA, USA). 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 5-30 % B over 66 min, followed by 10 min of 50 % B and a ramp from 50-80 % 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 2, ion transfer tube temperature at 180 ℃ 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 (350-2000 m/z range) was acquired in the FT-ICR cell at a resolution of 100,000 and a maximum ion accumulation time of 1000 msec. 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 CAD. The default AGC setting was used (full MS target at $3.0e^{+04}$, MS^n $1e^{+04}$) in the linear ion trap. The 10 most intense ions above a 500-count threshold were selected for fragmentation in CAD (MS2), which was performed concurrently with a 1 maximum ion accumulation time of 200 msec. Dynamic exclusion was activated for this process, with a repeat count of 1, exclusion duration of 20 s and ±5 ppm mass tolerance. For CAD, 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 BioworkBrowser (version 3.3, Thermo Scientific Inc.), X!Tandem [126] (Tornado edition, version 2010.01.01.4), and OMSSA (command line version 2.1.7) [127]. All searches were limited to a maximum of 2 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*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 [128] was also employed to predict the most likely kinases responsible for the phosphorylation sites in gastric cancer phosphoproteome.

Antibody array experiments

Protein lysates from stomach biopsies were probed for phosphorylated signaling proteins using Proteome Profiler™ antibody arrays (R&D Systems, Minneapolis, MN, USA). Manufactured in sets of 2, 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 ug 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 hours at room temperature. Phosphorylated proteins were detected on washed membranes using streptavidin-horse radish 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 analysed 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 Knowledgebase.

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 [129]. Multiple enrichment methods are recommended for comprehensive shotgun phosphoproteome analysis [54, 130]. We employed two methods, electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) and SCX-IMAC, to enrich phosphopeptides. The benefits of using two different enrichment methods are shown in Figure 2.1. SCX-IMAC and ERLIC increased the coverage of SNU5 phosphoproteome by 122% and 58%, respectively. Only 8% of non-redundant phosphopeptides identified were identified by both methods. Five cell lines, *i.e.* SNU5, SNU1, AGS, YCC1, and KatoIII, were included in phosphoproteomics 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 gastric cancer phosphoproteome from diverse gastric cancer cells.

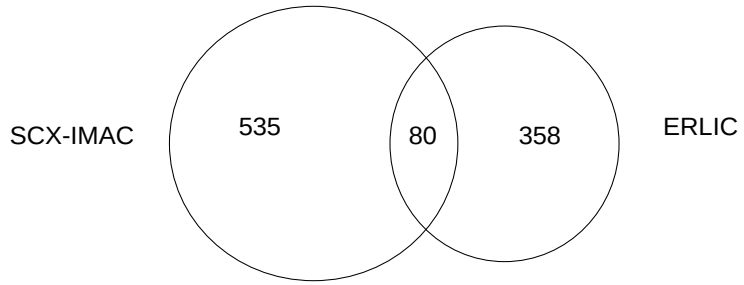


Figure 2.1. Comparison of SCX-IMAC and ERLIC in phosphoproteome coverage in SNU5 cells. Phosphoproteome of gastric cancer cell line, SNU5, was analyzed by SCX-IMAC and ERLIC. The number of non-redundant phosphopeptides identified by each method are shown in the Venn diagram.

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 [126] and OMSSA [127], 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 considerably enhance the sensitivity of shotgun proteomics [131]. In this study, we analyzed MS spectra using four different database search engines. False discovery rates (FDRs) were set at <1% for all searches (Table 2.1). The benefit of using multiple engines is illustrated in Figure 2.2 showing that spectra not identified in one engine *e.g.* Mascot, could be characterized by another *e.g.* Sequest.

Table 2.1. **Protein identification using four database search engines**. A total of 3600074 dta files were subject to target and decoy database search in Mascot, Sequest, X!Tandem and OMSSA. FDR for each search was set to <1%. The number of peptides, phosphopeptides and phosphorylation sites are shown.

Engine	Mascot	Sequest	X!Tandem	OMSSA
# Original peptide identifications	410565	742250	1062278	354978
Filtering criteria	pep expect<0.028	xc>1.8(+1), 2.3(+2), 2.8(+3 and above)	select only first hits; pepExp<0.03; ppm<4	select only first hits; MShits_ evalue<0.27
# of total target peptides	311677	140589	180325	65047
FDR	0.97%	0.92%	0.97%	0.92%
# of phosphopeptides	9939	7634	539	1679
# of unique phosphopeptides	1428	1135	294	341
# of phosphoproteins	718	441	210	227
# of unique phosphorylation sites	1576	1371	357	413
# of unique pS sites	1251	903	248	349
# of unique pT sites	259	346	88	58
# of unique pY sites	66	122	21	6

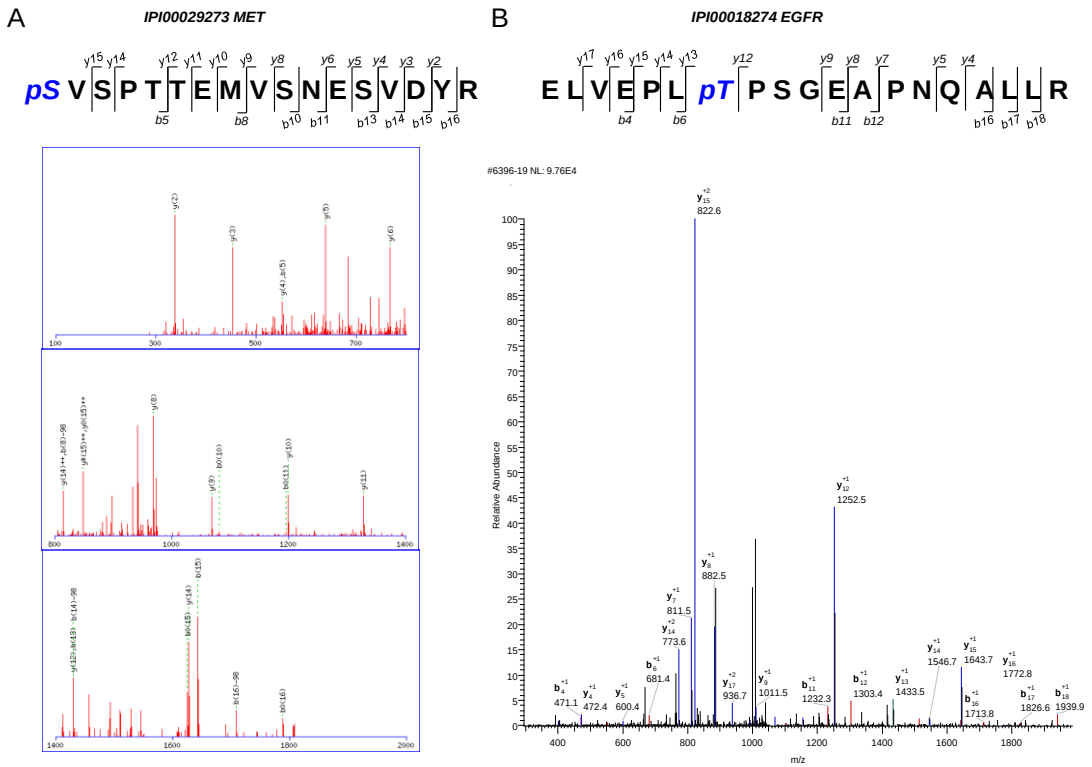


Figure 2.2. MS spectra interpreted by Mascot and Sequest. Annotated MS/MS spectra for peptides from (A) MET, identified via Mascot and (B) EGFR, identified via Sequest are shown. Spectrum annotated by Mascot is split to three parts according to mass range. Detected *b* ions and *y* ions are annotated.

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 Figure 2.3. 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 3021 unique phosphorylated peptides in 1211 phosphorylated proteins from gastric cancer cells. Among these, 547 (18%) phosphorylation sites and 295 (24%) phosphoproteins were identified by at least two search engines. Non-redundant phosphorylation sites comprised 2144 phosphorylated serines, 673 phosphorylated threonines and 204 phosphorylated tyrosines. The distribution of pS,

pT and pY was 71%, 22.3% and 6.8%, respectively. These results are consistent with other findings that in some proteins, phosphorylation sites with high occupancy are likely associated with serine while those of low occupancy involve threonine [132]. Compared to previous reports of global phosphoproteome profiling [53], 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.

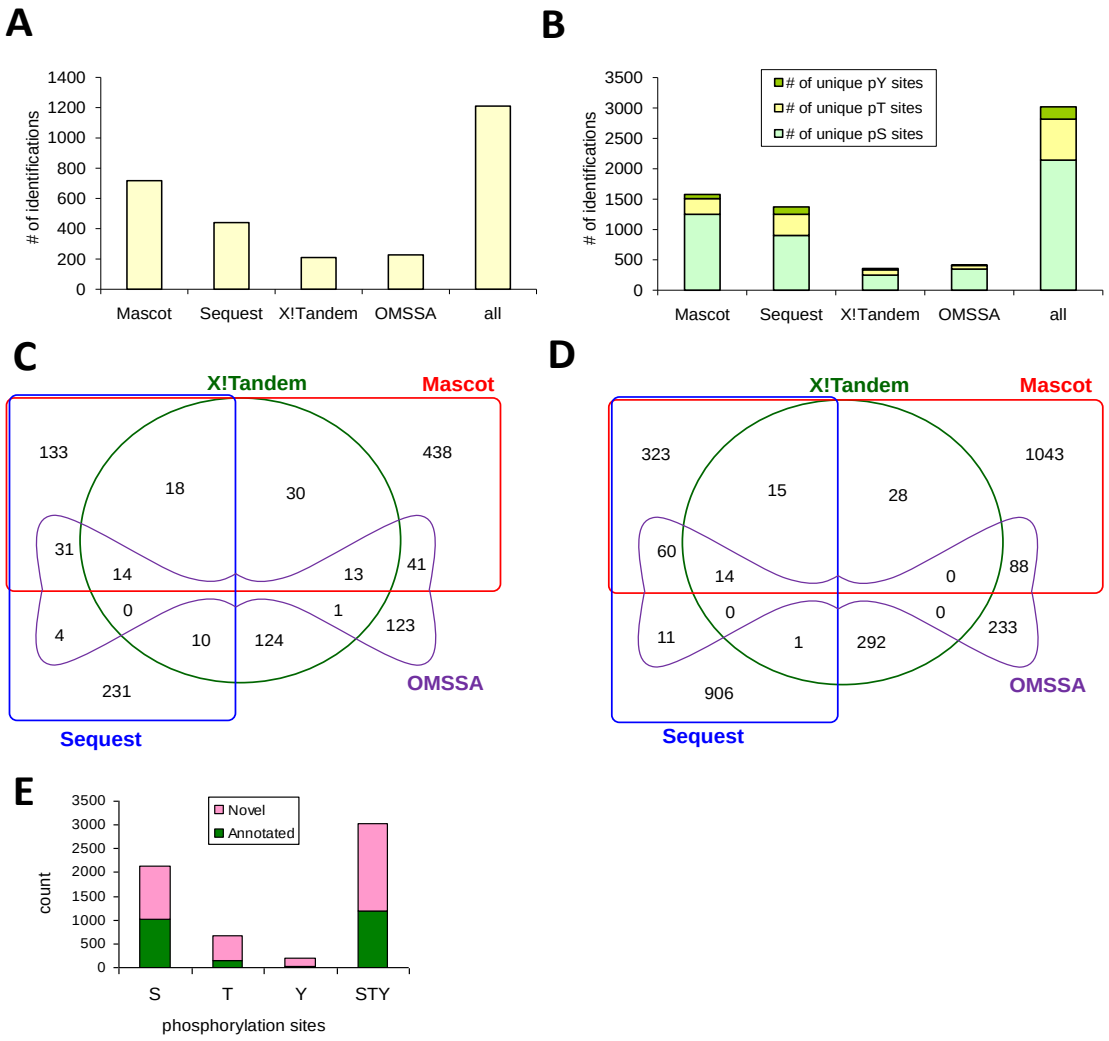


Figure 2.3 Identification and characterization of gastric cancer phosphoproteome. (A) Number of phosphorylated proteins identified from database search engines Mascot, Sequest, X!Tandem and OMSSA. Unique phosphoproteins identified by overlapping results from the four engines are shown. (B) Number of 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 1211 non-redundant phosphoproteins identified by four database search engines. (D) Venn diagram of 3021 non-redundant phosphorylation sites identified by four database search engines. (E) Summary of gastric cancer proteome 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 multiply phosphorylated peptides. To localize phosphorylation sites accurately, we first undertook a computational assessment of the phosphorylation site assignment using

the Ascore algorithm [133]. As shown in Figure 2.4, 64% of the localizations were assigned with >90% confidence ($p<0.01$) and 56% with >95% confidence ($p<0.05$). Near certainty (>99% confidence, $p<0.001$) of localization was achieved for 44% of the data set. It should be noted that 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 1056 to 915 (55% increment), 1117 (68% increment) and 1242 (75% increment), respectively. This indicated that the majority of phosphorylation assignments were of high confidence.

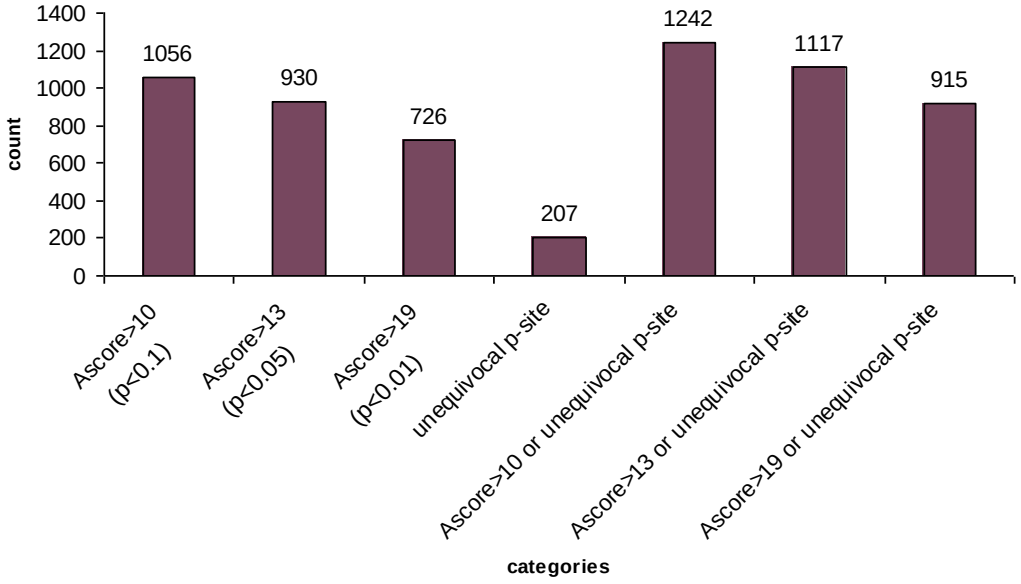


Figure 2.4. Confidence of phosphorylation site localization analyzed by Ascore. Counts of non-redundant phosphorylation sites with different Ascore cutoff values and/or unequivocal phosphorylation sites are shown.

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 Figure 2.5, 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 [53]. Phosphopeptides detected in MS were ionized with different charges, as shown in Figure 2.5.

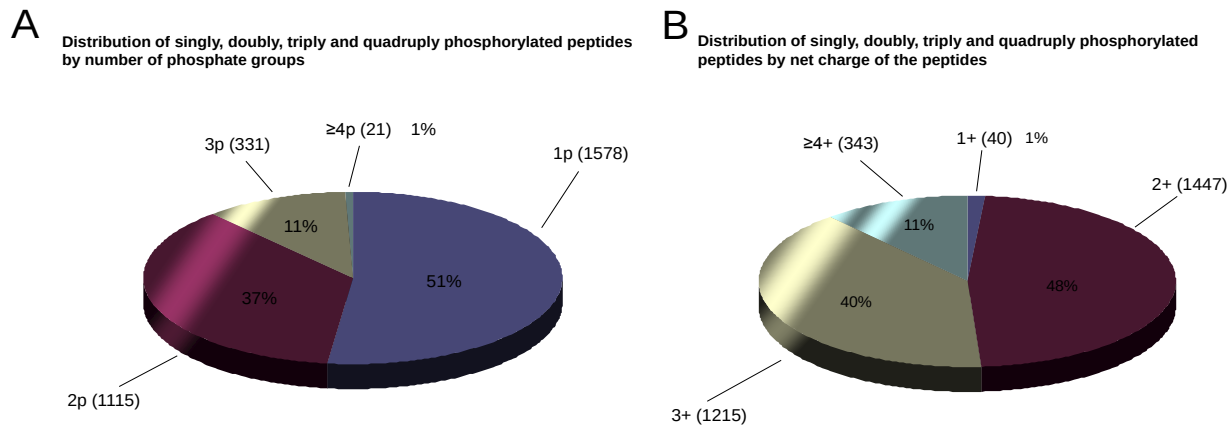


Figure 2.5. Distribution of gastric cancer phosphopeptides. (A) Distribution of singly (1p), doubly (2p), triply (3p) and quadruply (4p) phosphorylated peptides. (B) Distribution of singly, doubly, triply and quadruply charged phosphopeptides.

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 reference [134] (provided in link

http://www.springerlink.com/content/62ru85452070n021/supplementals/). In most cases, multiple spectra were interpreted as a single phosphopeptide sequence; only the spectrum with the highest identification score was manually inspected, and 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 3021 phosphorylation sites we identified, 1194 (40%) were annotated in the Uniprot database. Thus, our data revealed 1827 novel phosphorylation sites in gastric cancer (Figure 2.3E).

Subcellular localizations of gastric cancer phosphoproteins were annotated using Gene Ontology (Figure 2.6A). 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.

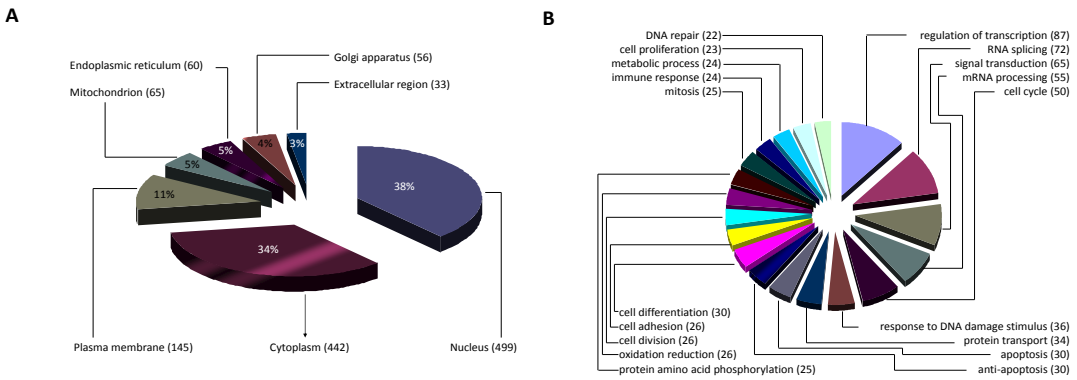


Figure 2.6. Classification of phosphoproteins based on Gene Ontology. Phosphorylated proteins in gastric cancer cells were classified according to subcellular localization (A) and biological process (B). Only biological processes with a hit number over 20 are shown.

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 (Figure 2.6B).

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 [135]. 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 [136]. Logo-like representations of the motifs are shown in Figure 2.7. Five acidic motifs associated with casein kinase 2 (CK2) and G

protein-coupled receptor kinase (GPCR kinase) were identified, and one basic motif 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 [136] using in-house programs. This revealed 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 (Figure 2.7G).

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 in reference [134], provided in link http://www.springerlink.com/content/62ru85452070n021/supplementals/)).

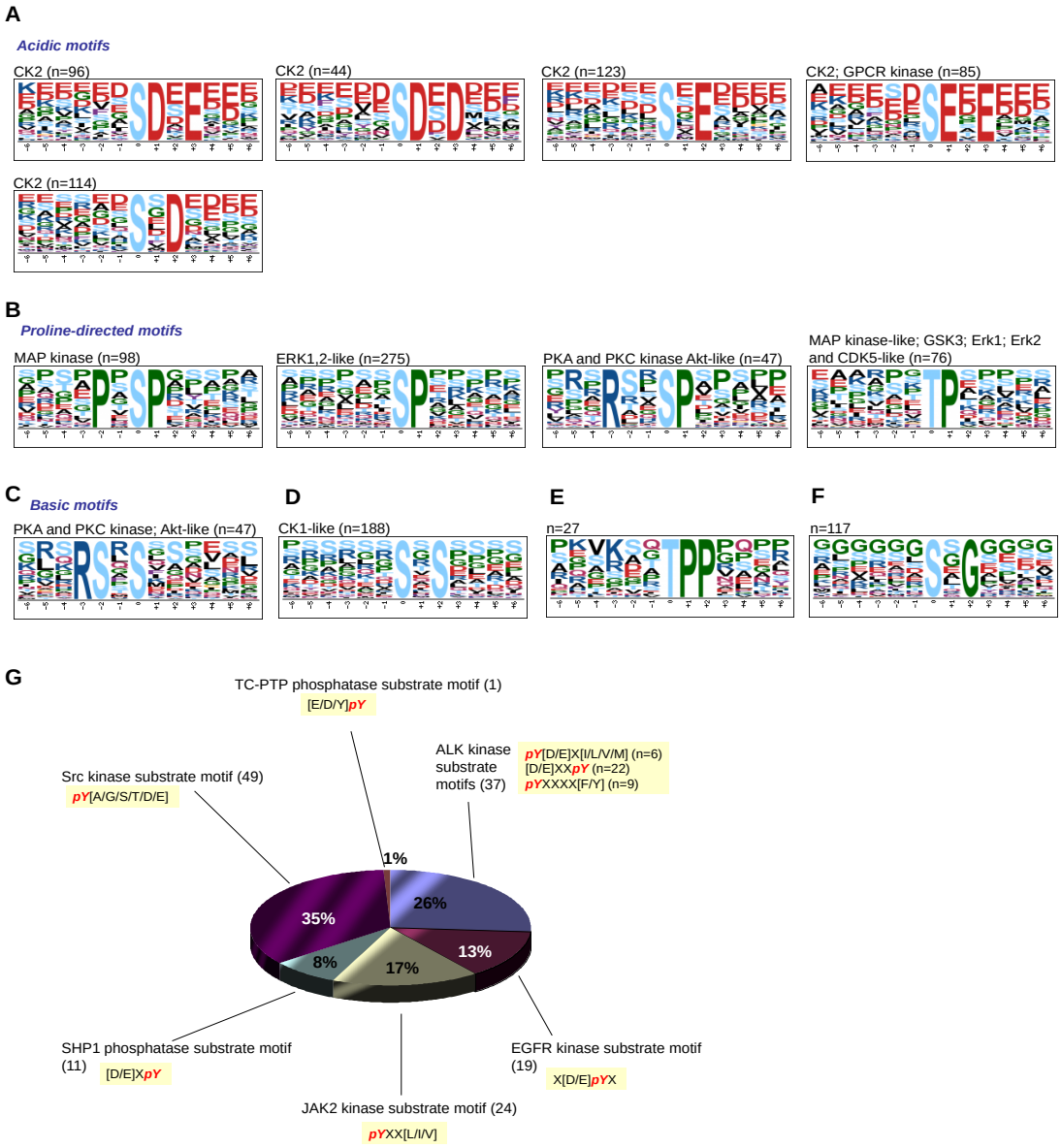


Figure 2.7. Motif analysis of gastric cancer phosphoproteome. Non-redundant 13-mer phosphorylated peptide sequences were analyzed in Motif-X. Motifs were classified according to annotations in Human Protein Reference Database. Logo-like representations of motifs are classified into acidic (A), proline-directed (B), basic (C) and others (D, E, 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).

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 2.2).

Table 2.2. Phosphorylated protein kinases in gastric cancer. List of protein kinases 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 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. # indicates the phosphorylation site was annotated previously in the literature. * indicates the protein kinase's role in gastric cancer was documented previously in the literature. § indicates the protein kinase's role in other cancers was documented previously in the literature. @ indicates the antibody detects both phosphorylation sites in the protein kinase.

protein/gene	protein kinase/ phosphatase family	Description	phosphorylation site	Identified By	Relative expression level in gastric cancer transcriptome	If annotated #	Documented in gastric cancer *	Documented in other cancers §
protein kinases								
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	Y273	MS	2.32	Yes	No	Yes

		1B						
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[@]	Antibody	0.79	Yes	Yes	Yes
ERK1/MAPK3	CMGC	mitogen-activated protein kinase 3	T202/Y204[@]	Antibody	0.79	Yes	Yes	Yes
ERK2/MAPK1	CMGC	mitogen-activated protein kinase 1	T185/Y187[@]	Antibody	1.46	Yes	Yes	Yes
ERK2/MAPK1	CMGC	mitogen-activated protein kinase 1	T202/Y204[@]	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[@]	Antibody	1.52	Yes	No	Yes
GSK3B	CMGC	glycogen synthase kinase 3 beta	S21/S9[@]	Antibody	1.48	Yes	No	Yes
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[@]	Antibody	1.28	Yes	Yes	Yes
JNK1/MAPK8	CMGC	mitogen-activated protein kinase 8	T221/Y223[@]	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-	S423	MS	1.28	Yes	Yes	Yes

		regulating kinase 2						
MEK1/MAP2K1	STE	mitogen-activated protein kinase kinase 1	S218/S222[@]	Antibody	1.64	Yes	Yes	Yes
MEK1/MAP2K1	STE	mitogen-activated protein kinase kinase 1	S222/S226[@]	Antibody	1.64	Yes	Yes	Yes
MEK2/MAP2K2	STE	mitogen-activated protein kinase kinase 2	S218/S222[@]	Antibody	1.64	Yes	Yes	Yes
MEK2/MAP2K2	STE	mitogen-activated protein kinase kinase 2	S222/S226[@]	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
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, 90kDa, polypeptide 5	S376/S360[@]	Antibody	0.29	Yes	No	Yes
MSK2/RPS6KA4	AGC	ribosomal protein S6 kinase,	S376/S36	Antibody	2.27	Yes	No	Yes

		90kDa, polypeptide 4	0[@]					
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[@]	Antibody	2.37	Yes	No	Yes
p70S6K/RPS6KB1	AGC	ribosomal protein S6 kinase, 70kDa, polypeptide 1	T229	Antibody	1.02	Yes	Yes	Yes
p70S6K/RPS6KB1	AGC	ribosomal protein S6 kinase, 70kDa, polypeptide 1	T389	Antibody	1.02	Yes	Yes	Yes
p70S6K/RPS6KB1	AGC	ribosomal protein S6 kinase, 70kDa, polypeptide 1	T421/S424[@]	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, 90kDa, polypeptide 1	S221	Antibody	1.06	Yes	No	Yes
RSK1/RPS6KA1	AGC	ribosomal protein S6 kinase, 90kDa, polypeptide 1	S380	Antibody	1.06	Yes	No	Yes
RSK2/RPS6KA3	AGC	ribosomal protein S6 kinase, 90kDa, polypeptide 3	S221	Antibody	2.46	Yes	No	Yes
RSK2/RPS6KA3	AGC	ribosomal protein S6 kinase, 90kDa, polypeptide 3	S380	Antibody	2.46	Yes	No	Yes
RSK3/RPS6KA2	AGC	ribosomal protein S6 kinase, 90kDa, polypeptide 2	S380	Antibody	0.39	Yes	No	Yes
SMG1	Atypical	phosphatidylinositol 3-kinase-related protein kinase	S2940	MS	0.93	No	No	Yes
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 phosphatases								
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	S171	MS	2.89	No	Yes	Yes

		phosphatase alpha						
PTPRA	PTP	cDNA FLJ56484, highly similar to Receptor-type tyrosine-protein phosphatase alpha	S172	MS	2.89	No	Yes	Yes
PTPRA	PTP	cDNA FLJ56484, highly similar to Receptor-type tyrosine-protein phosphatase alpha	T161	MS	2.89	No	Yes	Yes
MTMR7	DSP	Isoform 1 of Myotubularin-related protein 7	S213	MS	0.86	No	No	No

We also found ten phosphorylation sites in five protein phosphatases (phosphorylation sites in parenthesis), *i.e.* PTPN12 (S435), PTPN14 (S312, S314), PTPRA (S171, S172, T161), PTPRF (T1801, T1811, T1825) and MTMR7 (S213) in gastric cancer cells (Table 2.2). 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 over-expressed 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. These data showed over-expression 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 over-expressed 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 2.2) confirmed EGFR, MET and CDKs as over-expressed 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 activated 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 2 histologically normal antral biopsies, 7 cases of benign gastritis, and 3 pairs of gastric adenocarcinoma (2 intestinal histotype and 1 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 (Figure 2.8). Of these, 27 were phosphorylated protein kinases (Table 2.2).

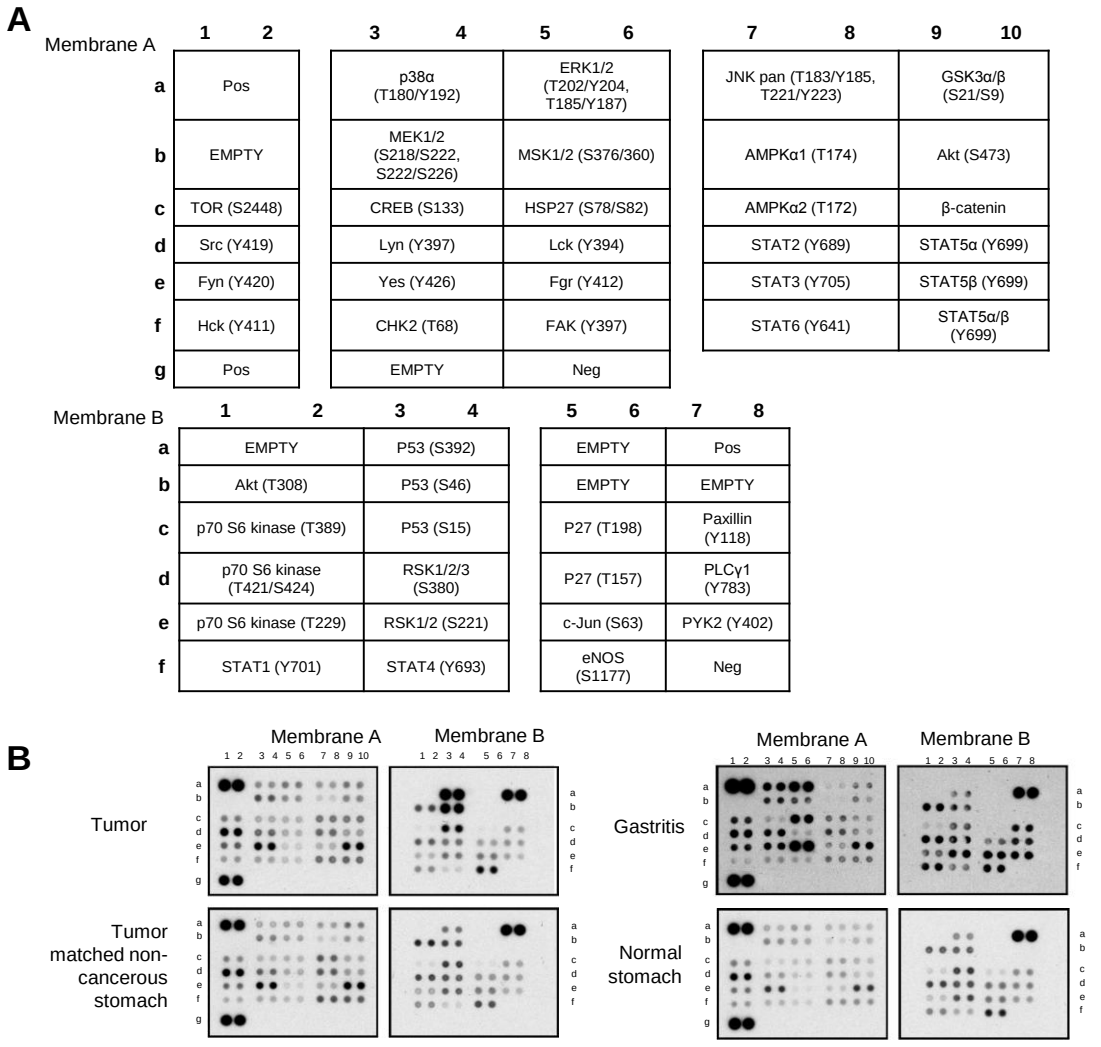


Figure 2.8. Representative antibody array images of primary gastric tissues. Proteome Profiler™ Human Phospho-Kinase Array Kit (R&D Systems) was used to simultaneously detect phosphorylation sites in a panel of protein kinases and key signaling proteins in fresh frozen primary endoscopic gastric tissues 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 Figure 2.9 for all images.

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).

Figure 2.9. Protein antibody array analysis of primary stomach tissues. R&D Systems Proteome Profiler™ Human Phospho-Kinase Array Kit was used to simultaneously detect phosphorylation sites in a variety of protein kinases and key signaling proteins in primary endoscopic gastric tissues from 12 human subjects: 2 histologically normal antral biopsies, 3 pairs of gastric adenocarcinoma with their cognate matched normal mucosa, and 7 cases of histologically benign gastritis. (A) Layout of protein antibody array. The array is composed of membrane A and membrane B. (B) Images of antibody arrays generated by each biopsy of primary stomach tissues.

A

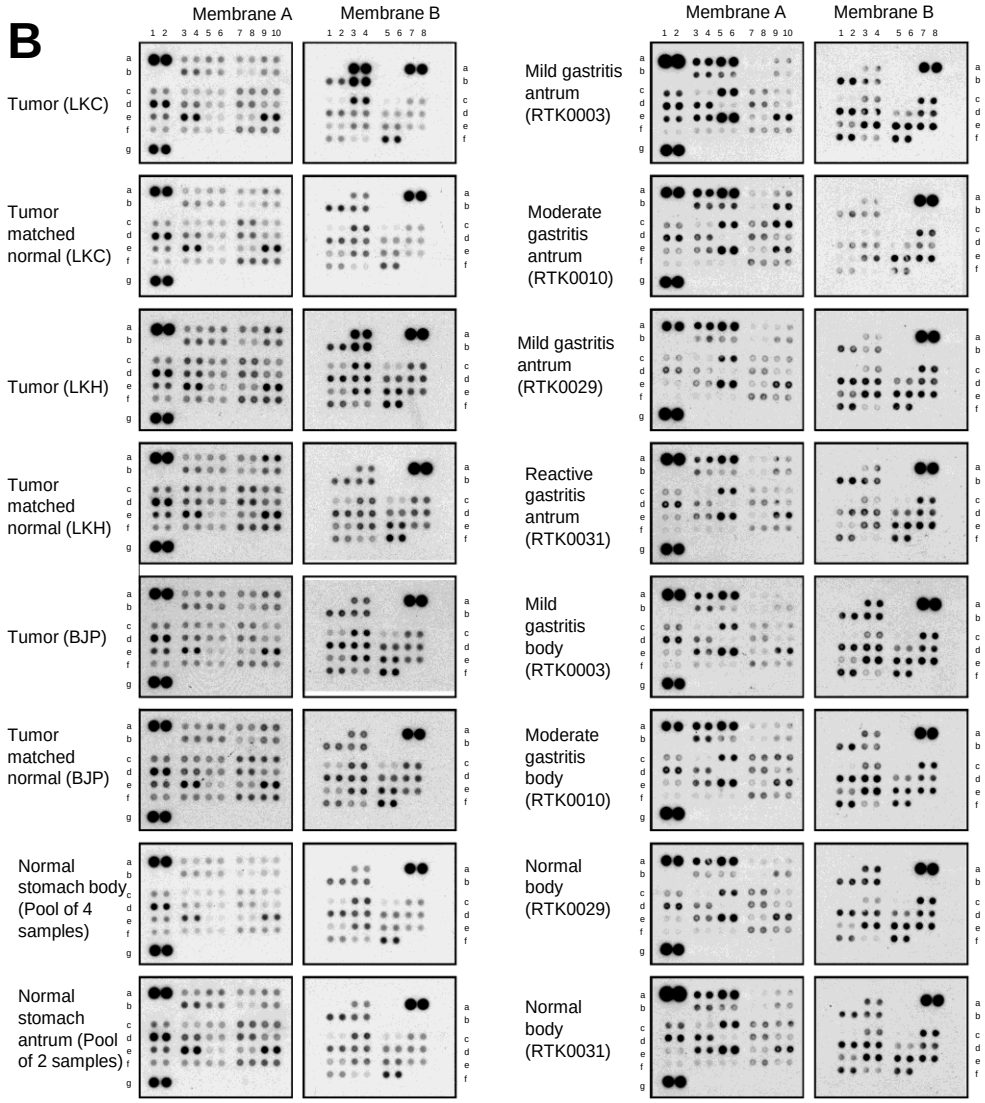
Membrane A

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

Membrane B

	1	2	3	4	5	6	7	8
a	EMPTY	EMPTY	p53 (S392)	p53 (S392)	EMPTY	EMPTY	Pos	Pos
b	Akt (T308)	Akt (T308)	p53 (S46)	p53 (S46)	EMPTY	EMPTY	EMPTY	EMPTY
c	p70 S6 kinase (T389)	p70 S6 kinase (T389)	p53 (S15)	p53 (S15)	p27 (T198)	p27 (T198)	Paxillin (Y118)	Paxillin (Y118)
d	p70 S6 kinase (T421/S424)	p70 S6 kinase (T421/S424)	RSK1/2/3 (S380)	RSK1/2/3 (S380)	p27 (T157)	p27 (T157)	PLCγ1 (Y783)	PLCγ1 (Y783)
e	p70 S6 kinase (T229)	p70 S6 kinase (T229)	RSK1/2 (S221)	RSK1/2 (S221)	c-Jun (S63)	c-Jun (S63)	PYK2 (Y402)	PYK2 (Y402)
f	STAT1 (Y701)	STAT1 (Y701)	STAT4 (Y693)	STAT4 (Y693)	eNOS (S1177)	eNOS (S1177)	Neg	Neg

B



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 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, 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 phosphoproteome patterns, consistent with field cancerization in this disease [137] (Figure 2.9 and Figure 2.10).

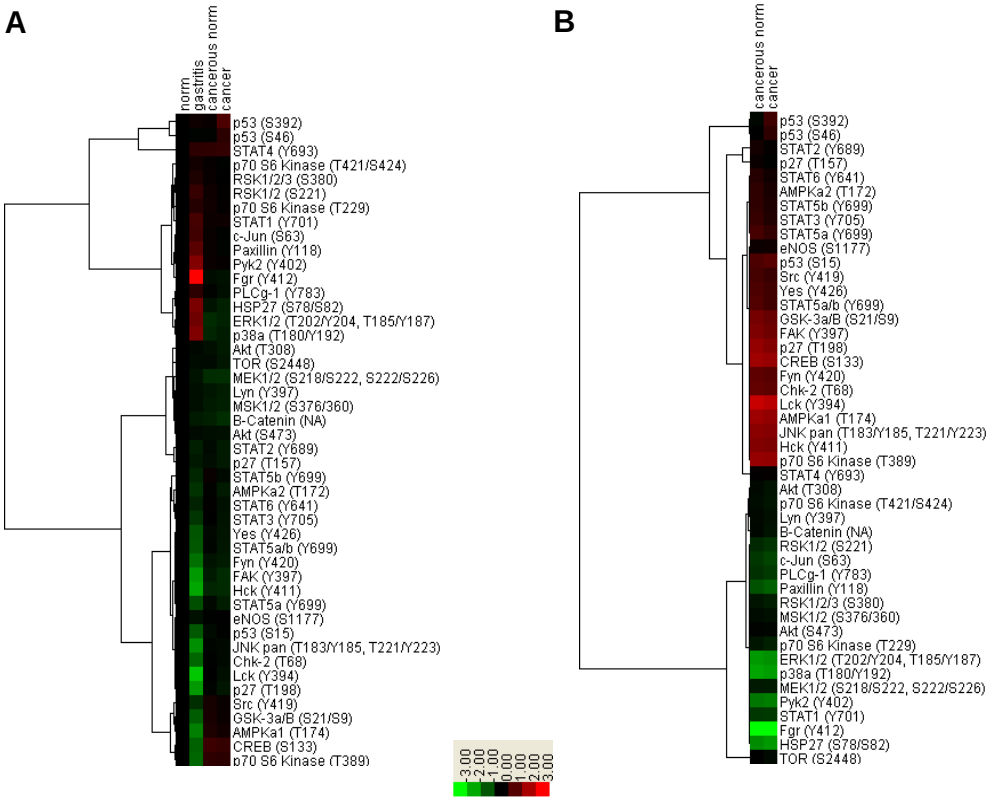


Figure 2.10. Phosphorylated signaling molecules determined by antibody array analysis of primary gastric tissues. One pooled sample of histologically normal stomach tissue from two individuals, 7 cases of histologically benign gastritis, as well as 3 pairs of gastric adenocarcinoma and matched normal 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.

LC-MS/MS-based and antibody array-based phosphoproteomics analysis jointly identified 74 phosphorylation sites in 41 protein kinases in gastric cancer cell lines and primary stomach tissues (Table 2.2). Eighteen of these phosphorylation sites (24%) are novel. Literature mining revealed that 37 of the 41 identified protein kinases (90%) have been implicated in different cancers, whereas only 19 (46%) have been associated with gastric cancer.

Discussion

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

Phosphoproteins are the most pervasive signaling molecules, whereas many over-expressed 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 phosphoproteomics 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 9 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 investigations are required.

It is noteworthy that the overlap between phosphoproteome from LC-MS/MS and phosphoproteome from protein antibody array is negligible. One reason is that the commercially available protein antibody array for probing phosphoproteins contains only 40 phosphoproteins. In addition, most of these proteins are low abundance signaling proteins that are rarely identified by LC-MS/MS approaches due to 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 which were >2-fold transcriptionally over-expressed, we identified 190 dysregulated phosphoproteins. 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 [108, 109] while our phosphoproteomics data set identified the presence of MET in its phosphorylated and active state. Selective inhibition of MET effectively is known to kill MET over-expressing gastric cancer cells [24, 112] and is the rationale for ongoing clinical trials of MET inhibitors for gastric cancer therapy. Our data also showed over-expression 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 [138], TOP2A [87], minichromosome maintenance 2 (MCM2) [87], erythropoietin-producing hepatocellular (Eph) A2 receptor [139], CTNNB1 [87], and hepatoma-derived growth factor (HDGF) [87]. 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 although some have been associated with other types of cancers. For instance, SLC7A5/CD98 is a proposed prognostic indicator of adult acute leukemia [140], breast cancer [87], lung cancer [141] and renal cancer [142]. 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 over-expressed phosphoproteins. DNA damage in the absence of physiological repair responses is the origin of many diseases, including cancers [143]. 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, ATM (ataxia telangiectasia mutated) 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 insulin-IGF-1-PI3K-AKT pathway [144]. As shown in Figure 2.11, our data sets identified over-expression 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 (Figure 2.9).

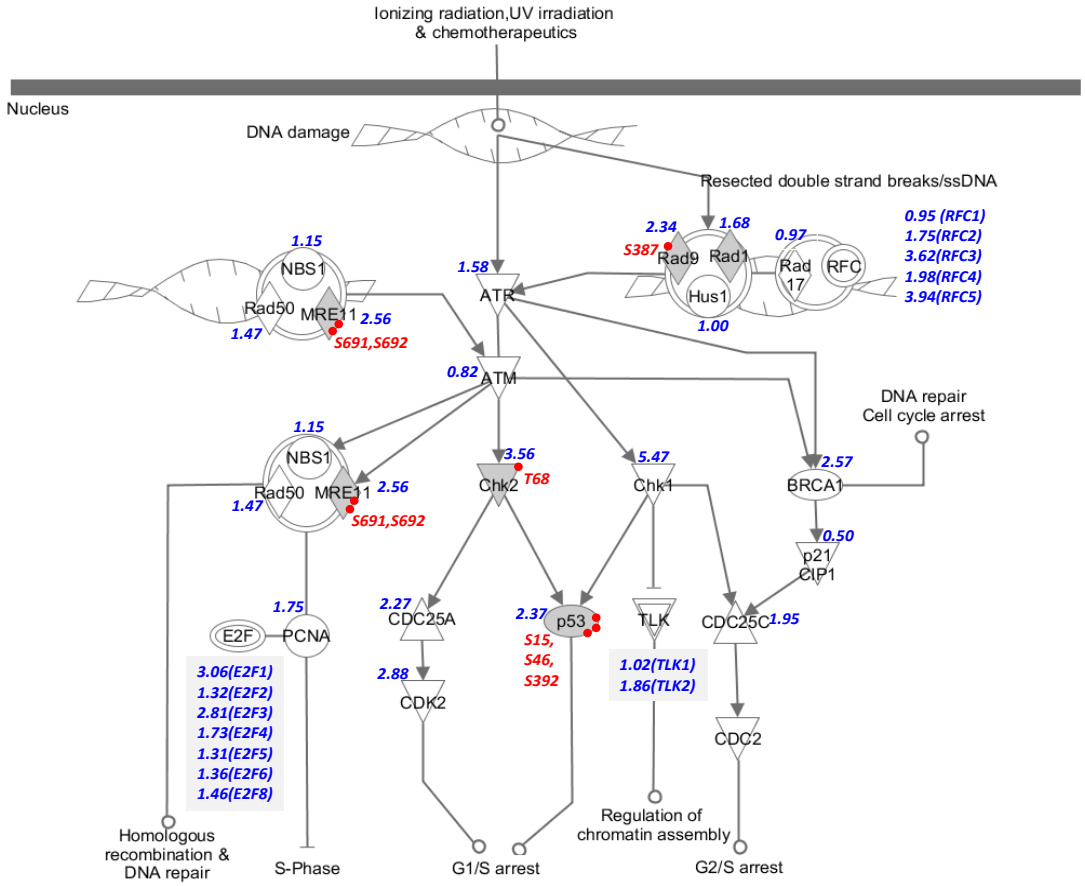


Figure 2.11. **DNA damage response pathway in gastric cancer.** Pathway is modified based on cell cycle checkpoint control pathway from Ingenuity Pathway Analysis. Over-expressed phosphoproteins are shaded in grey. Relative mRNA expression level of proteins is shown in blue. Phosphorylation sites are 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 [145]. Protein phosphatases play equally critical roles in setting the levels of protein phosphorylation in cells and in regulating many physiological processes [146]. 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) [146, 147]. Like protein kinases, phosphorylation of protein phosphatases is an important regulatory mechanism [146]. 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 2.2) from which key pathways in gastric oncogenesis may be discerned.

It is also worth noting that 30 over-expressed phosphoproteins (16%) were associated with mitochondria, implying critical roles for this organelle in gastric oncogenesis (Table 2.3). Mitochondria are pivotal in cell metabolism, survival and apoptosis. Several protein kinases and protein tyrosine phosphatases are known to reside in mitochondria while other mitochondrial proteins are themselves kinase substrates. As well as being the target of all major kinase signaling pathways, intramitochondrial signaling also occurs [148, 149]. Mitochondrial phosphoproteomes of mammalian cardiomyocytes [150], hepatocytes [151], pancreatic beta-cells[152], yeast[153] and *Arabidopsis thaliana* [154] have been reported. However, there is yet no systematic documentation of mitochondrial phosphoproteins in cancer cells. Our data demonstrated that TOMM20 (translocase of outer mitochondrial membrane 20) was over-expressed 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.

Table 2.3. Phosphorylated mitochondrial proteins in gastric cancer cells.

IPI	gene	description	site	Mascot	Omssa	Tandem	Sequest	If Annotated
IPI00000690	AIFM1	Isoform 1 of Apoptosis-inducing factor 1, mitochondrial	S56	N	N	Y	N	N
IPI00000858	RDBP	cDNA FLJ56180, highly similar to Negative elongation factor E	S58	Y	N	N	Y	N
IPI00003985	BCS1L	Mitochondrial chaperone BCS1	T308	N	N	N	Y	N
IPI00004902	ETFB	Isoform 1 of Electron transfer flavoprotein subunit beta	T172	N	N	N	Y	N
IPI00007001	MRPL11	39S ribosomal protein L11, mitochondrial	S88	N	N	N	Y	N
IPI00007001	MRPL11	39S ribosomal protein L11, mitochondrial	T86	N	N	N	Y	N
IPI00007001	MRPL11	39S ribosomal protein L11, mitochondrial	Y89	N	N	N	Y	N
IPI00007188	SLC25A5	ADP/ATP translocase 2	S127	N	N	Y	N	N
IPI00007979	MT-ND2	NADH-ubiquinone oxidoreductase chain 2	S301	N	N	N	Y	N
IPI00007979	MT-ND2	NADH-ubiquinone oxidoreductase chain 2	T20	N	N	N	Y	N
IPI00007979	MT-ND2	NADH-ubiquinone oxidoreductase chain 2	T45	N	N	N	Y	N
IPI00007979	MT-ND2	NADH-ubiquinone oxidoreductase chain 2	Y298	N	N	N	Y	N
IPI00008528	MT-ATP6	ATP synthase protein 8	T11	N	N	N	Y	N
IPI00008528	MT-ATP6	ATP synthase protein 8	T24	N	N	N	Y	N
IPI00008528	MT-ATP6	ATP synthase protein 8	T6	N	N	N	Y	N
IPI00008528	MT-ATP6	ATP synthase protein 8	T7	N	N	N	Y	N
IPI00009444	MRPL27	39S ribosomal protein L27, mitochondrial	T69	Y	N	N	N	N
IPI00009960	IMMT	Isoform 1 of Mitochondrial inner membrane protein	S746	N	N	N	Y	N
IPI00009960	IMMT	Isoform 1 of Mitochondrial inner membrane protein	T42	N	N	Y	N	N

IPI00009960	IMMT	Isoform 1 of Mitochondrial inner membrane protein	T43	N	N	Y	N	N
IPI00009960	IMMT	Isoform 1 of Mitochondrial inner membrane protein	T731	N	N	N	Y	N
IPI00009960	IMMT	Isoform 1 of Mitochondrial inner membrane protein	T753	N	N	N	Y	N
IPI00010906	CLN3	CLN3 protein	S12	Y	N	N	N	N
IPI00010906	CLN3	CLN3 protein	S14	Y	N	N	N	N
IPI00010906	CLN3	CLN3 protein	T19	Y	N	N	N	N
IPI00011062	CPS1	Isoform 1 of Carbamoyl-phosphate synthase [ammonia], mitochondrial	T1078	N	N	N	Y	N
IPI00011307	MTHFD2	Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, mitochondrial	S252	N	N	N	Y	N
IPI00011307	MTHFD2	Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, mitochondrial	T244	N	N	N	Y	N
IPI00011307	MTHFD2	Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, mitochondrial	T261	N	N	N	Y	N
IPI00011307	MTHFD2	Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, mitochondrial	T324	N	N	N	Y	N
IPI00011307	MTHFD2	Bifunctional methylenetetrahydrofolate dehydrogenase/cyclohydrolase, mitochondrial	Y304	N	N	N	Y	N
IPI00011635	BCL2L13	Isoform 2 of Bcl-2-like protein 13	S27	N	N	N	Y	N
IPI00011635	BCL2L13	Isoform 2 of Bcl-2-like protein 13	T15	N	N	N	Y	N
IPI00011635	BCL2L13	Isoform 2 of Bcl-2-like protein 13	Y13	N	N	N	Y	N
IPI00012728	ACSL1	Isoform 1 of Long-chain-fatty-acid--CoA ligase 1	T41	N	N	N	Y	N
IPI00012728	ACSL1	Isoform 1 of Long-chain-fatty-acid--CoA ligase 1	T46	N	N	N	Y	N
IPI00013146	MRPS22	28S ribosomal protein S22, mitochondrial	S77	N	N	N	Y	N
IPI00013146	MRPS22	28S ribosomal protein S22, mitochondrial	T90	N	N	N	Y	N
IPI00013146	MRPS22	28S ribosomal protein S22, mitochondrial	Y196	N	N	N	Y	N
IPI00013146	MRPS22	28S ribosomal protein S22, mitochondrial	Y250	N	N	N	Y	N
IPI00013623	SLC27A3	Isoform 1 of Long-chain fatty acid transport protein 3	S110	N	N	N	Y	N
IPI00013623	SLC27A3	Isoform 1 of Long-chain fatty acid transport protein 3	S128	N	N	N	Y	N

IPI00015833	CHCHD3	Coiled-coil-helix-coiled-coil-helix domain-containing protein 3, mitochondrial	S39	N	N	N	Y	N
IPI00018120	DAP3	28S ribosomal protein S29, mitochondrial	S252	N	N	N	Y	N
IPI00021016	TSFM	Isoform 1 of Elongation factor Ts, mitochondrial	S181	N	N	N	Y	N
IPI00021338	DLAT	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	S98	N	N	N	Y	N
IPI00021338	DLAT	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	T102	N	N	N	Y	N
IPI00021338	DLAT	Dihydrolipoyllysine-residue acetyltransferase component of pyruvate dehydrogenase complex, mitochondrial	T107	N	N	N	Y	N
IPI00022002	MRPS27	cDNA FLJ54536, highly similar to Mitochondrial 28S ribosomal protein S27	S232	N	N	N	Y	N
IPI00022002	MRPS27	cDNA FLJ54536, highly similar to Mitochondrial 28S ribosomal protein S27	S236	N	N	N	Y	N
IPI00022002	MRPS27	cDNA FLJ54536, highly similar to Mitochondrial 28S ribosomal protein S27	S237	N	N	N	Y	N
IPI00022002	MRPS27	cDNA FLJ54536, highly similar to Mitochondrial 28S ribosomal protein S27	S68	N	N	N	Y	N
IPI00022002	MRPS27	cDNA FLJ54536, highly similar to Mitochondrial 28S ribosomal protein S27	S69	N	N	N	Y	N
IPI00022002	MRPS27	cDNA FLJ54536, highly similar to Mitochondrial 28S ribosomal protein S27	Y242	N	N	N	Y	N
IPI00022585	AKAP1	cDNA FLJ56047, highly similar to A kinase anchor protein 1, mitochondrial	S76	N	N	N	Y	N
IPI00024145	VDAC2	Isoform 2 of Voltage-dependent anion-selective channel protein 2	S35	Y	N	N	N	N
IPI00024145	VDAC2	Isoform 2 of Voltage-dependent anion-selective channel protein 2	S37	Y	N	N	N	N
IPI00024145	VDAC2	Isoform 2 of Voltage-dependent anion-selective channel protein 2	S42	Y	N	N	N	N
IPI00024145	VDAC2	Isoform 2 of Voltage-dependent anion-selective channel protein 2	S44	Y	N	N	N	N
IPI00024145	VDAC2	Isoform 2 of Voltage-dependent anion-selective channel protein 2	T43	Y	N	N	N	N
IPI00024145	VDAC2	Isoform 2 of Voltage-dependent anion-selective channel protein 2	T57	Y	N	N	N	N
IPI00024145	VDAC2	Isoform 2 of Voltage-dependent anion-selective channel protein 2	T60	Y	N	N	N	N

IPI00026958	FDXR	Isoform Short of NADPH:adrenodoxin oxidoreductase, mitochondrial	S414	N	N	N	Y	N
IPI00026958	FDXR	Isoform Short of NADPH:adrenodoxin oxidoreductase, mitochondrial	S432	N	N	N	Y	N
IPI00026958	FDXR	Isoform Short of NADPH:adrenodoxin oxidoreductase, mitochondrial	T404	N	N	N	Y	N
IPI00028077	PSEN1	Isoform 1 of Presenilin-1	S366	Y	N	N	Y	N
IPI00076042	HSPD1	Short heat shock protein 60 Hsp60s2	S222	N	N	N	Y	N
IPI00076042	HSPD1	Short heat shock protein 60 Hsp60s2	T225	N	N	N	Y	N
IPI00076042	HSPD1	Short heat shock protein 60 Hsp60s2	T232	N	N	N	Y	N
IPI00096066	SUCLG2	Succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial	S408	N	N	N	Y	N
IPI00096066	SUCLG2	Succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial	T33	N	N	N	Y	N
IPI00096066	SUCLG2	Succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial	T396	N	N	N	Y	N
IPI00096066	SUCLG2	Succinyl-CoA ligase [GDP-forming] subunit beta, mitochondrial	T413	N	N	N	Y	N
IPI00170877	MRPL10	cDNA FLJ45232 fis, clone BRCAN2021718, highly similar to Homo sapiens mitochondrial ribosomal protein L10 (MRPL10), mRNA	Y63	N	N	N	Y	N
IPI00216932	ACSS1	Isoform 1 of Acetyl-coenzyme A synthetase 2-like, mitochondrial	S62	N	N	N	Y	N
IPI00216932	ACSS1	Isoform 1 of Acetyl-coenzyme A synthetase 2-like, mitochondrial	Y58	N	N	N	Y	N
IPI00218342	MTHFD1	C-1-tetrahydrofolate synthase, cytoplasmic	T712	N	N	Y	N	N
IPI00218342	MTHFD1	C-1-tetrahydrofolate synthase, cytoplasmic	T714	N	N	Y	N	N
IPI00219613	PITRM1	cDNA FLJ10321 fis, clone NT2RM2000504, highly similar to Homo sapiens pitrilysin metallopeptidase 1 (PITRM1), mRNA	T289	N	N	N	Y	N
IPI00219613	PITRM1	cDNA FLJ10321 fis, clone NT2RM2000504, highly similar to Homo sapiens pitrilysin metallopeptidase 1 (PITRM1), mRNA	T430	N	N	N	Y	N
IPI00219613	PITRM1	cDNA FLJ10321 fis, clone NT2RM2000504, highly similar to Homo sapiens pitrilysin metallopeptidase 1 (PITRM1), mRNA	T434	N	N	N	Y	N
IPI00219613	PITRM1	cDNA FLJ10321 fis, clone NT2RM2000504, highly similar to Homo sapiens pitrilysin metallopeptidase 1 (PITRM1), mRNA	Y433	N	N	N	Y	N
IPI00294398	HADH	Isoform 1 of Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	S103	N	N	N	Y	N
IPI00294398	HADH	Isoform 1 of Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial	T102	N	N	N	Y	N

IPI00296053	FH	Isoform Mitochondrial of Fumarate hydratase, mitochondrial	Y2	Y	N	N	N	N
IPI00303568	PTGES2	Prostaglandin E synthase 2	S55	N	N	Y	N	N
IPI00303722	FAM136A	Protein FAM136A	S43	N	N	Y	N	N
IPI00304435	NIPSNAP1	Protein NipSnap homolog 1	Y185	N	N	N	Y	N
IPI00304925	HSPA1A	Heat shock 70 kDa protein 1A/1B	S633	Y	N	N	Y	N
IPI00304925	HSPA1A	Heat shock 70 kDa protein 1A/1B	T636	Y	N	N	N	N
IPI00304925	HSPA1A	Heat shock 70 kDa protein 1A/1B	Y525	N	N	N	Y	N
IPI00306301	PDHA1	Mitochondrial PDHA1	S270	Y	N	N	Y	N
IPI00306301	PDHA1	Mitochondrial PDHA1	T269	N	N	N	Y	N
IPI00337541	NNT	NAD(P) transhydrogenase, mitochondrial	Y722	N	N	Y	N	N
IPI00383309	WHSC1L1	Putative uncharacterized protein pp14328	T6	Y	N	N	N	N
IPI00383309	WHSC1L1	Putative uncharacterized protein pp14328	T8	Y	N	N	N	N
IPI00440493	ATP5A1	ATP synthase subunit alpha, mitochondrial	T432	N	N	N	Y	N
IPI00465436	CAT	Catalase	T107	N	N	Y	N	N
IPI00604664	NDUFS1	NADH-ubiquinone oxidoreductase 75 kDa subunit	S493	Y	N	N	N	N
IPI00645805	IVD	Isovaleryl-CoA dehydrogenase, mitochondrial	S155	N	N	N	Y	N
IPI00645805	IVD	Isovaleryl-CoA dehydrogenase, mitochondrial	S165	N	N	N	Y	N
IPI00645805	IVD	Isovaleryl-CoA dehydrogenase, mitochondrial	Y158	N	N	N	Y	N
IPI00657692	COL4A3BP	alpha 3 type IV collagen binding protein isoform 3	S59	N	N	Y	N	N
IPI00790834	ACAD9	Acyl-Coenzyme A dehydrogenase family, member 9, isoform CRA_b	S14	N	N	N	Y	N
IPI00790834	ACAD9	Acyl-Coenzyme A dehydrogenase family, member 9, isoform CRA_b	Y2	N	N	N	Y	N
IPI00792673		24 kDa protein	T43	N	N	Y	N	N
IPI00853430	PRODH	Putative uncharacterized protein PRODH	S15	N	N	Y	N	N
IPI00908510	CDS2	Phosphatidate cytidylyltransferase	S5	Y	N	N	N	N
IPI00005792	PABPN1	Isoform 1 of Polyadenylate-binding protein 2	S52	N	N	Y	N	Y
IPI00011857	CHAF1B	Chromatin assembly factor 1 subunit B	S429	Y	N	N	N	Y

IPI00011857	CHAF1B	Chromatin assembly factor 1 subunit B	T433	Y	N	N	N	Y
IPI00016676	TOMM20	Mitochondrial import receptor subunit TOM20 homolog	S135	Y	N	N	N	Y
IPI00016676	TOMM20	Mitochondrial import receptor subunit TOM20 homolog	S138	Y	N	N	N	Y
IPI00018120	DAP3	28S ribosomal protein S29, mitochondrial	T272	N	N	N	Y	Y
IPI00022613	NOP14	Isoform 1 of Nucleolar protein 14	S96	Y	N	N	N	Y
IPI00023780	DNAJC5	Isoform 2 of DnaJ homolog subfamily C member 5	S10	Y	Y	N	Y	Y
IPI00024087	PDHA2	Pyruvate dehydrogenase E1 component subunit alpha, testis-specific form, mitochondrial	S291	Y	N	N	N	Y
IPI00024087	PDHA2	Pyruvate dehydrogenase E1 component subunit alpha, testis-specific form, mitochondrial	S298	Y	N	N	N	Y
IPI00024087	PDHA2	Pyruvate dehydrogenase E1 component subunit alpha, testis-specific form, mitochondrial	Y287	Y	N	N	N	Y
IPI00024087	PDHA2	Pyruvate dehydrogenase E1 component subunit alpha, testis-specific form, mitochondrial	Y299	Y	N	N	N	Y
IPI00024976	TOMM22	Mitochondrial import receptor subunit TOM22 homolog	S15	N	N	Y	N	Y
IPI00028077	PSEN1	Isoform 1 of Presenilin-1	S365	Y	N	N	Y	Y
IPI00028077	PSEN1	Isoform 1 of Presenilin-1	S367	N	N	N	Y	Y
IPI00032872	MRPS16	28S ribosomal protein S16, mitochondrial	T130	Y	N	Y	N	Y
IPI00166807	OXR1	Isoform 3 of Oxidation resistance protein 1	S112	Y	N	N	N	Y
IPI00166807	OXR1	Isoform 3 of Oxidation resistance protein 1	S113	Y	N	N	N	Y
IPI00166807	OXR1	Isoform 3 of Oxidation resistance protein 1	S115	Y	N	N	N	Y
IPI00171176	PANK2	Isoform 1 of Pantothenate kinase 2, mitochondrial	S189	Y	N	N	N	Y
IPI00171769	FUNDC2	FUN14 domain-containing protein 2	S151	Y	N	N	N	Y
IPI00257508	DPYSL2	Dihydropyrimidinase-related protein 2	S522	Y	N	N	N	Y
IPI00257508	DPYSL2	Dihydropyrimidinase-related protein 2	T509	Y	N	N	N	Y
IPI00257508	DPYSL2	Dihydropyrimidinase-related protein 2	T521	Y	N	N	N	Y
IPI00304925	HSPA1A	Heat shock 70 kDa protein 1A/1B	S631	N	N	N	Y	Y
IPI00334190	STOML2	Stomatin-like protein 2	T327	Y	N	N	N	Y

IPI00402231	DNAJC5	Isoform 1 of DnaJ homolog subfamily C member 5	S10	N	Y	N	N	Y
IPI00410079	FAM82A2	Isoform 1 of Regulator of microtubule dynamics protein 3	S44	Y	N	N	N	Y
IPI00410079	FAM82A2	Isoform 1 of Regulator of microtubule dynamics protein 3	S46	Y	N	N	N	Y
IPI00023780	DNAJC5	Isoform 2 of DnaJ homolog subfamily C member 5	S12	N	N	N	Y	Y
IPI00023780	DNAJC5	Isoform 2 of DnaJ homolog subfamily C member 5	S8	N	Y	N	Y	Y
IPI00023780	DNAJC5	Isoform 2 of DnaJ homolog subfamily C member 5	Y17	N	N	N	Y	Y
IPI00402231	DNAJC5	Isoform 1 of DnaJ homolog subfamily C member 5	S8	N	Y	N	N	Y

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. We also 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.

Chapter 3. Large-scale Analysis of Arginine and Lysine Methylated Proteins Reveals Proteins Involved in Metabolism Are Regulated by Methylation in Cancer Cells

Abstract

The critical roles of protein methylation in cellular signaling and human diseases are underestimated at present, partly for technical reasons. MS-based proteomics could offer a powerful approach for large scale investigation of protein methylation, but the lack of efficient chromatographic methods for enrichment of methylated peptides is a major hurdle. Here we report that exhaustive subfractionation of the whole proteome by various approaches, including cellular fractionation, peptide chromatography using strong cation exchange (SCX) and electrostatic repulsion-hydrophilic interaction chromatography (ERLIC), resulted in deep coverage of the methyl proteome. Heavy methyl SILAC experiment was performed to validate methylated peptides. We identified >3000 lysine- and arginine- methylation sites in a gastric cancer cell line, SNU5. Bioinformatic analysis revealed that lysines and arginines having a neighboring methionine are frequently methylated. Almost all enzymes involved in metabolic processes including glycolysis, citric acid cycle, and oxidative phosphorylation, were found to be methylated in SNU5 cells, suggesting tight regulation of aerobic glycolysis in cancer cells by protein methylation. In conclusion, this study demonstrates a workflow for large-scale analysis of methylated proteins, which could yield new insights such as the potential link between protein methylation and cancer metabolism.

Materials and Methods

Cell Culture

SNU5, a gastric cancer cell line, was obtained from the American Type Culture Collection (Manassas, Virginia, USA) and cultured in Iscove's modified Dulbecco's medium supplemented with 20% fetal bovine serum (FBS), 100 U penicillin, and 100 μg streptomycin per mL (Invitrogen, Carlsbad, California, USA). For heavy methyl SILAC experiments, SNU5 cells were cultured in RPMI-1640 medium supplemented with 10% dialyzed fetal bovine serum (Invitrogen). Both normal culture medium and medium with heavy L-methionine (1-13C, methyl-D3) (Cambridge Isotope Laboratories Inc, Andover, Massachusetts, USA) were used in the SILAC experiment. RPMI-1640 deficient in methionine, lysine and arginine was obtained from Biowest (Miami, Florida, USA). Normal methionine, lysine and arginine were purchased from Sigma-Aldrich (St. Louis, Missouri, USA).

Subcellular fractionation

SNU5 cells were washed twice with phosphate-buffered saline (PBS), and suspended in ice-cold cell homogenization medium (CHM) containing 105 mM $MgCl_2$, 10 mM KCl, 10 mM Tris-HCl, pH 6.7. Cells were homogenized until >90% cells were disrupted. Sucrose was then added to the cell homogenate to a final concentration of 0.25 M. The sample was next centrifuged for 5 min at 1000 ×g and 4 ^{o}C. The pellet was collected as the nuclear fraction. The supernatant was further centrifuged for 10 min at 5000 ×g and 4 ^{o}C. The resulting pellet was washed once with CHM

supplemented with 1 M sucrose. This final pellet was regarded as the mitochondrial fraction. Membrane fraction was obtained as described previously [131, 155]. Briefly, SNU5 whole cell lysate was cleared of debris and unbroken cells by centrifugation for 15 min at 1000 ×g and 4^{o}C. The supernatant was then diluted 10-fold with 0.1 M Na_2CO_3 and incubated for 1 hr at 4^{o}C with gentle shaking, followed by ultracentrifugation at 125000 ×g for 1 hr. The resulting pellet was washed with Milli-Q water twice, and considered as the membrane fraction.

In-gel digestion of proteins

About 200 μg protein from the cytosolic, nuclear, mitochondrial and membrane fractions was digested in-gel as described previously [156]. Briefly, proteins were reduced with 5 mM DTT and alkylated with 55 mM iodoacetamide before trypsin digestion at a protein to enzyme ratio of 1:100.

In-solution digestion of proteins

Proteins were lysed in 50 mM HEPES (pH 7.5), 8 M urea, 75 mM NaCl, protease inhibitor cocktail (COMPLETE, Roche Applied Science, Indianapolis, Indiana, USA) and phosphatase inhibitor cocktail (PHOSTOP, Roche Applied Science). Proteins were reduced by adding dithiothreitol (final concentration 10 mM) to the sample solution at 33 ℃ for 1 hr, and then alkylated by adding 55 mM iodoacetamide to a final concentration of 55 mM and incubating the samples at room temperature for 30 min, before diluting 8-fold with 50 mM HEPES (pH 7.5) and digestion with trypsin in a 1:100 (trypsin/protein) mass ratio. Peptide samples were desalted using SEP-PAK

C18 cartridges (Waters Corporation, Milford, Massachusetts, USA) and vacuum-dried prior to phosphopeptide enrichment.

ERLIC fractionation

Peptide digests was fractionated using ERLIC chromatography. Mobile phase was a gradient mixture of Buffer A (85% acetonitrile, 1% formic acid, 10 mM ammonium acetate) and Buffer B (30% acetonitrile, 0.1% formic acid). Digests were reconstituted in Buffer A and loaded into a PolyWAX LP™ column (4.6 × 200 mm, 5-µm particle size, 300-Å pore size; PolyLC, Columbia, Maryland, USA) on a Prominence™ HPLC unit (Shimadzu, Kyoto, Japan). The sample was fractionated using a gradient of 100% buffer A for 5 min, 0–30% buffer B for 25 min, 30–100% buffer B for 2 min, and finally 100% buffer B for 8 min at a constant flow rate of 1 ml/min for a total of 40 min. Finally the column was washed with 10 ml of 200 mM triethylamine phosphate (TEAP) with 60% acetonitrile, pH 2.0 to elute all peptides. The eluted fractions were monitored *via* a UV detector at 280 nm. Fractions were collected at 1-min intervals and vacuum dried.

Phosphoproteomic data set

Phosphopeptides were enriched using ERLIC [52] and SCX-IMAC as described [53]. Data published earlier [134] were re-analyzed in this study.

Rat liver data sets

Rat liver proteome data sets are described previously [157, 158]. Briefly, proteins from male Sprague Dawley rats (230-250g) were digested in solution, prior to LC-MS/MS analysis in LTQ FT Ultra. A total of 1068 MS analysis were included.

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, Massachusetts, USA) coupled with a ProminenceTM HPLC unit (Shimadzu), as described previously [159, 160] with some modifications. Briefly, the peptide samples were injected from an auto-sampler (Shimadzu) and concentrated in a Zorbax peptide trap (Agilent, Palo Alto, California, USA), and then resolved in a capillary column (200 µm ID x 10 cm) packed with C18 AQ (5-µm particle size, 100-Å pore size; Michrom BioResources, Auburn, California, USA). 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 5-30% B over 66 min, followed by 10 min of 50% B and a ramp from 50%-80% B over 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 2, ion transfer tube temperature at 180 ℃ 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 (350-1600 m/z range) was acquired in the FT-ICR cell at a

resolution of 100,000 and a maximum ion accumulation time of 1000 msec. 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 collisionally activated dissociation (CAD). The default AGC setting was used (full MS target at 3.0e+04, MSn 1e+04) in the linear ion trap. The 10 most intense ions above a 500-count threshold were selected for fragmentation by CAD (MS2), which was performed concurrently with a 1 maximum ion accumulation time of 200 msec. Dynamic exclusion was activated for this process, with a repeat count of 1, exclusion duration of 20 s and ±5 ppm mass tolerance. For CAD, 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 MS raw files were converted to mzXML format and mgf format using Trans-Proteome Pipeline [161]. Protein database search was performed by uploading mgf files to an in-house Mascot cluster server (version 2.2.07) (Matrix Science, Boston, MA) against a concatenated target and decoy version of manually annotated non-redundant UniProt Knowledgebase protein sequence database (40516 sequences, downloaded on 8 October 2010). The search was limited to a maximum of 2 missed trypsin cleavages; #13C of 2; mass tolerance of 20 ppm for peptide precursors; and 0.8 Da mass tolerance for fragment ions. Fixed modification at cysteine was set to methylthio for iTRAQ data sets, and carbamidomethyl for non-iTRAQ data sets. For all data sets, oxidation of methionine, mono- and di-methylation at lysine and arginine, as well as trimethylation at lysine were included as variable modifications.

For iTRAQ data sets, variable modifications also included iTRAQ labeling at N-terminal residues, lysine and tyrosine. Phosphorylation at serine, threonine and tyrosine was included as variable modifications for phosphoproteomic data sets. In heavy methyl SILAC experiments, variable modifications for these data sets included oxidation at light and heavy methionine (OxiM, +15.994915 Da; OxiM4, +20.017100), light and heavy monomethylation at lysine or arginine (methyl1kr, + 14.015650 Da; methyl1krH, + 18.037835 Da), light and heavy dimethylation at lysine or arginine (methyl2kr, + 28.031300 Da; methyl2krH, + 36.075670 Da), light and heavy trimethylation at lysine (methyl3k, + 42.046950 Da; methyl3kH, + 54.113505 Da), as well as heavy methionine (metH, + 4.022185 Da). For all searches, false discovery rate (FDR) was set below 1%. In heavy methyl SILAC experiments, only peptides with either light or heavy methylation were included for subsequent analysis. SILAC pairs were matched from the raw data, and signal intensities were measured for each light and heavy peak. SILAC pairs with peak intensity ratios between 0.70 ~ 1.43 were empirically considered as valid light and heavy methyl SILAC pairs.

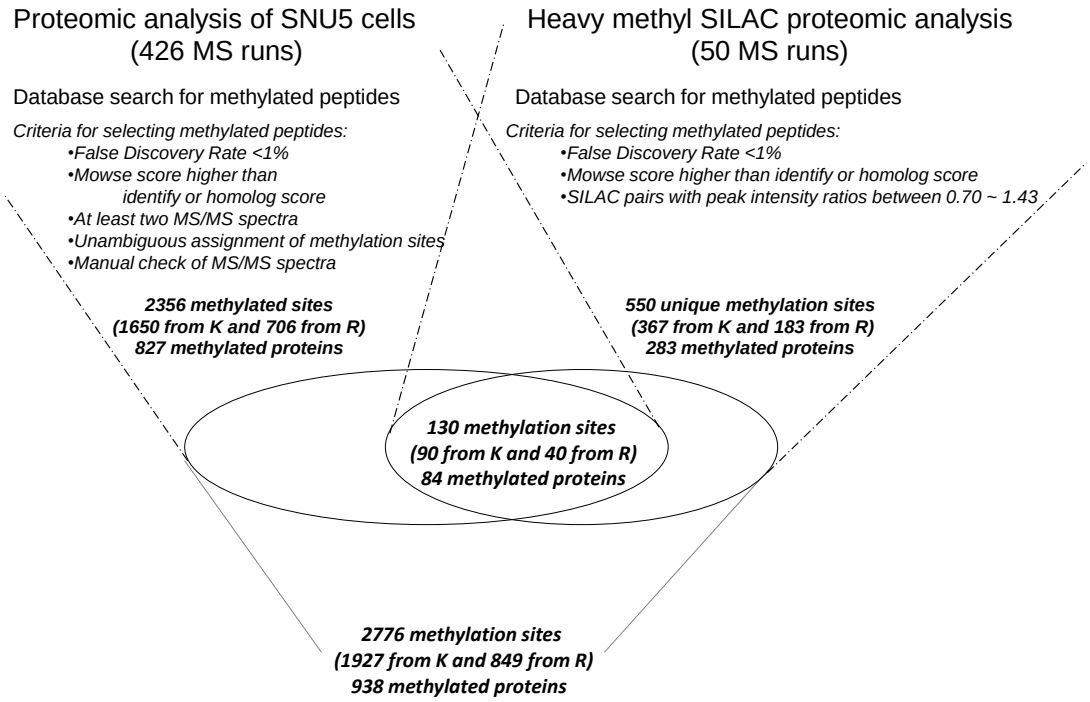
Results

Experimental design

Similar to other PTMs, methylated proteins/peptides are of low abundance in the proteome. But unlike phosphopeptides that can be enriched using various methods, enrichment of methylated peptides is challenging [57]. Proteomic methods based on immunoprecipitation of methylated proteins have identified only a limited number of methylation sites and methylated proteins [71, 73].

In this study (Figure 3.1), we reasoned that methylated proteins/peptides could be identified with confidence after extensive fractionation of the proteome using complementary proteomic approaches. The search space for methyl proteome is extremely large in that methylation occurs to multiple residues and in different degrees (*i.e.* monomethylation, dimethylation and trimethylation) [73]. Therefore, the chance of false positive and degree of ambiguity are large [74]. Although methylation occurs at multiple amino acids, here we focused on methylation at arginines and lysines which are of critical biological significance [62]. Through manual check of spectra, we found identifications with estimated FDR <1% by Mascot search against target-decoy databases still contains peptide-sequence matches that are unlikely correct and unambiguous. So, the identified methylation sites from large spectral data sets were not considered to be confident unless they met stringent criteria through manual check of all the spectra: a) Mowse score for the peptide identification is higher than identity score or homolog score; b) the peptide is identified from at least 2 MS/MS spectra; c) the assignment of post-translational modification sites is unambiguous; d) the majority of the strongest peaks of the spectra could be explained

by their corresponding peptide sequences; e) Noisy and weak spectra are removed. Even higher confidence of identification could be achieved by validating methylated peptides using heavy methyl SILAC [73].



Figuer 3.1 Workflow for methyl proteome. Methylated proteins were identified from an integrated analysis of 426 MS analysis of proteins from a gastric cancer cell SNU5. Heavy methyl SILAC experiment was used to add confidence to methylated proteins. Stringent criteria were used to ensure data quality.

Methylated proteins in gastric cancer

In order to identify methylated proteins in gastric cancer cells, we re-analyzed our mass spectral data sets acquired for a human gastric cancer cell line, SNU5. Subproteomes of SNU5 cells comprising nucleus proteome, membrane proteome, and mitochondrial proteome, were included. The collection also included gastric cancer phosphoproteome data sets. Phosphopeptides from SNU5 digests were enriched using two complementary phosphopeptides enrichment methods, *i.e.* SCX-IMAC and

ERLIC [52, 53]. Also included in this study were iTRAQ-labeled SNU5 proteomic data sets analyzed using pulsed-Q dissociation (PQD) and collisionally activated dissociation (CAD) [159]. All samples were analyzed in LTQ FT ultra. Together, this study is comprised of 426 raw files, representing a comprehensive investigation of SNU5 cell proteome.

We found methylation that was not rare in human gastric cancer cells. At a false discovery rate <1% level, 36253 methylated peptides were identified with high confidence. To ensure even better data quality, we performed manual check of the best MS/MS spectra for each unique peptides. A total of 1698 unique methylated peptides from 827 proteins satisfied the stringent criteria as described previously. In total, 2356 methylated sites (1650 from K and 706 from R) were uncovered.

To further add confidence to identification of methylated peptides by MS, we performed *in vivo* heavy methyl SILAC experiment as described by Ong, *et al* [73]. SNU5 cells were cultured in medium containing either light or heavy methionine, and digested, prior to separation into 25 fractions with ERLIC. Replicate samples were analyzed in LTQ FT Ultra. A total of 550 unique methylation sites were confidently confirmed from 1232 peptides in 283 proteins. We compared this list with the previous data set generated from 426 various MS analyses of the same cell and found 130 unique methylation sites that were identified from both data sets. All together, 2776 methylation sites from 938 proteins were identified from SNU5 gastric cancer cells.

In Uniprot database, monomethylation and dimethylation at R493 in polyadenylate-binding protein 1 (PABP1), an mRNA-binding protein involved in protein translation iniation, are documented. In our proteomic data sets, two high quality spectra were recorded (Figure 3.2). Consistently, heavy methyl SILAC experiments also identified methylation at this site. Characteristic isotopic peak pairs are displayed in Figure 3.2. The upper spectra also provide evidence for a novel methylation site, *i.e.* monomethylated R506.

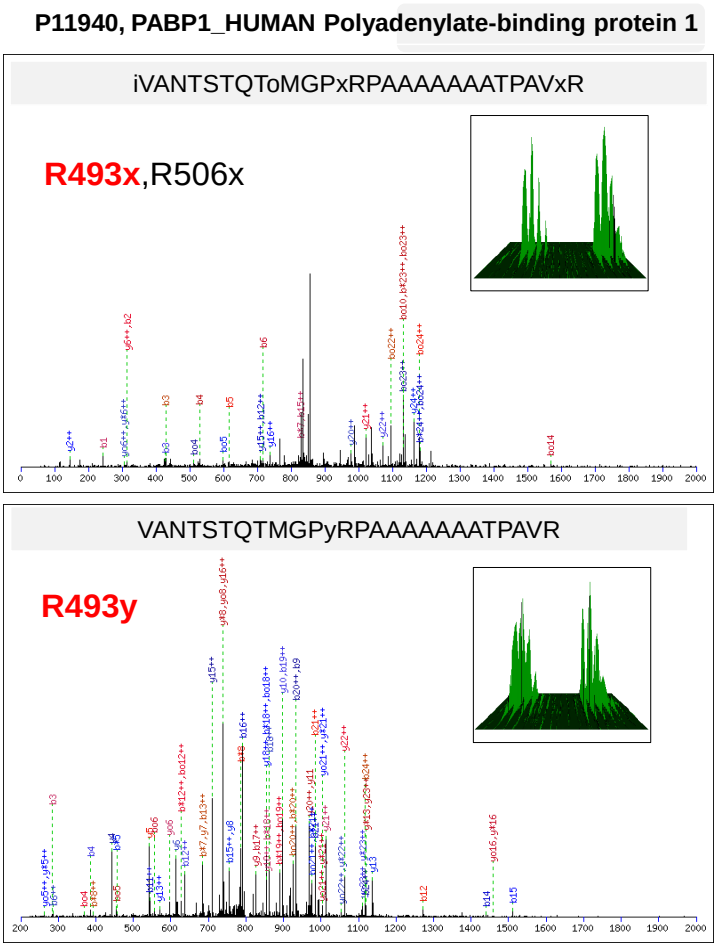


Figure 3.2. Methylation sites in protein PABP1. Methylation sites identified from protein PABP1 by proteomic and SILAC analysis. Monomethylation and dimethylation of arginine 493 are annotated in Uniprot database. Two representative spectra for these two sites from proteomic profiling are annotated here. Characteristic SILAC peak pairs for these two methylation sites are shown in the inserts. One novel methylation site at arginine 506 was identified, too. Lowercase letters in peptide sequence represent post-translational modifications: i,iTRAQ; o, oxidation; p, phosphorylation; x, monomehtylation; y, dimethylation.

Several technical limitations of MS-based methyl proteome analysis may account for methylation sites that were identified only in non-SILAC experiments. First, not all peptides are methylated from AdoMet as the methyl donor. Therefore these methylated peptides will not be labeled as heavy in the SILAC experiment. Second, SILAC pairs of some low abundance methylated peptide may not be obviously visible and detectable in the presence of strong peaks. Third, *in vitro* methylation may occur during protein sample preparation for MS analysis, as reported previously [74]. Fourth, endogenous amino acid substitution could result in an unusual peptide that shares the same mass with a methylated peptide [73]. Fifth, acetylation may be detected as trimethylation because their mass differences are very close [73].

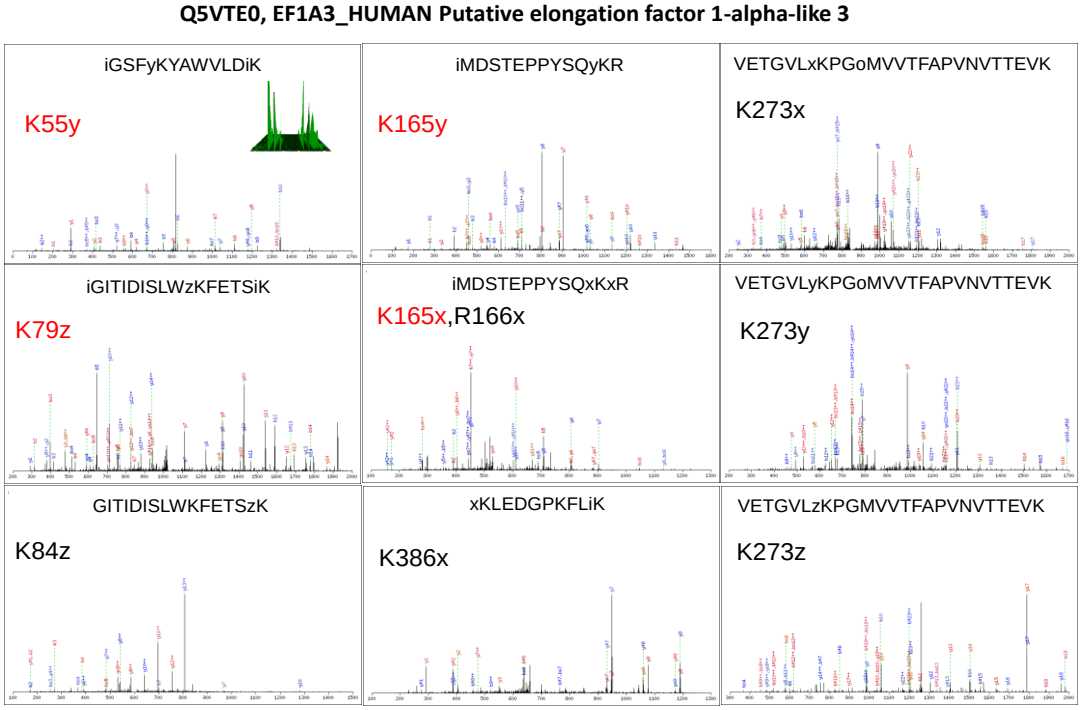


Figure 3.3. Methylation sites in protein EF1A. Methylation sites identified from protein EF1A by proteomic and SILAC analysis. Methylation sites labeled in red are annotated in UniProt database. Characteristic SILAC peaks are shown where available. Lowercase letters in peptide sequence represent post-translational modifications: i,iTRAQ; o, oxidation; p, phosphorylation; x, monomehtylation; y, dimethylation; z, trimethylation.

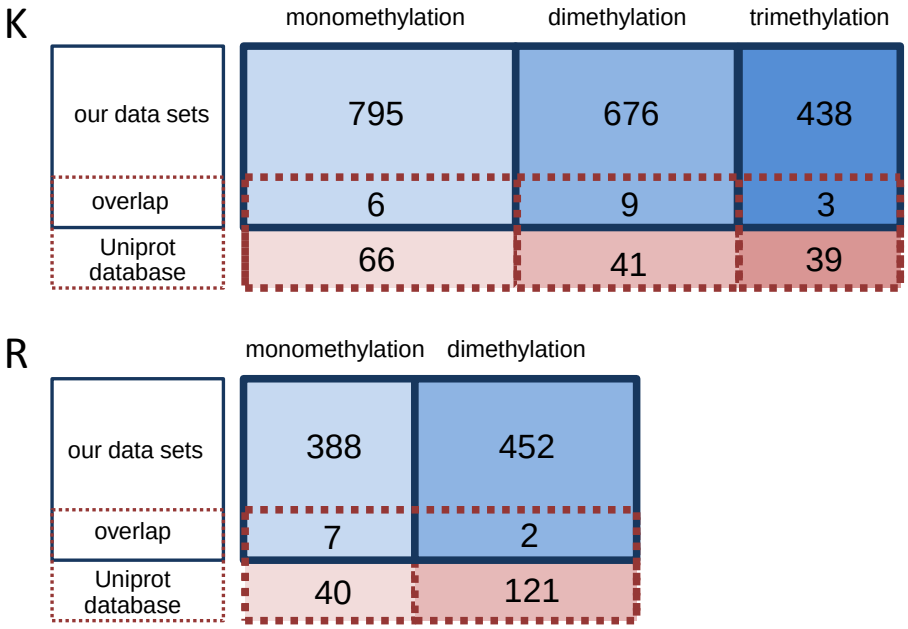


Figure 3.4. **Numbers of methylation sites on lysine and arginine identified in SNU5 gastric cancer cells.** Methylation sites on lysines and arginines were counted and compared to UniProt database. Only unique methylation sites were taken into account.

Comparison of our data set to UniProt Knowledgebase

To date, very few methylated proteins have been identified in the literature. Methylated arginines are more extensively studied, whereas lysine methylation is mainly focused on histone [62]. In this study, 2776 methylation sites were confidently identified from both proteome profiling and SILAC experiments. As shown in Figure 3.4, around 70% methylation occurs at lysine, indicating that lysines are even more likely to be methylated than arginines. The monomethylation is more likely to happen at lysines, whereas arginines are slightly more frequently dimethylated compared to those are monomethylated, in consistence with literature [56, 162]. Dimethylated arginine are divided into two types, *i.e.* asymmetric di-methylarginine and symmetric

dimethylated arginine [162]. However, MS-based approach does not tell the difference since their masses are the same.

We reviewed the UniProt Knowledgebase for annotated methylation sites. Only 334 unique methylation sites from 102 proteins were recorded as of 3 July, 2011. These methylation sites were almost equally distributed between arginines and lysines. Less than 1% of our identified methylation sites were annotated in the UniProt database. This is not surprising considering the small number of annotated methylated proteins in UniProt. Our data set thus greatly expanded the inventory and knowledge of methylated proteins (Figure 3.4).

Protein domain analysis

Arginine methylated is catalyzed by a series of protein arginine methyltransferases (PRMTs). The exact number of PRMTs, as well as their functions, is still uncertain [56]. Much less is known about lysine methyltransferases. Proteins containing the SET (suppressor of variegation, enhancer of *zeste* and trithorax) domain catalyzes lysine methylation [65]. Protein methyltransferases demonstrate specificity for amino acid sequences and the tertiary structure of protein substrates [62]. By analyzing the protein domains of methylated proteins, we sought to identify the enriched protein motifs that were frequently recognized by methyltransferases. Results of protein domain analysis of the 938 methylated proteins using Database for Annotation, Visualization and Integrated Discovery (DAVID, v6.7) are shown in Table 3.1. Our results showed domains involved in domain terms “RNA recognition”, “nucleotide-binding”, “DNA/RNA helicase” and “histone core” were highly represented in methylated proteins, in consistent with the fact that transcription is known to be

tightly regulated by protein methylation [63]. Our results also suggested potential roles of methylated proteins in the cytoskeleton since several protein domains were associated with "filament" and "actin binding". To our knowledge, this has not been reported previously. Interestingly, we found NAD(P)-binding domain to be highly enriched in our methyl proteome, suggesting that methylated proteins may be extensively involved in regulating metabolism.

Table 3.1. Protein domain analysis of methylated proteins. Gene symbols of 938 methylated proteins were uploaded to DAVID for protein domain analysis. INTERPRO database annotated >95% of our queries. *Homo sapiens* was set as background. Protein domains harbored by at least 20 gene products with significance $<10^{-6}$ are shown.

INTERPRO terms	Protein domain	Count	%	P value
IPR000504	RNA recognition motif, RNP-1	50	5.48	2.87E-19
IPR012677	Nucleotide-binding, alpha-beta plait	50	5.48	4.37E-19
IPR018039	Intermediate filament protein, conserved site	29	3.18	5.99E-18
IPR016044	Filament	29	3.18	5.99E-18
IPR001664	Intermediate filament protein	27	2.96	1.24E-15
IPR001715	Calponin-like actin-binding	21	2.30	2.38E-10
IPR016040	NAD(P)-binding domain	26	2.85	2.31E-07
IPR012335	Thioredoxin fold	20	2.19	9.73E-07

Motif analysis of gastric cancer methyl proteome

To gain functional insights into the methylated peptide sequences, we first employed Motif-X to analyze the 13-mer amino acid sequences centered on each methylated residue [135]. Of 2776 methylation sites, 1927 (69%) were at lysines (K), whereas 849 (31%) were at arginines (R). More motifs were matched from K-methylated peptides than those from R-methylated peptides (Figure 3.5). Our data show protein methylation is prone to occur in a motif containing charged residues, including

aspartic acid (D), glutamic acid (E), K and R. Lysines and arginines close to methionines are frequently methylated. Lysines close to the N-terminal of proteins were likely to be methylated, too, probably because most proteins begin with an N-terminal methionine. Amino acids with hydrophilic side chains such as alanine (A), isoleucine (I), leucine (L), tyrosine (Y) and valine (V), might also influence protein methylation. Glutamic acid appeared frequently in methylated peptides, in consistence with the literature which report that glycine and arginine-rich (GAR) motifs are often targets of protein arginine methyltransferases (PRMTs) [68, 69]. Our results showed lysine-methylated proteins were preferentially flanked by glycines, too.

Methyl proteome has not been investigated previously in large scale. We asked the question whether the motifs we found from human gastric cancer cell SNU5 are also found in other data sets. To this end, a methyl proteomic investigation of a different organism, *Rattus rattus*, was profiled for motif analysis. A total of 1068 MS analyses of rat liver were pooled for database search of methylated proteins. Majority of these rat liver samples were fractionated with ERLIC [157, 158]. At FDR<1%, 1605 methylated sites were identified from rat liver. Motifs identified from methylated proteins from rat liver are shown in Figure 3.5, bordered with blue. Interestingly, despite a small number of motifs were unique for human gastric cancer and rat liver, the majority of these methyl motifs were similar. The methyl motifs could be classified into several groups: a) motifs containing residues with charged side chain, b) motifs containing N-terminal methionine, c) motifs containing residues with hydrophilic side chain, and d) motifs containing glycine. The repeated motifs present in both human gastric cancer and rat liver methyl proteomes add confidence to our data sets, and indicate the motifs we identified might be evolutionarily conserved.

Figure 3.5. Sequence analysis of gastric cancer and rat liver methylproteome. Motif analysis of lists of 13-mer peptide sequences pre-aligned at the methylated lysine and arginine were performed using Motif-X software. Significance cutoff was set to 0.001, and minimum occurrence was 10. Letter 'X' indicates absence of amino acid. K and R in the center of each motif are methylated. IPI human or rat proteome was set as background to statistical inference. Motifs with red border are for human gastric cancer methyl proteome, whereas blue borders indicated motifs from rat liver methyl proteome.

Analysis of amino acid composition of gastric cancer methyl proteome

To further understand the frequency of each amino acid in methylated peptides, the number of each amino acid in the gastric cancer 13-mer peptide sequences were counted. The frequency of each amino acid was subsequently normalized with its frequency in whole human proteome (Figure 3.6a). Glycine is the most commonly appeared amino acid in R-methylated peptides, except R itself, consistent with

previous discovery of GAR motif [68, 69]. Interestingly, methionine is the second most frequent residue appeared in R-methylated peptides, and the most frequent residue in K-methylated peptides, except K itself. The analysis reinforces the observation that lysines and arginines are likely to be methylated if there is a nearby methionine.

Next we investigated the co-existence of every two amino acids in methylated peptide sequences. A matrix of combinations of two amino acids was generated from peptide sequences containing 13 amino acids centered on methylated K and R was used in this analysis. Each data point was normalized against the total number of peptide sequences (Figure 3.6b). Lysines in K-methylated peptides appeared to be prone to co-exist with multiple residues including A, D, E, G, I, K, L, P, S and V. Top four amino acids that co-exist with arginines in R-methylated peptides comprise L, E, A and G.

The co-existence analysis of methylated peptides was then normalized to a background of whole human proteome. The co-existence matrix of K-methylated peptides markedly changed (Figure 3.6c). Although the co-existence patterns of K- and R- methylated peptides in Figure 3.6b are dissimilar from each other, their pattern in Figure 3.6c are strikingly similar. For both K- and R- methylated peptides, methionine became the most dominant amino acid in the matrix. Two methionines are the most likely co-existed residues in methylated peptides after subtracting the background. Methylated peptides are also featured with co-existence of methionine with alanine, aspartic acid and glycine. This analysis is in good agreement with our

analysis in Figure 3.6a, and reinforces the finding that a nearby methionine may enhance the chance that a peptide is methylated at lysines.

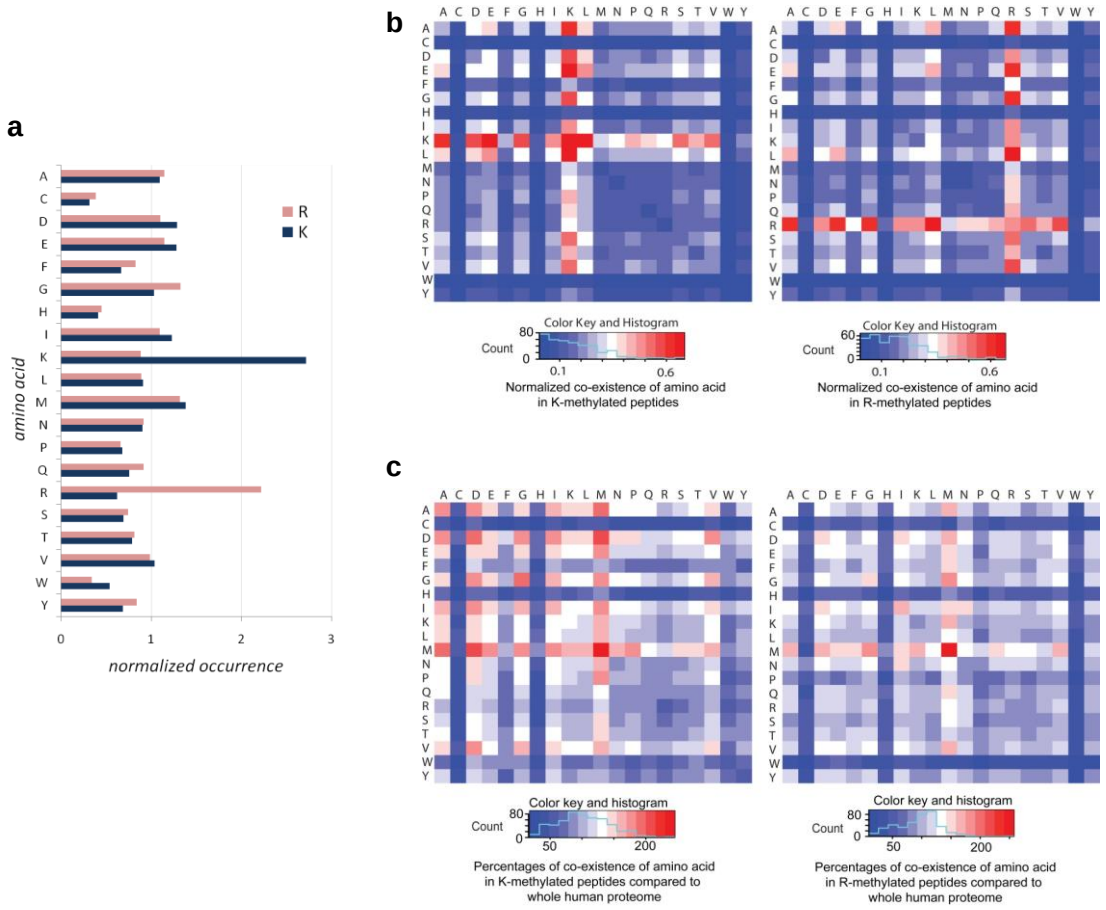


Figure 3.6. Amino acid composition analysis of gastric cancer methyl proteome. Normalized occurrence for each amino acid in the lists of 13-mer peptide sequences of methylated peptides is shown. Occurrence of each amino acid in the human Swissprot protein database was used for normalization. (b) Normalized co-existence matrix for each two amino acid combinations in K-methylated peptides (left panel) and R-methylated peptides (right panel). Color keys and histograms (in cyan) representing distribution of data points are shown below each matrix. (c) Data points from (b) were compared to normalized co-existence matrix generated from whole human proteome, and percentages of fold-change of methylated peptides against the whole human proteome are displayed. The left and right panel are the heatmaps for K-methylated and R-methylated peptides, respectively.

Methylated proteins in cancer metabolism

Consistent with the results of our protein domain analysis, pathway analysis using DAVID and KEGG databases showed glycolysis and citric acid cycle to be significantly processes enriched in the methyl proteome. We checked our methyl proteome and strikingly found that almost every key enzyme involved in glycolysis, citric acid cycle and oxidative phosphorylation to be methylated (Figure 3.7). Most of these methylation sites are novel, *i.e.* have not been reported previously. Several methylation sites were not only supported by high quality MS/MS spectra, but also evidenced with characteristic SILAC pairs (proteins accompanied with blue stars in Figure 3.7).

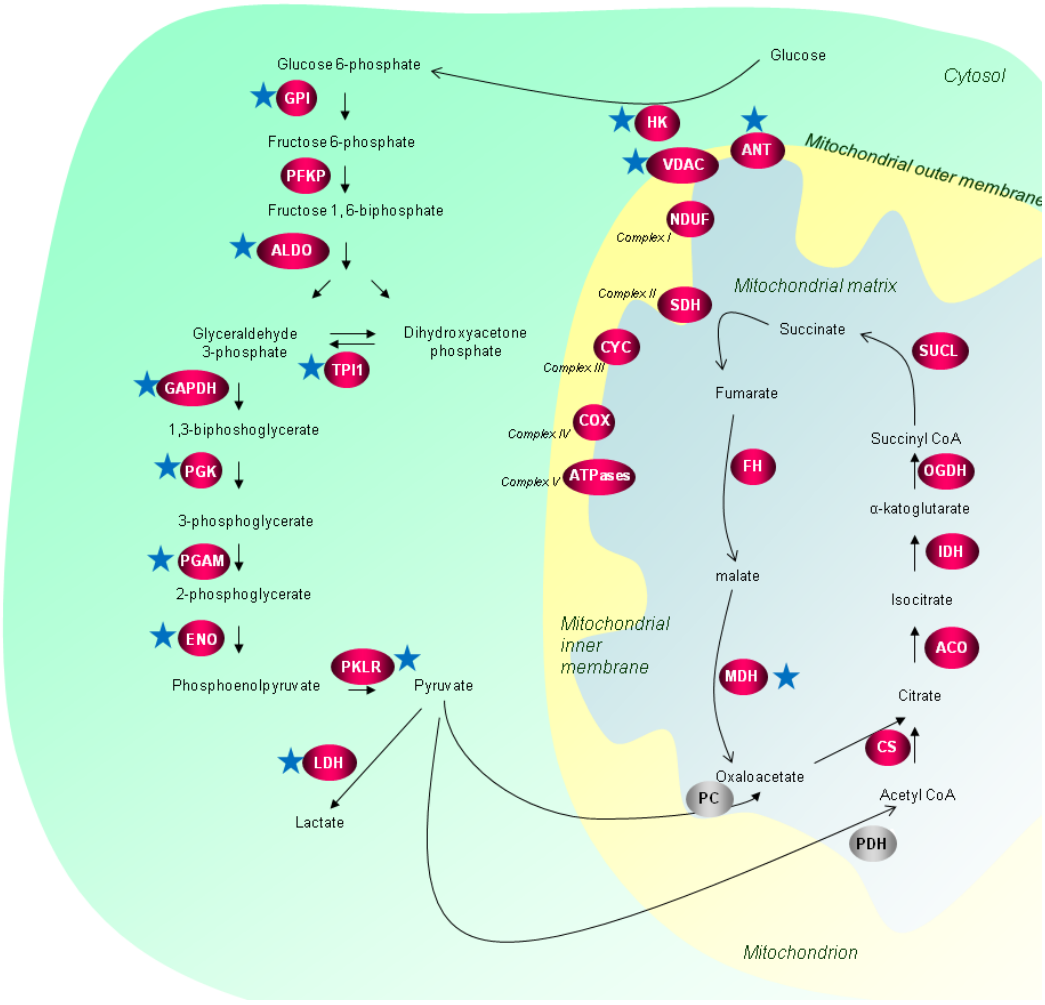


Figure 3.7. Methylated proteins involved in metabolism. This cartoon depicts metabolic processes including glycolysis, citric acid cycle and oxidative phosphorylation. Methylated proteins involved in these processes are shown in red ellipse, whereas proteins in grey ellipses, whereas proteins in grey ellipses were not detected as methylated. Proteins labeled with a blue star were characterized with SILAC pairs in heavy SILAC experiments. GPI, glucose-6-phosphate isomerase; PFKP, 6-phosphofructokinase; ALDO, aldolase; TPI1, triosephosphate isomerase 1; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; PGK, phosphoglycerate kinase; PGAM, phosphoglycerate mutase; ENO, enolase; LDH, lactate dehydrogenase; PK, Pyruvate kinase; COX, cytochrome c oxidase; CYC, cytochrome *c*; SDH, succinate dehydrogenase; NDUF, NADH dehydrogenase (ubiquinone); VDAC, voltage-dependent anion channel ; HK, hexokinase; ANT, adenine nucleotide translocator; FH, fumarate hydratase; MDH, malate dehydrogenase; PC, pyruvate carboxylase; PDH, pyruvate dehydrogenase; CS, citric synthase; ACO, acyl-CoA oxidase; IDH, isocitrate dehydrogenase; OGDH, oxoglutarate dehydrogenase; SUCL, succinate-CoA ligase.

Discussion

A robust system like a cellular biological system relies on responses to various endogenous and exogenous inputs. In cells, rapid responses (within seconds) mediated by protein phosphorylation have been extensively studied. Much less attention has been directed to responses mediated by another class of protein modification i.e. methylation, which is relatively more stable than phosphorylation.

Critical roles of protein methylation in multiple biological processes have been extensively elaborated in a few proteins such as histones. However, the breadth and depth of protein methylation is currently uncertain. In sharp contrast to thousands of phosphorylation sites documented in the literature, merely 334 unique methylation sites from 102 proteins were recorded in the UniProt database as of July 2011.

We have overcome the technical obstacles that have dogged mass spectrometry of methylated proteins on a large scale by extensively resolving the whole proteome into its components by means of subcellular organelle fractionation and multidimensional chromatography using both ERLIC and SCX. Thousands of methylated peptides were identified from re-analysis of published data sets where no enrichment of methylated proteins or peptides had been performed. Although the number of methylated peptides was not large from single MS analysis, the combination of more than 400 MS analyses resulted in unique K and R methylation sites from a gastric cancer cell line SNU5 in excess of 3000. Confidence of identification of some methylated peptides was further increased by experimental validation using heavy methyl SILAC in the same cells.

Although a very limited repertoire of methylated proteins has been developed over the past several decades, investigators have speculated that many more methylated proteins remain to be identified. Our data are the first to provide experimental evidence that protein methylation is not rare in cancer cells. Over 1700 proteins were found to be putatively methylated in our data sets. Less than 1% of these methylation sites are annotated in UniProt protein knowledgebase. Therefore this study provides an unprecedented rich reservoir for functional studies of protein methylation.

For the first time, large scale analysis of methylation allows sequence analysis of methylated proteins. A few motifs were significantly enriched in methylated proteins (Figure 3.3). Lysines and arginines that are close to methionines were frequently methylated. This has not been reported in literature. The underlying reason is unknown. Our data suggested cryptic roles of methionine in the biology of protein methylation which awaits further investigation.

In good agreement with the literature, we found that methylated proteins tend to contain protein domains that interact with nucleic acids. It may be speculated that nucleic acid turnover produces a rich pool of adenosines that could potentially produce AdoMet in the presence of methionine.

While involvement of methylated proteins in transcription, mRNA splicing and DNA repair have been reported [56, 62, 65], our data revealed an unexpected and intriguing association of the methyl proteome; namely the striking and extensive methylation of enzymes involved in metabolism. The different metabolism between cancer and normal cells was observed by Otto Warburg, and is known as the “Warburg effect” [163, 164]. Normal cells rely primarily on mitochondrial oxidative phosphorylation to generate ATP. In contrast, cancer cells tend to rely on aerobic glycolysis. The basis

for this phenomenon is not fully understood yet. In our study, we found almost all key enzymes which catalyze the processes of glycolysis, citric acid cycle and oxidative phosphorylation to be methylated. The majority have multiple putative methylation sites, suggesting tight regulation by methyltransferases. We speculate that methylation marks on these enzymes may have modulated their functions such that energy production was directed preferentially to glycolysis instead of mitochondrial oxidative phosphorylation.

Conclusion

This study presents a workflow for large-scale analysis of protein methylation. Extensive fractionation of the whole proteome resulted better coverage of the methyl proteome than immunopurification of methylated peptides/proteins. We catalogued the most comprehensive inventory of protein methylation in a gastric cancer cell line, SNU5. Sequence analysis of methylated proteins showed that lysine and arginine residues with a neighboring methionine residue are frequently methylated. Strikingly, we found that almost all enzymes involved in glycolysis, citric acid cycle and oxidative phosphorylation were methylated.

Chapter 4. An Integrative Membrane Proteome and Transcriptome Approach Defines the Surface Phenotype of Gastric Cancer Cells

Abstract

Cell surface proteins are a rich reservoir for diagnostic and therapeutic biomarker discovery; however, little is known about the cell surface proteome of gastric cancer, the fourth most common cancer and the second most lethal cancer globally. We present here a strategy for investigating the gastric cancer cell surface proteins by integrated bioinformatic analysis of membrane proteomic and transcriptomic data sets . Our data revealed wide but diverse expression of CD molecules by gastric cancer cells. A few receptor tyrosine kinases (RTKs) were highly expressed. Bioinformatic analysis highlighted a shortlist of proteins of potential pathobiological significance in gastric cancer. Flow cytometry analysis confirmed expression of selected surface proteins. Furthermore, immunohistochemical analysis of 49 pairs of gastric cancer tissues and their matched normal tissues in a tissue microarray format suggested that gastric cancer could be stratified by expression patterns of selected surface markers. Our data collectively define the gastric cancer surface phenotype for the first time. Integrative analysis of membrane proteomics and transcriptomics could be extended to mine cell surface biomarkers for other cancers.

Materials and Methods

Membrane protein enrichment and digestion

Each of six GC cells, *i.e.* SNU5, IM95, AGS, MKN7, KatoIII and SNU1 was lysed using HES buffer (20 mM HEPES, pH7.4, 1 mM EDTA, 250 mM sucrose) supplemented with protease inhibitors as described previously [131]. Cell lysates were diluted with 0.1 M Na_2CO_3, pH 11 and incubated at 4 ℃ with gentle rotation for 1 hour [155]. The suspension was centrifuged for 45 min at 250000 ×g and 4 ℃. The resulting membrane pellet was washed twice with Milli-Q water and centrifuged for 30 min at 250000 ×g. The washed pellet was dissolved in 2 % SDS. About 0.5 mg of membrane protein was resolved in SDS-PAGE. Each sample lane was cut into 15 bands, and proteins in each band were digested using trypsin, prior to LC-MS/MS analysis.

LC/MS/MS analysis

Membrane protein digests from each cell line were analyzed in an LTQ-FT Ultra mass spectrometer (Thermo Fisher, Waltham, Massachusetts, USA) coupled to a Prominence™ HPLC unit (Shimadzu, Kyoto, Japan), as described previously [131]. Briefly, peptide samples were injected from an auto-sampler (Shimadzu) and concentrated in a Zorbax peptide trap (Agilent Technologies, Santa Clara, California, USA), and subsequently resolved in a capillary column (200 μm ID x 10 cm) packed with C18 AQ (5-μm particle size, 100-Å pore size, Michrom BioResources, Auburn, California, USA). The samples were ionized in an ADVANCE™ CaptiveSpray™

Source (Michrom BioResources) with an electrospray potential of 1.5 kV. The LTQ-FT Ultra was set to perform data acquisition in the positive ion mode. The 10 most intense ions above a 500 count threshold were selected for fragmentation.

Protein identification by MS

The MS raw files were converted to mzXML format and mgf format using Trans-Proteome Pipeline. Protein database search was performed by uploading mgf files to an in-house Mascot cluster server (version 2.2.07) (Matrix Science, Boston, MA) against a concatenated target and decoy version of manually annotated non-redundant UniProt Knowledgebase protein sequence database (40516 sequences, downloaded on 8 October 2010). The search was limited to a maximum of 2 missed trypsin cleavages; $\#^{13}C$ of 2; mass tolerance of 20 ppm for peptide precursors; and 0.8 Da mass tolerance for fragment ions. Fixed modification was carbamidomethyl at Cys residues, while variable modification was oxidation at methionine residues. PeptideProphet [165] and ProteinProphet [166] from Trans-Proteome Pipeline (TPP) were employed to estimate false discovery rates at both peptide and protein levels. Only protein groups with a probability above 0.9 were considered as identifications. False discovery rate of protein identification was estimated at below 1% by receiver operating characteristic curves for each cell line. Membrane proteins were defined according to Gene Ontology annotation [167] and transmembrane topology using TMHMM (version 2.0) [168]. Relative abundance of proteins identified in GC cell lines was estimated by normalized spectral index (SI_N) as reported [169].

Subcellular classification

Term associations for each protein-encoding gene symbol were retrieved from Gene Ontology *Homo sapiens* annotation database updated on 30 January 2010. Membrane proteins were grouped according to annotations associated with membrane, plasma membrane, endoplasmic reticulum membrane, mitochondrial membrane, Golgi apparatus membrane, nuclear membrane, lysosomal membrane, endosome membrane, peroxisomal membrane, peroxisomal membrane and other membranes.

Gene expression analysis

Transcriptome data sets were described in Chapter 2.

Flow cytometry analysis

Fluorescence-conjugated antibodies against extracellular domains of CD molecules, CD13-PE, CD14-FITC, CD15-FITC, CD49e-PE, CD326-PE, CD44-APC, CD9-FITC, CD38-PE-cy5, CD59-FITC, and CD55-PE, were obtained from BD Biosciences (Franklin Lakes, New Jersey, USA). Mouse IgG_1-FITC and IgG_{2a}-PE were isotype controls. Antibodies were incubated with GC cells (cell density $5\sim10\times10^5$/ml) for 0.5 hour at 4℃ before analysis in a flow cytometer (FACSCalibur, BD Biosciences).

Immunostaining of tissue microarrays

Two or three 0.6 mm diameter disks were cored from tumor-rich paraffin blocks of each of 49 gastric adenocarcinomas (Beecher Instruments, Wisconsin, USA) and from adjacent histologically normal gastric epithelium, and arrayed on standard glass microscope sides. Use of archived tissues from the Pathology Department, Singapore General Hospital was approved by the SingHealth Institutional Review Board.

Histological evaluation and assignment of Lauren histotypes were reviewed by KHL and WKW. Immunostaining was performed using the basic IHC Kit with Antibody Amplifier™ (ProHisto, South Carolina, USA), EnVision Detection, Peroxidase/DAB, Rabbit/Mouse System (Dako, Denmark) and the following primary antibodies: anti-MET (C-12; sc-10; 1 μg/ml); anti-FGFR2 (ab58201; 0.2 μg/ml; Abcam, Cambridge, UK); anti-FGFR4 (sc-124; 0.2 μg/ml); anti-EPHA3 (ab5386; 1.25 μg/ml; Abcam); anti-EPHB2 (ab5418; 5 μg/ml; Abcam) and anti-ITGB4 (sc-9090; 0.6 μg/ml). Except where stated, all antibodies were from Santa Cruz Biotechnology, CA, USA. The recommended tissue for each antibody was stained as a positive control.

All stained tissue microarrays were scored independently by two pathologists. Each tumor and its adjacent normal epithelium were scored by the product of staining intensity and percentage of positively-stained cancer or normal epithelial cells. Staining intensity was scored on a scale of 0 – 3 (0 – no staining; 1 – weak; 2 – moderate and 3 – strong staining). The scores of each pair of cancer (T) and adjacent histologically normal (N) tissues were compared. T>N and T<N denote higher or lower staining, respectively, in GC compared to matched normal gastric epithelium. T=N denotes equal staining in tumor and matched normal tissues.

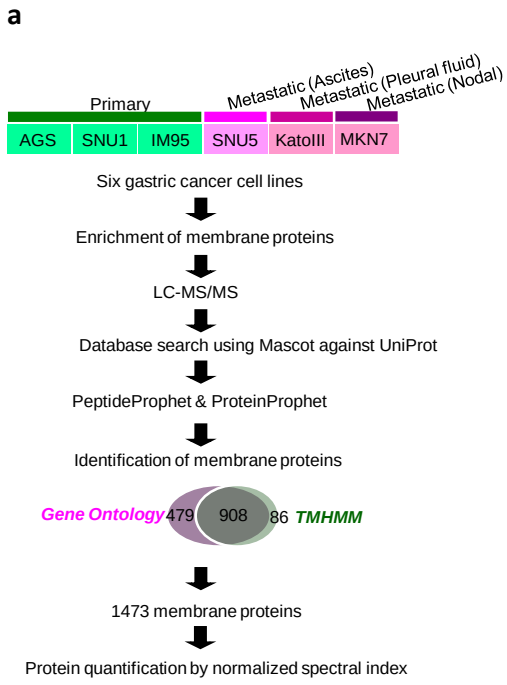
Statistical analyses

Statistical tests were performed in R (version 2.13.0). Fisher's exact 2-tail test was performed using the function "fisher.test". The function "cor" was used to calculate correlation coefficients.

Results

Gastric cancer membrane proteomes by LC-MS/MS

Membrane proteomes of six gastric cancer cell lines derived from both primary (AGS, SNU1 and IM95) and metastastic (SNU5, Kato III and MKN7) carcinoma were systematically investigated. The confidence of protein identifications was ensured by matching MS spectra to manually annotated, non-redundant UniProt Knowledgebase protein sequence database using Mascot search engine, followed by PeptideProphet [165] and ProteinProphet [166] qualification. Membrane proteins were specified both by sequence prediction by TMHMM computation and Gene Ontology annotations. Both approaches identified a total of 1473 membrane proteins, of which 86 and 479 membrane proteins were identified only by TMHMM and GO, respectively (Figure 4.1a).



b

Subcellular organelle	AGS	IM95	Kato III	MKN7	SNU1	SNU5	# total proteins
Plasma membrane	251	238	218	123	324	192	558
Endosome	24	17	21	12	29	22	53
Endoplasmic reticulum	135	138	113	76	161	95	235
Golgi	55	55	40	32	74	39	119
Lysosome	16	13	14	3	10	13	32
Mitochondrion	82	88	57	55	86	50	113
Nuclues	37	36	26	14	40	26	55
Peroxisome	10	11	8	7	14	10	20
Vesicles	7	7	5	2	11	5	14

c

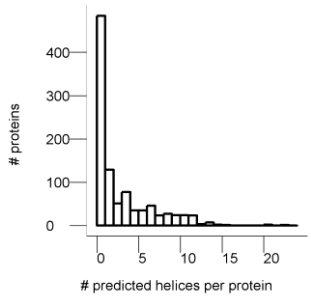


Figure 4.1 Gastric cancer membrane proteome characterized by MS-based proteomics. (a) Workflow of membrane proteome characterization. Cell lines from primary and metastatic gastric cancers were lysed before enrichment of membrane proteins and LC-MS/MS analysis. Database search results from Mascot were qualified by PeptideProphet and ProteinProphet, and quantified by normalized spectral index (SI_N). Membrane proteins were identified by Gene Ontology annotation and TMHMM prediction. (b) Summary of subcellular distribution of gastric cancer membrane proteins. (c) Distribution of predicted transmembrane helices in gastric cancer membrane proteins.

A summary of membrane proteins identified from various subcellular organelles in gastric cancer cells is shown in Figure 4.1b. About 38% of membrane proteins were plasma membrane proteins.

The robustness of our workflow was reflected not only by the large number of identified membrane proteins, but also by that fact that highly hydrophobic membrane proteins were identified. Over 51% of proteins were predicted to have multiple transmembrane helices (Figure 4.1c). Proteins with more than 20 transmembrane

helices, including PIEZO1 (Q92508) and voltage-dependent L-type calcium channel subunit alpha-1C (Q13936), were identified.

A total of 78 CD molecules were widely expressed on the surface of gastric cancer cells, although only 6 were identified in all six cell lines, indicating molecular heterogeneity among GC cell lines. Expression of 16 RTKs was detected also, among which EPHA2 and EGFR were ubiquitously expressed in all six cell lines and ERBB2 was expressed in 5 lines.

Transcriptome analysis of membrane proteins

We analyzed transcription levels of membrane proteins in a panel of 17 GC lines. Our previously published transcriptome data sets of six histologically benign gastric epithelial tissues from non-cancer subjects and 17 GC cell lines [170] were utilized in this analysis. Based on Gene Ontology, 4336 of 10307 annotated gene products were membrane proteins, including 336 CD molecules and 45 RTKs. Overexpressed genes were defined as those whose transcripts were ≥2-fold more abundant in GC cells than in non-cancerous tissues, whereas underexpressed genes were defined as those with cancer-to-control mRNA ratios ≤0.5. We performed cumulative curve analysis for genes of both CD molecules and RTKs, as shown in Figure 4.2. Inspection of the cumulative curves of CD gene expression revealed an interesting difference; namely, that the height of the starting point of the cumulative curve of overexpressed CD genes was higher than that of underexpressed CD genes (135 *versus* 30). This indicated that overexpression was a less frequent abnormality than underexpression with respect to CD molecules in GC (Figure 4.2a). In our data set, 201 of 336 (60 %) CD genes were overexpressed in at least 1 cell line, whereas 306 CD genes (91 %)

were underexpressed in at least 1 line. Therefore, underexpression at the mRNA level was a more frequent event in GC cells. The cumulative incremental steps of underexpressed CD genes was gradual across the 17 cell lines, suggesting that underexpression of CD genes was likely a random event. In contrast, only 29 of 201 (14 %) CD genes were overexpressed in ≥10 cell lines, compared to 144 of 201 (72 %) that were overexpressed in ≤5 lines, indicating that the majority of overexpressed CD genes were infrequent abnormalities. Strikingly, transcriptomic analysis of RTK genes (Figure 4.2b) displayed similar features as we observed for CD genes.

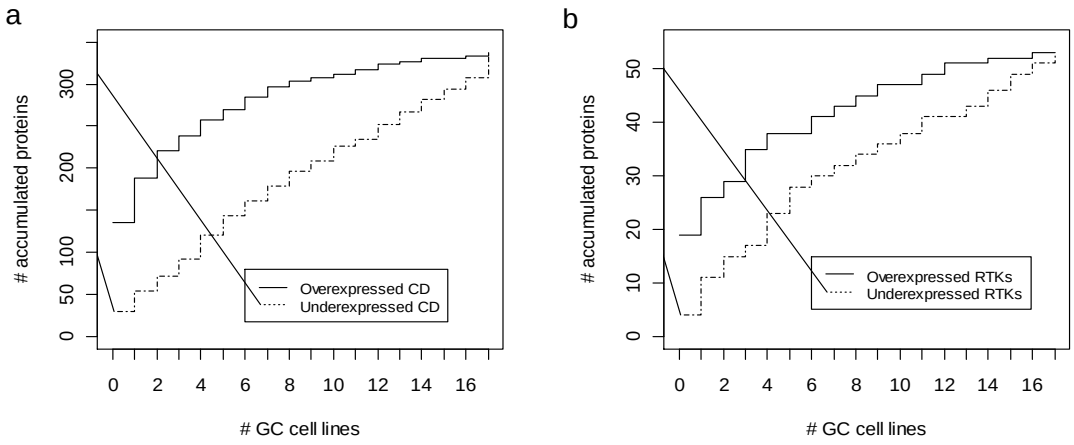


Figure 4.2. Cumulative curves of CD and RTK gene transcription in 17 GC lines. Cumulative numbers of overexpressed and underexpressed CD genes (a) and RTK genes (b) in a panel of 17 GC lines are shown.

GC cell surface proteins from integrated analysis of proteome and transcriptome data

Expression levels of mRNA do not correspond with protein abundance in many cases due to post-transcriptional and post-translational modifications, and variable protein stabilities [171, 172]. We reasoned that aberrant genes would be more confidently identified by selecting those that had both abnormally high expression of mRNA and

the cognate proteins. To this end, we extracted mRNA levels of 976 membrane proteins identified by LC-MS/MS from the transcriptome data sets. From this, we found 57 CD molecules and 16 RTKs thus characterized at both mRNA and protein levels. MS-detectable surface proteins, including CD molecules and RTKs, which were measured as transcriptionally overexpessed in six gastric cancer lines are shown in Figure 4.3 as proteins of potential pathological significance.

Proteins whose corresponding mRNAs were underexpressed, such as EPHA3, EPHA4, EPHA7, DDR2, INSRR and CSF1R, were rarely identified by LC-MS/MS. PDGFRA was underexpressed in all GC lines, except SNU1 in which MS data identified 14 peptides for PDGFRA in SNU1, but none in the other five cell lines.

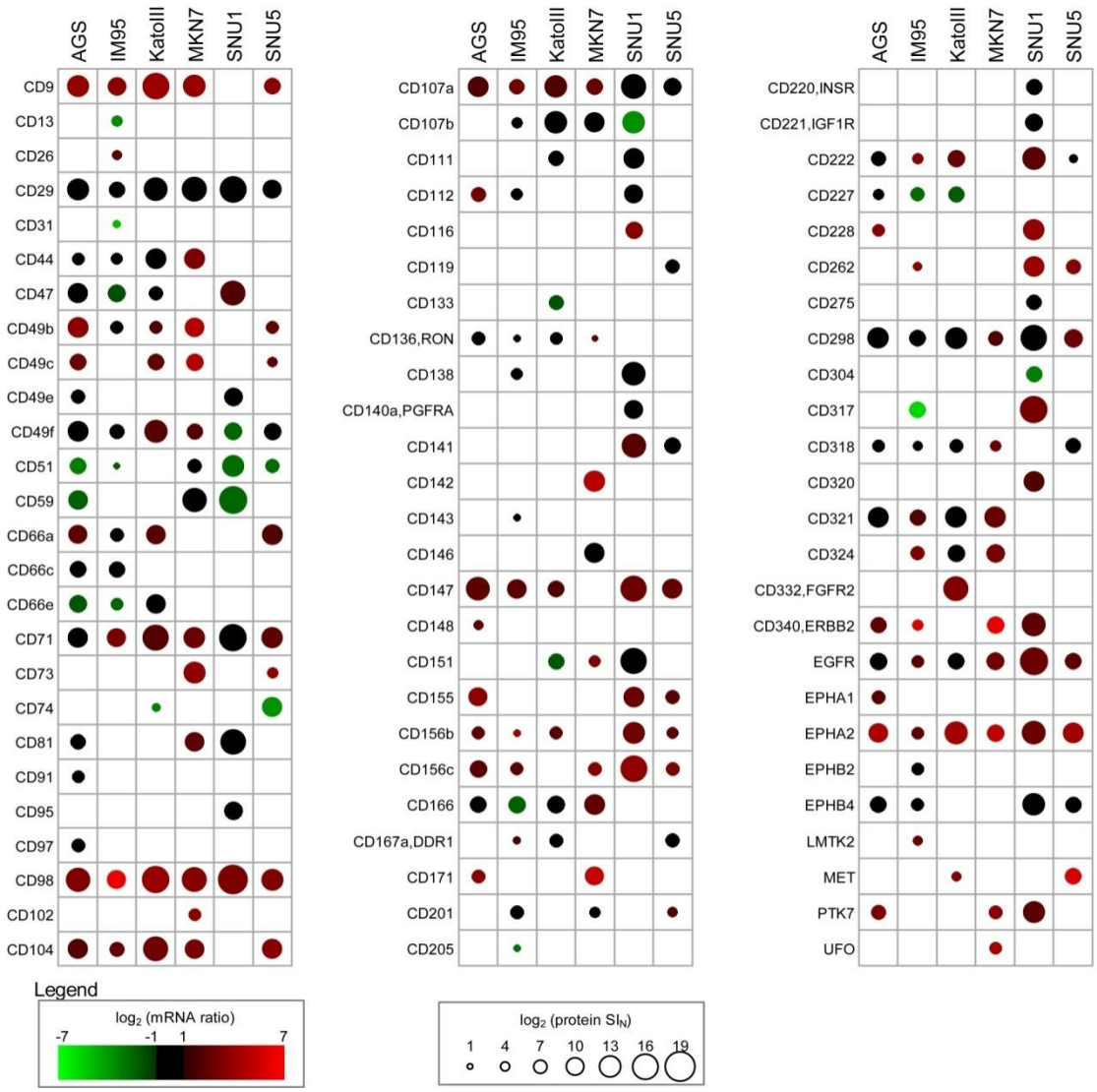


Figure 4.3. Bivariate map of aberrantly expressed cell surface proteins by integrative analysis of proteome and transcriptome data. Cell surface proteins identified by LC-MS/MS are shown with their mRNA expression for six gastric cancer lines. Spot size is proportional to logarithmic values of protein normalized spectral indices and colors indicate $\log_2$ ratios (cancer cells to benign gastric epithelium controls) of mRNA expression. Ratios >2 are shown in red, <0.5 in green. Ratios between 2 - 0.5 are in black.

Surface proteins on GC cells are organized in compact networks. Figure 4 shows bioinformatic analysis of GC surface protein networks based protein-protein interactions documented by Search Tool for the Retrieval of Interacting Genes/Proteins (STRING, version 8.3, string-db.org). The presence of these

extensively networked interactions point to detailed knowledge of molecular organizations as a critical step to understanding the pathobiology of GC.

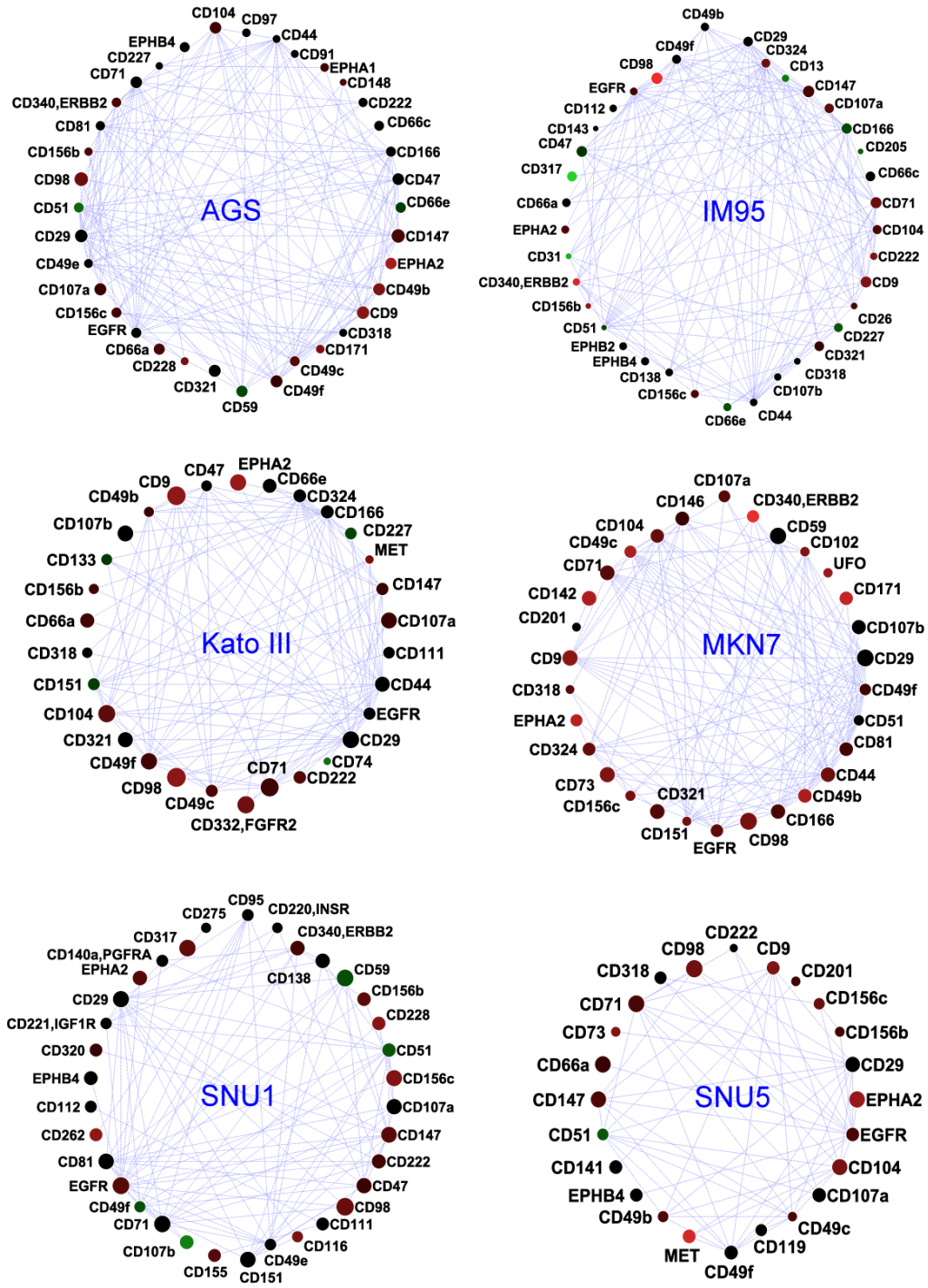


Figure 4.4. **Protein interaction networks of the surface proteomes of GC cells**. CD molecules and RTKs expressed on six gastric cancer cells are shown as interactome maps. The node size of each protein is proportional to the logarithmic transformed values of protein abundance as measured by its normalized spectral index. Node color denotes expression of the cognate mRNA relative to benign gastric epithelium. The color gradient from green to dark green indicates $\log_2$ transformed mRNA ratios < -1, whereas the color gradient from dark red to red represents $\log_2$ transformed mRNA ratios > 1. Ratios denoting expression levels less than ± 2-fold different from benign controls are in black. Interactions between proteins, as retrieved from STRING database, are shown in light blue dashed edges.

The sensitivity of our strategy was supported by re-identification of known biomarkers of established or potential pathological significance. Overexpression of CD340/ERBB2/HER2 has recently elicited great clinical interest as a therapeutic target in gastric cancer [91]. This protein was detected readily and its mRNA was overexpressed in four of the 6 GC cell lines investigated (Figure 4.3). MET, a candidate therapeutic target, was transcriptionally overexpressed in two cell lines. Most peptides from MET were identified in SNU5 cells, consistent with previous studies by us and others [103, 112, 170]. Bivariate analysis also identified a recently proposed gastric cancer target, CD332/FGFR2. Peptides from FGFR2 were detected only in Kato III. FGFR2 mRNA was >6 times higher in Kato III than in controls; FGFR2 peptides were detected 1611 times in this cell line but in none of the others, suggesting that FGFR2 is a highly aberrant feature in a subset of GC. Our data confirm another study which reported FGFR2 overexpression in three of 11 GC cell lines (including Kato III) [173].

Our integrative analysis also captured other emerging candidate biomarkers. LC-MS/MS detected as many as 110 peptides from EPHA2 in AGS cells, while transcriptome data likewise showed that EPHA2 mRNA was >19-fold more abundant in this cell line compared to benign controls. Similarly, EGFR was most highly transcribed in SNU1, and was also the cell line from which LC-MS/MS identified the highest number of EGFR peptides. Our data also highlighted aberrant high expression of CD98 in all six GC lines. This is in accordance with reports that CD98 expression in the gastrointestinal tract was among the highest compared to most other tissues [174]. CD98 is a widely expressed plasma membrane amino-acid transporter that also binds integrins, thereby regulating cell proliferation, survival, migration and adhesion.

Although CD98 has not been associated with GC, our data suggest that this protein merits attention as a potential GC biomarker.

The utility of integrating proteome and transcriptome data sets was validated by the appearance of two groups of proteins whose functions accurately reflect the clinical phenotype of gastric cancer i.e. invasiveness and altered cellular metabolism (Figure 5.3 and Figure 5.4). Four classes of proteins implicated in invasive and metastatic behavior were well expressed in gastric cancer cells. These were (a) integrins (CD49f/ITGA6, CD29/ITGB1, CD49b/ITGA2, CD51/ITGAV, CD104/ITGB4 and CD49e/ITGA5); (b) other adhesion proteins (CD171 and CD155); (c) cell surface proteins known to be associated with tumor invasion (CD147/basigin, CD156B/ADAM17 and CD156C/ADAM10); and (d) tetraspanins (CD9, CD81 and CD151). Metabolism-related cell surface proteins were highly expressed by gastric cancer cells. Overexpression of CD73/NT5E, CD98/SLC3A2, CD298/ATP1B3 and CD320/transcobalamin receptor was consistent with enhanced uptake and metabolism of extracellular metabolites (nucleotides, amino acids, calcium, cobalamin) and ATP hydrolysis. Surface proteins related to immune responses were also present in gastric cancer cells. CD55 (decay-accelerating factor) and CD59 (protectin) are cell surface proteins that protect tumor cells against complement-mediated cytolysis that were reported to be broadly expressed in gastrointestinal tumor cells [175].

Flow cytometry analysis confirms surface proteome

We compared the surface proteome developed by the integrative approach with immunophenotype data of 14 GC cell lines interrogated for the expression of selected surface proteins by flow cytometry (Figure 4.5). Flow cytometric analysis can

validate the presence of surface proteins, and quantify the percentage of the subpopulation of cells that express the surface protein. However owing to the sensitivity of fluorescence detection, signal strength from flow cytometry does not indicate the absolute amount of the protein of interest. Therefore, some proteins that are not identified by MS due to low abundance may still be detected by flow cytometry. The comparison validated the integrative approach. For example, CD14, CD36 and CD38 were not identified by MS and had very weak or no expression by flow cytometry. In contrast, flow cytometry confirmed high expression of CD9 in 10 of 14 GC cell lines, further suggesting a potential pathological role of this tetraspanin protein in GC.

cell	CD9	CD13	CD14	CD15	CD36	CD38	CD44	CD49e	CD55	CD59	CD133	CD326
AGS	85.9	0.2	0.0	93.8	0.1	0.1	44.2	82.9	84.6	97.5	0.0	56.1
Fu97	76.0	98.5	0.6	19.2	0.7	0.8	3.2	96.5	94.9	98.7	2.7	10.4
KatoIII	96.8	1.0	0.3	85.9	0.3	2.4	52.1	1.3	36.6	2.6	11.6	13.2
NCIN87	95.8	4.2	1.3	98.4	1.2	0.4	6.1	65.5	97.5	96.4	0.2	96.4
SNU1	0.0	0.5	0.3	0.1	0.0	0.0	1.3	51.6	99.5	99.2	0.1	0.0
SNU5	21.2	11.1	0.1	0.8	0.0	0.1	8.0	3.7	2.3	1.0	0.2	1.6
SNU16	72.8	0.9	0.1	38.0	0.1	0.4	64.5	1.4	99.4	79.0	0.4	0.0
YCC1	82.3	0.1	0.0	97.1	0.1	0.3	0.3	18.6	99.7	99.6	0.0	95.2
YCC2	98.8	93.5	1.2	63.6	1.2	1.0	33.7	13.7	99.6	98.6	0.4	96.9
YCC3	95.8	0.9	0.2	98.9	0.3	0.3	51.7	0.1	26.6	4.2	0.0	79.0
YCC6	99.1	11.0	0.1	48.1	0.1	4.6	1.6	15.9	42.7	1.7	1.5	95.8
YCC9	95.4	2.7	0.0	96.7	0.0	0.7	55.3	38.1	90.5	92.0	0.1	87.5
YCC11	39.4	1.1	0.0	45.1	7.9	0.1	37.6	98.5	99.6	98.3	0.0	0.8
YCC16	14.3	57.6	0.2	0.2	1.6	0.3	57.2	96.0	99.7	99.2	0.0	0.2

≥75%
≥50%
≥25%
≥1%
≥0%

Figure 4.5. The GC surface proteome by flow cytometry. The table is color coded according to the percentage of cells showing positive staining for each protein in 14 GC lines.

Overall, flow cytometric data showed good concordance with transcriptomic and proteomic data sets. The GC surface proteome included CD44 and CD133, putative markers of cancer stem cells, which were both identified by MS. CD44 was not detected by MS in SNU1 and SNU5 cells, and was also minimally expressed as evaluated by flow cytometry. The AGS cell line had a high percentage of CD44-positive cells and multiple CD44 peptides by LC-MS/MS. Only one cell line, Kato III,

among the six proteomic data sets in this study contained MS-detectable CD133 peptides. Flow cytometry again supported the MS results: 11% of Kato III cells expressed CD133, whereas the other five cell lines showed <1% CD133 positivity.

CD49e/ITGA5, an integrin, was widely but variably expressed in the surface proteome. It was detected in AGS and SNU1 cells, but not in Kato III and SNU5 cells. Flow cytometry data were in good agreement showing 83% and 52% CD49e positivity in AGS and SNU1 cells, respectively, but <4% positivity in Kato III and SNU5 cells.

CD59 was identified with high abundance in three of the 6 cell lines by MS (Figure 3). Flow cytometry data confirmed expression of CD59 in AGS and SNU1 (>97%). CD59 was not identified in SNU5 and Kato III. Consistent with this, fewer than 3% of SNU5 and Kato III cells were CD59-expressing by flow cytometry. Flow cytometry also revealed proteins that were not detected by MS in this study. For example, CD55 was highly expressed in gastric cancer cells and was co-expressed with CD59 in the majority of cell lines we investigated.

Tissue microarray analysis of gastric cancer verified disease-associated biomarkers

We next asked whether integrative analysis of gastric cancer membrane data sets would facilitate discovery of clinically relevant biomarkers. We investigated the expression of six surface proteins - MET, CD332/FGFR2, EPHA2, ITGB4 and EPHB2 - by immunostaining 49 pairs of primary gastric adenocarcinomas and their matched adjacent non-cancer tissues. Although FGFR4 was not identified in our bivariate analysis, we included it in our study of tissue microarrays as a control

because of its recently reported high expression in GC [176]. The clinicopathologic characterstics of these 49 cases are summarized in Table 4.1. Expression of the six proteins in primary GC tissues is summarized in Table 4.2 according to tumor histotypes (Lauren intestinal, diffuse or mixed types). In most tumor-normal tissue pairs, expression of the six surface proteins was different between tumors (T) and adjacent non-cancer tissues (N) *i.e.* N<T or N>T was the usual finding. It was uncommon for a tumor and its matched normal tissue to show similar expression (N=T); this pattern having occurred only once for FGFR2 and EPHB2.

Table 4. 1. Clinicopathologic characterstics of 49 cases of gastric cancers.

#	**Lauren histotype**	**Age**	**Gender**	**Ethnicity**
1	Diffuse	60	M	Chinese
2	Intestinal	58	F	Chinese
3	Intestinal	78	M	Chinese
4	Intestinal	66	M	Chinese
5	Intestinal	78	M	Chinese
6	Intestinal	61	M	Chinese
7	Intestinal	71	F	Malay
8	Intestinal	77	M	Chinese
9	Mixed	57	M	Malay
10	Intestinal	73	M	Chinese
11	Diffuse	39	M	Chinese
12	Mixed	45	F	Chinese
13	Mixed	52	M	Chinese
14	Intestinal	46	F	Chinese
15	Diffuse	77	M	Chinese
16	Intestinal	57	M	Chinese

#	Lauren histotype	Age	Gender	Ethnicity
17	Diffuse	64	M	Chinese
18	Diffuse	71	F	Chinese
19	Diffuse	65	F	Chinese
20	Diffuse	71	M	Chinese
21	Intestinal	76	M	Chinese
22	Diffuse	71	F	Chinese
23	Mixed	65	M	Indian
24	Diffuse	51	F	Malay
25	Intestinal	71	M	Chinese
26	Diffuse	69	M	Chinese
27	Diffuse	61	M	Chinese
28	Intestinal	79	M	Chinese
29	Diffuse	66	F	Chinese
30	Diffuse	70	M	Chinese
31	Intestinal	88	M	Chinese
32	Intestinal	63	M	Chinese
33	Intestinal	53	M	Indian
34	Intestinal	67	M	Chinese
35	Intestinal	69	M	Chinese
36	Intestinal	85	F	Chinese
37	Intestinal	71	F	Malay
38	Diffuse	68	F	Chinese
39	Intestinal	80	M	Chinese
40	Diffuse	42	F	Chinese
41	Intestinal	78	F	Chinese
42	Intestinal	76	M	Chinese
43	Diffuse	47	F	Chinese
44	Diffuse	70	M	Chinese

#	Lauren histotype	Age	Gender	Ethnicity
45	Intestinal	47	M	Chinese
46	Intestinal	62	M	Chinese
47	Intestinal	83	F	Chinese
48	Intestinal	71	M	Chinese
49	Intestinal	71	M	Chinese

EPHB2 expression was increased in GC tissues in 39 of 49 tumor-normal pairs (79.6%). Although EPHB2 expression has been reported in gastric adenocarcinoma [177], our data are the first to show increased expression in malignant compared to adjacent histologically non-malignant gastric epithelium. FGFR4 expression showed the opposite pattern, having decreased expression in GC tissues in 38 of 40 tumor-normal pairs. EPHA2, ITGB4 and MET were each overexpessed in 28 of 49 GC tissues (57.1%) compared to adjacent non-cancer tissues. Remarkably, 71% and 75% of intestinal type GCs overexpessed MET and EPHA2, respectively. There was a significantly higher proportion of the intestinal histotype compared to the diffuse histotype among both MET and EPHA2 overexpressing GC ($p=0.029$ and 0.0047, respectively, by Fisher's exact 2-tail test). There was also a trend for FGFR2 and ITGB4 to be more highly expressed in intestinal type GCs (≥60%) (Figure 4.6 a,b,c).

Expression of EPHB2 and ITGB4 were positively correlated in the 49 pairs of tissues ($r=0.52$), as were MET and EPHA2 ($r=0.50$). Interestingly, the correlation of MET and EPHA2 expression was stronger in the diffuse histotype ($r=0.60$). We also found significant correlation between expression of MET and FGFR2 ($r=0.54$) and of EPHA2 and FGFR4 ($r=0.57$) in diffuse GC tumors. No significant correlation was found in the intestinal histotype tumors (Figure 4.6d,e,f). Overall, stronger

correlations among the six surface proteins examined by immunostaining were found in diffuse than in intestinal GC (Figure 4.6g).

Table 4.2. Expression of six selected surface proteins by immunostaining microarrays of primary GC tissues.

		Lauren Histotype			total
		Diffuse	Intestinal	Mixed	
MET	T>N	6	20	2	28
	T=N	0	0	0	0
	T<N	11	8	2	21
FGFR2	T>N	8	17	3	28
	T=N	0	1	0	1
	T<N	9	10	1	20
FGFR4	T>N	2	8	1	11
	T=N	0	0	0	0
	T<N	15	20	3	38
EphA2	T>N	5	21	2	28
	T=N	0	0	0	0
	T< N	12	7	2	21
ITGB4	T>N	8	18	2	28
	T=N	0	0	0	0
	T<N	9	10	2	21
EphB2	T>N	12	25	2	39
	T=N	1	0	0	1
	T<N	4	3	2	9
total		17	28	4	49

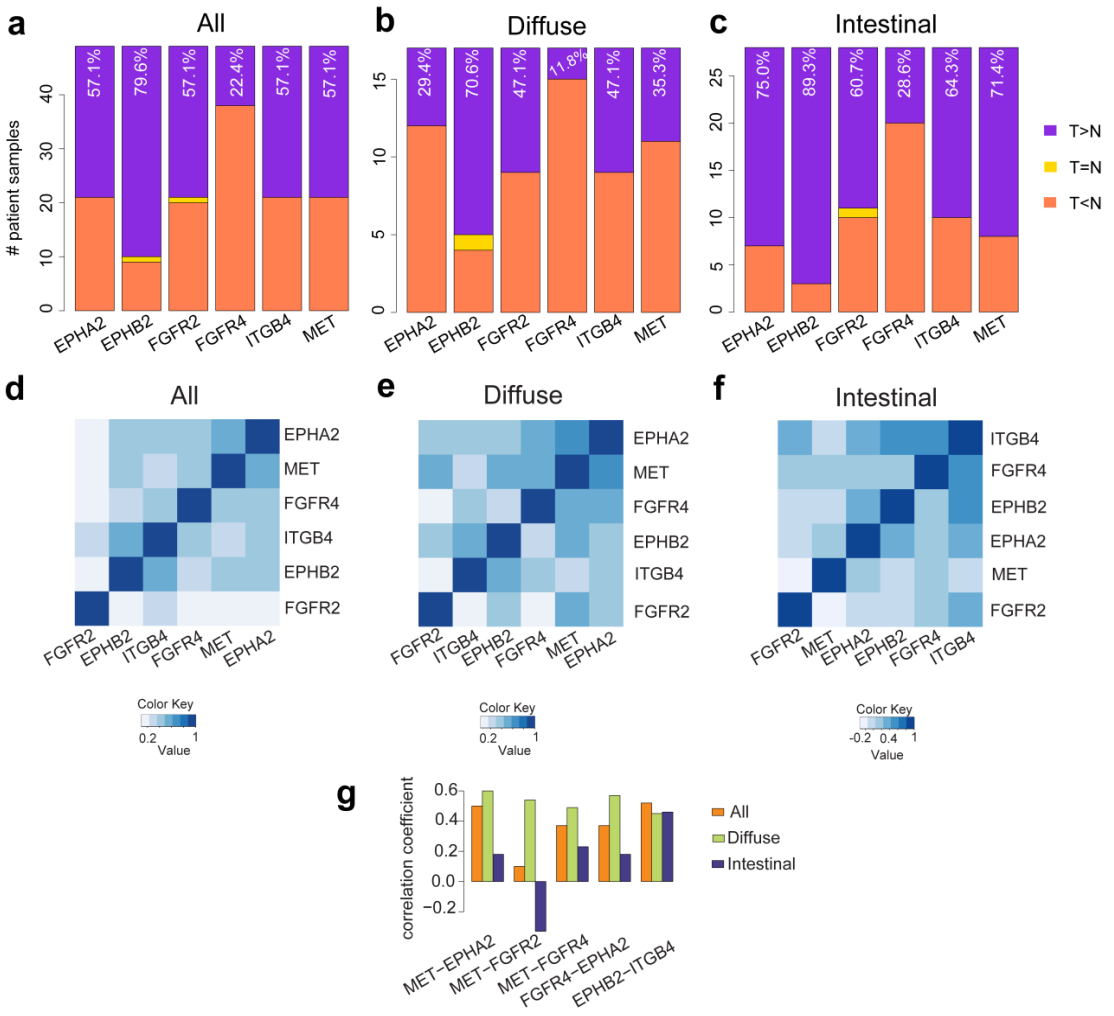


Figure 4.6. Expression of selected plasma membrane proteins using GC tissue microarrays. Summary of protein expression in (a) all gastric cancers; (b) diffuse type gastric cancer; and (c) intestinal type gastric cancer. Percentages of cases where protein expression was higher in cancerous tissues than in matched non-cancerous tissues are shown. Correlation coefficients between the expression of six plasma membrane proteins were calculated by R and is shown as a correlation heat map for (d) all gastric cancers; (e) diffuse type gastric cancers; and (f) intestinal type gastric cancers. Highly correlated protein pairs are shown in (g).

Discussion

Compared to other common cancers, gastric cancer lacks clinically useful diagnostic and therapeutic biomarkers. The cell surface proteome is a rich reservoir of protein biomarkers of pathological importance. However, this reservoir remains largely untapped because systematic investigation with comprehensive documentation of gastric cancer surface proteins has not hitherto been attempted.

Although recent progress in genomics and LC-MS/MS-based proteomics permits large-scale analysis of gene products, each approach alone has limitations in providing comprehensive qualitative and quantitative information of cell surface proteins. Here we demonstrate the power of an integrative analysis of cell surface proteins using membrane proteomics and transcriptomics.

Cell surface proteins of six gastric cancer cell lines were investigated in this study. This first systematic phenotype of the GC surface reveals that GC cells have various "faces" that are very different from benign gastric epithelium (Figure 4.4). This raises the possibility that GC tumors may be stratified by their surface phenotypes. A majority of cell surface proteins are expressed at lower levels in GC compared to benign gastric epithelium, but a few are present at extremely high levels on the plasma membrane. These may act as the "drivers" of gastric oncogenesis by transducing external signals. However, our data also point to different "driver" proteins acting on the surface of GC cells from different tumors.

mRNA expression is not always correlated with protein expression, mainly due to post-transcriptional and post-translational modifications [171]. Proteins that are highly transcribed, as quantified by transcriptomics, and highly expressed, as quantified by normalized spectral index, could be the focus for discovering biomarkers that are relevant to the pathogenesis of GC. As shown in Figure 4.3, a total of 57 CD molecules and 16 RTKs characterized at both mRNA and protein levels constitute a proposed shortlist for biomarker discovery in GC research.

The validity of this strategy was confirmed by flow cytometry and tissue microarray analyses. Surface proteins that were not detected by MS were also negative or weakly positive by flow cytometry while MS-detected proteins were confirmed by flow cytometry. Moreover, the proportion of gastric cancer cells expressing surface markers by flow cytometry in 14 different gastric cancer cells showed high concordance with proteomic and transcriptomic data sets. Immunostaining of primary GC tumor tissues was further validation of the strategy of integrative analysis for identification of disease-relevant biomarkers. Among 49 pairs of tumor-normal tissues, five surface proteins were highly overexpressed in tumor tissues compared to their matched normal controls. Furthermore, our data showed that diffuse and intestinal histotypes of GC had distinct patterns of surface protein expression. A signature comprising four surface proteins, *i.e.* MET, EphA2, FGFR4, CD104/ITGB4, was preferentially expressed in intestinal type GC, whereas diffuse type GC tended to underexpress these proteins.

Conclusion

We have demonstrated that integration of membrane proteomic and transcriptomic data provides valuable insights into proteins expressed on the surface of cancer cells. This analytical approach when applied to our data has uncovered a global view of gastric cancer surface proteins. This strategy could be readily extended to study the surface proteomes of other cancers.

Chapter 5. Quantitative Proteomics Discloses MET Expression in Mitochondria as a Direct Target of MET Kinase Inhibitor in Cancer Cells

Abstract

Cancer cells with MET overexpression are paradoxically more sensitive to MET inhibition than cells with baseline MET expression. The underlying molecular mechanisms are incompletely understood. Here, we have traced early responses of SNU5, a MET-overexpressing gastric cancer cell line, exposed to sublethal concentration of PHA-665752, a selective MET inhibitor, using iTRAQ-based quantitative proteomics. More than 1900 proteins were quantified, of which >800 proteins were quantified with at least five peptides. Proteins whose expression was perturbed by PHA-665752 included oxidoreductases, transfer/carrier proteins and signaling proteins. Strikingly, 38 % of proteins whose expression was confidently assessed to be perturbed by MET inhibition were mitochondrial proteins. Upon MET inhibition by a sublethal concentration of PHA-665752, mitochondrial membrane potential increased and mitochondrial permeability transition pore was inhibited concomitant with widespread changes in mitochondrial protein expression. We also showed the presence of highly activated MET in mitochondria, and striking suppression of MET activation by 50 nM PHA-665752. Taken together, our data indicate that mitochondria are a direct target of MET kinase inhibition, in addition to plasma membrane MET. Effects on activated MET in the mitochondria of cancer cells that are sensitive to MET inhibition might constitute a novel and critical non-canonical mechanism for the efficacy of MET-targeted therapeutics.

Materials and Methods

Chemicals

All chemicals were purchased from Sigma-Aldrich (St. Louise, Missouri, USA) unless otherwise stated. A selective MET inhibitor PHA-665752 [115] was from Pfizer Global Research and Development (La Jolla Laboratories, San Diego, California, USA). Stock solutions of this compound were prepared in DMSO, stored in -80℃ and diluted with fresh medium before use. In all experiments, the final concentration of DMSO was <0.1%.

Cell culture

Gastric cancer cell lines were cultured as described in Chapter 2.

Gene expression profiling

Transcriptomics data sets of gene expression has been described in Chapter 2.

MTT assay

Cell viability based on redox enzyme activity was quantified using 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide, the MTT assay as described [112].

iTRAQ protein sample preparation

Four experimental groups of SNU5 cells were prepared in the absence or presence of PHA-665752. Three groups were exposed to 50 nM PHA-665752 for three durations *i.e.* 4 hr, 24 hr and 72 hr. A parallel group without treatment served as the control. After treatment, proteins were extracted and three independent biological replicate flasks for each experimental condition were pooled and quantified by BCA protein assay kit as described previously [159].

Isobaric labeling

Two-hundred micrograms of protein from each experimental condition were tryptically digested and labeled with 4-plex iTRAQ reagents (Applied Biosystems, Foster City, California, USA) as follows: control, 114; 4 hr, 115; 24 hr, 116; 72 hr, 117. The labeled samples were pooled and resolved into 20 fractions using strong cation exchange (SCX) [159]. Eluted fractions were vacuum dried and desalted using SEP-PAK C18 cartridges (Waters, Milford, Massachusetts, USA). Dried peptides were stored at -80 ℃ before MS analysis.

LC-MS/MS analysis

The LC-MS/MS analysis was performed as previously described [159, 178] with some modifications. Briefly, dried iTRAQ-labeled peptide samples were dissolved in HPLC grade water (J.T.Baker, Phillipsburg, New Jersey, USA) acidified with 0.1% formic acid, and sequentially injected and separated in a home-packed nanobored C18

column with a picofrit nanospray tip (75 μm ID × 15 cm, 5 μm particles) (New Objectives, Woburn, Massachusetts, USA) on a TempoTM nano-MDLC system coupled with a QSTAR® Elite Hybrid LC-MS/MS system (Applied Biosystems). Each sample was divided into two equal aliquots and independently analyzed by the LC-MS/MS over a gradient of 120 min. The flow rate of LC system was set constantly at 300 nL/min. Data acquisition in QSTAR Elite was set to positive ion mode using Analyst® QS 2.0 software (Applied Biosystems). Precursors with a mass range of 300-2000 m/z and a calculated charge of +2 to +4 were selected for fragmentation. For each MS spectrum, a maximum of three most abundant peptides above 5 count threshold were selected for MS/MS. Each selected precursor ion was dynamically excluded for 30 s with a mass tolerance of 0.03 Da. Smart IDA was activated with automatic collision energy and automatic MS/MS accumulation. The fragment intensity multiplier was set to 20 and maximum accumulation time was 2 s.

MS spectrum analysis

Spectra acquired in LC-MS/MS system from the two independent runs were submitted in a batch to ProteinPilot (v2.0.1, Applied Biosystems) for peak-list generation, as well as protein identification and quantification against the International Protein Index (IPI) human database (version 3.34; 67758 sequences) supplemented with porcine trypsin. The Paragon algorithm in ProteinPilot software was configured as previously described [178] with some modifications. Briefly, default parameters including fixed and variable modifications for samples digested using trypsin and labeled with 4-plex iTRAQ reagents (peptide labeled) were employed. The search was done thoroughly where all cleavage variants were

considered. The confidence threshold for both peptide and protein identification was set to 70%. Default precursors and fragments mass tolerances for QSTAR ESI MS instrument were adopted by the software. A concatenated target-decoy database search strategy was also employed to estimate the false discovery rate (FDR) [179]. FDR was calculated as two folds of the percentage of decoy matches divided by the total matches. After stringent filtering as described in Results, FDR of the reported iTRAQ data set was <1%. ProteinPilot software employed the peak area of iTRAQ reporters for quantification. Details of the quantification algorithm can be found in the supplier's manual. Isoform-specific strategy was adopted to deal with quantification of isoforms. Quality control of the data set is addressed in Results.

Bioinformatics

Gene IDs of the proteins of interest were searched in a batch using PANTHER classification system [180] against NCBI (*H. sapiens*) dataset and the results were presented as genes. Most protein groups had more than one molecular function hit. Cellular localization information of the 50 proteins of interest was checked manually in Gene Ontology [167].

Western blotting

Western blotting was performed using primary antibodies at the dilutions indicated: 1:500 SDHB (clone 21A11), 1:500 NDUFS3 (clone 17D95), 1:1000 VDAC1 (clone 20B12), 1:1000 MET (clone C-12), 1:1000 phospho-MET (Y1234/1235), 1:1000 phospho-MET (Y1349), 1:1000 E-cadherin (G-10), 1:2500 actin (Clone C4), 1:2000

alpha-tubulin (clone B-7). Phospho-MET antibodies were from Cell Signaling (Danvers, Massachusetts, USA), actin antibody was from Millipore (Billerica, Massachusetts, USA), while the other primary antibodies were from Santa Cruz Biotechnology, Inc (Santa Cruz, California, USA). Antibody against integrin αL (1:500), MHM23, was kindly from Dr Alex Law (School of Biological Sciences, Nanyang Technological University, Singapore).

Mitochondrial membrane potential analysis

Cells with or without PHA-665752 treatment were washed with ice-cold PBS and incubated with 5 μg/ml rhodamine 123 for 1 hr, followed by flow cytometric analysis on FACS Calibur and CellQuest Pro software (Becton Dickinson, Franklin Lakes, New Jersey, USA).

Mitochondrial permeability transition pore analysis

The activity of mitochondrial transition pore was evaluated by the MitoProbe™ Transition Pore Assay Kit (Becton Dickinson) following the manufacturer's instruction. Briefly, cells were washed twice with ice-cold Hanks' balanced salt solution (HBSS) containing 1.3 mM calcium (Invitrogen) before incubation in the presence or absence of cobalt chloride at 37 ℃ for 15 min, followed by flow cytometry analysis as described above.

Confocal microscopy

SNU5 cells were washed with HEPES twice, before incubating with 500 nM Mito Tracker Red CMXRos (Invitrogen) for 15 min. Cells were then fixed in 3% paraformaldehyde for 20 min and permeabilized with 0.1% Triton X-100 for 2 min. After blocking nonspecific antibody binding sites with 1% BSA for 1 hr at 37 ℃, cells were probed with primary antibodies (1:500) overnight at 37 ℃ and Alexa 488-conjugated goat-anti-rabbit secondary antibodies (Invitrogen) for 1 hr at 37 ℃. Finally the cells were washed with PBS and counterstained with Vectashield mounting medium with DAPI (Vector Laboratories, Burlingame, California, USA). Images were captured with a Zeiss LSM 710 confocal microscope.

Mitochondria isolation

Mitochondria isolation kit (Miltenyi Biotec, Bergisch Gladbach, Germany) was employed to isolate mitochondria following the manufacturer's protocol. Briefly, 5×10^7 SNU5 cells with or without treatment were washed twice with PBS, and lysed in 2 ml of the provided lysis buffer supplemented with Complete Protease Inhibitor Cocktail Tablets and phosSTOP (Roche, Basel, Switzerland). The crude cell lysate was incubated with anti-TOM22 MicroBeads for 1 hr at 4 ℃ with gentle shaking. Subsequently, the suspension was loaded onto a pre-equilibrated MACS column, washed thrice with separation buffer before removing the column from the magnetic field and eluting the mitochondria.

Results

MET expression and susceptibility of gastric cancer cells to PHA-665752

As PHA-665752 is differentially cytotoxic in cancer cells depending on *MET* expression levels [112], we first evaluated *MET* expression data of a panel of 16 gastric cancer cell lines, (AGS, Kato III, SNU1, SNU5, SNU16, NCIN87, Hs746T, MKN7, IM95, YCC1, YCC2, YCC3, YCC6, YCC9, YCC11 and YCC16) in order to focus on a model cell line for systematic proteomics exploration. Our transcriptome data showed that SNU5 cells had markedly elevated levels of *MET* transcription (>40-fold compared with normal human stomach tissues), while *MET* expression of SNU1 cells was comparable to the controls (Figure 5.1). MET protein expression levels of these two cell lines were compared by immunoblotting (Figure 5.2). SNU5 and SNU1 cells showed, respectively, high and low expression of MET, in agreement with our transcriptome data as well as a previous study [112]. We determined cytotoxic responses of the two gastric cancer cell lines to PHA-665752 using MTT assay (Figure 5.3). The mean IC50 of PHA-665752 in SNU5 cells was approximately 77 nM, while SNU1 cells were relatively resistant to the compound (IC50>500 nM). SNU5 was selected as the model cell line in subsequent temporal quantitative proteomics analyses because it was highly sensitive to PHA-665752. Conversely, SNU1 was chosen as being representative of gastric cancer cells resistant to MET inhibition in functional studies.

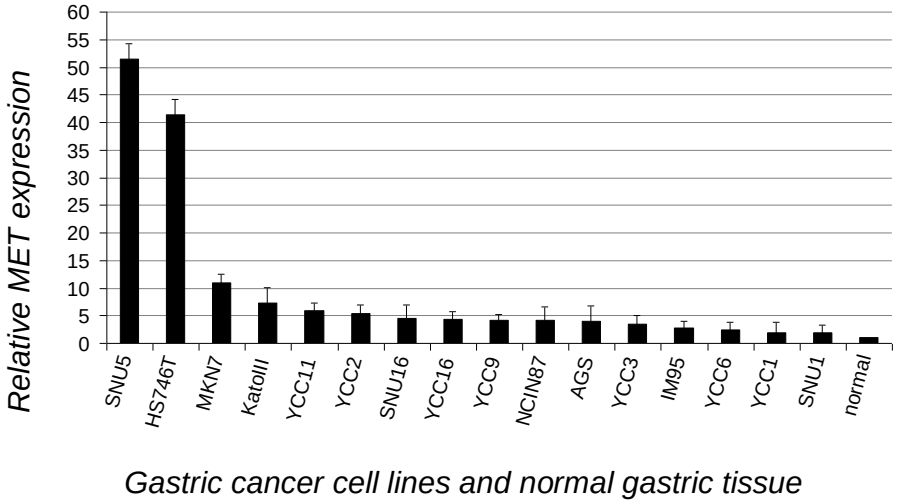


Figure 5.1. Relative *MET* expression and susceptibility of GC cells to PHA-665752. *MET* mRNA expression in 16 gastric cancer cell lines relative to pooled normal human gastric tissues was determined from transcriptome data sets.

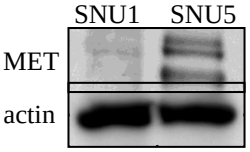


Figure 5.2. MET expression in SNU5 and SNU1. SNU5 and SNU1 cells were lysed for western blotting using anti-MET antibody. Actin immunostaining served as the loading control.

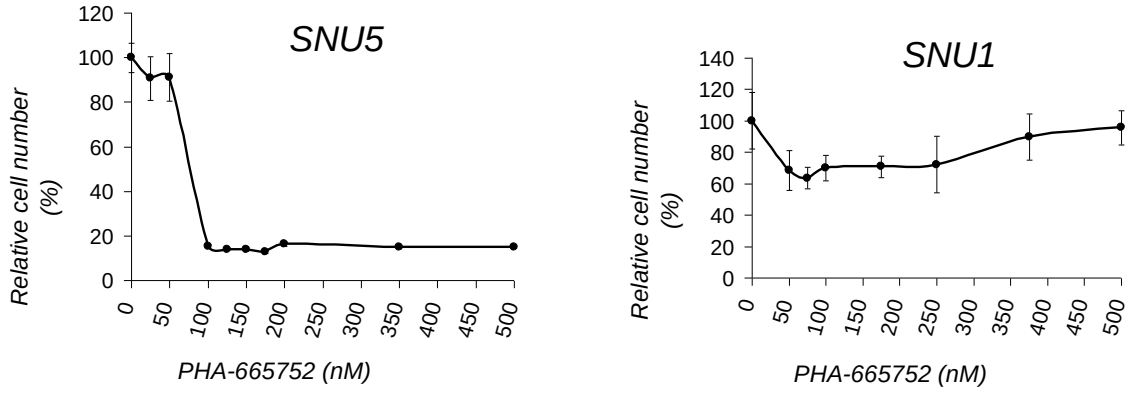
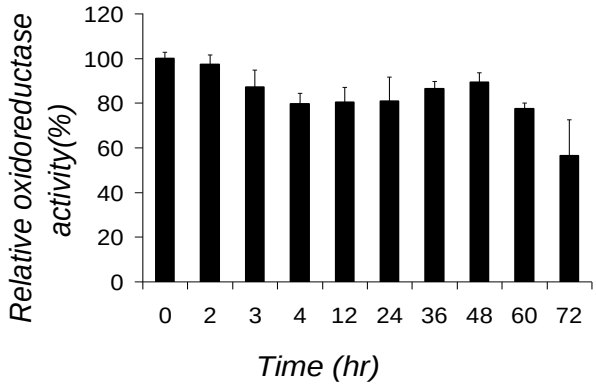


Figure 5.3. Effects of PHA-665752 on SNU5 and SNU1 cells. MTT assays were used to determine effects of PHA-665752 on viability of selected gastric cancer cell lines: (B) SNU5, (C) SNU1. Treatment duration with PHA-665752 was 72 hr. Each MTT data point was averaged from 4 replicates. Error bars at PHA-665752 concentrations ≥100 nM in panel B are not visible at this resolution.

Temporal quantitative proteomics analysis

We treated SNU5 cells with PHA-665752 and analyzed the temporal dynamics of the proteome. First, we sought to determine an appropriate concentration of PHA-665752 for quantitative proteomic investigation of SNU5 cells, in order to trace early cellular responses of the cells to MET inhibition. The ideal treatment conditions with PHA-665752 should suppress phosphorylation of MET without causing substantial cell death. SNU5 cells were exposed for varying durations to two concentrations of PHA-665752 around its IC50 (determined at 72 hr) and tested for viability using MTT assays. Our results showed that 50 nM PHA-665752 did not significantly impair cell viability, whereas 150 nM was rapidly cytotoxic (Figure 5.4). As such, we regarded 50 nM as a sublethal concentration for SNU5 cells, and adopted these conditions for the subsequent proteomics study.

A



B

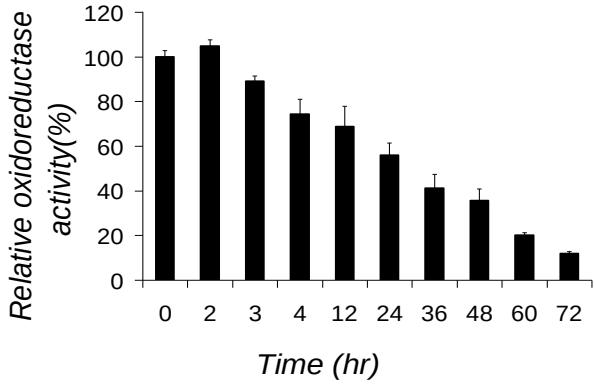


Figure 5.4. MTT assays of SNU5 cells in response to PHA-665752. (A) 50 nM and (B) 150 nM PHA-665752 were used to treat SNU5 cells. Each data point was averaged from four replicates.

It is worth noting that it remains a daunting challenge that some small molecule kinase inhibitors exhibit off-target effects that cannot be ignored [181]. To minimize off-target effects in this study, we selected one of the most potent and specific MET inhibitors PHA-665752, which is >20 times more selective for MET than other protein kinases [115]. Moreover, we applied it to a cell line SNU5 that overexpresses MET at unusually high levels *i.e.* >40 times higher than normal stomach tissue. In addition, we intentionally employed a low concentration of PHA-665752, i.e. 50 nM, which is sublethal to SNU5 cells but sufficient to inhibit MET activity. This further refined the data as arising from specific inhibition of MET since most off-target effects happen when inhibitors are used at high concentrations, such as >1 uM. We

believe, in this scenario, the probability of inhibiting other proteins with even comparable or higher affinity than MET is low or negligible. Furthermore, previous work has documented that the differential effects of PHA-665752 are truly attributed to its effect on MET using small interfering (si) RNA targeting the *MET* receptor transcript in SNU5 cells [112]. Finally, growth factor effectors in the downstream of MET signaling pathway, including ERK1/2, AKT, STAT3 and FAK, were effectively abrogated by 50 nM PHA-665752 [112]. Thus, MET could reasonably be considered the main target of 50 nM PHA-665752 in SNU5 cells in this study.

We used iTRAQ reagents to label the tryptically digested proteome, coupled with shotgun multi-dimensional liquid chromatography and tandem mass spectrometry [182] to profile the temporal proteome responses (Figure 5.5A). This approach allowed simultaneous comparison of the proteomes at four time points (0, 4 hr, 24 hr, and 72 hr) after PHA-665752 treatment, to capture both the early and late responses of the SNU5 proteome.

Quality control of quantitative MS data set

To ensure the reliability of the quantitative datasets, three independent biological replicates of SNU5 cells were pooled for the proteomics study (Figure 5.5A). Moreover, the iTRAQ-labeled samples were analyzed twice by LC-MS/MS to minimize technical variations. The ProteinPilot database search in a concatenated target and decoy strategy returned 26276 target matches and only one decoy match. A total of 1908 target proteins were identified and quantified with estimated FDR of <1%. We next employed stringent inclusion criteria to filter the data set. A total of 806 proteins quantified with high confidence *i.e.* quantified from at least five

peptides, of which there are at least two unique peptides, and having error factors <1.5, were advanced to the next phase of analysis (Supplemental Table 1 in reference [103], provided in the link http://www.mcponline.org/content/9/12/2629/suppl/DC1).

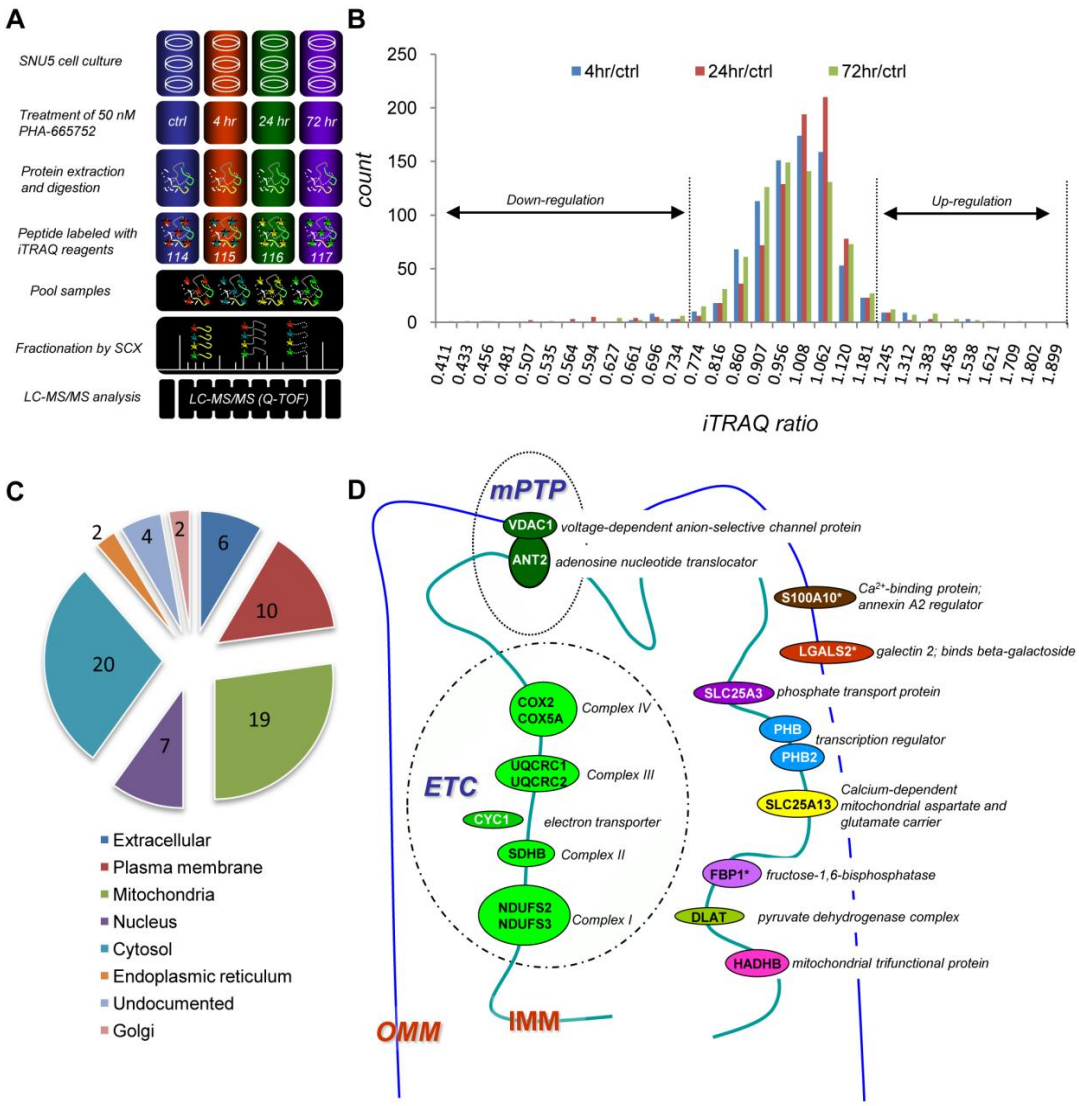


Figure 5.5. Quantitative proteomic analysis shows dominant effects on mitochondrial proteins. (A) Schematic diagram of the iTRAQ-LC-MS/MS experimental workflow. (B) Distribution of the relative expression levels of temporal proteomes. Ratios were calculated in log space before converting into linear space. (C) Subcellular classification of the differentially expressed proteins based on Gene Ontology. Details of the proteins are shown in Supplemental Table 2 (provided in the link http://www.mcponline.org/content/9/12/2629/suppl/DC1). Note that one protein may have >1 subcellular localization hits. (D) Schematic display of the functions and localizations of PHA-665752-perturbed expression of mitochondrial proteins. *, mitochondrial proteins whose exact localization within the mitochondrion is uncertain; OMM, outer mitochondrial membrane; IMM, inner mitochondrial membrane; mPTP, mitochondrial permeability transition pore; ETC, electron transport chain.

Estimation of cutoff for confidently defining perturbed proteins

The cutoff for defining perturbed and unperturbed protein expression in iTRAQ experiments depends on the characteristics of biological samples as well as MS instruments. To avoid setting the cutoffs arbitrarily, we examined the distribution of the expression levels of the 806 proteins (Figure 5.5B). The three relative expression levels of these proteins *i.e.* 4hr/control, 24hr/control and 72hr/control, were all normally distributed indicating that sublethal treatment of this compound only modulated a small percentage of the SNU5 proteome. Thus, we focused on the top 5% proteins whose expression was most perturbed by PHA-665752 treatment. With this criterion, protein ratios <0.774 were regarded as underexpressed, while ratios >1.181 were considered overexpressed, thereby narrowing the reliable differentially expressed proteins to a small number of 50 (Supplemental Table 2 in reference [103], provided in the link http://www.mcponline.org/content/9/12/2629/suppl/DC1), which reflected significant effects of MET inhibition in SNU5 cells.

Western blotting validation of iTRAQ ratios

To further evaluate the accuracy of iTRAQ ratios for the shortlisted 50 proteins, we examined the expression of three representative proteins, NDUFS3, SDHB, and VDAC1, by semi-quantitative western blot analysis. NDUFS3 and SDHB, proteins of the mitochondrial ETC, were quantified by MS with unique peptide numbers of 6 and 3, respectively, while VDAC1 a component of mPTP, was quantified with 10 unique peptides. As shown in Figure 5.6 and Table 5.1, western blot data showed similar trends corresponding iTRAQ ratios.

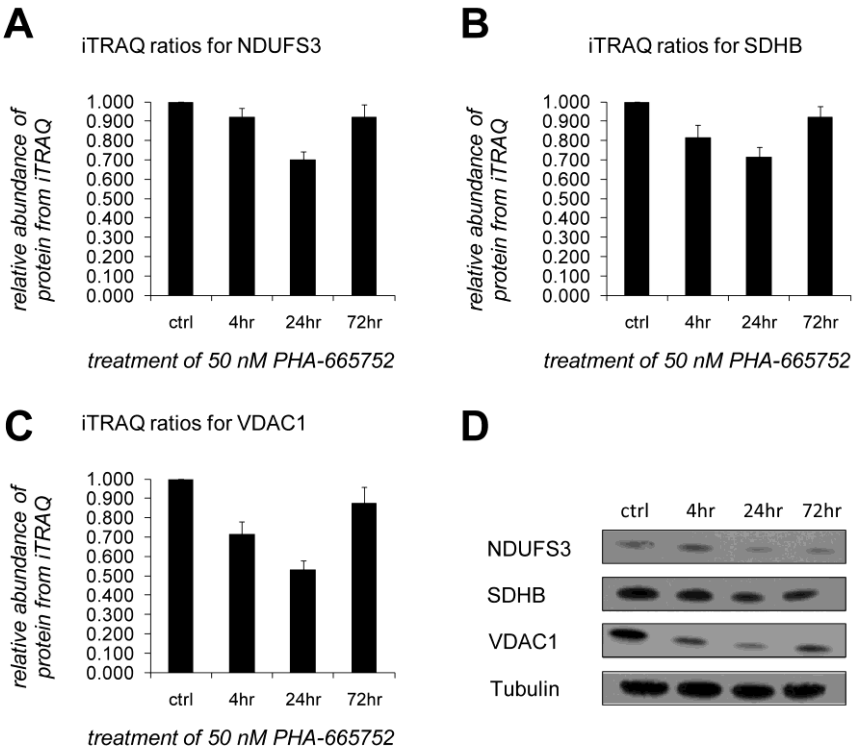


Figure 5.6. Western blotting validation of selected proteins in the iTRAQ data set. (A) DNUFS3 from ETC Complex I; (B) SDHB from ETC Complex II; and (C) VDAC1 were validated using western blot analysis. Matched results from western blots and iTRAQ (bar chart) are shown together for comparison.

Table 5.1. Differentially expressed mitochondrial proteins in MET-inhibited SNU5 cells as quantified by iTRAQ. Proteins reliably identified from the iTRAQ data set and documented to be localized in mitochondria are listed. *, proteins validated using western blotting.

Mitochondrial function	Accessions	Protein description	Gene symbol	% of sequence coverage	# of unique peptide	4h/ctrl		24hr/ctrl		72hr/ctrl	
						Ratio	N	Ratio	N	Ratio	N
ETC Complex I	IPI00025239.2	NADH dehydrogenase iron-sulfur protein 2	NDUFS2	26.1	3	0.838±0.116	6	0.613±0.058	6	0.772±0.059	6
ETC Complex I	IPI00025796.3	NADH dehydrogenase iron-sulfur protein 3	NDUFS3 *	41.7	6	0.813±0.070	5	0.712±0.055	5	0.922±0.054	5
ETC Complex II	IPI00294911.1	succinate dehydrogenase iron-sulfur subunit	SDHB *	29.6	3	0.921±0.045	5	0.701±0.041	5	0.922±0.062	5
ETC Complex III	IPI00013847.4	ubiquinol-cytochrome-c reductase complex core protein 1	UQCRC1	45.4	8	0.921±0.045	16	0.612±0.053	14	0.811±0.040	15
ETC Complex III	IPI00305383.1	ubiquinol-cytochrome-c reductase complex core protein 2	UQCRC2	42.8	4	0.863±0.066	8	0.551±0.032	8	0.802±0.062	8
ETC Complex IV	IPI00017510.3	cytochrome c oxidase subunit 2	COX2	23.3	3	0.691±0.034	13	0.431±0.029	13	0.772±0.052	13
ETC Complex IV	IPI00025086.3	cytochrome c oxidase subunit 5A	COX5A	64.7	3	0.784±0.074	8	0.615±0.080	8	0.894±0.085	8
ETC electron transporter	IPI00029264.3	cytochrome c1 heme protein	CYC1	44.3	4	0.761±0.044	6	0.603±0.057	6	0.856±0.104	6
mPTP	IPI00216308.5	voltage-dependent anion-selective channel protein 1	VDAC1 *	40.3	10	0.713±0.068	14	0.532±0.046	14	0.874±0.083	14
mPTP	IPI00007188.5	adenine nucleotide translocator 2	ANT2	72.1	10	0.818±0.114	6	0.563±0.059	6	0.814±0.085	6
others	IPI00604707.4	dihydrolipoamide S-acetyltransferase	DLAT	40.0	6	1.621±0.536	8	0.896±0.169	5	1.083±0.083	8
others	IPI00073772.5	fructose-1,6-bisphosphatase 1	FBP1	50.6	9	0.942±0.055	27	1.163±0.079	33	1.483±0.100	33
others	IPI00022793.4	trifunctional enzyme beta subunit	HADHB	45.9	5	0.784±0.074	7	0.722±0.055	6	0.926±0.104	7
others	IPI00007242.1	galectin-2	LGALS2	59.1	6	1.037±0.117	15	0.986±0.111	15	1.963±0.221	15
others	IPI00017334.1	prohibitin	PHB	58.1	6	0.844±0.080	8	0.523±0.054	8	0.766±0.156	9
others	IPI00027252.6	prohibitin-2	PHB2	63.2	6	0.862±0.058	11	0.502±0.048	11	0.642±0.049	11
others	IPI00183695.9	protein S100-A10	S100A10	70.1	4	0.661±0.038	28	0.712±0.048	28	1.062±0.062	29
others	IPI00007084.2	mitochondrial aspartate-glutamate carrier protein	SLC25A13	29.1	7	0.856±0.104	8	0.662±0.057	9	0.903±0.069	8
others	IPI00790115.1	CDNA FLJ90278 fis, mitochondrial precursor	SLC25A3	39.6	2	0.681±0.033	8	0.581±0.039	8	0.892±0.052	8

Quantitative proteomics data set reveals perturbed cellular responses after sublethal PHA-665752 treatment

The iTRAQ dataset provided identification and quantification of 806 proteins with various molecular functions as classified by PANTHER (Figure 5.7), including 50 proteins (Supplemental Table 2 in reference [103], provided in the link http://www.mcponline.org/content/9/12/2629/suppl/DC1) whose expression had been perturbed by PHA-665752-induced MET inhibition. These proteins represent various cellular responses as shown in Figure 5.7. Consistent with the specificity of MET inhibition, although proteins involved in nucleic acid binding and the cytoskeleton were the two most abundant groups, very few (<3%) were perturbed in expression by PHA-665752. Remarkably, sublethal concentration of this compound mainly affected the expression of several other groups of proteins, including oxidoreductases, calcium-binding proteins, transfer/carrier proteins involved in transport of specific substances and signaling proteins.

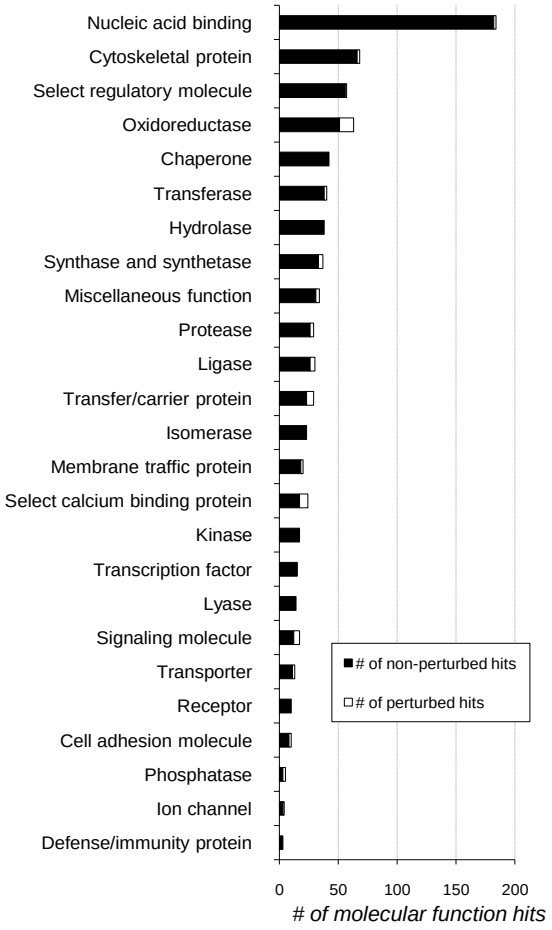


Figure 5.7. Classification of molecular functions of proteins quantified by iTRAQ. Gene IDs of the 806 proteins were batch searched using PANTHER classification system against NCBI (*H.sapiens*). 728 unique genes were mapped to a total of 891 hits of molecular functions. Fifty perturbed proteins mapped to 49 unique genes and a total of 59 hits of molecular functions. The numbers of non-perturbed and perturbed hits are shown.

Our data showed that 7 (29%) of the 24 molecular function hits were associated with calcium-binding proteins, including annexins (i.e. annexin A2 (IPI00455315.4), annexin A4 (IPI00793199.1), annexin A5 (IPI00329801.12)), calmodulin-related proteins comprising hippocalcin-like protein 1 (IPI00219344.4), protein S100-A4 (IPI00032313.1), protein S100-A10 (IPI00183695.9) and mitochondrial aspartate-glutamate carrier protein (IPI00007084.2). Modulation of calcium-binding proteins by PHA-665752 treatment suggested that calcium signaling might be a downstream response to inhibition of MET kinase activity. Calcium signaling has been extensively

documented to be tightly modulated e.g. by G protein-coupled receptors and tyrosine kinase receptors [183]. It is worth noting that the link between calcium and MET signaling is supported by the recent observation that MET, when stimulated by HGF, regulated calcium signals in human liver tumor cells [184].

Of the 29 proteins classified as having transfer/carrier function, 6 proteins (21%) responded to PHA-665752 at nanomolar concentration. Aside from the three annexin proteins and mitochondrial aspartate-glutamate carrier protein that were also classified in the calcium-binding protein category, another two phosphate carrier proteins i.e. adenine nucleotide translocator 2 (IPI00007188.5) and mitochondrial phosphate carrier protein (IPI00790115.1) were modulated by this compound.

Five of 17 (29%) proteins classified as signaling molecules, cytokine macrophage migration inhibitory factor (IPI00790382.1), myristoylated alanine-rich protein kinase C substrate (IPI00219301.7), COP9 constitutive photomorphogenic homolog subunit 8 (IPI00009480.1), galectin 2 (IPI00007242.1) and Rho GDP dissociation inhibitor alpha (IPI00794402.1), showed perturbed expression (Figure 5.7), as a probable consequence of altered signal transduction induced by PHA-665752.

As shown in Table 5.1, most of the mitochondrial proteins showed a similar trend of response to the compound. They were down-regulated at 4 hr, further down-regulated at 24 hr and partially recovered at 72 hr. This typical pattern of perturbed mitochondrial protein expression indicated a gradual but reversible effect of PHA-665752 on the SNU5 proteome.

Dominant roles of mitochondrial proteins in PHA-665752-induced MET inhibition

Interestingly, we found that many of the perturbed proteins in this data set were associated with mitochondria. Hence, we performed a bioinformatics classification based on subcellular localization information from Gene Ontology [167] to determine how many perturbed proteins were mitochondrial. Not surprisingly, a significant number of the perturbed proteins were cytosolic. However, it was noteworthy that 19 proteins, i.e. 38% of the 50 proteins whose expressions were altered by PHA-665752, were localized in mitochondria (Figure 5.5C), indicating a disproportionately dominant role of mitochondria in cellular responses to PHA-665752 treatment. Specifically, proteins from the two pivotal mitochondrial complexes *i.e.* electron transfer chain (ETC) and mitochondrial permeability transition pore (mPTP), were significantly perturbed by PHA-665752. Additionally, mitochondrial proteins involved in metabolism, signal transduction, survival and apoptosis were also affected by PHA-665752 treatment (Figure 5.5D).

Mitochondrial ETC is perturbed by PHA-665752

Eight (42%) of the 19 dysregulated mitochondrial proteins were components of the ETC in the inner mitochondrial membrane (IMM) (Figure 5.5D, Table 5.1). ETC is comprised of Complex I (NADH dehydrogenase), Complex II (succinate dehydrogenase), Complex III (cytochrome bc_1 complex) and Complex IV (cytochrome *c* oxidase). Proteins from all four ETC complexes had decreased by 4 hr after PHA-665752 treatment, were further inhibited at 24 hr and had partially recovered at 72 hr. These findings showed that PHA-665752 at a sublethal concentration rapidly modulated the expression of multiple ETC component proteins.

The ETC operates mainly through its constituent oxidoreductase enzyme activities. As the MTT assay measures cellular oxidoreductase enzyme activities, mainly of the ETC, we asked if oxidoreductase activity of the ETC was diminished by PHA-665752. By cross-referencing MTT assay data (Figure 5.4) with iTRAQ data (Table 5.1) at 24 hr of treatment with 50 nM PHA-665752, significant decreases in ETC protein expression levels were associated with only a small decrease of ETC oxidoreductase enzyme activity at the same time point, suggesting that despite inhibition of ETC protein expression, mitochondria retained substantial oxidoreductase enzyme activities required for cell survival.

One of the most important functions of ETC is to maintain the MMP [185]. The energy released by electron transport pumps protons across the IMM, generating the electrochemical and pH gradients. We examined the MMP using rhodamine 123 staining and flow cytometry [186]. Rhodamine 123, a cationic dye, has a strong emission at 529 nm that is quenched as it accumulates within mitochondrial intermembrane space, and then dequenched (with increased fluorescence) when released into the cytosol. Our data showed that rhodamine 123 fluorescence of SNU5 cells did not change upon treatment with nanomolar concentration of PHA-665752 until exposure was prolonged beyond 24 hr (Figure 5.8). The decrease in rhodamine 123 fluorescence at 48 hr and 72 hr was the evidence of an increase in MMP and hyperpolarization of the mitochondria [186]. However, MMP of the PHA-665752-resistent cell line, SNU1, did not change with the same treatment (Figure 5.8) and even at a higher concentration (300 nM, data not shown).

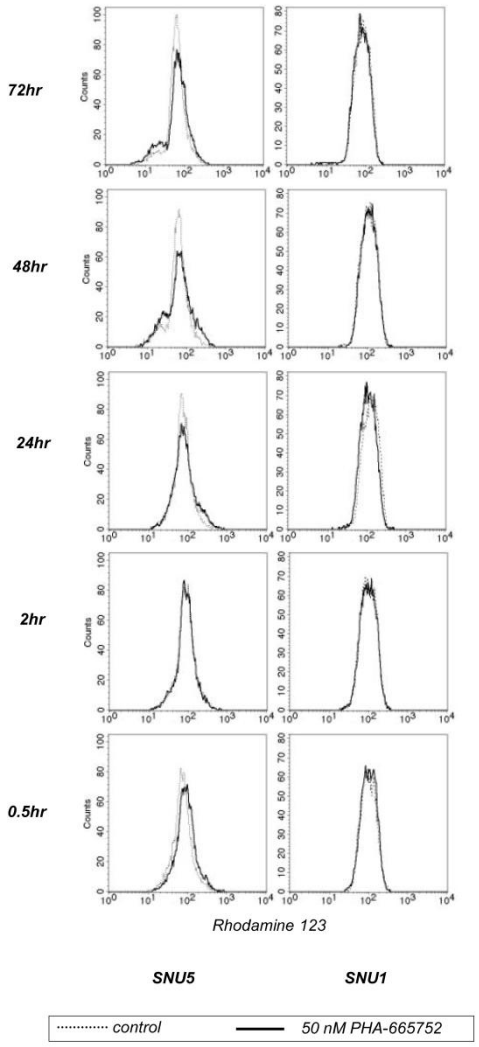


Figure 5.8. PHA-665752 effects on MMP of SNU5 and SNU1 cells. Mitochondrial membrane potentials were analyzed in SNU5 (left panel) and SNU1 (right panel) cells treated with 50 nM (bold solid line) PHA-665752 for 0.5, 2, 24, 48 and 72 hours. Control untreated cells were analyzed in parallel (dotted line). Flow cytometry data are presented as histograms of the FL1 channel.

mPTP responses to PHA-665752 treatment

We observed inhibited expression of two core components of mPTP *i.e.* VDAC1 and ANT2 [187-189], from the quantitative proteomics datasets (Table 5.1, Figure 5.5). The effect on VDAC1 was confirmed by western blot analysis (Figure 5.6). mPTP is generally regarded as a crucial channel which permits the exchange of metabolites

and ions [190]. Exchange of molecules is controlled by the flickering of mPTP between open and closed states [189]. Channel opening or an increase in the frequency of mPTP flickering is extensively documented as an event tightly associated with both necrotic and apoptotic cell death [191]. Therefore, we next asked whether mPTP was functionally associated with MET inhibition by PHA-665752. Flickering of mPTP was evaluated using the calcein AM/$CoCl_2$ method [192]. Strikingly, mPTP was rapidly and effectively inhibited by 50 nM PHA-665752 in SNU5 cells within 30 min (Figure 5.9). It is worth noting that inhibition of the mPTP was sustained for at least 72 hr. The resistant gastric cancer line SNU1 did not show inhibition of mPTP, but only rapid and transient activation of mPTP (Figure 5.9).

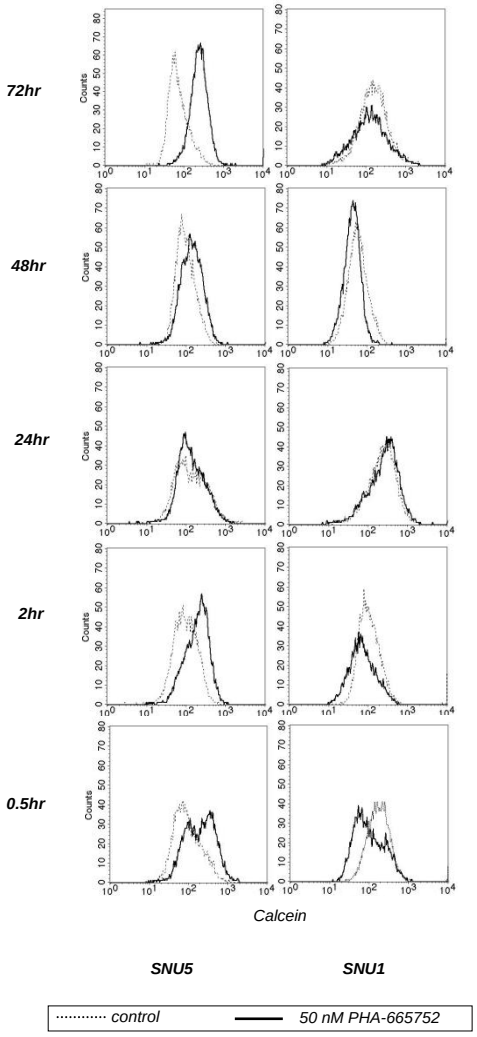


Figure 5.9. mPTP of SNU5 and SNU1 cells in response to PHA-665752 treatment. mPTP of SNU5 (left panel) and SNU1 (right panel) cells was determined after treatment with 50 nM PHA-665752 (bold solid line) for 0.5, 2, 24, 48 and 72 hours. Control untreated cells were analyzed in parallel (dotted line). Flow cytometry data are presented as histograms of the FL1 channel.

MET is present in mitochondria of SNU5 cells

The rapid effects of PHA-665752 on mitochondrial proteins and functions of SNU5 cells raised the possibility that this MET-selective inhibitor may act directly on mitochondria. Two reasons led us to hypothesize that MET might be present in mitochondria where it would be a target for PHA-665752. First, as PHA-665752 has

high specificity for MET kinase [115], the rapid mitochondrial responses could well be mediated by MET. Second, cells in which *MET* gene is highly amplified overexpress MET proteins that may be constitutively activated. High intracellular levels of activated MET could facilitate its translocation or localization to intracellular organelles such as mitochondria. Indeed EGFR, another oncogenic RTK, was found to translocate to mitochondria and nucleus when activated [193, 194]. Moreover, MET has been reported to be translocate to the nucleus when stimulated by HGF [184, 195]. However, no publication to date has reported the presence of MET in the mitochondria.

We employed two approaches to investigate whether MET is present in mitochondria of sensitive cells. Immunoblotting analysis could demonstrate the presence of MET in isolated mitochondria fraction in a semi-quantitative manner, although it may suffer from contamination of proteins from other organelles since no current technique is capable of enriching mitochondria to 100% purity. Confocal microscopic analysis provided additional complementary evidence visually.

Immunoblotting was used to probe MET and phospho-MET in SNU5 whole cell lysate and isolated mitochondrial lysate. In order to maximize mitochondria enrichment efficiency and minimize contaminations from other organelle, we employed a newly developed method for mitochondria isolation based on superparamagnetic microbeads conjugated to anti-TOM22 antibody [196]. The protocol is fast, reproducible and standardized, resulting in mitochondria of high purity, with minimal contamination from cytoskeleton, cytosol, Golgi apparatus, endosome, endoplasmic reticulum and nucleus [196]. We employed several controls

to further confirm the purity of isolated mitochondria in this study. Known plasma membrane proteins including E-calcium-dependent adhesion molecules (E-cadherin)/CD324 and integrin αL/CD18 were present in whole cell lysate, but were almost absent in mitochondria fractions, indicating minimal plasma membrane contamination (Figure 5.10A). The cytoskeleton protein, actin, displayed a similar pattern, whereas mitochondrial protein, SDHB, exhibited the opposite distribution, further proving the purity of isolated mitochondria (Figure 5.10A). Equal loading of proteins for whole cell lysate and mitochondria fraction was confirmed by Ponceau S staining [197].

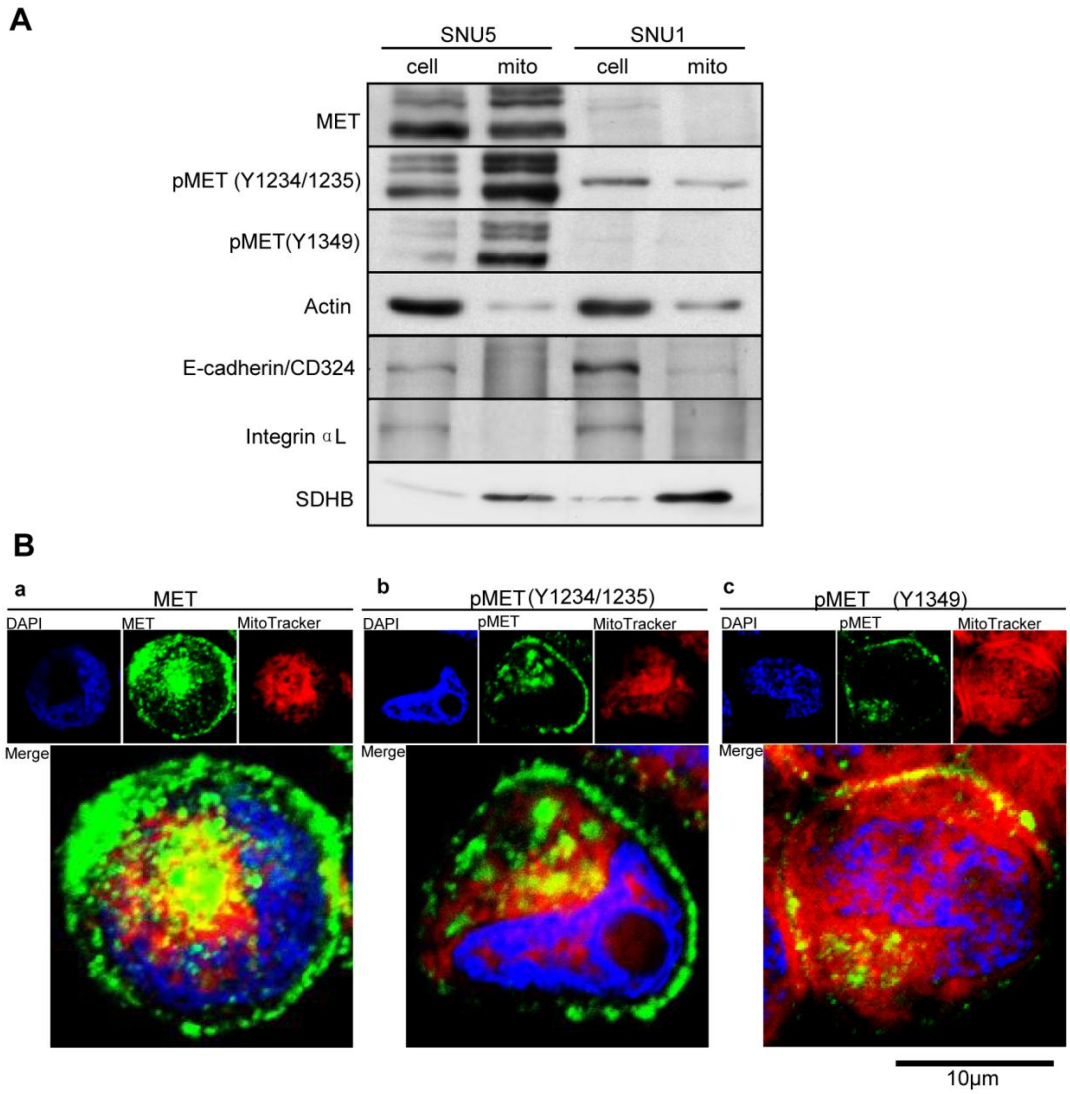


Figure 5.10. Expression of MET and phospho-MET in mitochondria. (A) Whole cell lysate and isolated mitochondrial lysate from untreated SNU5 and SNU1 were immonoblotted using antibodies as indicated. (B) Confocal images of SNU5 cells loaded with MitoTracker Red CMCROs, DAPI, and Alexa 488-conjugated antibodies against MET, phospho-MET (Y1234/1235), and phospho-MET (Y1349). Channels for DAPI, MET or pMET, MitoTracker, and merge are shown.

In striking contrast to E-cadherin, integrin αL and actin that were present in whole cell lysate but not in the mitochondrial fraction, MET was expressed at a high level in mitochondria of SNU5 cells (Figure 5.10A). Remarkably, phosphorylated MET appeared to be enriched in mitochondria. Signals of pMET (Y1234/1235) and pMET (Y1349) were higher in mitochondria than in whole cell lysate, indicating that

mitochondrial MET was highly activated. SNU1 cells demonstrated minimal expression of MET compared to SNU5 cells.

To further confirm the presence of MET in mitochondria of SNU5 cells visually, we employed confocal microscopy to determine if MET and mitochondria were co-localized. Figure 5.11B shows SNU5 cells fluorescently labeled with DAPI (blue channel), Mito Tracker Red CMXRos (red channel), and Alexa 488-conjugated antibodies against MET, pMET (Y1234/1235), and pMET (Y1349) (green channel). As expected, both MET and phosphorylated MET were found in high abundance in the plasma membrane of SNU5 cells. Remarkably, MET and phosphorylated MET also co-localized with mitochondria, as indicated by yellow colored areas in merged images. These results were consistent with immunoblotting experiments (Figure 5.10A). In contrast, SNU1 displayed very weak fluorescence for MET and phospho-MET using the same staining protocol (data not shown).

Taken together, these data provided convincing experimental evidence for the presence of MET in mitochondria of SNU5 cells.

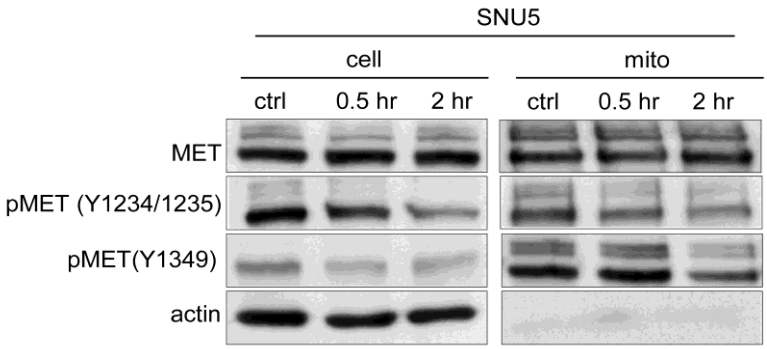


Figure 5.11. Mitochondrial MET is inhibited by MET inhibitor. Whole cell lysate and isolated mitochondrial lysate were prepared from control untreated SNU5 cells and SNU5 cells treated with 50 nM PHA-665752 for 0.5 hr and 2 hr. Antibodies against MET, phospho-MET (Y1234/1235), phospho-MET (Y1349) and actin were used in immunoblotting.

PHA-665752 inhibits phosphorylation of mitochondrial MET in SNU5 cells

Next we asked whether mitochondrial MET that was highly phosphorylated would respond to treatment with PHA-665752 for 0.5 hr and 2 hr. As shown in Figure 5.11, immunoblotting demonstrated that MET expression level in whole cell lysates was not altered by the compound at both time points but phosphorylation of MET was inhibited, consistent with previous reports [112, 115]. Importantly, phosphorylation of mitochondrial MET was inhibited in isolated mitochondrial lysates. Phosphorylation of MET at Y1234/Y1235 was suppressed by 50 nM PHA-665752 as early as 0.5 hr after treatment. Phosphorylation of Y1234/1235 was further inhibited when the treatment was prolonged for 2 hr. Phosphorylation of Y1349 in mitochondrial MET was inhibited by PHA-665752 at 2 hr.

Discussion

A curious observation in cancer research is that some cancers appear to be highly dependent for survival and progression on a single oncogene that is usually overexpressed [104, 106]. MET has essential functions in both normal and malignant cells [198]. It is not entirely clear why gastric cancer cells that overexpress MET are paradoxically much more sensitive to MET inhibition than cells whose MET expression is comparable to normal stomach [112].

In an effort to dissect the underlying molecular mechanisms, we have established a model to inhibit MET activity using a potent MET inhibitor PHA-665752 in a sensitive gastric cancer cell line SNU5. A resistant line SNU1 was utilized as control. A sublethal concentration of PHA-665752, sufficient to inhibit MET activation, was intentionally utilized to minimize off-target effects, and to investigate early cellular responses. Previously it has been demonstrated that effects of PHA-665752 are truly attributed to its effect on MET in SNU5 cells [112]. iTRAQ-based proteomics experiments allowed us to identify and quantify >1900 proteins with a FDR<1%. After stringent quality control of the data set, we narrowed down our focus to >800 proteins with at least five peptides quantified, and 50 proteins that were significantly deregulated upon MET inhibition. Our study revealed several cellular responses to PHA-665752 treatment in SNU5 cells, including modulation of oxidoreductases, calcium-binding proteins, transfer/carrier proteins involved in transport of specific substances and signaling proteins. Remarkably, we observed deregulation of mitochondrial proteins from ETC, mPTP, among others, formed a dominant protein group among the 50 deregulated proteins.

The observations prompted us to ask whether perturbation of mitochondrial functions is one of the early responses of PHA-665752 treated SNU5 cells. It is worth noting that, in the current literature, aberrant mitochondrial functions are mainly documented when cells are induced to apoptosis or other forms of cell death by lethal interventions as downstream consequences of cytotoxic factors [199-201]. Here we speculated that mitochondria might be a driving determinant of cellular death-survival decisions. Our cytometric data demonstrated that mitochondria of sensitive SNU5 cells rapidly responded to sublethal concentration of PHA-665752. In contrast, the resistant cell SNU1 displayed different patterns. These data indicated that mitochondrial deregulation appeared to be early response during MET inhibition.

To further understand the mechanisms, we hypothesized that MET might be present in mitochondria of sensitive cells as a direct target of PHA-665752. Currently, MET is known to be present in the plasma membrane, and MET inhibitors, including small molecular inhibitors and monoclonal antibodies, are considered to act through plasma membrane MET. However, it is possible that MET might reside in other organelles. Both MET and EGFR, another receptor tyrosine kinase that was once considered to be in the plasma membrane only, are capable of translocating to the nucleus [184, 194]. Activated EGFR translocates to mitochondria where it actively modulates mitochondrial proteins [193]. Our immunoblotting and confocal data both demonstrated that MET is present in mitochondria. Importantly, mitochondrial MET is hyperphosphorylated, indicating it is functionally highly active. In contrast, SNU1 cells display minimal expression of MET and phosphorylated MET. Furthermore, we demonstrated that PHA-665752 could effectively suppress the phosphorylation of

mitochondrial MET. These data suggest that mitochondrial MET is a direct target of PHA-665752.

Accumulating evidence suggests that many mitochondrial proteins, including ETC and mPTP components, that maintain the functions of this organelle, are tightly modulated by protein kinases including protein kinase A (PKA), PI3K/Akt/PKB, Raf-MEK-ERK, and MAPK [149, 202]. Phosphorylated mitochondrial MET may play active roles in modulating phosphorylation status of various mitochondrial proteins involved in mitochondrial functions. Deregulation of mitochondrial proteins and dysfunctional mitochondrial processes are likely consequences of inhibition of mitochondrial MET. Our findings also plausibly link RTK-targeted therapeutics with the Warburg effect [203]. Mitochondria are key regulators of glycolysis. We have found expression of several glycolysis-associated mitochondrial proteins, for example, VDAC1 and ANT2, was concurrently regulated rapidly in response to PHA-665752 treatment in SNU5 cells (Supplemental Table 2 in reference [103], provided in the link http://www.mcponline.org/content/9/12/2629/suppl/DC1).

Our observations have uncovered novel mechanisms through which a kinase inhibitor directly acts on mitochondrial targets and influences mitochondrial functions in sensitive cells. This may explain the rapidly lethal effects of some targeted therapies and advance understanding of how these anti-cancer agents work. As localization of MET to mitochondria was not found in resistant cells, this may be a hallmark of sensitive cells and have potential implications in personalized cancer therapeutics.

Our findings do not contradict but rather enhance the conventional paradigm that MET inhibitors including small molecule inhibitors and monoclonal antibodies act on plasma membrane MET [114]. As shown in Figure 5.11, phosphorylation of MET in whole cell lysate was rapidly inhibited by 50 nM PHA-665752, consistent with global suppression of MET molecules by the inhibitor as others have reported [112, 115]. As reported previously, baseline phosphorylation of downstream signaling effectors, such as ERK1/2, AKT, STAT3, and FAK, were effectively abrogated by 50 nM PHA-665752, confirming that canonical MET signaling pathway was inhibited by this compound [112]. However, this study provides evidence for an additional and novel locus of MET inhibition i.e. mitochondrial MET, that may be critical in contributing to the efficacy of MET inhibition in sensitive cells.

We are still unclear about the origin of the mitochondrial MET. One possibility is translocation of auto-phosphorylated plasma membrane MET to mitochondria. It is also possible that a fraction of synthesized MET is directly localized to mitochondria due to certain signal peptide sequence or post-translational modification. Localization of MET to mitochondria may correlate with the high genetic expression of MET in sensitive cells. The substrates of mitochondrial MET remain to be determined. Future work is also needed to demonstrate whether the failure of MET to localize to mitochondria affects the behavior and response of sensitive cells.

Conclusion

In an effort to understand molecular mechanisms of the curious sensitivity of MET-overexpressing cancer cells to MET inhibition, we have uncovered novel mechanisms of MET inhibition in sensitive cells. In response to low concentration of MET inhibitor, sensitive cells displayed substantial deregulation of mitochondrial proteins and functions. This study is the first to show the presence of activated MET in mitochondria of sensitive cancer cells which might be a direct target of MET inhibitor.

Chapter 6. Conclusions and Future Directions

Despite the global importance of gastric cancer as a major killer, the underlying molecular abnormalities of this dreaded disease are poorly understood. Although MS-based proteomics offers a powerful tool to systematically dissect the molecular mechanisms of gastric oncogenesis, this technology has not thus far been fully exploited as evidenced by the paucity of published studies to date. Proteomic techniques have evolved rapidly in recent years. Two dimensional gel electrophoresis (2-DE), which was once the method of choice for proteomic studies, suffers from limited sensitivity and bias towards soluble proteins. Current reports of the gastric cancer proteome were mainly from 2-DE-based investigations. Advanced LC-MS/MS-based shotgun proteomics has gained worldwide popularity in the field of proteomics, and has been applied to study a variety of diseases, including cancers. However, there is scarce information about gastric cancer from MS-based shotgun proteomics .

We started the thesis by profiling aberrantly modified proteins at post-translational level in various gastric cancer cell lines. Two post-translational modifications were investigated, *i.e.* phosphorylation and methylation.

Phosphorylated proteins are modified with a phosphate group which imposes negative charge(s) to serine, threonine or tyrosine residues in proteins, and thereby modulates the functions of proteins. Signal transduction is mediated by relays of phosphorylation events. Oncogenic signaling pathways are usually feature aberrant activation of protein kinases which catalyze phosphorylation of proteins that promote proliferation

and suppress apoptosis of cancer cells. The main challenge in investigating phosphorylated proteins is that they are present in substoichiometric amount. Enrichment of phosphorylated proteins/peptides is essential for their characterization in shotgun proteomics. In Chapter 2, we applied a complementary phosphoproteomics protocol, *i.e.* SCX-IMAC coupled with ERLIC [33] to comprehensively catalogue phosphoproteome of five gastric cancer cell lines. Stringent filtering criteria were applied to the data sets in order to generate a high-confidence inventory of the gastric cancer phosphoproteome. Phosphoproteins from primary GC tissues (cancerous tissue *versus* matched normal tissue) were also investigated using antibody arrays, which for the first time shed light on differential phosphoprotemic profiles between gastric cancer and matched histologically normal gastric epithelium. To gain more insights into the expression levels of phosphoproteins, we integrated proteomic data with transcriptomic data sets. One hundred and ninety phosphorylated proteins which were over-expressed at mRNA levels were highlighted in the integrative analysis. This has led to the discovery of overpresentation of DNA damage response (DDR) pathway in gastric cancer cells. Our data also provide the first global view of the gastric oncokinome and oncophosphatome. By integrating transcriptional expression levels of 221 protein kinases and 80 protein phosphatase genes in a panel of 17 gastric cacner cell lines with phosphoproteomics data sets, work in this thesis helps to define the dynamic molecular terrain of protein kinases and protein phosphatases from which key pathways in gastric concogenesis may be discerned. Our data also provide insights of phosphorylated mitochondrial proteins that were aberrantly expressed in gastric cancer cells.

Much less attention has been directed to cellular functions mediated by protein methylation, which is relatively more stable than phosphorylation. In Chapter 3, we performed the first large scale analysis of the gastric cancer methyl proteome. Through data mining from over 400 MS analyses, we confidently uncovered >3000 unique K and R methylation sites from a gastric cancer cell line, SNU5. Large scale analysis of protein methylation allows, for the first time, sequence analysis of methylated proteins. Lysine and arginine residues close to methionine residues were frequently methylated. This cryptic observation merits further investigations. In good agreement with the literature, we found that methylated proteins tend to contain protein domains that interact with nucleic acids. This is probably because there is a rich pool of adenosines in nucleic acids that could potentially produce AdoMet in the presence of methionine. Remarkably, our data showed that almost all the keys enzymes involved in glycolysis, citric acid cycle and oxidative phosphorylation were methylated in cancer cells. The methylation marks imposed to these enzymes may have regulated their functions and directed energy production preferentially to glycolysis instead of mitochondrial oxidative phosphorylation.

The cell surface proteome is a rich reservoir of protein biomarkers of pathological importance. However, this reservoir remains largely untapped because a systematic investigation with comprehensive documentation of gastric cancer surface proteins has not hitherto been attempted. We characterized cell surface proteome of six gastric cancer cells in Chapter 4. “Faces” of gastric cancer were depicted for the first time, which allows us to stratify gastric cancers by their surface proteins. The validity of the cell surface proteins identified by proteomics was further confirmed by flow cytometry and GC tissue microarray immunostaining. A signature comprising four

surface proteins, *i.e.* MET, EphA2, FGFR4, CD104/ITGB4, was preferentially expressed in intestinal type gastric cancers, whereas diffuse type gastric cancer cells tended to underexpress these proteins.

Of the aberrantly expressed proteins we identified from the phosphoproteome, methyl proteome and cell surface proteins, MET was one of the most dominant protein abnormalities in gastric cancer cells. MET-directed targeted therapy is considered to be a promising strategy to control gastric cancers. However, the molecular mechanisms for why MET inhibition is extremely effective in a subset of gastric cancers, but not in others, are not fully understood. Therefore, we attempted to elucidate possible mechanisms by quantitative proteomics as described in Chapter 5. Unexpectedly, we found that proteins significantly perturbed by a MET inhibitor, PHA-665752, included oxidoreductases, calcium-binding proteins, transfer/carrier proteins and signaling proteins. Bioinformatic analysis showed that 39% of the perturbed proteins were mitochondrial. Biochemical assays revealed that the mitochondrial permeability transition pore (mPTP) was rapidly inhibited by the compound, followed by altered mitochondrial protein expression and mitochondrial hyperpolarization in inhibitor-sensitive SNU5 cells but not in a resistant cell line, SNU1. This prompted us to speculate that MET inhibition by PHA-665752 in sensitive GC cells modulates mitochondrial functions. Further experiments with isolated mitochondria showed the presence of highly activated MET in mitochondria, and striking suppression of MET activation by 50 nM PHA-665752. Taken together, our data indicate that mitochondrial MET is a direct target of MET kinase inhibition, in addition to plasma membrane MET. Effects on activated MET in the mitochondria

of cancer cells that are sensitive to MET inhibition might constitute a novel and critical non-canonical mechanism for the efficacy of MET-targeted therapeutics.

In summary, this thesis has documented a few comprehensive proteomic studies of gastric cancers that enabled us to discover patterns of aberrantly expressed proteins and the likely importance of mitochondrial function in gastric cancer cells. These findings contribute to global efforts to deepen understanding of molecular mechanisms that initiate and perpetuate gastric oncogenesis.

The aberrant oncogenic signaling pathways uncovered in this thesis merit future mechanistic studies. Multiple proteins in the DDR pathway were up-regulated and modulated by protein kinases. It would be of interest to investigate and understand the underlying mechanisms of this up-regulation, and roles of protien kinases responsible for this modulation. Extensive methylation of metabolism-related enzymes suggests cancer metabolism may be tightly modulated by the corresponding enzymes. Based on these data, a hypothesis could be proposed that proteins involved in cancer metabolism are regulated by protein methylation. Identification of the enzymes responsible for these methylation events is of interest in future work to interrogate molecular mechanisms of altered metabolism in cancers. Our surface proteome shortlisted a few potential markers for gastric cancer. Although some have been validated using tissue microarray of a cohort of 49 sample pairs, validation in larger independent cohorts is required in the future.

Much of our efforts have focused on the discovery of mitochondrial MET. This non-classical localization of MET could be translocated from plasma membrane upon

auto-phosphorylation. It may also result from direct translocation to mitochondria immediately after protein synthesis, probably via a mitochondrial signal sequence, or via an endocytosis pathway. It is intriguing to ask what the source of mitochondrial MET might be and why MET is present in the mitochondria in some cancer cells but not in others. Another critical question is to uncover the functions of mitochondrial MET. A hypothesis could be proposed that mitochondrial MET phosphorylates some mitochondrial proteins which in turn influence mitochondrial functions including apoptosis and/or metabolism. Future experiments could be designed to interrogate the protein substrates of mitochondrial MET and their associations with mitochondrial functions.

Our data showed mitochondrial MET was plausibly associated with the sensitivity of cancer cells to a MET kinase inhibitor. However, the effect was indistinguishable from that of MET localized in the plasma membrane. It would be of interest to investigate the actions of mitochondrial MET and plasma membrane MET separately, and ask the question which localization of MET is the decisive factor that impacts drug sensitivity. If mitochondrial MET is proven to be a critical factor contributing to the drug sensitivity, we could generate a hypothesis that mitochondrial kinases are present in other types of cancer cells (*i.e.* is not limited to gastric cancer cells), and their presence determines responses to agents targeted at the particular kinase. Experiments could be designed to profile the repertoire of activated protein kinases in various cancer cells, and whether their presence affects the sensitivity of the cancer cells to molecularly-directed agents.

References

1. Raven, R.W., *The Theory and Practice of Oncology: Historical Evolution and Present Principles.* Informa Healthcare, New York, USA, 1990.
2. Forman, D. and V.J. Burley, *Gastric cancer: global pattern of the disease and an overview of environmental risk factors.* Best Pract Res Clin Gastroenterol, 2006. **20**(4): p. 633-49.
3. International Agency for Research on Cancer (IARC) Working Group, *IARC monographs on the evaluation of carcinogenic risk to humans.* Schistosomes, liver flukes and Helicobacter pylori, Volume 61, Lyon, France, 1994.
4. Parkin, D.M., et al., *Global cancer statistics, 2002.* CA Cancer J Clin, 2005. **55**(2): p. 74-108.
5. Kamangar, F., G.M. Dores, and W.F. Anderson, *Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world.* J Clin Oncol, 2006. **24**(14): p. 2137-50.
6. Mardis, E.R., *Cancer genomics identifies determinants of tumor biology.* Genome Biol, 2010. **11**(5): p. 211.
7. Lauren, P., *The Two Histological Main Types of Gastric Carcinoma: Diffuse and So-Called Intestinal-Type Carcinoma. An Attempt at a Histo-Clinical Classification.* Acta Pathol Microbiol Scand, 1965. **64**: p. 31-49.
8. Stratton, M.R., *Exploring the genomes of cancer cells: progress and promise.* Science, 2011. **331**(6024): p. 1553-8.
9. Dulbecco, R., *A turning point in cancer research: sequencing the human genome.* Science, 1986. **231**(4742): p. 1055-6.

10. McDermott, U., J.R. Downing, and M.R. Stratton, *Genomics and the continuum of cancer care.* N Engl J Med, 2011. **364**(4): p. 340-50.

11. Surinova, S., et al., *On the development of plasma protein biomarkers.* J Proteome Res, 2011. **10**(1): p. 5-16.

12. Sawyers, C., *Targeted cancer therapy.* Nature, 2004. **432**(7015): p. 294-7.

13. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.

14. Di Cosimo, S. and J. Baselga, *Management of breast cancer with targeted agents: importance of heterogeneity. [corrected]*. Nat Rev Clin Oncol, 2010. **7**(3): p. 139-47.

15. Esteva, F.J., et al., *Molecular predictors of response to trastuzumab and lapatinib in breast cancer.* Nat Rev Clin Oncol, 2010. **7**(2): p. 98-107.

16. Daub, H., K. Specht, and A. Ullrich, *Strategies to overcome resistance to targeted protein kinase inhibitors.* Nat Rev Drug Discov, 2004. **3**(12): p. 1001-10.

17. Xu, A.M. and P.H. Huang, *Receptor tyrosine kinase coactivation networks in cancer.* Cancer Res, 2010. **70**(10): p. 3857-60.

18. Engelman, J.A., et al., *MET amplification leads to gefitinib resistance in lung cancer by activating ERBB3 signaling.* Science, 2007. **316**(5827): p. 1039-43.

19. Nahta, R., et al., *Mechanisms of disease: understanding resistance to HER2-targeted therapy in human breast cancer.* Nat Clin Pract Oncol, 2006. **3**(5): p. 269-80.

20. Stommel, J.M., et al., *Coactivation of receptor tyrosine kinases affects the response of tumor cells to targeted therapies.* Science, 2007. **318**(5848): p. 287-90.

21. Cutsem, E.V., et al., *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).* Journal of Clinical Oncology, 2009. **27**: p. 18s (suppl. abstr LBA4509).

22. Dragovich, T. and C. Campen, *Anti-EGFR-Targeted Therapy for Esophageal and Gastric Cancers: An Evolving Concept.* J Oncol, 2009. **2009**: p. 804108.

23. Iwasaki, J. and S. Nihira, *Anti-angiogenic therapy against gastrointestinal tract cancers.* Jpn J Clin Oncol, 2009. **39**(9): p. 543-51.

24. Comoglio, P.M., S. Giordano, and L. Trusolino, *Drug development of MET inhibitors: targeting oncogene addiction and expedience.* Nature reviews. Drug discovery, 2008. **7**(6): p. 504-16.

25. Arkenau, H.-T., *Gastric cancer in the era of molecularly targeted agents: current drug development strategies.* Journal of Cancer Research and Clinical Oncology, 2009. **135**(7): p. 855-866.

26. Grabsch, H., et al., *HER2 expression in gastric cancer: Rare, heterogeneous and of no prognostic value - conclusions from 924 cases of two independent series.* Cell Oncol, 2010. **32**(1-2): p. 57-65.

27. Arkenau, H.T., *Gastric cancer in the era of molecularly targeted agents: current drug development strategies.* J Cancer Res Clin Oncol, 2009. **135**(7): p. 855-66.

28. Huang, P.H. and F.M. White, *Phosphoproteomics: unraveling the signaling web.* Molecular cell, 2008. **31**(6): p. 777-81.

29. Lim, Y.P., *Mining the tumor phosphoproteome for cancer markers.* Clinical cancer research : an official journal of the American Association for Cancer Research, 2005. **11**(9): p. 3163-9.

30. Yu, L.R., H.J. Issaq, and T.D. Veenstra, *Phosphoproteomics for the discovery of kinases as cancer biomarkers and drug targets.* Proteomics. Clinical applications, 2007. **1**(9): p. 1042-57.

31. Gembitsky, D.S., et al., *A prototype antibody microarray platform to monitor changes in protein tyrosine phosphorylation.* Molecular & cellular proteomics : MCP, 2004. **3**(11): p. 1102-18.

32. Metodiev, M.V., A. Timanova, and D.E. Stone, *Differential phosphoproteome profiling by affinity capture and tandem matrix-assisted laser desorption/ionization mass spectrometry.* Proteomics, 2004. **4**(5): p. 1433-8.

33. Ali, N.A. and M.P. Molloy, *Quantitative phosphoproteomics of transforming growth factor-beta signaling in colon cancer cells.* Proteomics, 2011. **11**(16): p. 3390-401.

34. Rexer, B.N., et al., *Phosphoproteomic mass spectrometry profiling links Src family kinases to escape from HER2 tyrosine kinase inhibition.* Oncogene, 2011. **30**(40): p. 4163-74.

35. Osinalde, N., et al., *Interleukin-2 signaling pathway analysis by quantitative phosphoproteomics.* Journal of proteomics, 2011: p. doi:10.1016/j.jprot.2011.06.007.

36. Stahl, S., et al., *Phosphoproteomic profiling of NSCLC cells reveals that ephrin B3 regulates pro-survival signaling through Akt1-mediated phosphorylation of the EphA2 receptor.* Journal of proteome research, 2011. **10**(5): p. 2566-78.

37. Bensimon, A., et al., *ATM-dependent and -independent dynamics of the nuclear phosphoproteome after DNA damage.* Science signaling, 2010. **3**(151): p. rs3.

38. Diella, F., et al., *Phospho.ELM: a database of experimentally verified phosphorylation sites in eukaryotic proteins.* BMC bioinformatics, 2004. **5**: p. 79.

39. http://www.phosphosite.org.

40. Bodenmiller, B., et al., *PhosphoPep--a database of protein phosphorylation sites in model organisms.* Nat Biotechnol, 2008. **26**(12): p. 1339-40.

41. Gnad, F., et al., *PHOSIDA (phosphorylation site database): management, structural and evolutionary investigation, and prediction of phosphosites.* Genome Biol, 2007. **8**(11): p. R250.

42. Chen, L., et al., *Investigation of phosphoprotein signatures of archived prostate cancer tissue specimens via proteomic analysis.* Electrophoresis, 2011. **32**(15): p. 1984-91.

43. Yan, G.R., et al., *Characterization of phosphoproteins in gastric cancer secretome.* Omics : a journal of integrative biology, 2011. **15**(1-2): p. 83-90.

44. Yu, G., et al., *Phosphoproteome profile of human lung cancer cell line A549.* Molecular bioSystems, 2011. **7**(2): p. 472-9.

45. Ge, F., et al., *Phosphoproteomic analysis of primary human multiple myeloma cells.* Journal of proteomics, 2010. **73**(7): p. 1381-90.

46. Wang, X., et al., *Characterization of the phosphoproteome in androgen-repressed human prostate cancer cells by Fourier transform ion cyclotron resonance mass spectrometry.* Journal of proteome research, 2011. **10**(9): p. 3920-8.

47. Chen, L., F. Giorgianni, and S. Beranova-Giorgianni, *Characterization of the phosphoproteome in LNCaP prostate cancer cells by in-gel isoelectric*

focusing and tandem mass spectrometry. Journal of proteome research, 2010. **9**(1): p. 174-8.

48. Tao, W.A., et al., *Quantitative phosphoproteome analysis using a dendrimer conjugation chemistry and tandem mass spectrometry.* Nature methods, 2005. **2**(8): p. 591-8.

49. Beausoleil, S.A., et al., *Large-scale characterization of HeLa cell nuclear phosphoproteins.* Proceedings of the National Academy of Sciences of the United States of America, 2004. **101**(33): p. 12130-5.

50. Villen, J., et al., *Large-scale phosphorylation analysis of mouse liver.* Proc Natl Acad Sci U S A, 2007. **104**(5): p. 1488-93.

51. Pinkse, M.W., et al., *Selective isolation at the femtomole level of phosphopeptides from proteolytic digests using 2D-NanoLC-ESI-MS/MS and titanium oxide precolumns.* Anal Chem, 2004. **76**(14): p. 3935-43.

52. Alpert, A.J., *Electrostatic repulsion hydrophilic interaction chromatography for isocratic separation of charged solutes and selective isolation of phosphopeptides.* Anal Chem, 2008. **80**(1): p. 62-76.

53. Gan, C.S., et al., *A comparative study of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) versus SCX-IMAC-based methods for phosphopeptide isolation/enrichment.* J Proteome Res, 2008. **7**(11): p. 4869-77.

54. Bodenmiller, B., et al., *Reproducible isolation of distinct, overlapping segments of the phosphoproteome.* Nature methods, 2007. **4**(3): p. 231-7.

55. Linding, R., et al., *NetworKIN: a resource for exploring cellular phosphorylation networks.* Nucleic Acids Res, 2008. **36**(Database issue): p. D695-9.

56. Bedford, M.T. and S.G. Clarke, *Protein arginine methylation in mammals: who, what, and why.* Molecular cell, 2009. **33**(1): p. 1-13.

57. Pahlich, S., R.P. Zakaryan, and H. Gehring, *Protein arginine methylation: Cellular functions and methods of analysis.* Biochimica et biophysica acta, 2006. **1764**(12): p. 1890-903.

58. Aletta, J.M., T.R. Cimato, and M.J. Ettinger, *Protein methylation: a signal event in post-translational modification.* Trends Biochem Sci, 1998. **23**(3): p. 89-91.

59. Lee, D.Y., et al., *Role of protein methylation in regulation of transcription.* Endocrine reviews, 2005. **26**(2): p. 147-70.

60. Ambler, R.P. and M.W. Rees, *Epsilon-N-Methyl-lysine in bacterial flagellar protein.* Nature, 1959. **184**: p. 56-7.

61. Teyssier, C., et al., *Protein arginine methylation in estrogen signaling and estrogen-related cancers.* Trends in endocrinology and metabolism: TEM, 2010. **21**(3): p. 181-9.

62. Paik, W.K., D.C. Paik, and S. Kim, *Historical review: the field of protein methylation.* Trends in biochemical sciences, 2007. **32**(3): p. 146-52.

63. Lee, D.Y., et al., *Role of protein methylation in regulation of transcription.* Endocr Rev, 2005. **26**(2): p. 147-70.

64. Dai, Q., et al., *Mammalian prenylcysteine carboxyl methyltransferase is in the endoplasmic reticulum.* J Biol Chem, 1998. **273**(24): p. 15030-4.

65. Huang, J. and S.L. Berger, *The emerging field of dynamic lysine methylation of non-histone proteins.* Current opinion in genetics & development, 2008. **18**(2): p. 152-8.

66. Huang, J., et al., *Repression of p53 activity by Smyd2-mediated methylation.* Nature, 2006. **444**(7119): p. 629-32.

67. Huang, J., et al., *p53 is regulated by the lysine demethylase LSD1.* Nature, 2007. **449**(7158): p. 105-8.

68. Bedford, M.T. and S. Richard, *Arginine methylation an emerging regulator of protein function.* Molecular cell, 2005. **18**(3): p. 263-72.

69. Najbauer, J., et al., *Peptides with sequences similar to glycine, arginine-rich motifs in proteins interacting with RNA are efficiently recognized by methyltransferase(s) modifying arginine in numerous proteins.* J Biol Chem, 1993. **268**(14): p. 10501-9.

70. Springer, M.S., M.F. Goy, and J. Adler, *Protein methylation in behavioural control mechanisms and in signal transduction.* Nature, 1979. **280**(5720): p. 279-84.

71. Boisvert, F.M., et al., *A proteomic analysis of arginine-methylated protein complexes.* Mol Cell Proteomics, 2003. **2**(12): p. 1319-30.

72. Wu, C.C., et al., *Organellar proteomics reveals Golgi arginine dimethylation.* Mol Biol Cell, 2004. **15**(6): p. 2907-19.

73. Ong, S.E., G. Mittler, and M. Mann, *Identifying and quantifying in vivo methylation sites by heavy methyl SILAC.* Nature methods, 2004. **1**(2): p. 119-26.

74. Jung, S.Y., et al., *Complications in the assignment of 14 and 28 Da mass shift detected by mass spectrometry as in vivo methylation from endogenous proteins.* Anal Chem, 2008. **80**(5): p. 1721-9.

75. Cao, X.J., et al., *High-coverage proteome analysis reveals the first insight of protein modification systems in the pathogenic spirochete Leptospira interrogans.* Cell Res, 2010. **20**(2): p. 197-210.

76. Speers, A.E. and C.C. Wu, *Proteomics of integral membrane proteins--theory and application.* Chem Rev, 2007. **107**(8): p. 3687-714.

77. Leth-Larsen, R., R.R. Lund, and H.J. Ditzel, *Plasma membrane proteomics and its application in clinical cancer biomarker discovery.* Mol Cell Proteomics, 2010. **9**(7): p. 1369-82.

78. Schiess, R., B. Wollscheid, and R. Aebersold, *Targeted proteomic strategy for clinical biomarker discovery.* Mol Oncol, 2009. **3**(1): p. 33-44.

79. Overington, J.P., B. Al-Lazikani, and A.L. Hopkins, *How many drug targets are there?* Nat Rev Drug Discov, 2006. **5**(12): p. 993-6.

80. Turtoi, A., E. De Pauw, and V. Castronovo, *Innovative proteomics for the discovery of systemically accessible cancer biomarkers suitable for imaging and targeted therapies.* Am J Pathol, 2011. **178**(1): p. 12-8.

81. Van Hoof, D., et al., *Proteomics and human embryonic stem cells.* Stem Cell Res, 2008. **1**(3): p. 169-82.

82. Ahn, S.M., R.J. Goode, and R.J. Simpson, *Stem cell markers: insights from membrane proteomics?* Proteomics, 2008. **8**(23-24): p. 4946-57.

83. de Tute, R.M., *Flow cytometry and its use in the diagnosis and management of mature lymphoid malignancies.* Histopathology, 2011. **58**(1): p. 90-105.

84. Melton, S.D., R.M. Genta, and R.F. Souza, *Biomarkers and molecular diagnosis of gastrointestinal and pancreatic neoplasms.* Nat Rev Gastroenterol Hepatol, 2010. **7**(11): p. 620-8.

85. http://www.expasy.org/cgi-bin/lists?cdlist.txt.

86. Woolfson, A., et al., *The application of CD antigen proteomics to pharmacogenomics.* Pharmacogenomics, 2006. **7**(5): p. 759-71.

87. !!! INVALID CITATION !!!

88. Robinson, D.R., Y.M. Wu, and S.F. Lin, *The protein tyrosine kinase family of the human genome.* Oncogene, 2000. **19**(49): p. 5548-57.

89. Hubbard, S.R. and W.T. Miller, *Receptor tyrosine kinases: mechanisms of activation and signaling.* Curr Opin Cell Biol, 2007. **19**(2): p. 117-23.

90. Cohen, P., *Protein kinases--the major drug targets of the twenty-first century?* Nat Rev Drug Discov, 2002. **1**(4): p. 309-15.

91. Bang, Y.J., et al., *Trastuzumab in combination with chemotherapy versus chemotherapy alone for treatment of HER2-positive advanced gastric or gastro-oesophageal junction cancer (ToGA): a phase 3, open-label, randomised controlled trial.* Lancet, 2010. **376**(9742): p. 687-97.

92. Griffin, N.M. and J.E. Schnitzer, *Overcoming key technological challenges in using mass spectrometry for mapping cell surfaces in tissues.* Mol Cell Proteomics, 2011. **10**(2): p. R110 000935.

93. Schiess, R., et al., *Analysis of cell surface proteome changes via label-free, quantitative mass spectrometry.* Mol Cell Proteomics, 2009. **8**(4): p. 624-38.

94. Hofmann, A., et al., *Proteomic cell surface phenotyping of differentiating acute myeloid leukemia cells.* Blood, 2010. **116**(13): p. e26-34.

95. Shin, B.K., *Global Profiling of the Cell Surface Proteome of Cancer Cells Uncovers an Abundance of Proteins with Chaperone Function.* Journal of Biological Chemistry, 2002. **278**(9): p. 7607-7616.

96. Alfonso, P., et al., *Proteome analysis of membrane fractions in colorectal carcinomas by using 2D-DIGE saturation labeling.* J Proteome Res, 2008. **7**(10): p. 4247-55.

97. Pshezhetsky, A.V., et al., *Subcellular proteomics of cell differentiation: quantitative analysis of the plasma membrane proteome of Caco-2 cells.* Proteomics, 2007. **7**(13): p. 2201-15.

98. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative.* Nature chemical biology, 2005. **1**(5): p. 252-62.

99. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.

100. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Molecular & cellular proteomics : MCP, 2004. **3**(12): p. 1154-69.

101. Griffin, T.J., et al., *iTRAQ reagent-based quantitative proteomic analysis on a linear ion trap mass spectrometer.* J Proteome Res, 2007. **6**(11): p. 4200-9.

102. Guo, T., et al., *Hybridization of pulsed-Q dissociation and collision-activated dissociation in linear ion trap mass spectrometer for iTRAQ quantitation.* Journal of proteome research, 2008. **7**(11): p. 4831-40.

103. Guo, T., et al., *Quantitative proteomics discloses MET expression in mitochondria as a direct target of MET kinase inhibitor in cancer cells.* Mol Cell Proteomics, 2010. **9**(12): p. 2629-41.

104. Sharma, S.V. and J. Settleman, *Oncogene addiction: setting the stage for molecularly targeted cancer therapy.* Genes Dev, 2007. **21**(24): p. 3214-31.

105. Weinstein, I.B., *Cancer. Addiction to oncogenes--the Achilles heal of cancer.* Science, 2002. **297**(5578): p. 63-4.

106. Weinstein, I.B. and A. Joe, *Oncogene addiction.* Cancer research, 2008. **68**(9): p. 3077-80; discussion 3080.

107. Weinstein, I.B. and A.K. Joe, *Mechanisms of disease: Oncogene addiction--a rationale for molecular targeting in cancer therapy.* Nature clinical practice. Oncology, 2006. **3**(8): p. 448-57.

108. Kuniyasu, H., et al., *Frequent amplification of the c-met gene in scirrhous type stomach cancer.* Biochem Biophys Res Commun, 1992. **189**(1): p. 227-32.

109. Heideman, D.A., et al., *Absence of tpr-met and expression of c-met in human gastric mucosa and carcinoma.* J Pathol, 2001. **194**(4): p. 428-35.

110. Catalano, V., et al., *Gastric cancer.* Critical reviews in oncology/hematology, 2005. **54**(3): p. 209-41.

111. Peek, R.M., Jr. and M.J. Blaser, *Helicobacter pylori and gastrointestinal tract adenocarcinomas.* Nat Rev Cancer, 2002. **2**(1): p. 28-37.

112. Smolen, G.A., et al., *Amplification of MET may identify a subset of cancers with extreme sensitivity to the selective tyrosine kinase inhibitor PHA-665752.* Proc Natl Acad Sci U S A, 2006. **103**(7): p. 2316-21.

113. Zou, H.Y., et al., *An orally available small-molecule inhibitor of c-Met, PF-2341066, exhibits cytoreductive antitumor efficacy through antiproliferative and antiangiogenic mechanisms.* Cancer research, 2007. **67**(9): p. 4408-17.

114. Comoglio, P.M., S. Giordano, and L. Trusolino, *Drug development of MET inhibitors: targeting oncogene addiction and expedience.* Nat Rev Drug Discov, 2008. **7**(6): p. 504-16.

115. Christensen, J.G., et al., *A selective small molecule inhibitor of c-Met kinase inhibits c-Met-dependent phenotypes in vitro and exhibits cytoreductive antitumor activity in vivo.* Cancer research, 2003. **63**(21): p. 7345-55.

116. Okamoto, W., et al., *Identification of c-Src as a potential therapeutic target for gastric cancer and of MET activation as a cause of resistance to c-Src inhibition.* Mol Cancer Ther. **9**(5): p. 1188-97.

117. Hov, H., et al., *A selective c-met inhibitor blocks an autocrine hepatocyte growth factor growth loop in ANBL-6 cells and prevents migration and adhesion of myeloma cells.* Clin Cancer Res, 2004. **10**(19): p. 6686-94.

118. Mukohara, T., et al., *Inhibition of the met receptor in mesothelioma.* Clin Cancer Res, 2005. **11**(22): p. 8122-30.

119. Puri, N., et al., *A selective small molecule inhibitor of c-Met, PHA665752, inhibits tumorigenicity and angiogenesis in mouse lung cancer xenografts.* Cancer Res, 2007. **67**(8): p. 3529-34.

120. Accornero, P., et al., *An in vivo model of Met-driven lymphoma as a tool to explore the therapeutic potential of Met inhibitors.* Clin Cancer Res, 2008. **14**(7): p. 2220-6.

121. Bachleitner-Hofmann, T., et al., *HER kinase activation confers resistance to MET tyrosine kinase inhibition in MET oncogene-addicted gastric cancer cells.* Mol Cancer Ther, 2008. **7**(11): p. 3499-508.

122. Lim, C.S. and R.S. Walikonis, *Hepatocyte growth factor and c-Met promote dendritic maturation during hippocampal neuron differentiation via the Akt pathway.* Cell Signal, 2008. **20**(5): p. 825-35.

123. Toschi, L. and P.A. Janne, *Single-agent and combination therapeutic strategies to inhibit hepatocyte growth factor/MET signaling in cancer.* Clin Cancer Res, 2008. **14**(19): p. 5941-6.

124. Yang, Y., et al., *A selective small molecule inhibitor of c-Met, PHA-665752, reverses lung premalignancy induced by mutant K-ras.* Mol Cancer Ther, 2008. **7**(4): p. 952-60.

125. Huang, P.H., et al., *Quantitative analysis of EGFRvIII cellular signaling networks reveals a combinatorial therapeutic strategy for glioblastoma.* Proc Natl Acad Sci U S A, 2007. **104**(31): p. 12867-72.

126. Fenyo, D. and R.C. Beavis, *A method for assessing the statistical significance of mass spectrometry-based protein identifications using general scoring schemes.* Anal Chem, 2003. **75**(4): p. 768-74.

127. Geer, L.Y., et al., *Open mass spectrometry search algorithm.* J Proteome Res, 2004. **3**(5): p. 958-64.

128. Obenauer, J.C., L.C. Cantley, and M.B. Yaffe, *Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs.* Nucleic Acids Res, 2003. **31**(13): p. 3635-41.

129. Macek, B., M. Mann, and J.V. Olsen, *Global and site-specific quantitative phosphoproteomics: principles and applications.* Annu Rev Pharmacol Toxicol, 2009. **49**: p. 199-221.

130. Gan, C.S., et al., *A comparative study of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) versus SCX-IMAC-based methods for phosphopeptide isolation/enrichment.* Journal of proteome research, 2008. **7**(11): p. 4869-77.

131. Zhang, H., et al., *Simultaneous characterization of glyco- and phosphoproteomes of mouse brain membrane proteome with electrostatic repulsion hydrophilic interaction chromatography.* Mol Cell Proteomics, 2010. **9**(4): p. 635-47.

132. Sarg, B., et al., *Histone H1 phosphorylation occurs site-specifically during interphase and mitosis: identification of a novel phosphorylation site on histone H1.* J Biol Chem, 2006. **281**(10): p. 6573-80.

133. Beausoleil, S.A., et al., *A probability-based approach for high-throughput protein phosphorylation analysis and site localization.* Nat Biotechnol, 2006. **24**(10): p. 1285-92.

134. Guo, T., et al., *Global molecular dysfunctions in gastric cancer revealed by an integrated analysis of the phosphoproteome and transcriptome.* Cell Mol Life Sci, 2011. **68**(11): p. 1983-2002.

135. Schwartz, D. and S.P. Gygi, *An iterative statistical approach to the identification of protein phosphorylation motifs from large-scale data sets.* Nat Biotechnol, 2005. **23**(11): p. 1391-8.

136. Amanchy, R., et al., *A curated compendium of phosphorylation motifs.* Nat Biotechnol, 2007. **25**(3): p. 285-6.

137. Subramanian, H., et al., *Nanoscale cellular changes in field carcinogenesis detected by partial wave spectroscopy.* Cancer Res, 2009. **69**(13): p. 5357-63.

138. Nicholson, R.I., J.M. Gee, and M.E. Harper, *EGFR and cancer prognosis.* Eur J Cancer, 2001. **37 Suppl 4**: p. S9-15.

139. Yuan, W., et al., *Expression of EphA2 and E-cadherin in gastric cancer: correlated with tumor progression and lymphogenous metastasis.* Pathol Oncol Res, 2009. **15**(3): p. 473-8.

140. Nikolova, M., et al., *Levels of expression of CAF7 (CD98) have prognostic significance in adult acute leukemia.* Leuk Res, 1998. **22**(1): p. 39-47.

141. Kaira, K., et al., *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, 2009. **100**(2): p. 249-54.

142. Prager, G.W., et al., *CD98hc (SLC3A2), a novel marker in renal cell cancer.* Eur J Clin Invest, 2009. **39**(4): p. 304-10.

143. Jackson, S.P. and J. Bartek, *The DNA-damage response in human biology and disease.* Nature, 2009. **461**(7267): p. 1071-8.

144. Harper, J.W. and S.J. Elledge, *The DNA damage response: ten years after.* Mol Cell, 2007. **28**(5): p. 739-45.

145. Manning, G., et al., *The protein kinase complement of the human genome.* Science, 2002. **298**(5600): p. 1912-34.

146. Alonso, A., et al., *Protein tyrosine phosphatases in the human genome.* Cell, 2004. **117**(6): p. 699-711.

147. Vintonyak, V.V., et al., *The therapeutic potential of phosphatase inhibitors.* Curr Opin Chem Biol, 2009. **13**(3): p. 272-83.

148. Thomson, M., *Evidence of undiscovered cell regulatory mechanisms: phosphoproteins and protein kinases in mitochondria.* Cellular and molecular life sciences : CMLS, 2002. **59**(2): p. 213-9.

149. Horbinski, C. and C.T. Chu, *Kinase signaling cascades in the mitochondrion: a matter of life or death.* Free radical biology & medicine, 2005. **38**(1): p. 2-11.

150. Deng, N., et al., *Phosphoproteome analysis reveals regulatory sites in major pathways of cardiac mitochondria.* Mol Cell Proteomics, 2010. **(DOI:10.1074/mcp.M110.000117)**.

151. Deng, W.J., et al., *Proteome, phosphoproteome, and hydroxyproteome of liver mitochondria in diabetic rats at early pathogenic stages.* Mol Cell Proteomics, 2010. **9**(1): p. 100-16.

152. Cui, Z., et al., *The Profile of Mitochondrial Proteins and Their Phosphorylation Signaling Network in INS-1 beta Cells.* J Proteome Res, 2010. **9**(6): p. 2898-908.

153. Reinders, J., et al., *Profiling phosphoproteins of yeast mitochondria reveals a role of phosphorylation in assembly of the ATP synthase.* Mol Cell Proteomics, 2007. **6**(11): p. 1896-906.

154. Ito, J., et al., *A survey of the Arabidopsis thaliana mitochondrial phosphoproteome.* Proteomics, 2009. **9**(17): p. 4229-40.

155. Fujiki, Y., et al., *Isolation of intracellular membranes by means of sodium carbonate treatment: application to endoplasmic reticulum.* J Cell Biol, 1982. **93**(1): p. 97-102.

156. Meng, W., et al., *One-step procedure for peptide extraction from in-gel digestion sample for mass spectrometric analysis.* Anal Chem, 2008. **80**(24): p. 9797-805.

157. Hao, P., T. Guo, and S.K. Sze, *Simultaneous analysis of proteome, phospho- and glycoproteome of rat kidney tissue with electrostatic repulsion hydrophilic interaction chromatography.* PLoS One, 2011. **6**(2): p. e16884.

158. Hao, P., et al., *Novel application of electrostatic repulsion-hydrophilic interaction chromatography (ERLIC) in shotgun proteomics: comprehensive profiling of rat kidney proteome.* J Proteome Res, 2010. **9**(7): p. 3520-6.

159. Guo, T., et al., *Hybridization of pulsed-Q dissociation and collision-activated dissociation in linear ion trap mass spectrometer for iTRAQ quantitation.* J Proteome Res, 2008. **7**(11): p. 4831-40.

160. Zhu, Y., et al., *Elucidating in vivo structural dynamics in integral membrane protein by hydroxyl radical footprinting.* Mol Cell Proteomics, 2009. **8**(8): p. 1999-2010.

161. Deutsch, E.W., et al., *A guided tour of the Trans-Proteomic Pipeline.* Proteomics, 2010. **10**(6): p. 1150-9.

162. Paik, W.K. and S. Kim, *Natural occurrence of various methylated amino acid derivatives in protein methylation.* New York: John Wiley & Sons, 1980: p. pp.8-25.

163. Warburg, O., *On the origin of cancer cells.* Science, 1956. **123**(3191): p. 309-14.

164. Vander Heiden, M.G., L.C. Cantley, and C.B. Thompson, *Understanding the Warburg effect: the metabolic requirements of cell proliferation.* Science, 2009. **324**(5930): p. 1029-33.

165. Keller, A., et al., *Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search.* Anal Chem, 2002. **74**(20): p. 5383-92.

166. Nesvizhskii, A.I., et al., *A statistical model for identifying proteins by tandem mass spectrometry.* Anal Chem, 2003. **75**(17): p. 4646-58.

167. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.

168. Krogh, A., et al., *Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.* J Mol Biol, 2001. **305**(3): p. 567-80.

169. Griffin, N.M., et al., *Label-free, normalized quantification of complex mass spectrometry data for proteomic analysis.* Nat Biotechnol, 2010. **28**(1): p. 83-9.

170. Guo, T., et al., *Global molecular dysfunctions in gastric cancer revealed by an integrated analysis of the phosphoproteome and transcriptome.* Cell Mol Life Sci, 2010.

171. Gygi, S.P., et al., *Correlation between protein and mRNA abundance in yeast.* Mol Cell Biol, 1999. **19**(3): p. 1720-30.

172. Greenbaum, D., et al., *Comparing protein abundance and mRNA expression levels on a genomic scale.* Genome Biol, 2003. **4**(9): p. 117.

173. Kunii, K., et al., *FGFR2-amplified gastric cancer cell lines require FGFR2 and Erbb3 signaling for growth and survival.* Cancer Res, 2008. **68**(7): p. 2340-8.

174. Nguyen, H.T., et al., *CD98 expression modulates intestinal homeostasis, inflammation, and colitis-associated cancer in mice.* J Clin Invest, 2011.

175. Schmitt, C.A., et al., *Expression and regulation by interferon-gamma of the membrane-bound complement regulators CD46 (MCP), CD55 (DAF) and CD59 in gastrointestinal tumours.* Eur J Cancer, 1999. **35**(1): p. 117-24.

176. Ye, Y.W., et al., *Fibroblast growth factor receptor 4 regulates proliferation and antiapoptosis during gastric cancer progression.* Cancer, 2011.

177. Lugli, A., et al., *EphB2 expression across 138 human tumor types in a tissue microarray: high levels of expression in gastrointestinal cancers.* Clin Cancer Res, 2005. **11**(18): p. 6450-8.

178. Datta, A., et al., *Phenotyping of an in vitro model of ischemic penumbra by iTRAQ-based shotgun quantitative proteomics.* J Proteome Res. **9**(1): p. 472-84.

179. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nat Methods, 2007. **4**(3): p. 207-14.

180. Thomas, P.D., et al., *PANTHER: a library of protein families and subfamilies indexed by function.* Genome Res, 2003. **13**(9): p. 2129-41.

181. Zhang, J., P.L. Yang, and N.S. Gray, *Targeting cancer with small molecule kinase inhibitors.* Nat Rev Cancer, 2009. **9**(1): p. 28-39.

182. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.

183. Clapham, D.E., *Calcium signaling.* Cell, 2007. **131**(6): p. 1047-58.

184. Gomes, D.A., et al., *c-Met must translocate to the nucleus to initiate calcium signals.* J Biol Chem, 2008. **283**(7): p. 4344-51.

185. Chen, L.B., *Mitochondrial membrane potential in living cells.* Annual review of cell biology, 1988. **4**: p. 155-81.

186. Foster, K.A., et al., *Optical and pharmacological tools to investigate the role of mitochondria during oxidative stress and neurodegeneration.* Prog Neurobiol, 2006. **79**(3): p. 136-71.

187. Bernardi, P., et al., *The mitochondrial permeability transition from in vitro artifact to disease target.* The FEBS journal, 2006. **273**(10): p. 2077-99.

188. Wallace, D.C. and W. Fan, *The pathophysiology of mitochondrial disease as modeled in the mouse.* Genes Dev, 2009. **23**(15): p. 1714-36.

189. Bernardi, P., *Mitochondrial transport of cations: channels, exchangers, and permeability transition.* Physiol Rev, 1999. **79**(4): p. 1127-55.

190. Lemasters, J.J. and E. Holmuhamedov, *Voltage-dependent anion channel (VDAC) as mitochondrial governator--thinking outside the box.* Biochimica et biophysica acta, 2006. **1762**(2): p. 181-90.

191. Halestrap, A.P., *What is the mitochondrial permeability transition pore?* Journal of molecular and cellular cardiology, 2009. **46**(6): p. 821-31.

192. Petronilli, V., et al., *Imaging the mitochondrial permeability transition pore in intact cells.* Biofactors, 1998. **8**(3-4): p. 263-72.

193. Boerner, J.L., et al., *Phosphorylation of Y845 on the epidermal growth factor receptor mediates binding to the mitochondrial protein cytochrome c oxidase subunit II.* Mol Cell Biol, 2004. **24**(16): p. 7059-71.

194. Lin, S.Y., et al., *Nuclear localization of EGF receptor and its potential new role as a transcription factor.* Nature cell biology, 2001. **3**(9): p. 802-8.

195. Matteucci, E., P. Bendinelli, and M.A. Desiderio, *Nuclear localization of active HGF receptor Met in aggressive MDA-MB231 breast carcinoma cells.* Carcinogenesis, 2009. **30**(6): p. 937-45.

196. Hornig-Do, H.T., et al., *Isolation of functional pure mitochondria by superparamagnetic microbeads.* Anal Biochem, 2009. **389**(1): p. 1-5.

197. Zhang, J., et al., *Altered proteome biology of cardiac mitochondria under stress conditions.* J Proteome Res, 2008. **7**(6): p. 2204-14.

198. Birchmeier, C., et al., *Met, metastasis, motility and more.* Nat Rev Mol Cell Biol, 2003. **4**(12): p. 915-25.

199. Vander Heiden, M.G., et al., *Outer mitochondrial membrane permeability can regulate coupled respiration and cell survival.* Proc Natl Acad Sci U S A, 2000. **97**(9): p. 4666-71.

200. Vander Heiden, M.G., et al., *Bcl-xL promotes the open configuration of the voltage-dependent anion channel and metabolite passage through the outer mitochondrial membrane.* J Biol Chem, 2001. **276**(22): p. 19414-9.

201. Nowak, G., *Protein kinase C-alpha and ERK1/2 mediate mitochondrial dysfunction, decreases in active Na+ transport, and cisplatin-induced apoptosis in renal cells.* J Biol Chem, 2002. **277**(45): p. 43377-88.

202. Salvi, M., A.M. Brunati, and A. Toninello, *Tyrosine phosphorylation in mitochondria: a new frontier in mitochondrial signaling.* Free radical biology & medicine, 2005. **38**(10): p. 1267-77.

203. Warburg, O., *On respiratory impairment in cancer cells.* Science, 1956. **124**(3215): p. 269-70.

Appendix A: Publications

1. **Tiannan Guo**, Yi Zhu, Chee Sian Gan, Sze Sing Lee, Jiang Zhu, Haixia Wang, Xin Li, James Christensen, Shiang Huang, Oi Lian Kon* and Siu Kwan Sze*. Quantitative Proteomics Discloses MET Expression in Mitochondria as a Direct Target of MET Kinase Inhibitor in Cancer Cells. ***Molecular & Cellular Proteomics***. ***2010***. 9(12):2629-41.
2. **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*. Global molecular dysfunctions in gastric cancer revealed by an integrated analysis of the phosphoproteome and transcriptome. ***Cellular and Molecular Life Sciences. 2011***. 68 (11): 1983-2002.
3. **Tiannan Guo**, Chee Sian Gan, Huoming Zhang, Yi Zhu, Oi Lian Kon, Siu Kwan Sze. Hybridization of Pulsed-Q Dissociation and Collision-Activated Dissociation in Linear Ion Trap Mass Spectrometer for iTRAQ Quantitation. ***Journal of Proteome Research***, ***2008***, 7(11):4831-4840.
4. Huoming Zhang, **Tiannan Guo**, Xin Li, Arnab Datta, Jung Eun Park, Jie Yang, Sai Kiang Lim, James P. Tam, and Siu Kwan Sze. Simultaneous characterization of glyco- and phospho-proteomes of mouse brain membrane proteome with electrostatic repulsion hydrophilic interaction chromatography (ERLIC). ***Molecular & Cellular Proteomics. 2010.*** *Apr;9(4):635-47.*
5. Yi Zhu, **Tiannan Guo**, Jung Eun Park, Xin Li, Wei Meng, Arnab Datta, Marshall Bern, Sai Kiang Lim, Siu Kwan Sze. Elucidating in vivo structural dynamics in integral membrane protein by hydroxyl radical footprinting. ***Molecular & Cellular Proteomics***. ***2009***. 8:1999-2010.

6. Piliang Hao, **Tiannan Guo**, Siu Kwan Sze. Simultaneous Analysis of Proteome, Phospho- and Glycoproteome of Rat Kidney Tissue with Electrostatic Repulsion Hydrophilic Interaction Chromatography. ***PLoS ONE 2011***. 6(2): e16884.
7. Piliang Hao, **Tiannan Guo**, Xin Li, Adav Sunil, Jie Yang, Wei Meng, Siu Kwan Sze. Novel Application of Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) in Shotgun Proteomics: Comprehensive Profiling of Rat Kidney Proteome. ***Journal of Proteome Research***. ***2010***, 9 (7), 3520-3526.
8. Chee Sian Gan, **Tiannan Guo**, Huoming Zhang, Sai Kiang Lim, Siu Kwan Sze. A Comparative Study of Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) versus SCX-IMAC-Based Methods for Phosphopeptide Isolation/Enrichment. ***Journal of Proteome Research***, ***2008***. 7(11): 4869-4877.
9. Wei Meng, Huoming Zhang, **Tiannan Guo**, Chhiti Pandey, Yi Zhu, Oi Lian Kon, Siu Kwan Sze. A one-step precedure for peptide extraction from in-gel digestion sample for mass spectrometric analysis. ***Analytical Chemistry***, ***2008***. 80(24): 9797–9805.
10. Siu Kwan Sze*, Wei Wang, Randong Yuan, Wei Meng, **Tiannan Guo**,Yi Zhu, James P. Tam*. Elucidating structure of cyclotide by partial acid hydrolysis and LC-MS/MS analysis. ***Analytical Chemistry***. ***2009***. 81(3):1079-88. (*corresponding authors)

Appendix B: Poster Presentations

1. **Tiannan Guo**, Yi Zhu, Wai Har Ng, Oi Lian Kon* and Siu Kwan Sze*. Large-Scale Proteome Analysis Reveals Complex Molecular Abnormalities in Gastric Cancer. ***American Society for Cancer Research (AACR) 102nd Annual Meeting 2011***. (April 2-6, 2011, Orlando, Florida, USA).
2. **Tiannan Guo**, Oi Lian Kon and Siu Kwan Sze. Roles of Mitochondrial Protein Kinases in Targeted Cancer Therapy. ***American Society for Cancer Research (AACR) 102nd Annual Meeting 2011***. (April 2-6, 2011, Orlando, Florida, USA).
3. **Tiannan Guo**. Quantitative proteomics discloses MET expression in mitochondria as a direct target of MET kinase inhibitor in cancer cells. ***Duke-NUS Early Career Scientists Association (DUNES) Symposium 2010***.
4. **Tiannan Guo**, Yi Zhu, Chee Sian Gan, Sze Sing Lee, Jiang Zhu, Haixia Wang, Xin Li, James Christensen, Shiang Huang, Oi Lian Kon and Siu Kwan Sze. Quantitative Proteomics Discloses MET Expression in Mitochondria as a Direct Target of a MET Kinase Inhibitor in Cancer Cells. ***Human Proteome Organisation's (HUPO) 9th Annual World Congress***. (19-23 Sep 2010) Organized by Human Proteome Organisation. Syndey, Australia.
5. **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. Integrated Phosphoproteome and Transcriptome Analyses Synergistically Uncover Oncogenic Signaling Networks in Gastric Cancer. ***Human Proteome Organisation's (HUPO) 9th Annual World Congress***. (19-23 Sep 2010) Organized by Human Proteome Organisation. Syndey, Australia.

6. **Tiannan Guo**, Lingling Fan, Wai Har Ng, Kiat Hon Lim, Yi Zhu, Shiang Huang, Oi Lian Kon, Siu Kwan Sze. In-depth Analysis of Gastric Cancer Membrane Proteome. ***Human Proteome Organisation's (HUPO) 9th Annual World Congress***. (19-23 Sep 2010) Organized by Human Proteome Organisation. Syndey, Australia.

7. **Tiannan Guo**, Yi Zhu, Chee Sian Gan, Sze Sing Lee, Jiang Zhu, Haixia Wang, Xin Li, James Christensen, Shiang Huang, Oi Lian Kon and Siu Kwan Sze. Quantitative Proteomics Discloses MET Expression in Mitochondria as a Direct Target of a MET Kinase Inhibitor in Cancer Cells. ***SINGHEALTH DUKE-NUS SCIENTIFIC CONGRESS 2010***. (15-16 Oct 2010) Organized by SingHealth cluster of institutions and Duke-NUS Graduate Medical School. (Abstract reference number: SHSC20100330a0004)

8. **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. Integrated Phosphoproteome and Transcriptome Analyses Synergistically Uncover Oncogenic Signaling Networks in Gastric Cancer. ***SINGHEALTH DUKE-NUS SCIENTIFIC CONGRESS 2010***. (15-16 Oct 2010) Organized by SingHealth cluster of institutions and Duke-NUS Graduate Medical School. (Abstract reference number: SHSC20100330a0003)**.**

9. **Tiannan Guo**, Lingling Fan, Wai Har Ng, Kiat Hon Lim, Yi Zhu, Shiang Huang, Oi Lian Kon, Siu Kwan Sze. In-depth Analysis of Gastric Cancer Membrane Proteome. ***SINGHEALTH DUKE-NUS SCIENTIFIC CONGRESS 2010***. (15-16 Oct 2010) Organized by SingHealth cluster of institutions and Duke-NUS Graduate Medical School. (Abstract reference number: SHSC20100330a0002)

10. **Tiannan Guo**, Oi Lian KON, Siu Kwan SZE. Global Investigation of Oncogenic Signaling Pathways in Gastric Cancer by Proteomics Approaches. ***3rd Singapore Gastric Cancer Consortium (SGCC) Annual Scientific Meeting (ASM)***. (2010.7.15-16.) Organized by Singapore Gastric Cancer Consortium.
11. **Tiannan Guo**, Oi Lian KON, Siu Kwan SZE. Quantitative Proteomics Discloses MET Expression in Mitochondria as a Direct Target of a MET Kinase Inhibitor in Cancer Cells. ***2nd SSMS (Singaporean Society for Mass Spectrometry) Seminar Day***. (2009.11.4) Organized by Singaporean Society for Mass Spectrometry.
12. **Tiannan Guo**, Chee Sian Gan, Huoming Zhang, Yi Zhu, Oi Lian Kon, Siu Kwan Sze. Hybridization of Pulsed-Q Dissociation and Collision-Activated Dissociation in Linear Ion Trap Mass Spectrometer for iTRAQ Quantitation. ***Joint 5th Structural Biology & Functional Genomics and 1st Biophysics International Conference***. 2008. Singapore.

Appendix C: Oral Presentations

1. Comprehensive phosphoproteomic and transcriptomic analysis of gastric cancer: an informatic challenge. ***3rd SSMS (Singaporean Society for Mass Spectrometry) Seminar Day***. (2010.11.24-25) Organized by Singaporean Society for Mass Spectrometry.
2. Mitochondrial MET is a Therapeutic Target of Gastric Cancer. ***3rd SSMS (Singaporean Society for Mass Spectrometry) Seminar Day***. (2010.11.24-25) Organized by Singaporean Society for Mass Spectrometry.
3. Global molecular dysfunctions in gastric cancer revealed by an integrated analysis of the phosphproteome and transcriptome. ***SINGHEALTH DUKE-NUS SCIENTIFIC CONGRESS 2010***. (2010.10.16) Organized by SingHealth and Duke-NUS.
4. Global Investigation of Oncogenic Signaling Pathways in Gastric Cancer by Proteomics Approaches. ***Singapore Proteome Forum 2010***. (2010.5.14) Organized by Agilent Technologies Singapore and National Cancer Centre Singapore.
5. Challenges and opportunities in proteomics. ***Singapore Proteomics Forum. 2009***. Singapore. Organized by Agilent Technologies Singapore and National University of Singapore.
6. Systematic investigation of the molecular network of hepatocyte growth factor receptor, Met, in gastric cancer cells. ***Duke-NUS CSCB & NCC RIP/ JC Conference. 2009***. Singapore.

7. Applications of Mass Spectrometry Techniques in Biological and Biomedical Sciences. ***The 3rd Korea-Singapore International Conference on Bioscience & Biotechnology***. 2008. Singapore.