DEVELOPMENT OF QUANTITATIVE PROTEOMIC STRATEGIES TO IDENTIFY TYROSINE PHOSPHATASE SUBSTRATES

by

Peipei Zhu

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

Department of Chemistry

West Lafayette, Indiana

December 2021

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. W. Andy Tao, Chair

Department of Chemistry

Dr. Kavita Shah

Department of Chemistry

Dr. Chang-Deng Hu

Department of Medicinal Chemistry and Molecular Pharmacology

Dr. **Andrew D Mesecar**

Department of Biochemistry

Approved by:

Dr. Christine A Hrycyna

This dissertation is dedicated to my family. Thank you so much for your unwavering love and support. To my grandfather and grandmother for teaching me the value of hard work and integrity; To my father and mother for instilling in me the significance of keeping optimistic and hopeful; and to my sister for assisting me in caring for my family.

ACKNOWLEDGMENTS

First and foremost, I want to express my gratitude to my advisor, Dr. Weiguo Andy Tao, for his guidance and support in assisting me in finding projects. During my stay in his lab, I learned a lot about science and establishing collaborations, writing, and presenting my work. I owe thanks to my many collaborators for supplying materials, encouraging me, and offering me new skills. I particularly appreciate Dr. Jian-Kang Zhu for multiple projects related to plant AP-MS projects; Dr. Ruoyu Zhang and Dr. Zhong-Yin Zhang for phosphatase projects, Dr. Feng Yue and Dr. Shihuan Kuang for the Fam210A project, Dr. Alan Hsu and Dr. Qing Deng for the CDK2 project, and many other projects that I collaborated during my time at Purdue. I appreciate the chance to work with them and learned from them. I would also like to thank my committee members, Dr. Chang-Deng Hu, Dr. Kavita Shah, and Dr. Andrew D. Mesecar, who gave me useful advice and suggestions. I want to give special thanks to my MS advisor, Dr. Suzanne O'Handley, for mentorship. Thank you for being such a good mentor and friend in my Ph. D. journey.

I would like to express my appreciation to Tao Lab members. Thank you, Dr. Anton Iliuk. He has been and will continue to be my role model both in work and life. I have benefited from him tremendously by teaching me how to fix HPLC leaks, clean MS, give suggestions to my manuscript, etc. I also want to thank Dr. Chuan-Chih Hsu, who guided me to research at the very beginning of my Ph.D. journey, taught me how to prep MS samples, design my experiments, and maintain and clean MS. I want to thank all past and present members, Justin Arrington, Sarah Elders, Sebastian Paez, Blair (I-Hsuan) Chen, Mayank Srivastava, Hillary Andaluz, Xiao Feng Wu, Leo (Der-Xiang) Kao, Marco Hadisurya, Andrew DeMarco, Michelle Lihon, Pooja Saklani for making enjoyable environment to come to the lab and having all good memories my time here.

Finally, I would like to thank my friends and family, without your support and understanding, I would not be able to achieve anything. Min Chen, Yan Li, Ei Sandi New, and Mavreen Tuvilla, thank you for being such great listeners all along and always supportive, and I am grateful for all your encouragement. Thank you to my parents and sister for their constant support and unconditional love.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CAA	2-Chloroacetamide
DEPOD	Human Dephosphorylation Database
DM	Double Mutant
DOK1	Docking Protein 1
DTT	Dithiothreitol
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
HA	Hemagglutinin
HGF	Hepatocyte Growth Factor
IP	Immunoprecipitation
KEA	Kinase Enrichment Analysis
KO	Knockout
LC-MS	Liquid Chromatography Mass Spectrometry
MS	Mass Spectrometry
PEI	Polyethyleneimine
PolyMAC	Polymer-based Metal-ion Affinity Capture
PTM	Post Translational Modifications
PTP	Protein Tyrosine Phosphatase
PTP1B	Protein Tyrosine Phosphatase 1B
PTS	Phase Transfer Surfactants
pTyr	Phosphotyrosine
SDB-XC	Poly Styrene Divinyl Benzene Extraction Disk
SDC	Sodium Deoxycholate
SLS	N-Lauroylsarcosine Sodium Salt
SHP2	Src Homology 2 (SH2) Containing Protein Tyrosine Phosphatase 2
TFA	Tri-fluoric acid
TCEP	Tris-(2-Carboxy Ethyl) Phosphine Hydrochloride
TEAB	Triethylammonium Bicarbonate

TMT	Tandem Mass Tag
WT	Wildtype

ABSTRACT

Protein tyrosine phosphorylation is an essential posttranslational modification that controls cell signaling involving various biological processes, including cell growth, proliferation, migration, survival, and death. Balancing tyrosine phosphorylation levels is necessary for normal and pathological states, and this reversible mechanism occurs through protein tyrosine kinases and phosphatases. Advancements in instrumentation and applying conventional biochemical and genetic methods have led to cell signaling studies and pharmaceutical development discoveries. However, there is still a lack of understanding of tyrosine phosphatases' mechanisms, substrates, and activities within complex networks. The challenges remain in the tyrosine phosphatase field due to the low abundance and dynamic nature, sample preparation steps, and sensitivity to detect tyrosine phosphorylation events. Although mass spectrometry (MS)-based phosphoproteomics has allowed the identification of thousands of phosphotyrosine sites in a single run, protein phosphorylation poses another analysis caveat of dissecting complex phosphorylation signaling pathways involved in healthy cellular processes similarly in disease pathogenesis. This dissertation discusses strategies for improving tyrosine phosphatase sample preparation and identifying the tyrosine phosphatases' direct substrates. Chapter one is an overview of current techniques to study tyrosine phosphatases. In contrast, chapters two and three highlight the work that has been done to identify the direct substrates of phosphatase SHP2 and PTP1B, respectively, whose dysregulation leads to the development of cancers.

In chapter 2, we describe a novel method that incorporated three separate MS-based experiments to identify the direct substrates of phosphatase SHP2: immunoprecipitation of substrate trapping mutants complex, *in vivo* global phosphoproteomics, and *in vitro* dephosphorylation of SHP2 phosphatase substrates. With immunoprecipitation of substrate trapping mutant experiment, weak and transient phosphatase-substrate interactions were detected by mass spectrometry after being stabilized by substrate trapping mutant method. This experiment not only identified the interactions between phosphatase and substrates but also revealed phosphotyrosine sites that are potentially protected in the substrate trapping mutant. We identified 80 phosphotyrosine proteins that showed upregulated in SHP2 mutant samples, and GAB1, GAB2, IRS1, SIRPA, and MPZL1 were examined in our list, which are reported SHP2 substrates. In the second experiment in parallel, we explored the global phosphorylation in HEK293 cells stimulated

by epidermal growth factor. Peptides containing phosphotyrosine residues were captured by immobilized anti-pY PT-66 antibody and analyzed by LC-MS/MS. The results provided information on how SHP2 regulates downstream protein tyrosine phosphorylation and global phosphotyrosine response initiated by EGF. We used SHP2 substrate trapping mutant to isolate phosphotyrosine-containing proteins to serve as a SHP2 substrate pool for an *in vitro* phosphatase assay, then analyzed by LC-MS/MS. Finally, the overlap of the three separate MS-based experiments gave us the final list of high-confidence SHP2 substrates. DOK1 was validated to be a direct SHP2 substrate.

Chapter 3 describes a novel method that integrates *in vivo* global phosphoproteomics perturbed by PTP1B inhibitor and stimulated by insulin with *in vitro* kinetic profile of PTP1B phosphatase to identify its substrates. We were able to identify 114 phosphotyrosine proteins that showed upregulated in PTP1B inhibitor and insulin-treated sample in *in vivo* global phosphoproteomics experiment. CTTN, EGFR, FER, IRS1, PTPN11, SRC, TYK2, PKM, GAB1, GAB2, and INSR were examined, which are PTP1B reported substrates. In *in vitro* kinetic profile of the PTP1B phosphatase experiment, we utilized dimethyl labeling to quantify the PTP1B dephosphorylation rate. No PTP1B substrate motif consensus was observed in the labeling experiments. We finally overlapped *in vivo* and *in vitro* experiments to identify PTP1B *bona fide* substrates with high confidence.

CHAPTER 1. ADVANCES IN PHOSPHOPROTEOMICS FOR THE STUDY OF TYROSINE PHOSPHATASES

1.1 Summary

Despite the low cellular abundance of phosphotyrosine (pTyr), it modulates various cell signaling pathways in normal physiology and pathology. pTyr phosphorylation of oncoproteins is reversibly controlled by opposing enzymes kinases and phosphatases. Unlike kinases, phosphatases remain challenging and understudied due to the misconception of unregulated and non-specificity activities. Traditional genetic and biochemical methods have contributed significantly to our understanding, and now many phosphatases are known that their activity is highly specific and tightly regulated in the cell. Moreover, innovative pharmacological approaches for targeting phosphatases have recently emerged, and as a result, phosphatases may reveal therapeutic insights critical to combating diseases. Even the phosphatase field has thrived in the past years. It has set new insights for therapeutic development; our understanding of their mechanisms, substrates, and functions within highly intricated signaling networks remains elusive due to the limitations of methodologies. We describe here advances in phosphoproteomics for the study of tyrosine phosphatases.

1.2 Introduction

The reversible phosphorylation of proteins allows for highly dynamic and coordinated regulation of almost every aspect of cellular processes, including metabolic homeostasis, stress response, cell cycle transitions, and many other essential biological mechanisms[1]. It is not surprising that phosphorylation is one of the most studied posttranslational modifications due to its importance in signal transduction. In eukaryotes, serine, threonine, and tyrosine account for more than one-third of protein phosphorylation events[2]. Only 1.8 % of tyrosine residues are phosphorylated in particular[3]. Despite the rarity of tyrosine phosphorylation, it is essential for cellular functions, ranging from cell growth, proliferation, immunity to cell death[4]. A reversible mechanism mediates the balancing of tyrosine phosphorylation through protein tyrosine kinases and phosphatases. It is estimated about 90 protein tyrosine kinases (PTK) and 108 protein tyrosine

phosphatases (PTP) occur in humans[2,5]. A dysregulation of tyrosine phosphorylation signaling crosstalk plays a role in developing diseases, including malignant development and progression[6,7].

Tyrosine phosphatases, like kinases, can act as either oncogenes or tumor suppressors depending on the function of their target phosphosites. *PTPN11* gene encodes SHP2; a tyrosine phosphatase is required to fully activate the central signaling cascade in cancer biology, like RAS/ERK1/2, PI3K/AKT, and JAK/STAT signaling, which are strongly associated with various human cancers[8]. In addition, SHP2 overexpression is common in many types of carcinomas and various cancers[9]. Similarly, PTP1B plays a key role in regulating growth factor and hormone signaling pathways, is also found to be oncogenic in mouse studies with human epidermal growth factor receptor 2-induced breast cancer[10]. On the contrary, the receptor protein tyrosine phosphatase delta (PTPRD)[11] and protein tyrosine phosphatase receptor-T (PTPRT)[12] are recognized as tumor suppressors. Hence, dissecting their signaling interaction events in tyrosine phosphorylation is essential in a new cancer treatment strategy.

Nevertheless, despite significant progress, our understanding of phosphatases remains limited compared to kinases. There are several causes for this, to name two leading causes here. First, there is a persisting misperception that phosphatases are less selective and tightly regulated. Second, some phosphatases require regulatory proteins for substrate recognition and their highly transient interaction between most phosphatases and their substrates, which poses a challenge in predicting substrates. In this chapter, we will summarize how phosphatases are characterized and recent advances in sample preparation techniques.

1.2.1 Classification of phosphatases

Protein phosphatases have derived from a diverse and independent ancestor that are structurally and functionally distinct from each other compared to the evolution of kinases based on a common ancestor. Phosphatases have been grouped based on their similarity in catalytic domain sequence; there are six superfamilies in this classification. 1)The protein-tyrosine phosphatase domain (PTP), 2) the serine/threonine-specific protein phosphatase domain (PPP), 3) the protein phosphatase 2C-like domain (PPM), 4) the Haloacid dehalogenase-like hydrolase domain (HAD), 5) the NUDIX hydrolase domain (NUDT), and 6) LP phosphatase which contains one of phosphatidic acid or inositol monophosphatase or inositol polyphosphate-related phosphatase[13,14]. The superfamilies were further divided into classes based on their substrate

specificity, catalytic domain, and literature annotation. For instance, the PTP family is defined by its catalytic signature motif of CX5R, and there are four subgroups of PTPs: 1) the classical PTPs, 2) VH1-like, 3) LM-PTP, and 4) CDC25C[13,14]. The classical PTPs, also known as phosphotyrosine-specific phosphatases, are categorized as receptor-like PTPs and non-transmembrane PTPs. The receptor-like PTPs have a transmembrane portion and can regulate signaling through a ligand-induced system. Whereas non-transmembrane PTPs are cytoplasmic enzymes that directly or indirectly control activity by mediating the substrates or interactors, respectively. In this chapter, we will focus only on the non-transmembrane PTPs.

1.3 Identification of phosphatase protein substrates

Both protein tyrosine kinases and tyrosine phosphatases play critical roles in balancing tyrosine phosphorylation, they are equally attractive for studying. However, there is a significant discrepancy in our knowledge of tyrosine phosphatases. There are about an equal number of tyrosine kinases and tyrosine phosphatase[2,5], but more than 1800 tyrosine kinases substrates are identified in KEA[15]; up to date, less than 400 tyrosine phosphatase substrates are found in the DEPOD[16] database. Many reasons caused this knowledge gap, partly due to the lack of suitable methodology, to name a few. First, most tyrosine phosphatases are regulatory enzymes that require interactors to activate their activity. For example, the SHP2 phosphatase has 2 SH2 domains. In its basal state, the N-SH2 domain is wedged into the PTP domain, and this conformation blocks the active site of SHP2 and prevents substrate access. However, when both SH2 domain binds to appropriate binding proteins, it will change its conformation to an "open" state where it exposes the active site for the substrate binding. Second, unlike kinases, phosphatases remove a phosphate group from a substrate, leading to a loss of signal for detection. Even if there is a way to save the phosphate group in the substrate, this phosphate group will be on a tyrosine residue for tyrosine phosphatases. The low abundance of pTyr adds another layer of difficulty isolating this residue in the sample preparation step. Hence, tyrosine phosphatases cannot be studied straightforwardly. Last, the interaction of phosphatases and their substrates are highly transient, and they often require a specific context for the events to happen. Therefore, it is still a challenge to identify tyrosine phosphatases substrates.

We are in the "omics" era with the advancement of mass spectrometry (MS) allows large-scale studies of genes, RNA transcripts, proteins, and metabolites. So, it is named genomics,

transcriptomics, proteomics, and metabolomics. Phosphoproteomics, in particular, is a branch of proteomics that studies a protein with a phosphate group as a PTM; it permits thousands of phosphorylation sites identification. It will be beneficial to apply a high-throughput strategy to discover novel substrates of tyrosine phosphatases. In recent years, various approaches have been developed to improve the selectivity and sensitivity of identifying tyrosine phosphatase substrates using MS in recent years.

1.3.1 Phosphatase-substrate interaction stabilization

The interaction between phosphatase and a substrate in a cell depends on intrinsic enzymatic specificity and their subcellular colocalization *in situ*. This interaction is highly transient and weak when a phosphatase catalyzes a phosphomonoester's hydrolysis from the substrate. Thus, stabilization of the interaction of phosphatase-substrate is the first step to identifying the substrate. The Tonks laboratory named the substrate-trapping mutants[17], a method that serves as a valuable tool for PTPs to stabilize the association of phosphatase-substrate and then pull-out substrates from a heterogeneous mixture. The notion is based on site-directed mutants of phosphatases that do not maintain any catalytic activities but still recognize and bind tightly to their pTyr substrates. The cysteine in the catalytic site and the aspartate in the WPD loop are two common mutation sites for substrate-trapping mutants that abolish nucleophilic attacks on the substrate phosphate. Studies show that the efficiency of substrate-trapping varies in combinations of their mutant variants. For example, the association between the D181A mutant of PTP1B and its substrates was sufficiently stable to allow isolation of the complex in the proceeding step[17]. However, SHP2 requires both Cys to Ser and Asp to Ala double mutations for equal efficient isolation of its substrate[18]. Substrate-trapping mutants are a robust and effective method to isolate pTyr phosphatase substrates, but their capturing efficiency differs from each PTPs.

1.3.2 Phosphatase-substrate isolation

Once the phosphatase-substrate interaction is stabilized, the substrate is ready to pull out from the heterogeneous mixture. Substrate trapping mutant phosphatase only recognizes the pTyr substrate. Hence, to maintain pTyr integrity, the endogenous phosphatases are either inhibited or

knockdown in cell lysate. Pervanadate is a common phosphatase inhibitor to maintaining global phosphorylation levels in the cells.

The exogenous substrate trapping mutant phosphatase is often introduced into the mammalian cells to eliminate the isolation complication of phosphatase by fusing protein tags such as FLAG or HA to facilitate the pullout of the corresponding phosphatase with its interacting substrates. However, this will also isolate other non-specific proteins interacting with phosphatase or substrate, but not the genuine substrate. An additional step is required to differentiate between the *bonafide* substrate and interacting proteins. The conventional method uses a competitive inhibitor like pervanadate in PTPs. Others will need to isolate the substrate and validate it in further analysis.

1.3.3 Phosphatase-substrate identification

The fused protein tags are a common strategy for isolating phosphatase substrate since this can be purified by coimmunoprecipitation. After substrate candidates are isolated, they can be analyzed by biochemical methods like SDS-PAGE or Western Blot, or most recently, MS can yield the complete proteome profile. MS can quantify the sample's large scale of peptides or phosphopeptides by labeling-free or metabolic- or chemical-labeling.

With the low abundances of phosphopeptides and low degrees of phosphorylation, it is necessary to enrich and concentrate phosphopeptides before MS analysis. There are several phosphopeptides enrichment methods. Immobilized metal affinity chromatography (IMAC) relies on the interaction between positively charged metal-ligand complexes like Fe^{3+}, Ga^{3+}, Al^{3+}, Ti^{4+}, and Zr^{4+} and negatively charged phosphopeptides[17]. Even IMAC is less specific than other affinity methods, and this lower specificity makes it suitable for the separation of phosphopeptides in a complex mixture. To overcome the issue of lower specificity, polymer-based metal ion affinity capture (PolyMAC) employs polyamidoamine dendrimers with titanium ions to capture phosphopeptides with excellent selectivity and sensitivity[19]. Metal oxide affinity chromatography (MOAC) is based on selective metal oxides for phosphopeptides enrichment, such as titanium dioxide, zirconium dioxide, aluminum oxide, and niobium oxide[20,21]. Antibody-based enrichment for pTyr is also well established, like PT66, 4G10, and anti-pY for probing and isolation. Phosphopeptides can also be separated in different chromatography like hydrophilic interaction liquid chromatography (HILIC)[22], strong cationic ion-exchange chromatography (SCX)[23], strong

anionic ion-exchange chromatography (SAX)[24]. Phosphoproteomes can determine a possible substrate by quantifying the phosphopeptides stoichiometry or intensity/abundance changes in the presence of phosphatase interest with the appropriate control.

After several putative PTP phosphatase substrates have been identified, they need further validation to confirm if they are cognate substrates or not. To ensure the cognate substrate, they need to qualify for three criteria[25]. First, the isolated phosphatase-substrate by substrate trapping mutant should form a stable interaction with the endogenous pTyr substrate. On the other hand, the wild-type correspondent phosphatase should not interact or have a lesser association. Second, changing the phosphatase activity can control the endogenous pTyr levels of the putative substrate. Identifying the net substrate pTyr level is difficult, but it is necessary to identify the dephosphorylation site on the PTP substrate. However, the identification of the site of pTyr by PTP phosphatase is not the main requirement. Finally, wild-type phosphatase can directly dephosphorylate the putative substrate *in vitro* is also required. The successful assignment of PTP phosphatase substrate is by providing convincing evidence of one or a combination of the criteria mentioned above.

1.4 Sample preparation for mass spectrometry-based phosphoproteomics

Top-down mass spectrometry takes advantage of analyzing intact proteins and their PTMs. Still, it lacks efficient separation and provides limited coverage of the protein of interest due to many technical challenges[26]. The middle-down approach provides longer peptides (3.0 kDa <peptides<10 kDa) for better detection and identification than the top-down MS method but is restricted by its proteolytic enzyme digestion in the workflow[27]. Bottom-up, also known as shotgun named by the Yates lab, has few advantages because it is a reasonably straightforward and reliable method of determining the protein composition of a wide range of samples from cells, tissues to human biofluids and so on. However, only a limited and variable fraction of a protein can be retrieved, resulting in a low percentage coverage of the protein sequence. Despite the disadvantage, most phosphoproteome studies are done in a bottom-up approach due to the accessibility of instrumentation and software.

In a typical shotgun experiment, the sample preparation consists of several critical steps:

1. Extract proteins from the studied biological sample.
2. Protein denaturation, reduction, and alkylation.

3. Protein digestion.
4. Cleaning up of peptides by desalting.
5. Peptide separation and injection to MS.
6. Database search for protein identification.

Protein extraction is the first and most important step in a shotgun experiment; if this step fails to perform, the downstream steps are pointless to analyze. The initial step often involves crude mechanical disruption such as cutting, smashing, or shearing tissue into smaller pieces. Then cells or tissue lysis are performed directly in a buffer with optimum pH and ionic strength; protein stability is also maintained to prevent proteolysis. With these considerations, a particular lysis buffer is optimally designed to extract the proteins of interest with appropriate denaturants or detergents that contain 1) ionic detergents like sodium dodecyl sulfate, deoxycholate, or cetyltrimethylammonium bromide that disrupt the cell membrane and solubilize proteins[28–30], or 2) nonionic zwitterionic detergents such as Triton X-100, NP-40, or CHAPS can be used for less denaturing conditions for maintaining proteins in their biologically active form, or 3) strong chaotropic reagents like guanidine[31] or urea[32] disrupt the native beta-sheets hydrophobic interaction or backbone-backbone hydrogen bonds, respectively. The latter reagent is also efficient at breaking protein-protein interactions; hence any studies related to protein-protein interactions should avoid this type of reagent. Sonication is also used to disrupt the membrane further and help to solubilize proteins. Often, protease or phosphoprotease inhibitors are added to the mixture to prevent protein degradation.

The most common reducing reagents used in proteomics experiments are dithiothreitol (DTT) and tris-2(-carboxyethyl)-phosphine (TCEP), reduce disulfide bridges and unfold proteins. The alkylation of the free SH-groups is usually done by iodoacetamide (IAA) or chloroacetamide (CAA). Many studies have shown the combination of DTT and IAA is the most frequent[33–36], but TCEP and CAA have shorter incubation times and eliminate unwanted side reaction[37].

After reduction and alkylation, protein digestions are either enzymatic, involve proteolytic enzymes, or nonenzymatic digestion using chemicals or reagents. Nonenzymatic digestion has potential advantages over enzyme-based digestion in simplicity, robustness, cost, and speed[38]. A few common nonenzymatic approaches include cyanogen bromide cleaves at the N-terminus of methionine[39], acid hydrolysis at aspartate residue[40], hydroxylamine cleaves at asparagine and glycine bonds[41], etc. On the other hand, trypsin is considered the gold standard in proteomics. It

cleaves the carboxyl group of arginine and lysine residues, resulting in short, positively charged peptides for MS analysis[42]. In addition, trypsin oftentimes is accompanied by using other enzymes for complementary digestion to enhance proteomic identification and PTM localization rates, such as pepsin, chymotrypsin, Lys-C, Asp-N, and Glu-C[43–48]. Aside from protease choice, the digestion needs to perform under optimum temperature, buffer pH, and time.

Clean-up after protein digestion is necessary for removing detergent and salt to eliminate the signal suppression issue in MS analysis and increase the lifetime of the columns. The clean-up procedure will depend on the properties of interfering species and reagents that used in protein digestions. For example, detergents can be removed using a specialized detergent removal spin-column or filtered assisted sample preparation column with an appropriate molecular weight size cutoff membrane. Other charged-neutral contaminants, buffer salts, or small molecules can be separated by ion-exchange chromatography, reverse phase chromatography C18 column, or copolymer-based styrene-divinylbenzene sorbent column/membrane.

Peptide separation is performed in two different fashions, online and offline approaches, and either one is used to simplify complex samples prior to MS analysis. Online chromatography is coupled with MS, whereas offline is either StageTip-based, column-based, or gel-based. Both online or offline can be done in different multidimensional chromatographic approaches such as reversed-phase, ion exchange, or hydrophilic interaction chromatography[49–53] to provide high-speed, high resolution, high sensitivity, and specificity separation of the complex mixture and facilitate MS detection and quantitation for in-depth coverage of peptides/phosphopeptides identification. As peptides are ready for electrospray ionization, they will directly spray into the MS, taking a tandem measurement of samples.

MS enables high-throughput proteomics experiments by producing thousands and thousands of tandem-MS spectra. Database search engines are spectra annotation methods for peptide sequence based on the calculation of every peptide spectrum match to reflect the quality of the experimental peptide spectrum in comparison to theoretical ones. The most widely used being 1) Andromeda, a probability calculation for the scoring of peptide spectrum matches[54], 2) SEQUEST, algorithm-based scoring[55], and 3) Mascot, which is performing machine learning method[56], and 4) X!Tandem also is algorithm-based scoring[57]. Each search engine has its percolator or cutoffs based on the estimated false discovery rates to confidently assign spectra to peptides/proteins.

1.5 Quantitation methodologies for bottom-up LC-MS

Over the last few decades, MS-based phosphoproteomics has become an increasingly powerful tool to detect and quantify thousands of phosphopeptides. Multiple methods have been developed to increase the accuracy, comprehension, and sensitivity to identify and quantify phosphopeptides in a bottom-up fashion. For this reason, there are two categories in the field: discovery and targeted proteomics. Discovery is also known as data-dependent acquisition (DDA); it enables the identification of thousands of proteins or phosphopeptides per run[58,59]. Label-free quantitation is in DDA mode, and quantitation is performed in relative or absolute using either peak intensity/area or spectral counts. This strategy is ideal for large sample analyses in measuring changes in protein expression, screening, and discovering drug/biomarker experiments. However, DDA lacks reproducibility and consistency of quantitation. Targeted proteomics has emerged to overcome these limitations, selected/multiple/parallel reaction monitoring[60–63]; they provide consistent and accurate quantification and are suitable for hypothesis-driven studies with small sample size. Hence, another MS acquisition method has arisen, data-independent acquisition (DIA)[64]. The significant improvement of DIA is its coverage, where every peptide is fragmented multiple times in smaller window size. This method is favored for low-abundant protein/peptide identification. Recently SureQuant was introduced to increase the pTyr profiling in human tumors[65]. The only caveat is that it requires building a library which is a very time-consuming step. Other quantitation methods involve isobaric or isotopic labeling, including metabolic labeling (SILAC[66]), chemical isotopic labeling (ICAT[67], dimethyl labeling[68]), isobaric tagging (TMT[65,69–72], iTRAQ[73]), and labeled spike-in peptides of known concentration in samples[74]. Currently, isobaric tag-based TMT can accurately and sensitively quantify up to 16 samples[65,72]. While label-based methods are common and increased accuracy and sensitivity, their major limitations remain. I will focus on TMT and dimethyl labeling strategy in the following sections.

1.5.1 Isobaric labeling tandem mass tag

The central focus of proteomics studies is to discover new biomarkers/drugs or gain mechanistic understandings in different biological states[75,76]. Hence, isobaric tag labeling has become a popular method for comparative or quantitative analysis of protein expression in different biological states such as healthy versus disease patients, non-treated versus treated

conditions in different time points[77–82]. Isobaric labeling with tandem mass tags (TMT) allows for quantitation in many samples simultaneously in one experiment. TMT tags are composed of four groups, namely an amine-reactive group, a cleavable linker region, a mass normalization group, and a reporter ion group. The amine-reactive group can label peptides at N-termini and the ε-amino groups of lysine. The chemical structure and nominal mass of each amine-reactive TMT tag are identical. But each tag varies due to different isotopes located in the reporter ion and normalization groups. The difference in reporter ion mass is necessary for quantification among different tags, the mass normalization group offsets this mass difference. In this way, the combined groups of TMT tags have the same total molecular weights and structure. So that labeled peptides coelute as a single composite peak with the same m/z value in an MS1 scan. Upon fragmentation of the labeled peptides during the subsequent MS2 or higher-energy collisional dissociation (HCD) event, the sequence information is obtained from reporter ion peaks, and quantitation data are also simultaneously gained of the tags across samples[83]. TMT reagent family consists of TMTzero, TMT duplex (up to 2 samples), and TMT multiplex (ranging from 6 to 16 samples)[72].

While TMT tag labeling has been shown to improve phosphopeptides quantification[71,84–86] due to its robustness and efficiency, its problem of ratio compression remains. Several researchers have shown that TMT improved phosphoproteome studies. For example, Friedrich et al. showed comprehensive micro-scaled proteome and phosphoproteome from lung squamous cell and adenocarcinoma with more than 14,000 phosphosites and more than 8,000 quantified proteins with TMT strategy[85]. Jiang et al. proposed a new acquisition method for analyzing TMT-labeled multiplexed phosphoproteome samples, resulting in more than 12,000 phosphopeptides without prefractionation. They applied the technique in A549 cells treated with insulin or insulin growth factor 1, and as a result, they showed less ratio distortion as well[86]. Ratio distortion/compression is a concept that when isolating the target peptide in the MS1 spectrum for MS2 analysis, the presence of interfering peptides or peptides with the same m/z value within the isolation window is used for selection will also co-isolated. Since both the target peptide and the contaminating peptides carry the same reporter groups, in subsequent fragmentation events, these reporter ion signals for that particular isolation will be a summation of peptide ions, leading to a detrimental effect on the accuracy of isobaric tagging-based quantification in a complex sample. As a result, an underestimation of actual protein/peptide abundance is calculated. This distortion effect is more significant for low-abundance peptides, where the interfering signal is relatively more powerful.

For qualitative studies, ratio compression is not substantial. However, it is detrimental for quantitative analysis. Several solutions have been proposed for this problem. Some suggested reducing sample complexity through fractionation as well as narrowing the isolation window[87,88]. Ting et al. proposed MS3 almost eliminates interfering ions in complex samples[89]. Whereas Wenger demonstrated a new MS method named QuantMode to remove the co-isolated peptide ions of different charge states than the precursor to improving the quantification accuracy[90]. While all proposed solutions have some benefits, they either partially solve the problem or solve the problem at the cost of decreased throughput or rely on specialized MS instrumentation[87–90].

1.5.2 Dimethyl labeling

Dimethyl labeling is another common chemical labeling strategy that emerged in 2003 as a quantitative proteomics technology by Hsu and co-workers[91]. Briefly, a peptide forms a Schiff base by reacting formaldehyde with the N-termini or an ε-amino group at lysine residue, reduced by sodium cyanoborohydride to yield a secondary amine. The reaction is fast and completes in a few minutes without giving any significant side products, and has no adverse effects on MS2 peptide identification[91]. Because dimethyl labeling is relatively fast, specific, and mild, it is perfectly suitable for many biological applications. Heck and co-workers standardized in-solution, online and on-column protocols for dimethyl labeling to meet the requirements of different sample amounts for quantitative proteomics[68]. Although dimethyl labeling is commonly used for comparative quantification of two samples, it can be extended for multiplex analysis in a combination of different isotopic forms of the reagents with 28 Da, 30 Da, 32 Da, 34 Da, and 36 Da shift per site in MS analysis. Various isotopic forms of formaldehyde such as deuterated formaldehyde (CD_2O) or C13 formaldehyde ($^{13}CH_2O$) and sodium cyanoborohydride $NaBD_3CN$ or $NaBH_3CN$ are commercially available at low cost. Using an extra isotopic reagent combination such as $^{13}CD_2O/NaBD_3CN$, the multiplex can be expanded to a five-plex. Currently, multiplex labeling may be achieved by different combinations of reagents to generate an array of mass differences and apply to compare to multiple samples taken from other time points or different dosages.

Owing to the low cost, specificity, and simple strategy to label samples, stable isotope dimethyl labeling has been a better choice than other labeling reagents for PTM studies. For instance, Lemmer et al. demonstrated comparative phosphoproteomics of wild-type zebrafish

embryos and zebrafish embryos with morpholino-mediated knockdown of the Fyn/Yes kinases by investigating signaling pathways and phosphorylation events in convergence and extension cell movements[92]. Heck and co-workers improved the identification of phosphopeptides by combining low pH SCX and Ti^{4+}-IMAC with a triple-dimethyl labeled sample. They identified more than 9,000 unique phosphorylation sites within a 400 µg sample from a single experiment[93]. Overall, dimethyl labeling accompanies other phosphopeptides enrichment methods generating excellent quantitative results, and its workflow applies to every kind of biological sample.

1.6 Future perspectives

PTPs field is still in its early stage of development. Identifying their substrates remains a challenge, as even with many new methods explicitly developed for phosphatase or the different areas have entered the field of phosphatase research. MS-based high throughput enables the identification of tens of thousands of phosphorylation events and speeds up the discovery of phosphatase substrates. However, the complexity of PTPs and the issues related to MS experiments, such as sample preparation and quantitation strategies, still need further methods to address the problems related to the study and deepen the understanding of phosphatase biology. New insights into therapeutic development are expected to emerge as more progress is made in characterizing the signaling role of PTPs and their linkages to human disease, at either the level of the PTPs or from targets within the pathways they regulate.

1.7 References

(1) Ubersax, J. A.; Ferrell Jr, J. E. Mechanisms of Specificity in Protein Phosphorylation. *Nature* **2007**, *8*, 530–541.

(2) Roskoski, R. ERK1/2 MAP Kinases: Structure, Function, and Regulation. *Pharmacol. Res.* **2012**, *66* (2), 105–143.

(3) Ardito, F.; Giuliani, M.; Perrone, D.; Troiano, G.; Lo Muzio, L. The Crucial Role of Protein Phosphorylation in Cell Signaling and Its Use as Targeted Therapy (Review). *Int. J. Mol. Med.* **2017**, *40* (2), 271–280.

(4) Aschner, Y.; Downey, G. P. The Importance of Tyrosine Phosphorylation Control of Cellular Signaling Pathways in Respiratory Disease: PY and PY Not. *Am. J. Respir. Cell Mol. Biol.* **2018**, *59* (5), 535.

(5) Hunter, T. Tyrosine Phosphorylation: Thirty Years and Counting. *Curr. Opin. Cell Biol.* **2009**, *21* (2), 140–146.

(6) Qi, X. M.; Wang, F.; Mortensen, M.; Wertz, R.; Chen, G. Targeting an Oncogenic Kinase/Phosphatase Signaling Network for Cancer Therapy. *Acta Pharm. Sin. B* **2018**, *8* (4), 511–517.

(7) Vainonen, J. P.; Momeny, M.; Westermarck, J. Druggable Cancer Phosphatases. *Sci. Transl. Med.* **2021**, *13* (588), 1–14.

(8) Matozaki, T.; Murata, Y.; Saito, Y.; Okazawa, H.; Ohnishi, H. Protein Tyrosine Phosphatase SHP-2: A Proto-Oncogene Product That Promotes Ras Activation. *Cancer Sci.* **2009**, *100* (10), 1786–1793.

(9) Frankson, R.; Yu, Z.-H. H.; Bai, Y.; Li, Q.; Zhang, R.-Y. Y.; Zhang, Z.-Y. Y. Therapeutic Targeting of Oncogenic Tyrosine Phosphatases. *Cancer Res.* **2017**, *77* (21), 5701–5705.

(10) Julien, S. G.; Dubé, N.; Read, M.; Penney, J.; Paquet, M.; Han, Y.; Kennedy, B. P.; Muller, W. J.; Tremblay, M. L. Protein Tyrosine Phosphatase 1B Deficiency or Inhibition Delays ErbB2-Induced Mammary Tumorigenesis and Protects from Lung Metastasis. *Nat. Genet.* **2007**, *39* (3), 338–346.

(11) Veeriah, S.; Brennan, C.; Meng, S.; Singh, B.; Fagin, J. A.; Solit, D. B.; Paty, P. B.; Rohle, D.; Vivanco, I.; Chmielecki, J.; et al. The Tyrosine Phosphatase PTPRD Is a Tumor Suppressor That Is Frequently Inactivated and Mutated in Glioblastoma and Other Human Cancers. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106* (23), 9435–9440.

(12) Scott, A.; Wang, Z. Tumour Suppressor Function of Protein Tyrosine Phosphatase Receptor-T. *Biosci Rep* **2011**, *31* (5), 303–307.

(13) Sacco, F.; Perfetto, L.; Castagnoli, L.; Cesareni, G. The Human Phosphatase Interactome: An Intricate Family Portrait. *FEBS Lett.* **2012**, *586* (17), 2732–2739.

(14) Tonks, N. K. Protein Tyrosine Phosphatases: From Genes, to Function, to Disease. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (11), 833–846.

(15) Damle, N. P.; Köhn, M. The Human DEPhOsphorylation Database DEPOD: 2019 Update. *Database (Oxford).* **2019**, *2019*.

(16) Lachmann, A.; Ma'ayan, A. KEA: Kinase Enrichment Analysis. *Bioinforma. Appl. NOTE* **2009**, *25* (5), 684–686.

(17) Flint, A. J.; Tiganis, T.; Barford, D.; Tonks, N. K. Development of "Substrate-Trapping" Mutants to Identify Physiological Substrates of Protein Tyrosine Phosphatases. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (5), 1680–1685.

(18) Agazie, Y. M.; Hayman, M. J. Development of an Efficient "Substrate-Trapping" Mutant of Src Homology Phosphotyrosine Phosphatase 2 and Identification of the Epidermal Growth Factor Receptor, Gab1, and Three Other Proteins as Target Substrates. *J. Biol. Chem.* **2003**, *278* (16), 13952–13958.

(19) Iliuk, A. B.; Martin, V. A.; Alicie, B. M.; Geahlen, R. L.; Tao, W. A. In-Depth Analyses of Kinase-Dependent Tyrosine Phosphoproteomes Based on Metal Ion-Functionalized Soluble Nanopolymers. *Mol. Cell. Proteomics* **2010**, *9* (10), 2162–2172.

(20) Wolschin, F.; Weckwerth, W. Combining Metal Oxide Affinity Chromatography (MOAC) and Selective Mass Spectrometry for Robust Identification of in Vivo Protein Phosphorylation Sites. *Plant Methods* **2005**, *1* (1), 1–10.

(21) Aryal, U. K.; Ross, A. R. S. Enrichment and Analysis of Phosphopeptides under Different Experimental Conditions Using Titanium Dioxide Affinity Chromatography and Mass Spectrometry. *Rapid Commun. Mass Spectrom.* **2010**, *24* (2), 219–231.

(22) McNulty, D. E.; Annan, R. S. Hydrophilic Interaction Chromatography Reduces the Complexity of the Phosphoproteome and Improves Global Phosphopeptide Isolation and Detection. *Mol. Cell. Proteomics* **2008**, *7* (5), 971–980.

(23) Villén, J.; Gygi, S. P. The SCX_IMAC Enrichment Approach for Global Phosphorylation. *Nat. Protoc.* **2008**, *3* (10), 1630–1638.

(24) Wang, F.; Han, G.; Yu, Z.; Jiang, X.; Sun, S.; Chen, R.; Ye, M.; Zou, H. Fractionation of Phosphopeptides on Strong Anion-Exchange Capillary Trap Column for Large-Scale Phosphoproteome Analysis of Microgram Samples. *J. Sep. Sci.* **2010**, *33* (13), 1879–1887.

(25) Tiganis, T.; Bennett, A. M. M. Protein Tyrosine Phosphatase Function: The Substrate Perspective. *Biochem. J.* **2007**, *402* (1), 1–15.

(26) Catherman, A. D.; Skinner, O. S.; Kelleher, N. L. Top Down Proteomics: Facts and Perspectives. *Biochem Biophys Res Communr* **2014**, *445* (4), 683–693.

(27) Cristobal, A.; Marino, F.; Post, H.; Mohammed, S.; Heck, A. J. R. Toward an Optimized Work Flow for Middle-Down Proteomics. *Anal. Chem.* **2017**, *89*, 3318–3325.

(28) Rotunda, A. M.; Suzuki, H.; Moy, R. L.; Kolodney, M. S. Detergent Effects of Sodium Deoxycholate Are a Major Feature of an Injectable Phosphatidylcholine Formulation Used for Localized Fat Dissolution. *Dermatologic Surg.* **2004**, *30* (7), 1001–1008.

(29) Akins, R. E.; Levin, P. M.; Tuan, R. S. Cetyltrimethylammonium Bromide Discontinuous Gel Electrophoresis: Mr-Based Separation of Proteins with Retention of Enzymatic Activity. *Anal. Biochem.* **1992**, *202* (1), 172–178.

(30) Bhuyan, A. K. On the Mechanism of SDS-Induced Protein Denaturation. *Biopolymers* **2010**, *93* (2), 186–199.

(31) Huerta-Viga, A.; Woutersen, S. Protein Denaturation with Guanidinium: A 2D-IR Study. *J. Phys. Chem. Lett.* **2013**, *4* (20), 3397–3401.

(32) Hua, L.; Zhou, R.; Thirumalai C, D.; Berne, B. J. Urea Denaturation by Stronger Dispersion Interactions with Proteins than Water Implies a 2-Stage Unfolding. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105* (44), 16928–16933.

(33) Müller, T.; Winter, D. Systematic Evaluation of Protein Reduction and Alkylation Reveals Massive Unspecific Side Effects by Iodine-Containing Reagents. *Mol. Cell. Proteomics* **2017**, *16* (7), 1173–1187.

(34) Schmid, P. W. N.; Lim, N. C. H.; Peters, C.; Back, K. C.; Bourgeois, B.; Pirolt, F.; Richter, B.; Peschek, J.; Puk, O.; Amarie, O. V.; et al. Imbalances in the Eye Lens Proteome Are Linked to Cataract Formation. *Nat. Struct. Mol. Biol.* **2021**, *28* (2), 143–151.

(35) Washburn, M. P.; Wolters, D.; Yates, J. R. Large-Scale Analysis of the Yeast Proteome by Multidimensional Protein Identification Technology. *Nat. Biotechnol.* **2001**, *19* (3), 242–247.

(36) Li, Z.; Tremmel, D. M.; Ma, F.; Yu, Q.; Ma, M.; Delafield, D. G.; Shi, Y.; Wang, B.; Mitchell, S. A.; Feeney, A. K.; et al. Proteome-Wide and Matrisome-Specific Alterations during Human Pancreas Development and Maturation. *Nat. Commun.* **2021**, *12* (1), 1–12.

(37) Goodman, J. K.; Zampronio, C. G.; Jones, A. M. E.; Hernandez-Fernaud, J. R. Updates of the In-Gel Digestion Method for Protein Analysis by Mass Spectrometry. *Proteomics* **2018**, *18* (23).

(38) Basile, F.; Hauser, N. Rapid Online Non-Enzymatic Protein Digestion Combining Microwave Heating Acid Hydrolysis and Electrochemical Oxidation. *Anal. Chem* **2011**, *83* (1), 359–367.

(39) Rodríguez, J. C.; Wong, L.; Jennings, P. A. The Solvent in CNBr Cleavage Reactions Determines the Fragmentation Efficiency of Ketosteroid Isomerase Fusion Proteins Used in the Production of Recombinant Peptides. *Protein Expr. Purif.* **2003**, *28* (2), 224–231.

(40) Schultz, J. [28] Cleavage at Aspartic Acid. *Methods Enzymol.* **1967**, *11* (C), 255–263.

(41) Bornstein, P.; Balian, G. Cleavage at Asn-Gly Bonds with Hydroxylamine. *Methods Enzymol.* **1977**, *47* (C), 132–145.

(42) Dau, T.; Bartolomucci, G.; Rappsilber, J. Proteomics Using Protease Alternatives to Trypsin Bene Fi Ts from Sequential Digestion with Trypsin. *Anal. Chem.* **2020**, *92* (14), 9523–9527.

(43) Swaney, D. L.; Wenger, C. D.; Coon, J. J. The Value of Using Multiple Proteases for Large-Scale Mass Spectrometry-Based Proteomics. *J. Proteome Res.* **2010**, *9* (3), 1323.

(44) Dau, T.; Bartolomucci, G.; Rappsilber, J. Proteomics Using Protease Alternatives to Trypsin Benefits from Sequential Digestion with Trypsin. *Anal. Chem.* **2020**, *92* (14), 9523–9527.

(45) Vermachova, M.; Purkrtova, Z.; Santrucek, J.; Jolivet, P.; Chardot, T.; Kodicek, M. Combining Chymotrypsin/Trypsin Digestion to Identify Hydrophobic Proteins from Oil Bodies. *Methods Mol. Biol.* **2014**, *1072*, 185–198.

(46) Gershon, P. D. Cleaved and Missed Sites for Trypsin, Lys-C, and Lys-N Can Be Predicted with High Confidence on the Basis of Sequence Context. *J. Proteome Res.* **2013**, *13* (2), 702–709.

(47) Fu, Z.; Akula, S.; Thorpe, M.; Hellman, L. Marked Difference in Efficiency of the Digestive Enzymes Pepsin, Trypsin, Chymotrypsin, and Pancreatic Elastase to Cleave Tightly Folded Proteins. *Biol. Chem.* **2021**, *402* (7), 861–867.

(48) Switzar, L.; Giera, M.; Niessen, W. M. A. Protein Digestion: An Overview of the Available Techniques and Recent Developments. *J. Proteome Res.* **2013**, *12* (3), 1067–1077.

(49) Washburn, M. P.; Yates, J. R. Analysis of the Microbial Proteome. *Curr. Opin. Microbiol.* **2000**, *3* (3), 292–297.

(50) Hennrich, M. L.; Toorn, H. W. P. van den; Groenewold, V.; Heck, A. J. R.; Mohammed, S. Ultra Acidic Strong Cation Exchange Enabling the Efficient Enrichment of Basic Phosphopeptides. *Anal. Chem.* **2012**, *84* (4), 1804–1808.

(51) Dong, M.; Ye, M.; Cheng, K.; Song, C.; Pan, Y.; Wang, C.; Bian, Y.; Zou, H. Depletion of Acidic Phosphopeptides by SAX To Improve the Coverage for the Detection of Basophilic Kinase Substrates. *J. Proteome Res.* **2012**, *11* (9), 4673–4681.

(52) Batth, T. S.; Francavilla, C.; Olsen, J. V. Off-Line High-PH Reversed-Phase Fractionation for In-Depth Phosphoproteomics. *J. Proteome Res.* **2014**, *13* (12), 6176–6186.

(53) Zarei, M.; Sprenger, A.; Gretzmeier, C.; Dengjel, J. Combinatorial Use of Electrostatic Repulsion-Hydrophilic Interaction Chromatography (ERLIC) and Strong Cation Exchange (SCX) Chromatography for In-Depth Phosphoproteome Analysis. *J. Proteome Res.* **2012**, *11* (8), 4269–4276.

(54) Cox, J.; Neuhauser, N.; Michalski, A.; Scheltema, R. A.; Olsen, J. V.; Mann, M. Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment. *J. Proteome Res.* **2011**, *10* (4), 1794–1805.

(55) Diament, B.; Noble, W. S. Faster SEQUEST Searching for Peptide Identification from Tandem Mass Spectra. *J. Proteome Res.* **2011**, *10* (9), 3871.

(56) Brosch, M.; Yu, L.; Hubbard, T.; Choudhary, J. Accurate and Sensitive Peptide Identification with Mascot Percolator. *J. Proteome Res.* **2009**, *8* (6), 3176.

(57) Craig, R.; Beavis, R. C. TANDEM: Matching Proteins with Tandem Mass Spectra. *Bioinformatics* **2004**, *20* (9), 1466–1467.

(58) Bateman, N. W.; Goulding, S. P.; Shulman, N. J.; Gadok, A. K.; Szumlinski, K. K.; MacCoss, M. J.; Wu, C. C. Maximizing Peptide Identification Events in Proteomic Workflows Using Data-Dependent Acquisition (DDA). *Mol. Cell. Proteomics* **2014**, *13* (1), 329.

(59) Mann, M.; Hendrickson, R. C.; Pandey, A. Analysis of Proteins and Proteomes by Mass Spectrometry. *http://dx.doi.org/10.1146/annurev.biochem.70.1.437* **2003**, *70*, 437–473.

(60) Picotti, P.; Aebersold, R. Selected Reaction Monitoring–Based Proteomics: Workflows, Potential, Pitfalls and Future Directions. *Nat. Methods 2012 96* **2012**, *9* (6), 555–566.

(61) Wolf-Yadlin, A.; Hautaniemi, S.; Lauffenburger, D. A.; White, F. M. Multiple Reaction Monitoring for Robust Quantitative Proteomic Analysis of Cellular Signaling Networks. *Proc. Natl. Acad. Sci.* **2007**, *104* (14), 5860–5865.

(62) Cox, D. M.; Zhong, F.; Du, M.; Duchoslav, E.; Sakuma, T.; McDermott, J. C. Multiple Reaction Monitoring as a Method for Identifying Protein Posttranslational Modifications. *J. Biomol. Tech.* **2005**, *16* (2), 83.

(63) Rauniyar, N. Parallel Reaction Monitoring: A Targeted Experiment Performed Using High Resolution and High Mass Accuracy Mass Spectrometry. *Int. J. Mol. Sci.* **2015**, *16* (12), 28566.

(64) Hu, A.; Noble, W. S.; Wolf-Yadlin, A. Technical Advances in Proteomics: New Developments in Data-Independent Acquisition. *F1000Research* **2016**, *5* (419), 1–12.

(65) Thompson, A.; Wölmer, N.; Koncarevic, S.; Selzer, S.; Böhm, G.; Legner, H.; Schmid, P.; Kienle, S.; Penning, P.; Höhle, C.; et al. TMTpro: Design, Synthesis, and Initial Evaluation of a Proline-Based Isobaric 16-Plex Tandem Mass Tag Reagent Set. *Anal. Chem.* **2019**, *91* (24), 15941–15950.

(66) Mann, M. Functional and Quantitative Proteomics Using SILAC. *Nat. Rev. Mol. Cell Biol. 2006 712* **2006**, *7* (12), 952–958.

(67) Shiio, Y.; Aebersold, R. Quantitative Proteome Analysis Using Isotope-Coded Affinity Tags and Mass Spectrometry. *Nat. Protoc. 2006 11* **2006**, *1* (1), 139–145.

(68) Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J. R. Multiplex Peptide Stable Isotope Dimethyl Labeling for Quantitative Proteomics. *Nat. Protoc. 2009 44* **2009**, *4* (4), 484–494.

(69) L, Z.; JE, E. Relative Protein Quantification Using Tandem Mass Tag Mass Spectrometry. *Methods Mol. Biol.* **2017**, *1550*, 185–198.

(70) Mertins, P.; Tang, L. C.; Krug, K.; Clark, D. J.; Gritsenko, M. A.; Chen, L.; Clauser, K. R.; Clauss, T. R.; Shah, P.; Gillette, M. A.; et al. Reproducible Workflow for Multiplexed Deep-Scale Proteome and Phosphoproteome Analysis of Tumor Tissues by Liquid Chromatography–Mass Spectrometry. *Nat. Protoc. 2018 137* **2018**, *13* (7), 1632–1661.

(71) Zecha, J.; Satpathy, S.; Kanashova, T.; Avanessian, S. C.; Kane, M. H.; Clauser, K. R.; Mertins, P.; Carr, S. A.; Kuster, B. TMT Labeling for the Masses: A Robust and Cost-Efficient, In-Solution Labeling Approach *[S]. *Mol. Cell. Proteomics* **2019**, *18* (7), 1468–1478.

(72) Li, J.; Van Vranken, J. G.; Pontano Vaites, L.; Schweppe, D. K.; Huttlin, E. L.; Etienne, C.; Nandhikonda, P.; Viner, R.; Robitaille, A. M.; Thompson, A. H.; et al. TMTpro Reagents: A Set of Isobaric Labeling Mass Tags Enables Simultaneous Proteome-Wide Measurements across 16 Samples. *Nat. Methods 2020 174* **2020**, *17* (4), 399–404.

(73) S, W.; KA, R.; HE, M.; B, W. Protein Labeling by ITRAQ: A New Tool for Quantitative Mass Spectrometry in Proteome Research. *Proteomics* **2007**, *7* (3), 340–350.

(74) Tuli, L.; Tsai, T.-H.; Varghese, R. S.; Xiao, J. F.; Cheema, A.; Ressom, H. W. Using a Spike-in Experiment to Evaluate Analysis of LC-MS Data. *Proteome Sci.* **2012**, *10* (1), 13.

(75) Graves, P. R.; Haystead, T. A. J. Molecular Biologist's Guide to Proteomics. *Microbiol. Mol. Biol. Rev.* **2002**, *66* (1), 39.

(76) Aslam, B.; Basit, M.; Nisar, M. A.; Khurshid, M.; Rasool, M. H. Proteomics: Technologies and Their Applications. *J. Chromatogr. Sci.* **2017**, *55* (2), 182–196.

(77) Cheung, C. H. Y.; Juan, H.-F. Quantitative Proteomics in Lung Cancer. *J. Biomed. Sci. 2017 241* **2017**, *24* (1), 1–11.

(78) Nusinow, D. P.; Szpyt, J.; Ghandi, M.; Rose, C. M.; McDonald, E. R.; Kalocsay, M.; Jané-Valbuena, J.; Gelfand, E.; Schweppe, D. K.; Jedrychowski, M.; et al. Quantitative Proteomics of the Cancer Cell Line Encyclopedia. *Cell* **2020**, *180* (2), 387-402.e16.

(79) Morisaki, T.; Yashiro, M.; Kakehashi, A.; Inagaki, A.; Kinoshita, H.; Fukuoka, T.; Kasashima, H.; Masuda, G.; Sakurai, K.; Kubo, N.; et al. Comparative Proteomics Analysis of Gastric Cancer Stem Cells. *PLoS One* **2014**, *9* (11), e110736.

(80) Xu, H.-B.; Zhang, R.-F.; Luo, D.; Zhou, Y.; Wang, Y.; Fang, L.; Li, W.-J.; Mu, J.; Zhang, L.; Zhang, Y.; et al. Comparative Proteomic Analysis of Plasma from Major Depressive Patients: Identification of Proteins Associated with Lipid Metabolism and Immunoregulation. *Int. J. Neuropsychopharmacol.* **2012**, *15* (10), 1413–1425.

(81) Zhang, S.; He, D.; Yang, Y.; Lin, S.; Zhang, M.; Dai, S.; Chen, P. R. Comparative Proteomics Reveal Distinct Chaperone–Client Interactions in Supporting Bacterial Acid Resistance. *Proc. Natl. Acad. Sci.* **2016**, *113* (39), 10872–10877.

(82) Al-wajeeh, A. S.; Salhimi, S. M.; Al-Mansoub, M. A.; Khalid, I. A.; Harvey, T. M.; Latiff, A.; Ismail, M. N. Comparative Proteomic Analysis of Different Stages of Breast Cancer Tissues Using Ultra High Performance Liquid Chromatography Tandem Mass Spectrometer. *PLoS One* **2020**, *15* (1), e0227404.

(83) Thompson, A.; Schäfer, J.; Kuhn, K.; Kienle, S.; Schwarz, J.; Schmidt, G.; Neumann, T.; Hamon, C. Tandem Mass Tags: A Novel Quantification Strategy for Comparative Analysis of Complex Protein Mixtures by MS/MS. *Anal. Chem.* **2003**, *75* (8), 1895–1904.

(84) Hogrebe, A.; von Stechow, L.; Bekker-Jensen, D. B.; Weinert, B. T.; Kelstrup, C. D.; Olsen, J. V. Benchmarking Common Quantification Strategies for Large-Scale Phosphoproteomics. *Nat. Commun. 2018 91* **2018**, *9* (1), 1–13.

(85) Friedrich, C.; Schallenberg, S.; Kirchner, M.; Ziehm, M.; Niquet, S.; Haji, M.; Beier, C.; Neudecker, J.; Klauschen, F.; Mertins, P. Comprehensive Micro-Scaled Proteome and Phosphoproteome Characterization of Archived Retrospective Cancer Repositories. *Nat. Commun. 2021 121* **2021**, *12* (1), 1–15.

(86) X, J.; R, B.; J, B.; DL, D.; AM, R.; R, V.; AR, H.; Jiang, X.; Bomgarden, R.; Brown, J.; et al. Sensitive and Accurate Quantitation of Phosphopeptides Using TMT Isobaric Labeling Technique. *J. Proteome Res.* **2017**, *16* (11), 4244–4252.

(87) SY, O.; M, S.; J, N.; C, E.; PC, W. Minimising ITRAQ Ratio Compression through Understanding LC-MS Elution Dependence and High-Resolution HILIC Fractionation. *Proteomics* **2011**, *11* (11), 2341–2346.

(88) Savitski, M. M.; Sweetman, G.; Askenazi, M.; Marto, J. A.; Lang, M.; Zinn, N.; Bantscheff, M. Delayed Fragmentation and Optimized Isolation Width Settings for Improvement of Protein Identification and Accuracy of Isobaric Mass Tag Quantification on Orbitrap-Type Mass Spectrometers. *Anal. Chem.* **2011**, *83* (23), 8959–8967.

(89) Ting, L.; Rad, R.; Gygi, S. P.; Haas, W. MS3 Eliminates Ratio Distortion in Isobaric Labeling-Based Multiplexed Quantitative Proteomics. *Nat. Methods* **2011**, *8* (11), 937.

(90) CD, W.; MV, L.; AS, H.; GC, M.; DH, P.; MS, W.; JJ, C. Gas-Phase Purification Enables Accurate, Multiplexed Proteome Quantification with Isobaric Tagging. *Nat. Methods* **2011**, *8* (11), 933–935.

(91) Jue-Liang Hsu, †; Sheng-Yu Huang, †; Nan-Haw Chow, ‡ and; Shu-Hui Chen*, †; Hsu, J.-L.; Huang, S.-Y.; Chow, N.-H.; Chen, S.-H. Stable-Isotope Dimethyl Labeling for Quantitative Proteomics. *Anal. Chem.* **2003**, *75* (24), 6843–6852.

(92) Lemeer, S.; Jopling, C.; Gouw, J.; Mohammed, S.; Heck, A. J. R.; Slijper, M.; Hertog, J. den. Comparative Phosphoproteomics of Zebrafish Fyn/Yes Morpholino Knockdown Embryos. *Mol. Cell. Proteomics* **2008**, *7* (11), 2176–2187.

(93) Zhou, H.; Low, T. Y.; Hennrich, M. L.; Toorn, H. van der; Schwend, T.; Zou, H.; Mohammed, S.; Heck, A. J. R. Enhancing the Identification of Phosphopeptides from Putative Basophilic Kinase Substrates Using Ti (IV) Based IMAC Enrichment *. *Mol. Cell. Proteomics* **2011**, *10* (10).

CHAPTER 2. AN INTEGRATED PROTEOMIC STRATEGY TO IDENTIFY SHP2 SUBSTRATES

2.1 Summary

Protein phosphatases play an essential role in normal cell physiology and the development of diseases such as cancer. The challenges associated with studying protein phosphatases have limited our understanding of their substrates, molecular mechanisms, and unique functions within highly complicated networks. Here, we introduce a novel strategy using substrate-trapping mutant coupled with quantitative proteomics strategies to identify physiological substrates of Src homology 2 containing protein tyrosine phosphatase 2 (SHP2) in a high-throughput manner. The method integrates three parallel mass spectrometry-based proteomics experiments including affinity isolation of substrate-trapping mutant complex using wild-type and HEK293SHP2KO cells, *in vivo* global quantitative phosphoproteomics, and *in vitro* phosphatase reaction. We confidently identified eighteen direct substrates of SHP2 in EGFR signaling pathways, including both known and novel SHP2 substrates. Docking protein 1 (DOK1) was further validated using biochemical assays as a novel SHP2 substrate. This advanced workflow improves the systemic identification of direct substrates of phosphatases, facilitating our understanding of the equivalently important roles of protein phosphatase in cellular signaling.

2.2 Introduction

Protein phosphorylation is a critical and reversible protein post-translational modification mediated by a network of protein kinases and phosphatases. Balancing the dynamics of protein phosphorylation is vital in many biological events such as growth, proliferation, and cell death[1,2]. Perturbation of protein phosphorylation has been implicated in many human diseases, including different types of cancers[1,3–6], diabetes[7,8], autoimmune diseases[9,10], etc. It is estimated that approximately 1 million phosphorylation sites are present in the human proteome, most of which are phosphorylated on serine and threonine residues, with less than 2% on tyrosine residues[11–13].

Despite the rarity of phosphotyrosine signaling, the role of protein tyrosine kinases as protein oncogenes is widely recognized. For instance, a mutation in the receptor tyrosine kinase, EGFR, leads to its overexpression in 35-70% of glioblastomas[14,15] and more than 43% of lung

cancers[16,17]. Both BCR-ABL and TEL-ABL tyrosine kinases are shown to increase activation due to reciprocal translocation between chromosomes 9 and 12 in all cases of chronic myeloid leukemias[18,19]. Constitutive activation of the ALK kinase is also shown to be involved in inflammatory myofibroblastic tumors[20]. Similarly, HER2 gene amplification or overexpression is found in 15-30% of invasive breast tumors and is used as a prognostic and predictive biomaker[21]. Along the same lines, decreasing expression or activity of several protein tyrosine phosphatases (PTPs) is associated with cancer cell growth and survival. It can be linked to different diseases, like PTEN in breast cancer[22,23]. PTPRO in hepatocellular carcinoma[24,25], DUSP6/MKP3 in pancreatic cancer[26], or PTP1B in mammary tumor latency[27].

Both protein kinases and phosphatases offer important information for the study of the control of protein phosphorylation levels in signaling pathways and pathogenesis. However, there has been much more progress toward the understanding of kinase signaling[28–30], while our knowledge of phosphatases lags significantly behind. This discrepancy is partly caused by the experimental methods limiting the effective study of phosphatases. Investigation of phosphatases is not as straightforward as kinases, which could be examined by radiolabeling ATP as one of the techniques. In contrast, protein phosphatases remove a phosphate group from a substrate, resulting in losing the positive signal for detection and labeling. In addition, different phosphatase families work in a diverse and complex manner, which adds another layer of difficulty when examining phosphatase mechanisms. Hence, the discovery of physiological substrates for specific tyrosine phosphatases has remained challenging. Over time, various approaches have been developed to circumvent these restrictions to uncover the missing knowledge of phosphatase networks[31]. Compared to other protein-protein interactions, phosphatase-substrate interactions are transient and weak. Therefore, a direction affinity-based isolation is less effective. Thus, stabilizing the phosphatase-substrate interaction will be an essential start. The most common method here is a substrate-trapping mutant where an inactive phosphatase is created to recognize the substrate without removing the phosphate group from it. Hence, the mutated phosphatase can “trap” its substrate. This is followed by the isolation of the trapped phosphatase-substrate complex through affinity-based immunoprecipitation. Finally, the isolated complex and the phosphate group is detected either by western blotting or mass spectrometry.

Here, we have devised a novel mass spectrometry (MS)-based strategy by integrating substrate-trapping mutant-based immunoprecipitation, *in vivo* global phosphoproteomics, and *in*

vitro phosphatase reaction experiments to identify direct SHP2 substrates in the EGFR signaling pathway in a high throughput manner with high confidence. SHP2 phosphatase is an evolutionarily conserved protein with 593 amino acid residues. It contains three domains with a hydrophobic C-tail, two tandem SH2 domains (N-SH2 and C-SH2), and a catalytic PTP domain[32]. The x-ray structure of SHP2 demonstrates that in the inactive conformation, SHP2 adopts a closed and autoinhibited state in which the N-SH2 domain occupies the catalytic pocket of the PTP domain and sterically blocks the active site[33]. Hence, engagement of adaptor protein with the SHP2 tandem SH2 domains is required to overcome autoinhibition and turns SHP2 into its active form[34]. SHP2 is a *bona fide* proto-oncogene[35], and mutations of this gene not only occur in several types of cancer, like leukemia but also cause Noonan and LEOPARD syndromes[36–38]. As an essential component in several other oncogene signaling pathways, understanding the events underlying how SHP2 evokes cell transformation may provide new insights into the oncogenic mechanisms and novel targets for anti-cancer therapy. We are particularly interested in SHP2 substrates in the EGFR signaling pathway since SHP2 acts upstream of Ras in the EGFR pathway involved in the pathogenesis and progression of different carcinoma types. We confidently identified several novel proteins and three known SHP2 substrates by overlapping the potential substrate pool generated by the three MS-based experiments. Of these, docking protein 1 (DOK1) was further confirmed using biochemistry assays.

2.3 Materials and Methods

2.3.1 Cell culture, transfection, and substrate trapping

HEK293 SHP2 KO cell line was cultured at 37 °C and 5% CO2 in Dulbecco's modified Eagle's medium (Coring) supplemented with 10% fetal bovine serum (HyClone).

Wildtype or SHP2 DM (C459S and D425A) was cloned in the mammalian expression vector pCN-HA or pCN-FLAG. The plasmids were transfected into SHP2 KO HEK293 cells using polyethyleneimine (PEI) according to the manufacturer's protocol. *For immunoprecipitation and phosphatase assay experiments*, 48 hours after transfection, the cells were treated with 1 mM pervanadate, a powerful PTP inhibitor, for 30 min to accumulate c in cell phosphorylated proteins, then pervanadate was removed and cells were recovered in 10% FBS containing medium for another half an hour. Then Cells were washed twice with PBS and lysed in lysis buffer (20 mM

Tris-HCl, pH 7.5, 100 mM NaCl, 1 % Triton x-100, 10 % glycerol, 5 mM IAA). The lysate was sonicated in the ice-water bath at 15 W output with 3 bursts of 10 sec each and cooled on ice for 20 sec between each burst. The cell debris was cleared by centrifugation at 16,000 × g for 10 min at 4 °C, and the supernatant containing the soluble proteins was collected. Protein concentration was measured by BCA assay (Pierce) according to the manufacturer's protocol, and 8-10 mg of total protein from each sample was subjected to IP by the mouse anti-HA-agarose monoclonal beads (Sigma) for 3 hr. at 4 ℃. Then bead-bound protein complexes were washed with lysis buffer and PBS three times, respectively. The samples were subjected PTS digestion for MS analysis.

2.3.2 Methanol-Chloroform precipitation

The protein precipitation for the global phosphoproteomics experiment was performed as previously described39. Briefly, wildtype or SHP2DM (C459S and D425A) was cloned in the mammalian expression vector pCN-HA or pCN-FLAG. The plasmids were transfected into SHP2 KO HEK293 cells using polyethyleneimine (PEI) according to the manufacturer's protocol. 24 hours after transfection, the cells were starved overnight with serum free DMEM media, and then stimulated with 5 ng/ml EGF or 10 ng/ml HGF (Western blot) for 30 min. Then cells were washed twice with PBS and cell pellets were lysed in Gnd-HCl lysis buffer (6 M guanidine, 10 mM TCEP, 40 mM CAA, 1X protease inhibitor, 500 mM NaF, 100 mM Tris-HCl pH 8.5). The lysate was sonicated in the ice-water bath at 15 W output with 5 bursts of 10 sec each. The samples were cooled on ice for 20 sec between each burst. The sonicated lysate was boiled at 95 °C for 5 min. The cell debris was cleared by centrifugation at 16,000 × g for 10 min at 4 °C, and the supernatant containing the soluble proteins was collected. 150 μL of soluble protein was transferred into a new tube and methanol:chloroform:ddH2O with a volume ratio of 4:1:3, respectively, was added. The solution was mixed gently and then centrifuged at 16,000 x g for 3 min. The upper aqueous layer was removed, the protein pellet was washed with four volumes of methanol, and the precipitated protein pellet was air-dried for 10 min. The dried protein pellet was digested using the PTS protocol.

2.3.3 Phase transfer surfactant aided digestion and desalting

Protein pellets or bead-bound proteins were denatured using the phase-transfer surfactant-aided method (PTS) with few modifications. In brief, protein pellets or bead-bound proteins were resuspended in 12 mM SLS/12 mM SDC without or with 10 mM TCEP and 40 mM CAA in 100 mM Tris-HCl pH 8.5. The samples were boiled at 95 °C for 5 min and diluted fivefold with 50 mM TEAB. Lys-C was added at a 1:100 ratio by mass (enzyme: substrate) and the samples incubated for 3 hr at 37 °C. Proteomics-grade trypsin was added subsequently to a final 1:100 (enzyme: substrate) ratio and incubated overnight at 37 °C. On the next day, samples were acidified with 10% trifluoroacetic acid (TFA), detergents were removed by adding the equal volume of ethyl acetate. The aqueous phase was collected and desalted with either Sep-Pak 18 cartridges (Waters) or in-house packed SDB-XC (Empore) StageTips. The capacity for each cartridge is 1-2 mg of digested peptides, and for each SDB-XC StageTip is 10 µg. Each cartridge was activated with 1 mL methanol, equilibrated with 1 mL 80%ACN/0.1% TFA (Buffer B), and then washed with 1 mL of 5%ACN/0.1% TFA (Buffer A). Following sample loading, the cartridge was washed with 1mL of 5% Buffer A, and samples were eluted with 1 mL Buffer B. Each SDB-XC StageTip was equilibrated with 20 µL Buffer B, and washed with the same volume of Buffer A. Following sample loading, the tip was washed with 20 µL Buffer A, and samples were eluted with 20 µL Buffer B. Desalted samples were dried to completion in a vacuum concentrator. Dried peptide samples were stored at -80°C for further use.

2.3.4 Immunoprecipitation of phosphotyrosine containing peptides

Peptide samples for the global phosphoproteomics experiment were resuspended in 100 mM Tris-HCl buffer, pH 7.5. Anti-phosphotyrosine antibodies (clone PT66) conjugated to agarose beads (Sigma) were added at a ratio of 30 µL of bead slurry per mg of starting protein and incubated 12-16 h at 4 °C with rotation. On the next day, the beads with peptide solutions were loaded into a 10 µm filter tube (Mo Bi Tec), flow-through was collected for further PolyMAC enrichment. The beads were washed thrice with 550 µL of 100 mM Tris-HCl pH 7.5 and centrifuged at 350 x g for 1 min at 4 °C after each wash. The washed beads were transferred into a new 1.7 mL tube with 500 µL of water and the water removed with a fine gel-tip. Phosphopeptides were eluted at room temperature once with 100 µL of 0.1% TFA for 10 min with vigorous agitation, and twice

more with 100 μL of 0.1% TFA in 50% acetonitrile, 10 min each. All eluates were combined and dried in a vacuum concentrator.

2.3.5 PolyMAC enrichment of phosphopeptides

PolyMAC phosphopeptides enrichment was performed according to the manufacturer's instructions (Tymora Analytical, IN)[40]. Peptides were dissolved in 200 μL of the loading buffer and incubated with 50 μL of the PolyMAC bead slurry for 20 min with gentle agitation. The beads were loaded with samples into the tip with frit and the flowthrough collected for subsequent proteomics analysis after desalting. The beads were washed in the tip with 200 μL of loading buffer twice, then 200 μL of wash buffer once at 20 x g for 2 min at room temperature, followed by 100 x g spin down for 1 min or until the buffers were gone. The phosphopeptides were then eluted twice from the beads by adding 50 μL of elution buffer at the same centrifuge settings. The eluates were collected and dried in a vacuum centrifuge.

2.3.6 Phosphatase reaction in filtered aided sample preparation tube

The affinity bead-bound protein complexes were loaded into a Microcon 10 kDa (Millipore Sigma) filter tube using the FASP protocol[41]. SHP2 substrates were eluted by 25 μM sodium orthovanadate in 100 mM Tris-HCl pH 7.5 from the SHP2 complex. The eluted SHP2 substrate pool was divided into 2 portions and excessive sodium orthovanadate was removed by washing with 7 volume of 100 mM Tris-HCl before proceeding with the SHP2 dephosphorylation assay. One portion was incubated with WT SHP2 reaction buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 20 μg WT SHP2) overnight at 4 °C. Another portion was treated with SHP2-dead mutant reaction buffer (50 mM Tris-HCl pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, 20 μg SHP2DM) overnight at 4 °C. Both samples were reduced and alkylated with 10 mM TCEP/40 mM CAA at 37 °C for 30 min. The samples were washed with 200 μL of 50 mM TEAB once, then directed to Lys-C and trypsin digestion. The digests were collected with two washes of 200 μL of 50 mM TEAB buffer, and the peptides were then acidified with 10% TFA to pH ~3 and desalted using in-house packed SDB-XC (Empore) Stage-Tip with > 100 μg peptide capacity.

2.3.7 LC-MS/MS data acquisition

The phosphopeptides samples were resuspended in 15 μL of MS-grade 0.3% FA with 3% ACN. 4 μL of phosphopeptides were then injected into an Easy-NanoLC 1000 system (Thermo Fisher Scientific). Phosphopeptides were separated on a 45 cm in-house packed column (360 μm OD ×75 μm ID) containing C18 resin (2.2 μm bead size, 100Å pore size; Bischoff) with a column heater (Analytical Sales and Services) set at 60 °C. The mobile phase buffer consisted of 0.1% FA in MS-grade ultra-pure water (buffer A); the elution buffer (buffer B) consisted of 0.1% FA in 80% ACN. Samples were analyzed using a shallow gradient of 3% B to 35% B, followed by an increase to 95% B, and then a return to 6% B. The LC flow rate was set to 250 nL/min and total analysis times ranged from 60-90 min. The LC system was coupled online with a high-resolution hybrid dual-cell linear ion trap orbitrap mass spectrometer (LTQ-Orbitrap Velos Pro; Thermo Fisher). The mass spectrometer was operated in the data-dependent mode in which a full MS scan from mz 350–1,500 was followed by MS/MS scans of the 15 most intense ions. Ions with a charge state of +1 were excluded, and the mass exclusion time was 60 s.

2.3.8 Experimental design and statistical analysis

The raw files were searched against the *Homo sapiens* database with no redundant entries using the SEQUEST search engine built into Proteome Discoverer (version 2.2). Peptide precursor mass tolerance for the main search was set to 10 ppm, and MS/MS tolerance was set to 0.6 Da. Searches were performed with full tryptic digestion, and peptides were allowed a maximum of two missed cleavages. Search criteria included a static modification of cysteine residues of +57.021 Da to account for alkylation and a variable modification of +15.995 Da for potential oxidation of methionine residues. A variable modification of +79.996 Da was also set on serine, threonine, and tyrosine residues for the identification of phosphorylation sites. The false discovery rate (FDR) for PSMs, proteins, and phosphorylation sites was set to 1% for each analysis using a reverse decoy database, and proteins matching the reverse decoy database, or the common laboratory contaminant database were discarded.

For determination of IP-, global-, and dephospho- SHP2-dependent phosphorylation, label-free quantitation was also done with Proteome Discoverer, and samples were normalized based on total peptide amount. Both unique and razor peptides were used for protein quantitation. All three

experiments were performed three times, so tyrosine phosphorylation sites unique to the dephosphorylation samples, as well as sites with at least 2-fold increased phosphorylation relative to the control, were used for further analysis. Volcano plots and heat maps illustrating which tyrosine phosphorylation sites were up or down-regulated generated using Perseus 1.6.15.0[42]. First, the phosphorylation site intensities were transformed into their log2 counterparts, and any missing values were imputed according to a normal distribution using the default settings (width 0.3, down-shift 1.8). A two-sample Student's t-test (FDR $\leq$ 0.05) was then used to determine which phosphorylation sites were significantly different in the SHP2 dead mutant samples versus the SHP2-WT samples for each replicate. Significantly increased pTyr sites intensity in global phosphoproteome experiment were further analyzed using DAVID[43], STRING v11[44] and visualized using Cytoscape v3.7[45] with enriched pathways, biological processes, and molecular function (FDR *q*-value <0.05). The sequences containing the phosphosites identified in three separate experiments were aligned according to the phosphotyrosine residue with six amino acids on either side of the phosphosite and were analyzed with PhosphoSitePlus[46].

2.3.9 Wester blotting

Proteins were separated on 10% SDS-PAGE gels and transferred to nitrocellulose membrane. After blocking with 5% milk or bovine serum albumin in Tris-buffered saline with 0.1% Tween-20, the membranes were probed against the proteins of interest using the appropriate primary antibodies: Anti-Phosphotyrosine antibody 4G10 (EMD Millipore), anti-HA (Santa Cruz), anti-pY398-Dok1 (Thermo Fisher), anti-RasGAP (Cell Signaling Technologies), anti-GAPDH (Santa Cruz), anti-SHP2 (Santa Cruz) and anti-FLAG (Sigma) antibodies. These were followed by incubation with horseradish peroxidase-conjugated secondary antibodies (Cell Signaling). The blots were developed using the SuperSignal™ West Pico Chemiluminescent Substrate (Thermo Fisher Scientific).

2.3.10 Data availability

Mass spectrometry derived phosphoproteomics data from this study has been deposited to the ProteomeXchange Consortium (http://www.proteomexchange.org/) through the PRIDE partner repository with the dataset identifier PXD28404.

For reviewer account details:
Username: reviewer_pxz028404@ebi.ac.uk
Password: KFx3JCcp

2.4 Results

2.4.1 An integrated proteomic strategy to identify direct substrates of protein phosphatases

Tyrosine phosphorylation is a crucial component of many signaling pathways that regulate cell growth and survival, and its perturbation is a hallmark of many diseases, including cancers. Although a similar number of genes encode for tyrosine kinases and phosphatases in the human genome[47] to balance phosphorylation level, our understanding of phosphatases compared to kinases is still limited. According to the KEA dataset, there are 1830 known tyrosine kinase substrates up to date[48], but only 367 known tyrosine phosphatase substrates are reported in the DePOD database[49] (Supplemental Information Table S1 and S2). The challenge with the identification of tyrosine phosphatase substrates leads to a considerable knowledge gap compared to kinases.

The traditional strategies for identifying phosphatase substrates either suffer from weak phosphatase-substrate interactions, like a typical substrate-trapping mutant method with a single site mutation[50,51], interference of the neighboring proteins in chemical cross-linking[52] or BioID[53], or those that cannot distinguish between direct substrates and downstream protein-protein interactions[54–57]. Our novel strategy consists of three sets of MS experiments run in parallel, including pull-down of substrate-trapping double mutants complex, *in vivo* global phosphoproteomics, and *in vitro* phosphatase reaction, to improve the specificity and identify direct substrates of a phosphatase. We reason that overlapping the results from these three separate MS phosphoproteomics experiments allow us to identify the most biologically relevant substrate candidates with high confidence (Fig. 1). In the first experiment to isolate the SHP2 phosphatase-substrate complex, SHP2-WT and SHP2DM- substrate complexes were affinity-purified, digested, and the resulting peptides were further phospho-enriched using polymer-based metal ion affinity capture method (PolyMAC)[40], followed by LC-MS analysis. SHP2-WT activity causes a loss of phosphate group signal on tyrosine residues in peptides compared to the SHP2DM sample. Hence, comparing the phosphotyrosine levels in the two samples could identify potential SHP2 substrates

(Fig 1A). In the second experiment in parallel, we used an optimized global phosphoproteomics workflow based on label-free quantification to unravel SHP2-modulated downstream receptor tyrosine kinase signaling networks driven by epidermal growth factor (EGF) in SHP2-WT and SHP2KO HEK293 cells. SHP2DM or SHP2-WT were separately overexpressed in SHP2KO HEK293 cells. peptides containing phosphotyrosine residues were captured by immobilized anti-pY PT-66 antibody, further enriched by PolyMAC, and analyzed by LC-MS. The results provide information how SHP2 regulates downstream protein tyrosine phosphorylation and global pTyr response initiated by EGF (Fig 1B). In our third experiment, *in vitro* phosphatase assay, we used endogenous tyrosine phosphoproteins in the phosphatase substrate complex trapped by the SHP2DM mutant as the SHP2 substrate pool. Sodium orthovanadate, a competitive PTP inhibitor, was utilized to elute pTyr proteins from the complex to increase the specificity and reduce the background generated by the affinity pull down[58,59]. Purified recombinant SHP2-WT (active) or SHP2DM (inactive) phosphatase was subsequently added to the substrate pool in the dephosphorylation assay. Proteins after the phosphatase treatment were subjected to Lys-C/trypsin digestion, phosphopeptides enrichment by PolyMAC, and LC-MS/MS analysis (Fig 1C). Lastly, the overlap of the three separate MS-based experiments gave us the final list of high-confidence SHP2 substrates in the EGFR signaling pathway (Fig 1D).

2.4.2 Identification of SHP2 interacting phosphoproteins by immunoprecipitation of substrate trapping mutant complex

In the design of the first experiment, we reason that SHP2 substrates are likely SHP2-interacting proteins and at certain point tyrosine phosphorylated. A kinase adds a phosphate group which can be directly used to identify its substrates. Offering similar benefits, the PTP substrate-trapping mutants provide a valuable tool to protect a phosphate group from removal from potential substrates. A typical trapping mutant, with a serine replacing the catalytic cysteine residue of the PTP signature motif (C/S), recognizes and binds the substrates but has no PTP activity. Subsequently, the catalytic pocket of the phosphatase traps that substrate and forms a more stable phosphatase-substrate complex. SHP2 double mutant (SHP2 DM) with C459S and D425A mutations was demonstrated to be superior to the single C/S mutant by reinforcing substrate-phosphatase interaction and improving sensitivity for substrates detection[60]. Hence, we used the

double mutant SHP2 in our approach. To identify SHP2-interacting phosphoproteins as potential substrates, we introduced the HA tag fused with either SHP2 wildtype (SHP2-WT) or SHP2 DM into the SHP2 CRISPR deleted (SHP2KO) HEK293 cells. SHP2-WT and SHP2DM-substrate complexes were immunoprecipitated by anti-HA-agarose beads. The isolated proteins were digested, followed by isolation of phosphopeptides for MS analysis. MS analysis of substrate-trapping samples across three replicate experiments resulted in the identification of 1,798 peptides, of which 1,343 were phosphorylated, and 644 of the phosphosites were on phosphotyrosine (Fig. 2A). Among them, 606 phosphotyrosine peptides were identified in the SHP2DM overexpressed samples while 345 phosphotyrosine peptides in the SHP2-WT overexpressed samples. As expected, SHP2DM showed an approximately 2-fold increase in overall phosphotyrosine intensity (Fig. 2B). We defined the fold-change cutoff values in phosphotyrosine peptides and the statistical significance for discriminating differences in the quantitative results. We set the cutoff value of fold change based on the log2 ratio of phosphotyrosine peptides and student t-test with a false discovery rate (FDR) less than 0.05 across the two sample groups. Based on these two thresholds, we screened for phosphotyrosine sites with significant increases in the overexpressed SHP2DM group and found 383 such pTyr sites (Fig. 2C and Supplementary Table S3). SHP2 is ubiquitously expressed in cytosol due to its primary function regulating the intracellular signaling of RTKs in the cytoplasm. However, unconventional nuclear localization of SHP2 has also been identified by several studies[61–68]. Hence, SHP2 might be shuttling between different subcellular localizations to execute biological processes. Thus, we interrogated the subcellular localization of these upregulated phosphotyrosine proteins. To evaluate the subcellular localization of phosphotyrosine proteins that interacted with SHP2DM, we utilized Yloc subcellular prediction tool[69]. We found that 50% of the upregulated phosphotyrosine proteins were found in the cytoplasm, 30% in the nucleus, and 10% were shuffling between nucleus and cytoplasm (Fig. 2D). Furthermore, we investigated the relationships between the upregulated phosphotyrosine proteins in biological processes by examining known protein-protein interactions. We found extensive reported physical interactions between many of the upregulated phosphotyrosine proteins in biological processes, with protein tyrosine kinase binding (34%) and regulation of mRNA processing (25%) being the most over-represented (Fig. 2E & Supplementary Table S4).

2.4.3 Profiling SHP2-modulated tyrosine phosphorylation in EGFR signaling

We next profiled the global pTyr signaling response initiated by the receptor tyrosine kinases in SHP2-WT or SHP2DM HEK293 cells after stimulation with EGF. In doing so, we overexpressed SHP2DM or SHP2-WT in SHP2KO HEK293 cells. After starvation overnight, the cells were stimulated with EGF for 10 minutes. Cells were collected, followed by cell lysis, protein extraction, and digestion with Lys-C and trypsin. Tyrosine phosphopeptides (pTyr) were enriched first by immobilized anti-pY PT-66 antibody, further enriched by PolyMAC to improve the selectivity and analyzed by LC-MS (Fig 1B). We identified 4,537 total peptides and 4,242 total phosphopeptides, of which 1,965 were phosphorylated on tyrosine residues (Fig. 3A), among which 1,900 phosphotyrosine peptides were found in the SHP2DM sample, and 1,660 were identified in SHP2-WT (Fig. 3B). The ligand EGF induced intracellular signaling cascade, which was attenuated by the SHP2 activity, as was evidenced by the increased overall phosphotyrosine peptide intensity in SHP2DM compared to the SHP2-WT cells (Fig. 3C). Label-free quantitation was performed to distinguish SHP2-modulated tyrosine phosphorylation in the EGFR pathway. In total, 345 upregulated phosphotyrosine sites in 219 phosphoproteins were identified. Details of the phosphopeptides identified, including the peptide sequence and the PTM scores, are provided in Supplementary Table S5.

To better understand the phosphorylation event in 219 phosphoproteins, we performed the Gene Ontology (GO) functional annotation analysis according to the biological process and molecular function using DAVID tool[43] (Supplementary Table S6 and Supplementary Table S7) to compare to known SHP2 activities. Knowing that EGF was used to stimulate cells, it was not surprising that the top hit is epidermal growth factor receptor signaling pathways (Fig. 3D). Owing to the importance of SHP2 in biological processes, most phosphoproteins in the list were enriched in cell signaling transduction[70] and cell adhesion[70,71] in the top 15 hits (Fig. 3D). Consistent with our first experiment and previous studies[61,65,72], the global analysis also suggested that SHP2 might play a role in mRNA processing, influencing cellular transcription. GO enrichment analysis of molecular function has consistently proved that our candidate list of substrates was linked closely to the binding proteins and enzyme activity (Fig. 3E). To gain further insight into the protein-protein interaction network and functional category of phosphoproteins, we further analyzed in String[44] and visualized by Cytoscape[45]. In our list, we performed an MCODE cluster analysis in the Cytoscape on the entire 219 upregulated pTyr and showed in Fig. 3F. Several proteins formed

noticeable clusters and revealed significantly enriched cellular processes and functions. These included focal adhesion, transmembrane RTK signaling pathway, protein binding, cell communication that regulate the biological process. In addition, several proteins were related to mRNA processing, mRNA splicing, and translation.

To compare the known SHP2 dependent signaling pathways, we further mapped 219 upregulated pTyr proteins in three different signaling pathway databases, Kyoto Encyclopedia of Genes and Genomes (KEGG) (Fig. 3G), Wikipathways (Supplementary Fig. 1A), and Reactome (Supplementary Fig 1B) databases. There is no surprise that EGF/EGFR related signaling pathway was one of the top hits. In addition, cancer-related signaling pathways such as ERBB, RAS, VEGF, and GTPase are also highlighted in the top 15 hits (Fig. 3G). The results are consistent with our understanding that SHP2 plays a prominent role in regulating RAS/MAPK signaling pathways. SHP2 binds to and dephosphorylates RAS to increase its association with Raf and activate downstream proliferative RAS/ERK/MAPK signaling[73,74]. We identified a large proportion of phosphoproteins involved in RAS/MAPK signaling, as shown in Fig. 3H. In addition, we also identified a large number of phosphoproteins involved in focal adhesion (Fig. 3I). The previous report has demonstrated that SHP2 plays a role in FAK tyrosine dephosphorylation and promotes cell apoptosis[75].

2.4.4 Detection of SHP2 substrate *in vitro* through phosphatase reaction

To discover SHP2's direct substrate *in vitro* in our third experiment, we performed the SHP2 phosphatase assay using endogenous pTyr peptides isolated from cells as the substrate pool. FLAG tagged SHP2DM was firstly overexpressed in HEK293 SHP2KO cells. pretreatment of cells with pervanadate led to the maximum accumulation of phosphorylated SHP2 substrates, which could be recognized by FLAG-SHP2DM. The formed phosphatase-substrate complex was then isolated with anti-FLAG agarose beads and eluted competitively with sodium orthovanadate to increase the specificity and reduce the background[58,59]. Recombinant SHP2-WT (active) and SHP2DM (inactive as the control) phosphatase were purified and subsequently added to the substrate pool separately for the *in vitro* phosphatase reaction. After Lys-C and trypsin digestion, peptides were subjected to phosphopeptides enrichment by PolyMAC, followed by LC-MS/MS analysis (Fig 1C). We identified 366 pTyr peptides in SHP2DM, and 28 were in SHP2-WT (Fig.

4A). We compared the occurrences of each amino acid, six amino acid residues upstream and downstream of the pTyr site, to determine if *in vitro* phosphatase substrates reveal any sequence specificity. The generated results in Fig. 4B and motif analysis in Fig. 4C show that the substrate candidates do not appear to have strong consensus amino acid patterns as most kinases do. However, there are some apparent trends. On the N-terminal side of pTyr, there is an overrepresentation of acidic (Asp, Glu, pSer), with no basic residues. The great majority of the substrates contain at least one acidic residue at positions +1 and +2 in the C-terminal side of pTyr, whereas basic residues are rare. This shows good agreement between our SHP2 substrate list and the moderate SHP2 sequence specificity profiles reported previously in the literature[76–78].

2.4.5 Identification of direct substrates of SHP2 through integrating three proteomic strategies

Finally, we reason that overlapping the three substrate lists generated in the three parallel MS-based experiments would give us the best chance to uncover direct substrates of SHP2 in the EGFR signaling pathway with high confidence (Fig. 1). We overlapped three lists of the phosphoproteins that showed significant changes compared to the controls in each experiment and identified 18 tyrosine phosphoproteins (Fig. 4D) as direct substrates of SHP2 in EGFR signal pathways. Among them, src substrate cortactin (CTTN), microtubule-associated protein 1B (MAP1B), and tyrosine-protein phosphatase non-receptor type substrate 1 (SIRPA) (Table 1 & Supplementary Table S9) are contributed to the core cell migration. GRB2 associated binding protein 1 and 2 (GAB1/GAB2), and SIRPA proteins are critical nodes in the PI3K/AKT pathway that is a part of the EGFR signaling pathway. MPZL1 has been reported to be a SHP2 substrate that plays an important role in signaling pathways related to cell adhesion[79–81]. GAB1, GAB2, SIPRA, MPZL1 are known SHP2 substrates[60,82–85], highlighting overall high sensitivity and specificity of this integrated strategy.

2.4.6 SHP2 dephosphorylates DOK1 at Y398

Among the newly identified potential SHP2 substrates, we prioritized Docking protein 1 (DOK1) for further validation, as DOK1 is a negative regulator of multiple oncogenic signaling pathways, including ERK, AKT, STAT, and is a critical regulator of cell proliferation and cell

migration[86]. DOK1 was originally identified as a 62 kDa protein that binds with RasGAP. It has N-terminal pleckstrin homology (PH), phosphotyrosine binding (PTB) domains, and a C-terminal SH2-binding motif. The phosphotyrosine residues at its C-terminus play many important roles. For example, DOK1 pY362 and pY398 are essential for RasGAP binding and Ras, AKT activation, and DOK1 pY449 recruits Csk to active Src[87]. Cumulative evidence has shown that DOK1 exhibits tumor-suppressive activity in both non-hematopoietic and hematopoietic malignancies[87–91].

To further validate DOK1 as SHP2 substrate, we performed co-precipitation of DOK1 with SHP2 in HEK293 cells. We co-expressed FLAG-tagged SHP2-WT or SHP2DM with HA-tagged DOK1 in SHP2KO 293 cells using lipofectamine 3000. FLAG SHP2 was then immunoprecipitated for Wester blotting analysis. As shown in Fig 5A, upon EGF and HGF stimulation, SHP2DM trapping mutant could pull down DOK1, while SHP2-WT resulted in a much weaker binding with DOK1, indicating that the SHP2 substrate-trapping mutant could capture DOK1. When HA-tagged DOK1 was immunoprecipitated from the cell lysates (Fig. 5B), DOK1 showed a higher affinity with SHP2DM. More importantly, the overall pTyr level of DOK1, as detected by the anti-pY 4G10 antibody, was significantly lower in the SHP2-WT sample than both SHP2DM and vector only overexpressed cells, suggesting that SHP2-WT dephosphorylated DOK1 in the cells. DOK1 pY362 and pY398 were reported to be essential for RasGAP binding. Consistent with the literature report, RasGAP binding with DOK1 was much weaker in SHP2-WT cells than in the trapping mutant or vector-overexpressed cells, indicating that dephosphorylation by SHP2 impairs RasGAP binding with DOK1 (Fig. 5B).

The SHP2-WT or DOK1-SHP2DM complex was further isolated for MS analysis. The coverage for DOK1 in this experiment reached 80 % by identifying 26 phosphotyrosine sites that belonged to DOK1 protein (13 unique peptides) (Table 2 & Supplementary Table S10). The highlighted phosphotyrosine sites in Table 2 showed a significant increase in the intensity with at least a 1.5-fold change compared to the control sample group, suggesting that SHP2 might dephosphorylate DOK1 on these sites.

RasGAP, Ras GTPase activating protein, inactivates Ras by hydrolysis of Ras-GTP to Ras-GDP and plays a role in proliferation, migration, and anti- and pro-apoptosis. Given that MS results suggested that DOK1 pY398 is a potential dephosphorylation site by SHP2, and this site is essential for the binding for RasGAP, thus we next carried out a western blot analysis using a specific antibody for DOK1 pY398 to further confirm this finding. As expected, the

phosphorylation level of DOK1 Y398 was significantly reduced in the SHP2KO 293 cells with overexpressed SHP2-WT protein compared to the vector control (Fig. 5C), which suggests that SHP2 dephosphorylates this site *in situ*. To further confirm that SHP2 can directly dephosphorylate DOK1 Y398, FLAG-tagged DOK1 was overexpressed in SHP2KO 293 cells, and phosphorylated DOK1 was pulled down by anti-FLAG antibody after pervanadate treatment. An *in vitro* dephosphorylation assay was carried out with phosphorylated DOK1 and the recombinant SHP2 PTP domain at varying concentrations for 1 hour. The products were analyzed by western blot with anti-pY398-DOK1 antibody. As shown in Fig. 5D, the phosphorylation of DOK1 Y398 was reduced in an SHP2 concentration-dependent manner, indicating that SHP2 can directly dephosphorylate DOK1 *in vitro*. In aggregate, our present results indicate that DOK1 is a *bona fide* physiological substrate of SHP2.

2.5 Discussion and Conclusion

Here we illustrated an integrated phosphoproteomics strategy combined with three existing approaches (substrate trapping mutant pull down, *in vivo* global phosphoproteomics, and *in vitro* phosphatase) based MS assay for screening and identifying SHP2 phosphatase substrates. Our study identified several known SHP2 substrates and binding proteins and validated DOK1 Y398 as a physiological substrate of SHP2. Our strategy generated a high-confidence list of potential SHP2 substrates that reveal the enzyme's substrate specificity. However, while the method includes three analyses that provide three individual potential substrate pools, each has its advantages and disadvantages. Substrate-trapping mutant immunoprecipitation experiment can identify substrates based on the binding with the substrate trapping-mutant. Nonetheless, the tyrosine-phosphorylated binding proteins could also be identified due to their higher affinity to SHP2DM than SHP2-WT. While the global phosphoproteomics assay would reveal the proteins affected by the SHP2 activity *in vivo*, other phosphorylated proteins downstream of SHP2 could not be differentiated from the direct substrates. The *in vitro* phosphatase assay can recognize the potential direct substrates of SHP2. However, *in vitro* assay results do not always translate to physiological conditions *in vivo*. We find that combining these three MS-based experiments helps remove false-positive identification and discover/confirm the *bona-fide* direct physiological substrates of SHP2. Our strategy generates a list with the highest possibility of SHP2 substrates

that revealed the enzyme's substrate specificity. We identified 18 proteins (Table 2) through our strategy, in which GAB1[60,92,93], GAB2[83,94], SIRPA[84,95], MPZL1[80,81] are known SHP2 substrates, proving the feasibility of our approach. Meanwhile, we also identified 14 new SHP2 substrates, in which DOK1 drew most of our interest because of its essential role in ERK, AKT, STAT signal pathways, and some physiological processes like cell proliferation and cell migration. Therefore, we focused on DOK1 and validated its Y398 site as a *bona-fide* physiological substrate of SHP2. SHP2 is known to activate Ras, which is involved in a wide range of important cellular processes. Interestingly its underlying mechanism still has not been very clear. DOK1 pY398 is responsible for the binding of RasGAP, a negative regulator of Ras, the dephosphorylation of DOK1 pY398 by SHP2 might lead to reduced recruitment of RasGAP, and consequently overactivated of Ras/ERK MAPK signal pathway. Our work linking SHP2 to RAS through DOK1 might shed some light on the mechanism of how SHP2 plays a critical role in the MAPK signal pathway. We hope that the newly identified SHP2 substrates will help further define the biochemical and functional roles of SHP2 in various cellular processes.

2.6 References

(1) Östman, A.; Hellberg, C.; Böhmer, F. D. Protein-Tyrosine Phosphatases and Cancer. *Nat. Rev. Cancer* **2006**, *6* (4), 307–320.

(2) Bollu, L. R.; Mazumdar, A.; Savage, M. I.; Brown, P. H. Molecular Pathways: Targeting Protein Tyrosine Phosphatases in Cancer. *Clin. Cancer Res.* **2017**, *23* (9), 2136–2142.

(3) Cohen, P. The Role of Protein Phosphorylation in Human Health and Disease. *Eur. J. Biochem.* **2001**, *268* (19), 5001–5010.

(4) Arora, A.; Scholar, E. M. Role of Tyrosine Kinase Inhibitors in Cancer Therapy. *Journal of Pharmacology and Experimental Therapeutics*. American Society for Pharmacology and Experimental Therapeutics December 1, 2005, pp 971–979.

(5) Motiwala, T.; Jacob, S. T. Role of Protein Tyrosine Phosphatases in Cancer. *Prog. Nucleic Acid Res. Mol. Biol.* **2006**, *81*, 297–329.

(6) Hunter, T. The Genesis of Tyrosine Phosphorylation. *Cold Spring Harb. Perspect. Biol.* **2014**, *6* (5), 1–14.

(7) Louvet, C.; Szot, G. L.; Lang, J.; Lee, M. R.; Martinier, N.; Bollag, G.; Zhu, S.; Weiss, A.; Bluestone, J. A. Tyrosine Kinase Inhibitors Reverse Type 1 Diabetes in Nonobese Diabetic Mice. *Proc. Natl. Acad. Sci.* **2008**, *105* (48), 18895–18900.

(8) Schmelzle, K.; Kane, S.; Gridley, S.; Lienhard, G. E.; White, F. M. Temporal Dynamics of Tyrosine Phosphorylation in Insulin Signaling. *Diabetes* **2006**, *55* (8), 2171–2179.

(9) Weinblatt, M. E.; Kavanaugh, A.; Genovese, M. C.; Musser, T. K.; Grossbard, E. B.; Magilavy, D. B. An Oral Spleen Tyrosine Kinase (Syk) Inhibitor for Rheumatoid Arthritis. *N. Engl. J. Med.* **2010**, *363* (14), 1303–1312.

(10) Aschner, Y.; Downey, G. P. The Importance of Tyrosine Phosphorylation Control of Cellular Signaling Pathways in Respiratory Disease: PY and PY Not. *Am. J. Respir. Cell Mol. Biol.* **2018**, *59* (5), 535.

(11) Boersema, P. J.; Foong, L. Y.; Ding, V. M. Y.; Lemeer, S.; Van Breukelen, B.; Philp, R.; Boekhorst, J.; Snel, B.; Hertog, J. Den; Choo, A. B. H.; et al. In-Depth Qualitative and Quantitative Profiling of Tyrosine Phosphorylation Using a Combination of Phosphopeptide Immunoaffinity Purification and Stable Isotope Dimethyl Labeling. *Mol. Cell. Proteomics* **2010**, *9* (1), 84–99.

(12) Yoshizaki, H.; Okuda, S. Large-Scale Analysis of the Evolutionary Histories of Phosphorylation Motifs in the Human Genome. *Gigascience* **2015**, *4* (1), 1–6.

(13) Olsen, J. V; Blagoev, B.; Gnad, F.; Macek, B.; Kumar, C.; Mortensen, P.; Mann, M. Global, In Vivo, and Site-Specific Phosphorylation Dynamics in Signaling Networks. *Cell* **2006**, *127*, 635–648.

(14) Du, Z.; Lovly, C. M. Mechanisms of Receptor Tyrosine Kinase Activation in Cancer. *Mol. Cancer* **2018**, *17* (1), 1–13.

(15) Lopez-Gines, C.; Gil-Benso, R.; Ferrer-Luna, R.; Benito, R.; Serna, E.; Gonzalez-Darder, J.; Quilis, V.; Monleon, D.; Celda, B.; Cerdá-Nicolas, M. New Pattern of EGFR Amplification in Glioblastoma and the Relationship of Gene Copy Number with Gene Expression Profile. *Mod. Pathol.* **2010**, *23* (6), 856–865.

(16) Bethune, G.; Bethune, D.; Ridgway, N.; Xu, Z. Epidermal Growth Factor Receptor in Lung Cancer: An Overview and Update. *J Thorac Dis* **2010**, *2* (1), 48–51.

(17) M, S.; H, S.; K, H.; T, I.; Y, N.; JD, M.; AF, G.; T, F. Epidermal Growth Factor Receptor Expression Status in Lung Cancer Correlates With Its Mutation. *Hum. Pathol.* **2005**, *36* (10), 1127–1134.

(18) Paul, M. K.; Mukhopadhyay, A. K. Tyrosine Kinase – Role and Significance in Cancer. *Int. J. Med. Sci.* **2004**, *1* (2), 101.

(19) Million, R. P.; Aster, J.; Gilliland, D. G.; Etten, R. A. Van. The Tel-Abl (ETV6-Abl) Tyrosine Kinase, Product of Complex (9;12) Translocations in Human Leukemia, Induces Distinct Myeloproliferative Disease in Mice. *Blood* **2002**, *99* (12), 4568–4577.

(20) Scheijen, B.; Griffin, J. D. Tyrosine Kinase Oncogenes in Normal Hematopoiesis and Hematological Disease. *Oncogene* **2002**, *21* (21), 3314–3333.

(21) Iqbal, N. N.; Iqbal, N. N. Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications. *Mol. Biol. Int.* **2014**, *2014*, 1–9.

(22) Cully, M.; You, H.; Levine, A. J.; Mak, T. W. Beyond PTEN Mutations: The PI3K Pathway as an Integrator of Multiple Inputs during Tumorigenesis. *Nat. Rev. Cancer* **2006**, *6* (3), 184–192.

(23) Li, S.; Shen, Y.; Wang, M.; Yang, J.; Lv, M.; Li, P.; Chen, Z.; Yang, J. Loss of PTEN Expression in Breast Cancer: Association with Clinicopathological Characteristics and Prognosis. *Oncotarget* **2017**, *8* (19), 32043–32054.

(24) Tasneem, M.; Ghoshal, K.; Das, A.; Majumder, S.; Weichenhan, D.; Wu, Y.-Z.; Holman, K.; James, S. J.; Jacob, S. T.; Plass, C. Suppression of the Protein Tyrosine Phosphatase Receptor Type O Gene (PTPRO) by Methylation in Hepatocellular Carcinomas. *Oncogene* **2003**, *22* (41), 6319–6331.

(25) Hou, J.; Xu, J.; Jiang, R.; Wang, Y.; Chen, C.; Deng, L.; Huang, X.; Wang, X.; Sun, B. Estrogen-Sensitive PTPRO Expression Represses Hepatocellular Carcinoma Progression by Control of STAT3. *Hepatology* **2013**, *57* (2), 678–688.

(26) Furukawa, T.; Sunamura, M.; Motoi, F.; Matsuno, S.; Horii, A. Potential Tumor Suppressive Pathway Involving DUSP6/MKP-3 in Pancreatic Cancer. *Am. J. Pathol.* **2003**, *162* (6), 1807–1815.

(27) Julien, S. G.; Dubé, N.; Read, M.; Penney, J.; Paquet, M.; Han, Y.; Kennedy, B. P.; Muller, W. J.; Tremblay, M. L. Protein Tyrosine Phosphatase 1B Deficiency or Inhibition Delays ErbB2-Induced Mammary Tumorigenesis and Protects from Lung Metastasis. *Nat. Genet.* **2007**, *39* (3), 338–346.

(28) Frankson, R.; Yu, Z.-H. H.; Bai, Y.; Li, Q.; Zhang, R.-Y. Y.; Zhang, Z.-Y. Y. Therapeutic Targeting of Oncogenic Tyrosine Phosphatases. *Cancer Res.* **2017**, *77* (21), 5701–5705.

(29) Chan, R. J.; Feng, G.-S. PTPN11 Is the First Identified Proto-Oncogene That Encodes a Tyrosine Phosphatase. *Blood* **2007**, *109* (3), 862–867.

(30) Xue, L.; Wang, W.-H.; Iliuk, A.; Hu, L.; Galan, J. A.; Yu, S.; Hans, M.; Geahlen, R. L.; Tao, W. A. Sensitive Kinase Assay Linked with Phosphoproteomics for Identifying Direct Kinase Substrates. *Proc. Natl. Acad. Sci.* **2012**, *109* (15), 5615–5620.

(31) Zeng, L.; Wang, W. H.; Arrington, J.; Shao, G.; Geahlen, R. L.; Hu, C. D.; Tao, W. A. Identification of Upstream Kinases by Fluorescence Complementation Mass Spectrometry. *ACS Cent. Sci.* **2017**, *3* (10), 1078–1085.

(32) Arrington, J.; Xue, L.; Wang, W. H.; Geahlen, R. L.; Tao, W. A. Identification of the Direct Substrates of the ABL Kinase via Kinase Assay Linked Phosphoproteomics with Multiple Drug Treatments. *J. Proteome Res.* **2019**, *18* (4), 1679–1690.

(33) Fahs, S.; Lujan, P.; Köhn, M. Approaches to Study Phosphatases. *ACS Chem. Biol.* **2016**, *11* (11), 2944–2961.

(34) Neel, B. G.; Gu, H.; Pao, L. The 'Shp'ing News: SH2 Domain-Containing Tyrosine Phosphatases in Cell Signaling. *Trends Biochem. Sci.* **2003**, *28* (6), 284–293.

(35) Hof, P.; Pluskey, S.; Dhe-Paganon, S.; Eck, M. J.; Shoelson, S. E. Crystal Structure of the Tyrosine Phosphatase SHP-2. *Cell* **1998**, *92* (4), 441–450.

(36) Pluskey, S.; Wandless, T. J.; Walsh, C. T.; Shoelson, S. E. Potent Stimulation of SH-PTP2 Phosphatase Activity by Simultaneous Occupancy of Both SH2 Domains. *J. Biol. Chem.* **1995**, *270* (7), 2897–2900.

(37) Grossmann, K. S.; Rosário, M.; Birchmeier, C.; Birchmeier, W. The Tyrosine Phosphatase Shp2 in Development and Cancer. *Adv. Cancer Res.* **2010**, *106*, 53–89.

(38) Mohi, M. G.; Neel, B. G. The Role of Shp2 (PTPN11) in Cancer. *Curr. Opin. Genet. Dev.* **2007**, *17* (1), 23–30.

(39) Zhang, J.; Zhang, F.; Niu, R. Functions of Shp2 in Cancer. *J. Cell. Mol. Med.* **2015**, *19* (9), 2075–2083.

(40) Hsu, C.-C.; Zhu, Y.; Arrington, J. V; Paez, J. S.; Wang, P.; Zhu, P.; Chen, I.-H.; Zhu, J.-K.; Tao, W. A. Universal Plant Phosphoproteomics Workflow and Its Application to Tomato Signaling in Response to Cold Stress. *Mol. Cell. Proteomics* **2018**, *17* (10), 2068–2080.

(41) Iliuk, A. B.; Martin, V. A.; Alicie, B. M.; Geahlen, R. L.; Tao, W. A. In-Depth Analyses of Kinase-Dependent Tyrosine Phosphoproteomes Based on Metal Ion-Functionalized Soluble Nanopolymers. *Mol. Cell. Proteomics* **2010**, *9* (10), 2162–2172.

(42) Millipore Sigma. Microcon-10kDa Centrifugal Filter Unit with Ultracel-10 membrane | MRCPRT010 https://www.emdmillipore.com/US/en/product/Microcon-10kDa-Centrifugal-Filter-Unit-with-Ultracel-10-membrane,MM_NF-MRCPRT010#documentation (accessed Jun 18, 2020).

(43) Tyanova, S.; Temu, T.; Sinitcyn, P.; Carlson, A.; Hein, M. Y.; Geiger, T.; Mann, M.; Cox, J. The Perseus Computational Platform for Comprehensive Analysis of (Prote)Omics Data. *Nat. Methods* **2016**, *13* (9), 731–740.

(44) Dennis, G.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol. 2003 49* **2003**, *4* (9), 1–11.

(45) Szklarczyk, D.; Gable, A. L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N. T.; Morris, J. H.; Bork, P.; et al. STRING V11: Protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. *Nucleic Acids Res.* **2019**, *47* (D1), D607–D613.

(46) Paul Shannon, 1; Andrew Markiel, 1; Owen Ozier, 2 Nitin S. Baliga, 1 Jonathan T. Wang, 2 Daniel Ramage, 2; Nada Amin, 2; Benno Schwikowski, 1, 5 and Trey Ideker2, 3, 4, 5. Cytoscape: A Software Environment for Integrated Models. *Genome Res.* **1971**, *13* (22), 426.

(47) Hornbeck, P. V.; Zhang, B.; Murray, B.; Kornhauser, J. M.; Latham, V.; Skrzypek, E. PhosphoSitePlus, 2014: Mutations, PTMs and Recalibrations. *Nucleic Acids Res.* **2015**, *43* (Database issue), D512.

(48) Lachmann, A.; Ma'ayan, A. KEA: Kinase Enrichment Analysis. *Bioinforma. Appl. NOTE* **2009**, *25* (5), 684–686.

(49) Damle, N. P.; Köhn, M. The Human DEPhOsphorylation Database DEPOD: 2019 Update. *Database (Oxford).* **2019**, *2019*.

(50) Blanchetot, C.; Chagnon, M.; Dubé, N.; Hallé, M.; Tremblay, M. L. Substrate-Trapping Techniques in the Identification of Cellular PTP Targets. *Methods* **2005**, *35* (1), 44–53.

(51) Flint, A. J.; Tiganis, T.; Barford, D.; Tonks, N. K. Development of "Substrate-Trapping" Mutants to Identify Physiological Substrates of Protein Tyrosine Phosphatases. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (5), 1680–1685.

(52) Herzog, F.; Kahraman, A.; Boehringer, D.; Mak, R.; Bracher, A.; Walzthoeni, T.; Leitner, A.; Beck, M.; Hartl, F.-U.; Ban, N.; et al. Structural Probing of a Protein Phosphatase 2A Network by Chemical Cross-Linking and Mass Spectrometry. *Science* **2012**, *337* (6100), 1348–1352.

(53) Roux, K. J.; Kim, D. I.; Burke, B.; May, D. G. BioID: A Screen for Protein-Protein Interactions. *Curr. Protoc. protein Sci.* **2018**, *91*, 19.23.1-19.23.15.

(54) Batth, T. S.; Papetti, M.; Pfeiffer, A.; Tollenaere, M. A. X. X.; Francavilla, C.; Olsen Correspondence, J. V; Olsen, J. V.; Olsen Correspondence, J. V. Large-Scale Phosphoproteomics Reveals Shp-2 Phosphatase-Dependent Regulators of Pdgf Receptor Signaling. *CellReports* **2018**, *22* (10), 2601–2614.

(55) Rusin, S. F.; Schlosser, K. A.; Adamo, M. E.; Kettenbach, A. N. Quantitative Phosphoproteomics Reveals New Roles for the Protein Phosphatase PP6 in Mitotic Cells. *Sci. Signal.* **2015**, *8* (398), rs12.

(56) Hilger, M.; Bonaldi, T.; Gnad, F.; Mann, M. Systems-Wide Analysis of a Phosphatase Knock-down by Quantitative Proteomics and Phosphoproteomics. *Mol. Cell. Proteomics* **2009**, *8* (8), 1908–1920.

(57) Kanshin, E.; Michnick, S.; Thibault, P. Sample Preparation and Analytical Strategies for Large-Scale Phosphoproteomics Experiments. *Semin. Cell Dev. Biol.* **2012**, *23* (8), 843–853.

(58) Agazie, Y. M.; Hayman, M. J. Development of an Efficient "Substrate-Trapping" Mutant of Src Homology Phosphotyrosine Phosphatase 2 and Identification of the Epidermal Growth Factor Receptor, Gab1, and Three Other Proteins as Target Substrates. *J. Biol. Chem.* **2003**, *278* (16), 13952–13958.

(59) Gordon, J. A. Use of Vanadate as Protein-Phosphotyrosine Phosphatase Inhibitor. *Methods Enzymol.* **1991**, *201*, 477–482.

(60) Chen, C.; Cao, M.; Zhu, S.; Wang, C.; Liang, F.; Yan, L.; Luo, D. Discovery of a Novel Inhibitor of the Protein Tyrosine Phosphatase Shp2. *Sci. Rep.* **2015**, *5*, 17626.

(61) Huang, Y.; Wang, J.; Cao, F.; Jiang, H.; Li, A.; Li, J.; Qiu, L.; Shen, H.; Chang, W.; Zhou, C.; et al. SHP2 Associates with Nuclear Localization of STAT3: Significance in Progression and Prognosis of Colorectal Cancer. *Sci. Reports 2017 71* **2017**, *7* (1), 1–10.

(62) Idrees, M.; Kumar, V.; Joo, M.-D.; Ali, N.; Lee, K.-W.; Kong, I.-K. SHP2 Nuclear/Cytoplasmic Trafficking in Granulosa Cells Is Essential for Oocyte Meiotic Resumption and Maturation. *Front. Cell Dev. Biol.* **2021**, *0*, 1623.

(63) Ran, H.; Kong, S.; Zhang, S.; Cheng, J.; Zhou, C.; He, B.; Xin, Q.; Lydon, J. P.; DeMayo, F. J.; Feng, G.-S.; et al. Nuclear Shp2 Directs Normal Embryo Implantation via Facilitating the ERα Tyrosine Phosphorylation by the Src Kinase. *Proc. Natl. Acad. Sci.* **2017**, *114* (18), 4816–4821.

(64) Zhu, C.; Lindsey, S.; Konieczna, I.; Eklund, E. A. Constitutive Activation of SHP2 Protein Tyrosine Phosphatase Inhibits ICSBP-Induced Transcription of the Gene Encoding Gp91PHOX during Myeloid Differentiation. *J. Leukoc. Biol.* **2008**, *83* (3), 680.

(65) Aceto, N.; Sausgruber, N.; Brinkhaus, H.; Gaidatzis, D.; Martiny-Baron, G.; Mazzarol, G.; Confalonieri, S.; Quarto, M.; Hu, G.; Balwierz, P. J.; et al. Tyrosine Phosphatase SHP2 Promotes Breast Cancer Progression and Maintains Tumor-Initiating Cells via Activation of Key Transcription Factors and a Positive Feedback Signaling Loop. *Nat. Med. 2012 184* **2012**, *18* (4), 529–537.

(66) Zuo, C.; Wang, L.; Kamalesh, R. M.; Bowen, M. E.; Moore, D. C.; Dooner, M. S.; Reginato, A. M.; Wu, Q.; Schorl, C.; Song, Y.; et al. SHP2 Regulates Skeletal Cell Fate by Modifying SOX9 Expression and Transcriptional Activity. *Bone Res. 2018 61* **2018**, *6* (1), 1–13.

(67) Ran, H.; Kong, S.; Zhang, S.; Cheng, J.; Zhou, C.; He, B.; Xin, Q.; Lydon, J. P.; DeMayo, F. J.; Feng, G.-S.; et al. Nuclear Shp2 Directs Normal Embryo Implantation via Facilitating the ERα Tyrosine Phosphorylation by the Src Kinase. *Proc. Natl. Acad. Sci.* **2017**, *114* (18), 4816–4821.

(68) Jakob, S.; Schroeder, P.; Lukosz, M.; Büchner, N.; Spyridopoulos, I.; Altschmied, J.; Haendeler, J. Nuclear Protein Tyrosine Phosphatase Shp-2 Is One Important Negative Regulator of Nuclear Export of Telomerase Reverse Transcriptase. *J. Biol. Chem.* **2008**, *283* (48), 33155–33161.

(69) Briesemeister, S.; Rahnenführer, J.; Kohlbacher, O. YLoc-an Interpretable Web Server for Predicting Subcellular Localization. *Nucleic Acids Res.* **2010**, *38* (SUPPL. 2), 497–502.

(70) QU, C. K. The SHP-2 Tyrosine Phosphatase: Signaling Mechanisms and Biological Functions. *Cell Res. 2000 104* **2000**, *10* (4), 279–288.

(71) Agazie, Y. M.; Hayman, M. J. Molecular Mechanism for a Role of SHP2 in Epidermal Growth Factor Receptor Signaling. *Mol. Cell. Biol.* **2003**, *23* (21), 7875.

(72) Huang, W.-Q.; Lin, Q.; Zhuang, X.; Cai, L.-L.; Ruan, R.-S.; Lu, Z.-X.; Tzeng, C.-M. Structure, Function, and Pathogenesis of SHP2 in Developmental Disorders and Tumorigenesis. *Curr. Cancer Drug Targets* **2014**, *14* (6), 567–588.

(73) Dance, M.; Montagner, A.; Salles, J. P.; Yart, A.; Raynal, P. The Molecular Functions of Shp2 in the Ras/Mitogen-Activated Protein Kinase (ERK1/2) Pathway. *Cell. Signal.* **2008**, *20* (3), 453–459.

(74) Bunda, S.; Burrell, K.; Heir, P.; Zeng, L.; Alamsahebpour, A.; Kano, Y.; Raught, B.; Zhang, Z.-Y.; Zadeh, G.; Ohh, M. Inhibition of SHP2-Mediated Dephosphorylation of Ras Suppresses Oncogenesis. *Nat. Commun.* **2015**, *6* (1), 1–12.

(75) Rafiq, K.; Kolpakov, M. A.; Abdelfettah, M.; Streblow, D. N.; Hassid, A.; Dell'Italia, L. J.; Sabri, A. Role of Protein-Tyrosine Phosphatase SHP2 in Focal Adhesion Kinase Down-Regulation during Neutrophil Cathepsin G-Induced Cardiomyocytes Anoikis *. *J. Biol. Chem.* **2006**, *281* (28), 19781–19792.

(76) Bunda, S.; Burrell, K.; Heir, P.; Zeng, L.; Alamsahebpour, A.; Kano, Y.; Raught, B.; Zhang, Z.-Y.; Zadeh, G.; Ohh, M. Inhibition of SHP2-Mediated Dephosphorylation of Ras Suppresses Oncogenesis. *Nat. Commun. 2015 61* **2015**, *6* (1), 1–12.

(77) Ren, L.; Chen, X.; Luechapanichkul, R.; Selner, N. G.; Meyer, T. M.; Wavreille, A.-S.; Chan, R.; Iorio, C.; Zhou, X.; Neel, B. G.; et al. Substrate Specificity of Protein Tyrosine Phosphatases 1B, RPTPα, SHP-1, and SHP-2. *Biochemistry* **2011**, *50* (12), 2339.

(78) Rafiq, K.; Kolpakov, M. A.; Abdelfettah, M.; Streblow, D. N.; Hassid, A.; Dell'Italia, L. J.; Sabri, A. Role of Protein-Tyrosine Phosphatase SHP2 in Focal Adhesion Kinase Down-Regulation during Neutrophil Cathepsin G-Induced Cardiomyocytes Anoikis. *J. Biol. Chem.* **2006**, *281* (28), 19781–19792.

(79) Hadari, Y. R.; Kouhara, H.; Lax, I.; Schlessinger, J. Binding of Shp2 Tyrosine Phosphatase to FRS2 Is Essential for Fibroblast Growth Factor-Induced PC12 Cell Differentiation. *Mol. Cell. Biol.* **1998**, *18* (7), 3966–3973.

(80) Zhao, R.; Guerrah, A.; Tang, H.; Joe Zhao, Z. Cell Surface Glycoprotein PZR Is a Major Mediator of Concanavalin A-Induced Cell Signaling. *J. Biol. Chem.* **2002**, *277* (10), 7882–7888.

(81) Zhao, Z. J.; Zhao, R. Purification and Cloning of PZR, a Binding Protein and Putative Physiological Substrate of Tyrosine Phosphatase SHP-2. *J. Biol. Chem.* **1998**, *273* (45), 29367–29372.

(82) Fujioka, Y.; Matozaki, T.; Noguchi, T.; Iwamatsu, A.; Yamao, T.; Takahashi, N.; Tsuda, M.; Takada, T.; Kasuga, M. A Novel Membrane Glycoprotein, SHPS-1, That Binds the SH2-Domain-Containing Protein Tyrosine Phosphatase SHP-2 in Response to Mitogens and Cell Adhesion. *Mol. Cell. Biol.* **1996**, *16* (12), 6887–6899.

(83) Wheadon, H.; Paling, N. R. D.; Welham, M. J. Molecular Interactions of SHP1 and SHP2 in IL-3-Signalling. *Cell. Signal.* **2002**, *14* (3), 219–229.

(84) Ohnishi, H.; Murata, T.; Kusakari, S.; Hayashi, Y.; Takao, K.; Maruyama, T.; Ago, Y.; Koda, K.; Jin, F. J.; Okawa, K.; et al. Stress-Evoked Tyrosine Phosphorylation of Signal Regulatory Protein α Regulates Behavioral Immobility in the Forced Swim Test. *J. Neurosci.* **2010**, *30* (31), 10472–10483.

(85) Noguchi, T.; Matozaki, T.; Fujioka, Y.; Yamao, T.; Tsuda, M.; Takada, T.; Kasuga, M. Characterization of a 115-KDa Protein That Binds to SH-PTP2 , a Protein-Tyrosine Phosphatase with Src Homology 2 Domains , in Chinese Hamster Ovary Cells *. *J. Biol. Chem.* **1996**, *271* (44), 27652–27658.

(86) Lee, S.; Roy, F.; Galmarini, C. M.; Accardi, R.; Michelon, J.; Viller, A.; Cros, E.; Dumontet, C.; Sylla, B. S. Frameshift Mutation in the Dok1 Gene in Chronic Lymphocytic Leukemia. *Oncogene* **2004**, *23* (13), 2287–2297.

(87) M, Z.; JA, J.; M, N.; PP, P.; L, V. A. Dok-1 Independently Attenuates Ras/Mitogen-Activated Protein Kinase and Src/c-Myc Pathways to Inhibit Platelet-Derived Growth Factor-Induced Mitogenesis. *Mol. Cell. Biol.* **2006**, *26* (7), 2479–2489.

(88) Lee, S.; Huang, H.; Niu, Y.; Tommasino, M.; Lenoir, G.; Sylla, B. S. Dok1 Expression and Mutation in Burkitt's Lymphoma Cell Lines. *Cancer Lett.* **2007**, *245* (1–2), 44–50.

(89) Berger, A. H.; Niki, M.; Morotti, A.; Taylor, B. S.; Socci, N. D.; Viale, A.; Brennan, C.; Szoke, J.; Motoi, N.; Rothman, P. B.; et al. Identification of DOK Genes as Lung Tumor Suppressors. *Nat. Genet.* **2010**, *42* (3), 216–223.

(90) Mashima, R.; Honda, K.; Yang, Y.; Morita, Y.; Inoue, A.; Arimura, S.; Nishina, H.; Ema, H.; Nakauchi, H.; Seed, B.; et al. Mice Lacking Dok-1, Dok-2, and Dok-3 Succumb to Aggressive Histiocytic Sarcoma. *Lab. Invest.* **2010**, *90* (9), 1357–1364.

(91) M, N.; A, D. C.; M, Z.; H, H.; H, H.; L, V. A.; C, C.-C.; PP, P. Role of Dok-1 and Dok-2 in Leukemia Suppression. *J. Exp. Med.* **2004**, *200* (12), 1689–1695.

(92) Koncz, G.; Tóth, G. K.; Bökönyi, G.; Kéri, G.; Pecht, I.; Medgyesi, D.; Gergely, J.; Sármay, G. Co-Clustering of Fcγ and B Cell Receptors Induces Dephosphorylation of the Grb2-Associated Binder 1 Docking Protein. *Eur. J. Biochem.* **2001**, *268* (14), 3898–3906.

(93) Kapoor, G. S.; Zhan, Y.; Johnson, G. R.; O'Rourke, D. M. Distinct Domains in the SHP-2 Phosphatase Differentially Regulate Epidermal Growth Factor Receptor/NF-KB Activation through Gab1 in Glioblastoma Cells. *Mol. Cell. Biol.* **2004**, *24* (2), 823–836.

(94) Nishida, K.; Yoshida, Y.; Itoh, M.; Fukada, T.; Ohtani, T.; Shirogane, T.; Atsumi, T.; Takahashi-Tezuka, M.; Ishihara, K.; Hibi, M.; et al. Gab-Family Adapter Proteins Act Downstream of Cytokine and Growth Factor Receptors and T- and B-Cell Antigen Receptors. *Blood* **1999**, *93* (6), 1809–1816.

(95) Peng, H. Y.; Chen, G. Den; Lai, C. Y.; Hsieh, M. C.; Lin, T. Bin. Spinal SIRPα1-SHP2 Interaction Regulates Spinal Nerve Ligation-Induced Neuropathic Pain via PSD-95-Dependent NR2B Activation in Rats. *Pain* **2012**, *153* (5), 1042–1053.

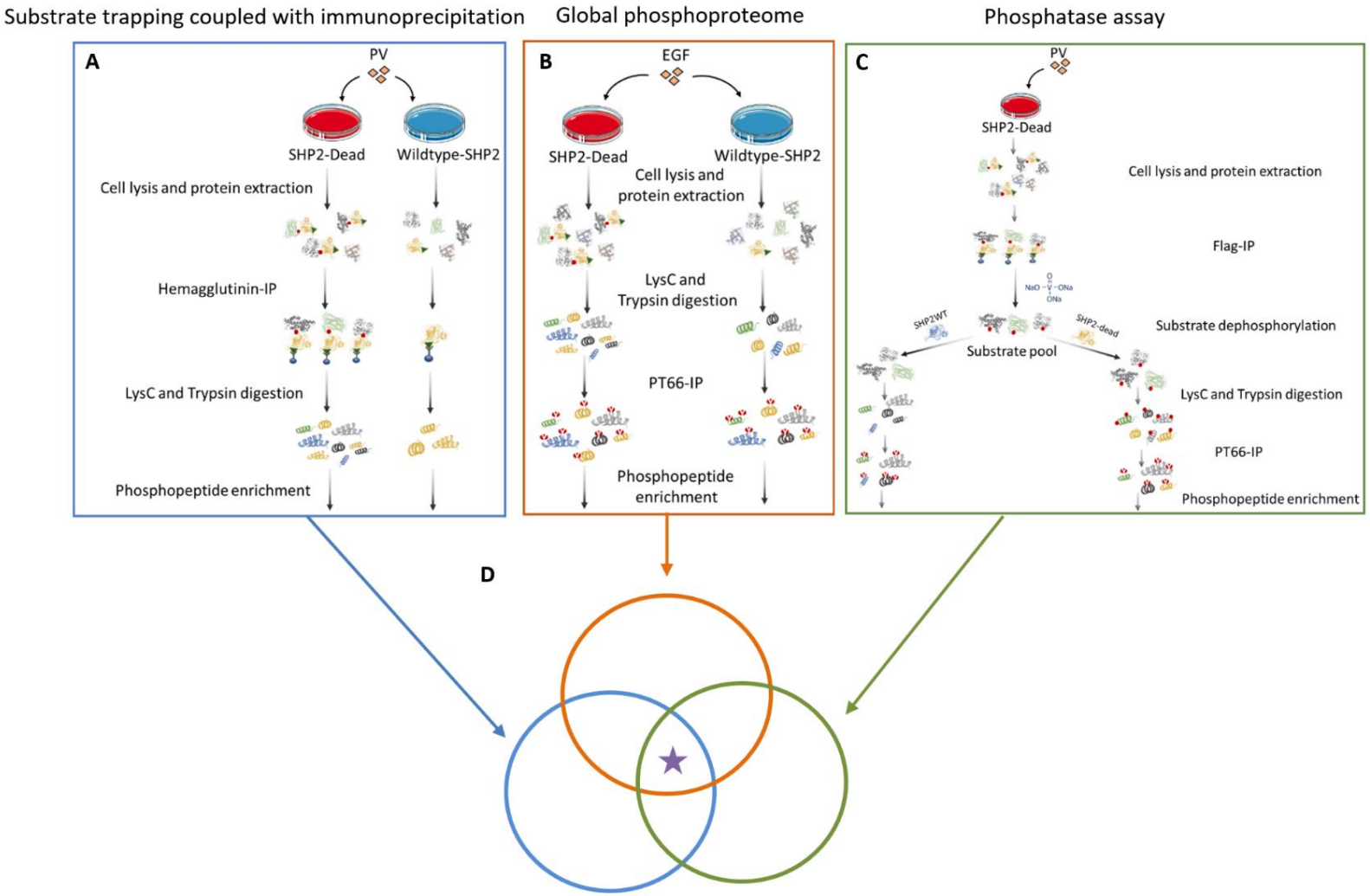


Figure 2-1: Experiment design for phosphoproteomics workflow to identify potential substrates of SHP2. A) IP procedure with substrate trapping mutant; B) Global phosphoproteomics procedure; C) Phosphatase assay procedure; D) The resulting phosphopeptides from the three MS experiments were overlapped together to determine the direct substrate candidates.

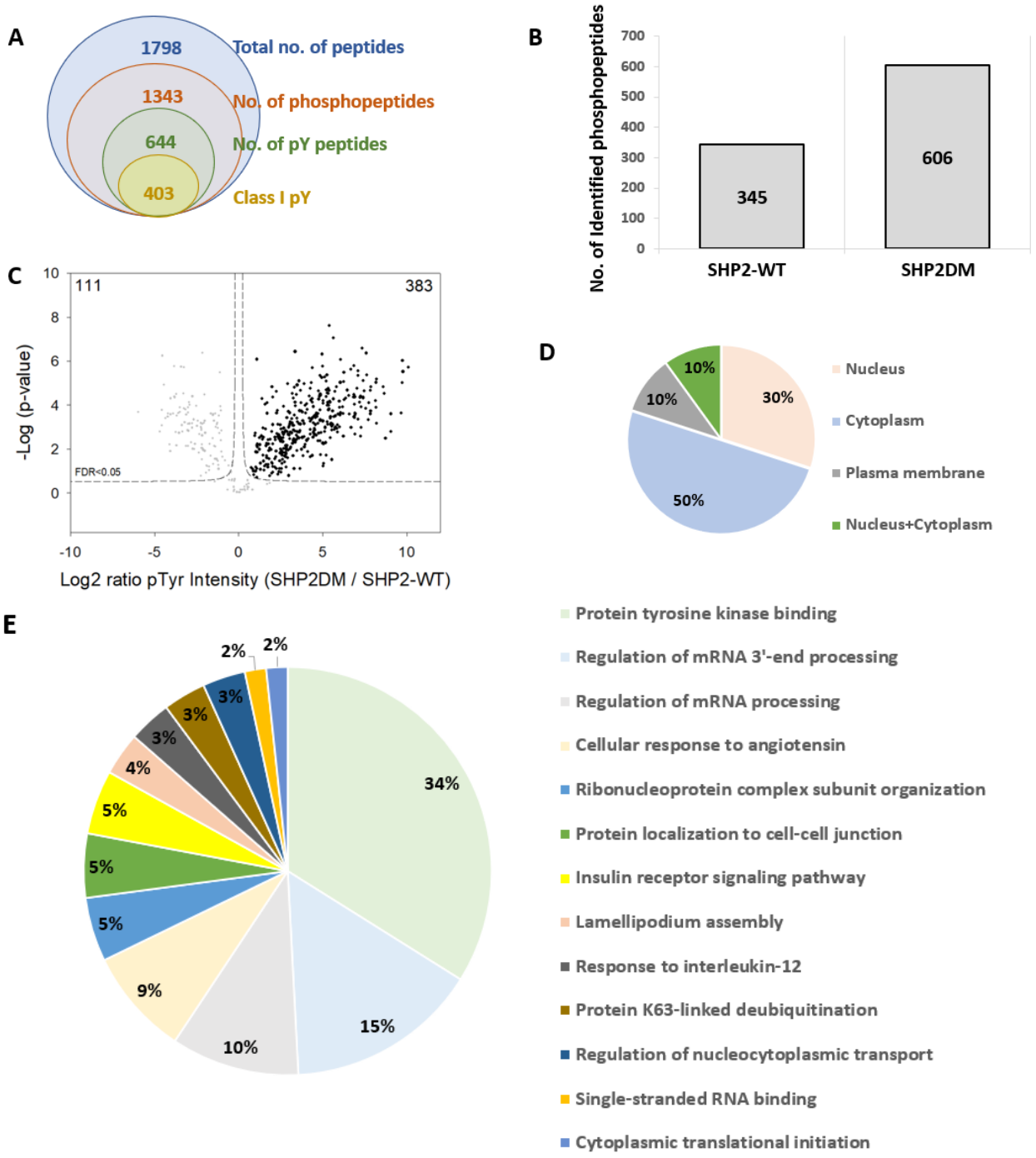
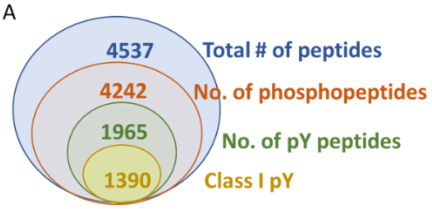


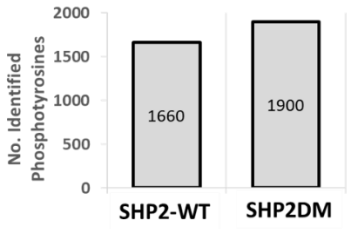
Figure 2-2: Phosphotyrosine sites identified in the immunoprecipitation experiment. A) Contribution of total identified peptides and phosphopeptides; B) Number of unique phosphopeptides identified in SHP2-WT, or phosphatase-dead SHP2 samples; C) Volcano plot of the significantly changed pTyr sites effected by the SHP2 activity; D) Subcellular localization of the upregulated phosphotyrosine proteins; E) Gene Ontology analysis of the upregulated phosphotyrosine proteins.

Figure 2-3: Phosphotyrosine sites identified in the global phosphoproteomics experiment. A) Total number of unique peptides and phosphopeptides identified; B) Contribution of phosphopeptides in each sample; C) Box plot of the distribution of pTyr site intensities in the SHP2-WT, and phosphatase-dead SHP2 samples; GO enrichment of top 15 hits using DAVID tool, as categorized according to D) biological processes and E) molecular functions; F) Protein-protein interaction network and functional category of upregulated pTyr proteins in STRING database; G) Pathway analysis of upregulated pTyr proteins by the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (FDR<0.05) with top 15 hits; H) Protein-protein interaction in focal adhesion and I) RAS/MAPK signaling pathway.

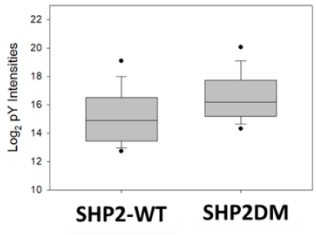
A


4537
Total # of peptides
4242
No. of phosphopeptides
1965
No. of pY peptides
1390
Class I pY

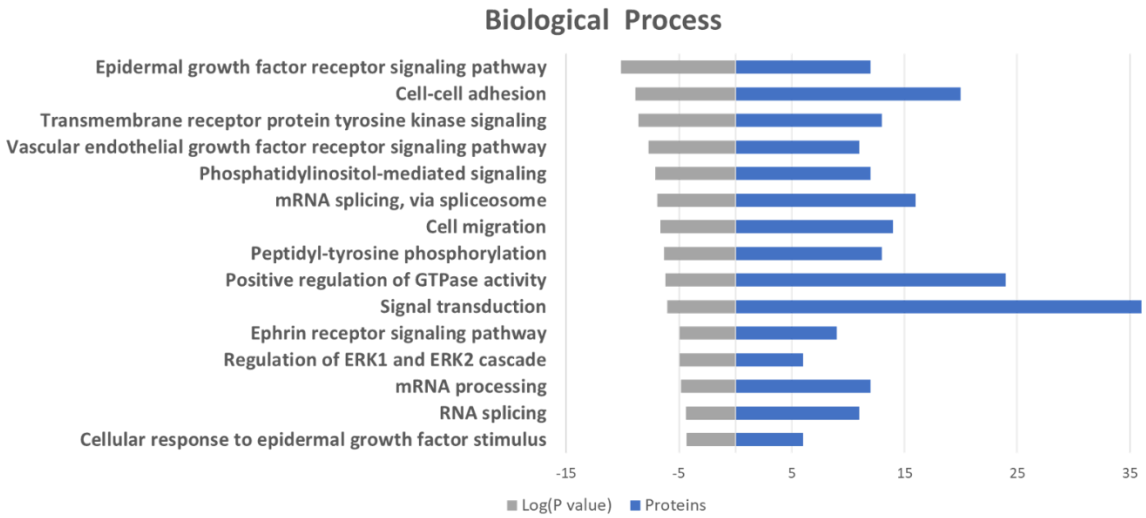
B


No. Identified Phosphotyrosines
2000
1500
1000
500
0
1660
1900
SHP2-WT
SHP2DM

C


Log2 pY Intensities
22
20
18
16
14
12
10
SHP2-WT
SHP2DM

D


Biological Process
Epidermal growth factor receptor signaling pathway
Cell-cell adhesion
Transmembrane receptor protein tyrosine kinase signaling
Vascular endothelial growth factor receptor signaling pathway
Phosphatidylinositol-mediated signaling
mRNA splicing, via spliceosome
Cell migration
Peptidyl-tyrosine phosphorylation
Positive regulation of GTPase activity
Signal transduction
Ephrin receptor signaling pathway
Regulation of ERK1 and ERK2 cascade
mRNA processing
RNA splicing
Cellular response to epidermal growth factor stimulus
-15
-5
5
15
25
35
Log(P value)
Proteins

E

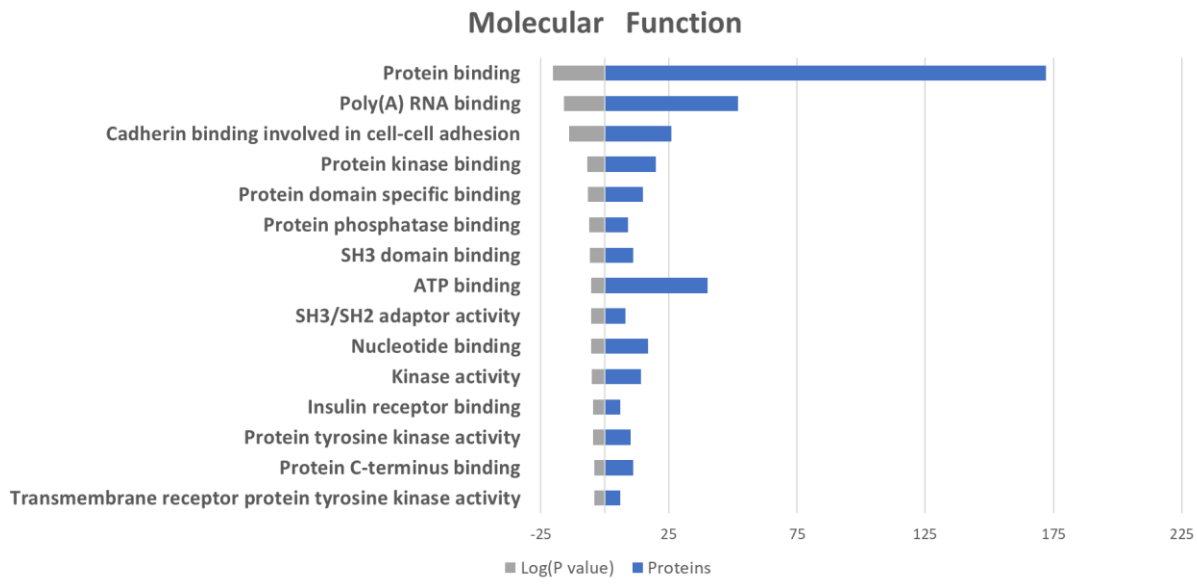
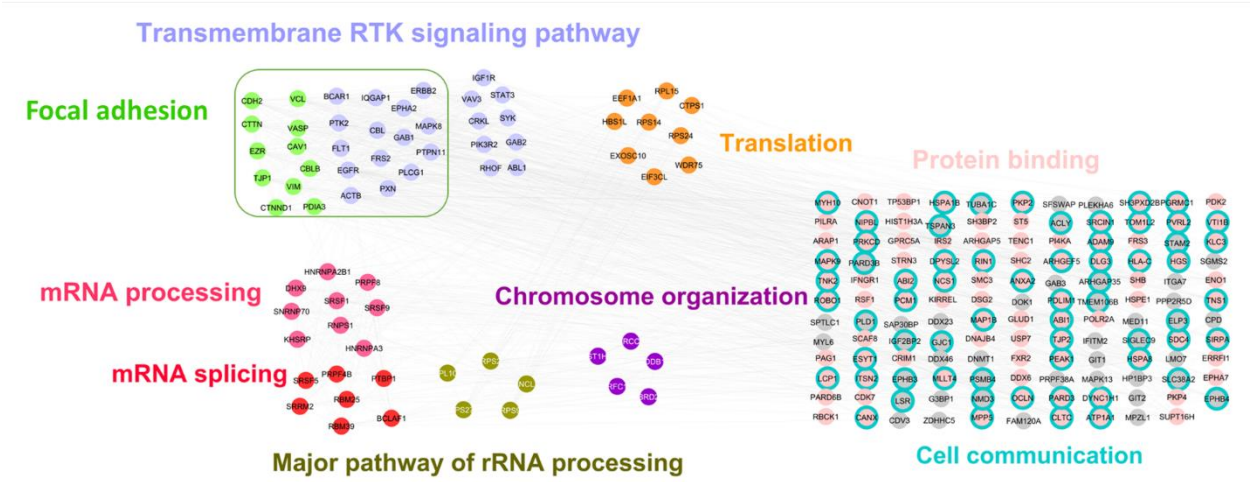

Molecular Function
Protein binding
Poly(A) RNA binding
Cadherin binding involved in cell-cell adhesion
Protein kinase binding
Protein domain specific binding
Protein phosphatase binding
SH3 domain binding
ATP binding
SH3/SH2 adaptor activity
Nucleotide binding
Kinase activity
Insulin receptor binding
Protein tyrosine kinase activity
Protein C-terminus binding
Transmembrane receptor protein tyrosine kinase activity
-25
25
75
125
175
225
Log(P value)
Proteins

Figure 2-3 continued

F



G

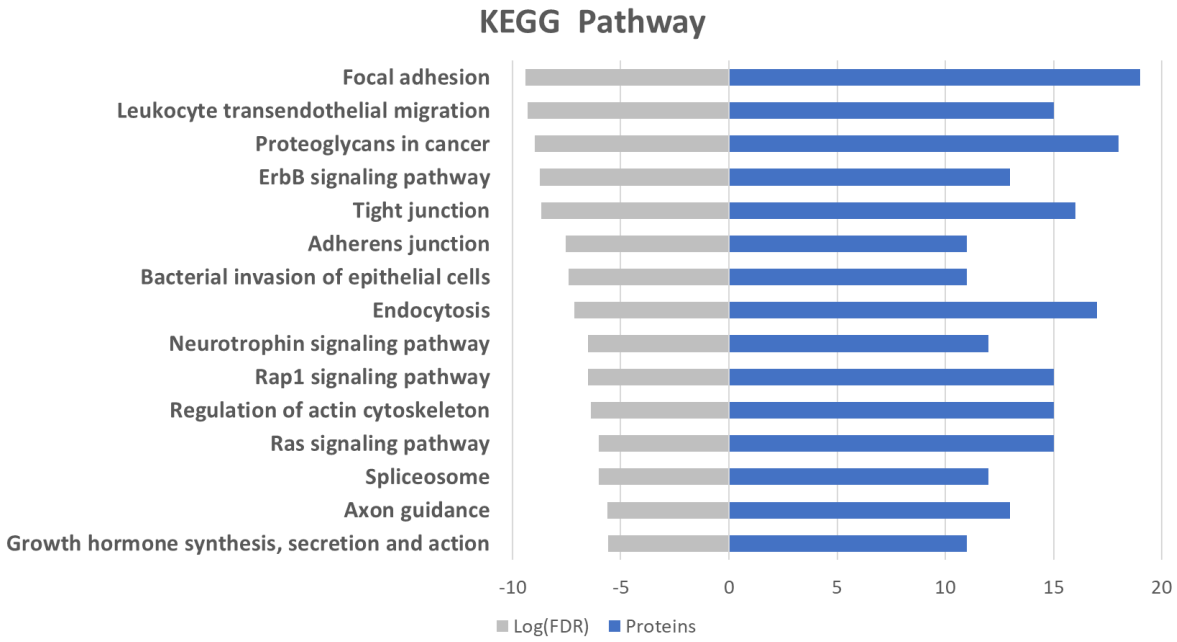
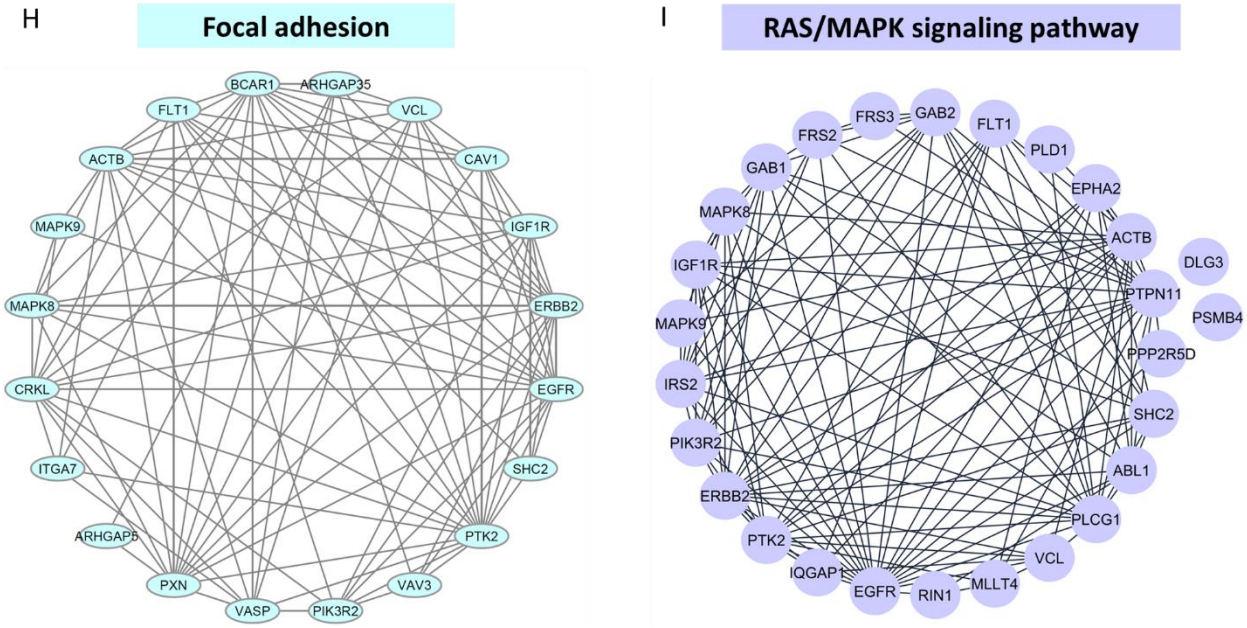


Figure 2-3 continued



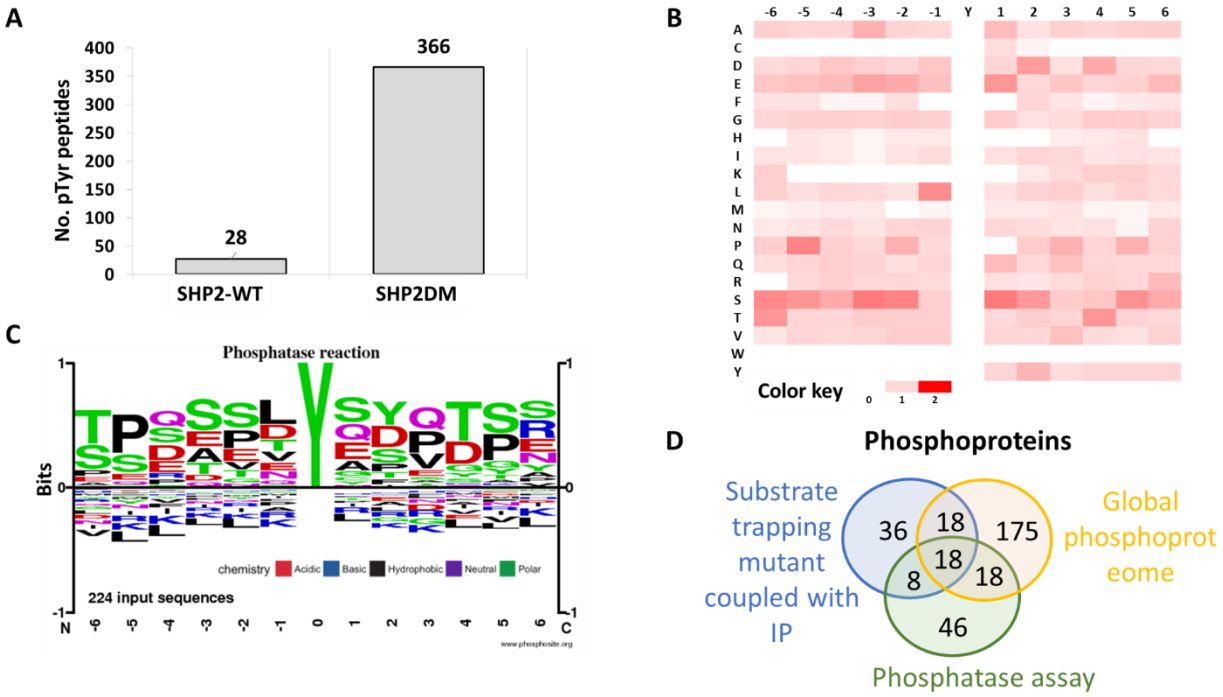


Figure 2-4: Phosphotyrosine sites identified in the SHP2-dependent dephosphorylation assay experiment. A) Number of unique pTyr peptides identified in the dephosphorylation assay using SHP2DM and SHP2-WT; B)Motif analysis of *in vitro* phosphatase reaction for SHP2 substrate specificity showing in heat map for the distribution of amino acid residues; C) Sequence logo visualization generated from PhosphoSitePlus[47] for SHP2-dependent dephosphorylation sites of candidate substrates; D) Overlap based on the phosphoproteins across the three MS phosphoproteomics experiments.

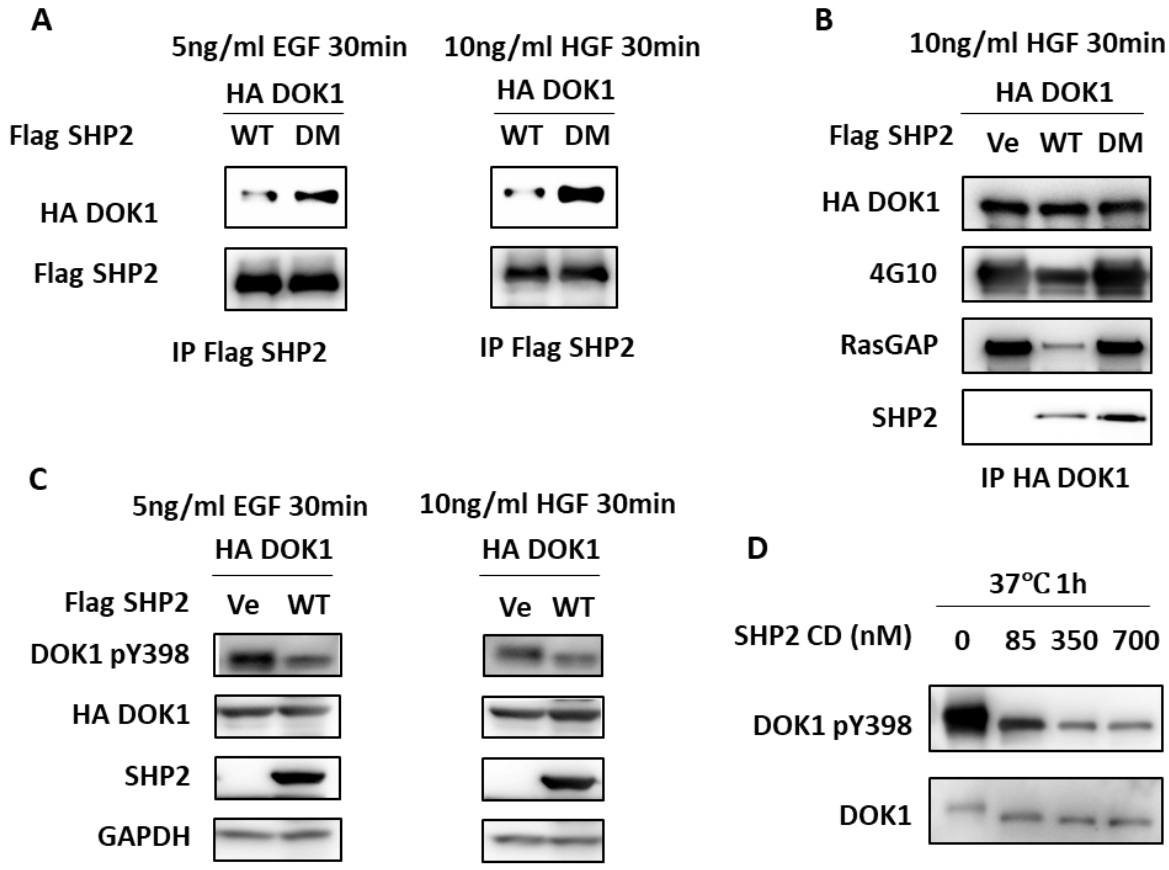


Figure 2-5: Validation of DOK1 as a SHP2 substrate. A) Western blot-based detection of the SHP2 and DOK1 pull-down experiments; B) Western blot-based detection of DOK1 phosphorylation and RasGAP levels after the SHP2 and DOK1 pull-down experiments; C) Western blot-based detection of DOK1 pY398 levels after the SHP2 and DOK1 pull-down experiments; D) Western blot-based detection of DOK1 pY398 *in vitro* dephosphorylation assay by SHP2 catalytic domain (SHP2 CD).

Table 2-1: Final list from overlapping of phosphoproteins of 3-MS experiments.

Overlap list of 3-MS experiments from unique protein				
CTNND1	TJP1	CTTN	PKP4	DOK1
ARHGAP12	MAP1B	IRS2	HNRNPA2B1	PTPN11
BCAR1	KIRREL	POLR2A	HBS1L	
GAB1	GAB2	MPZL1	SIRPA	

Table 2-2: DOK1-APMS experiment with annotated sequence.

DOK1-APMS	
No. identified DOK1 pTyr sites	26
No. identified DOK1 unique peptides	13
Coverage of DOK1 protein	80 %
Total PSM	2086
Identified DOK1 pTyr sites	
SHNSALYSQVQK	Y449
LTDPKEDPIYDEPEGLAPVPPQGLYDLPR	Y362, Y377
PLYWDLYEHAQQQLLK	Y337, Y341
EEGYELPYNPATDDYAVPPPR	Y398, Y402, Y409
EDPIYDEPEGLAPVPPQGLYDLPR	Y362, Y377
KPLYWDLYEHAQQQLLK	Y337, Y341
IAPCPSQDSLYSDPLDSTSAQAGEGVQR	Y315
LPSPPGPQELLDSPPALYAEPLDSLR	Y296
LSALEMLENSLYSPTWEGSQFWVTVQR	Y146

CHAPTER 3. DEVELOPMENT OF KINETIC ASSAY LINKED PHOSPHOPROTEOMICS TO IDENTIFY DIRECT SUBSTRATES OF PROTEIN TYROSINE PHOSPHATASE PTP1B

3.1 Summary

Tyrosine phosphorylation is critical for signal transduction, and dysregulation is a hallmark of many diseases, including numerous cancers. Even though both protein tyrosine kinases and tyrosine phosphatases equally contribute to balancing tyrosine phosphorylation levels, there is a significant gap in knowledge understanding phosphatases compared to kinases, which is partly due to the limited methods for identifying substrates. As a possible solution, we developed a novel strategy termed kinetic assay linked phosphoproteomics to identify phosphatase substrates. Briefly, the method incorporates two parallel MS-based experiments: *in vivo* PTP1B phosphoproteomics perturbed by PTP1B inhibitor and insulin treatment and *in vitro* kinetic profiles of PTP1B phosphatase utilizing dimethyl labeling to quantify enzyme reaction. To demonstrate the concept, we applied it to protein tyrosine phosphatase 1B (PTP1B) in the insulin signaling pathway. As results, eight potential substrates of PTP1B were discovered in our method, including the known substrates such as CTTN, CTTND1, and PXN.

3.2 Introduction

Protein tyrosine phosphatase 1B (PTP1B) is a non-transmembrane tyrosine-phosphatase that plays a significant role in multiple signaling networks, particularly insulin cascades that directly dephosphorylates the insulin receptor[1–4]. Accumulating evidence also indicates that PTP1B is involved in human disorders such as obesity[3], diabetes[2,4], cancers[5–9], and most recently in neurodegenerative diseases[10]. PTP1B has long been recognized to inhibit insulin signaling, which has led to its identification as a prospective therapeutic target for treating type 2 diabetes and obesity[2]. Trodusquemine (MSI-146) is a PTP1B inhibitor that showed very high selectivity for treatment of obesity and diabetes had made its way through to Phase II of the clinical trials[11]. Similarly, DPM-1001, an analog of the trodusquemine, also has gone through its clinical trial but has not reached Phase II yet[12]. Another clinical approach to inhibiting PTP1B is to apply antisense oligonucleotides. This has shown higher selectivity toward PTP1B than smaller molecular

inhibitors, which are currently in Phase II as well[13]. The clinical trial of MSI-1436C for the treatment of metastatic breast cancer was terminated at Phase I for undefined reasons[14]. Overall, few compounds have reached early-stage clinical trials, but none have yet achieved the clinical trials later stage or registration. PTP1B-targeting drugs have a promising future ahead of them, and many researchers are still working on improving PTP1B inhibitors' pharmacological properties. There are several challenges to overcome when it comes to choosing PTP1B as a therapeutic target. As a result, future research into this area to develop PTP1B inhibitors to treat the human disorders linked with it will be required.

PTP1B is widely accepted as a therapeutic target, although its exact mechanism of action remains elusive. The enzyme has a high substrate selectivity, but new substrates for the enzyme are still being discovered. One reason for the continuous discovery of PTP1B substrate is PTP1B can operate as both a tumor suppressor and a tumor promotor depending on the substrate it encounters in the cell. For instance, PTP1B overexpression has been demonstrated to increase tumorigenesis in colon cancer[15] and the pathogenesis in breast cancer[16]. In contrast, deficiency has been shown to increase the invasiveness of prostate cancer[17], and the absence of myeloid PTP1B in mice results in a shortened lifespan due to the late development of acute leukemia[7]. Hence, a better picture of how PTP1B interacts mechanistically with its substrates in the signaling cascades is required.

This study aims to develop a high throughput strategy to systematically identify PTP1B phosphatase substrates in the insulin signaling cascade. This new method relies on two MS-based phosphoproteome experiments; PTP1B substrate candidates were determined by overlapping the two data, specifically 1) PTP1B inhibitor-treated phosphoproteome and 2) PTP1B enzyme kinetic dephosphorylation assay. Briefly, in the PTP1B inhibitor phosphoproteome, we focused on the pTyr that shown upregulated in PTP1B inhibitor and insulin-treated sample. In kinetic dephosphorylation assay, the enriched phosphotyrosine peptides were dephosphorylated with active PTP1B phosphatase with a different time course. The overlap of these two lists will give high confident hits of potential PTP1B substrate candidates. The project is intended to reveal previously unknown substrates as a generic discovery tool for unresolving PTP1B phosphatase biology.

3.3 Methods and materials

3.3.1 Cell culture, transfection, stimulation, and protein extraction

HEK293 cells were maintained in Dulbecco's modified Eagle's medium (Corning) supplemented with 10% fetal bovine serum (Corning). Cells (~90% confluence) were transfected with *PTP1BD181A-FLAG* plasmids using polyethyleneimine (PEI) as in Longo et al (2013)[1]. 48 hours after transfection, the cells were treated with 1 mM pervanadate for 30 min. The stimulation reactions were stopped by washing with ice-cold phosphate-buffered saline (PBS). Cells were lysed in lysis buffer (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 % Triton x-100, 10 % glycerol, 1X phosphatase inhibitor cocktail 2 & 3 (Sigma)) and sonicated in the ice-water bath at 15 W output with 3 bursts of 10 sec each and cooled on ice for 20 sec between each burst. The extracts were cleared by centrifugation at 16,000 × g for 20 min at 4 ℃. Protein concentration was measured by BCA assay (Pierce) according to the manufacturer's protocol.

3.3.2 Peptide preparation

For Lys-C digestion before PT-66 enrichment. 1 mg of proteins denatured in 12 mM SDC/SLS buffer to a final volume of 200 µL, reduced and alkylated with 10 mM TCEP and 40 mM CAA at 95 ℃ for 5 min. Diluted 4-fold 50 mM TEAB and digested with Lys-C (Wako) (1 µg/100 µg sample protein overnight at 37 ℃). The Lys-C activity was quenched by acidification using TFA to a final concentration of 1%, and detergents were removed by ethyl acetate. The digests were desalted using Sep-Pak 100 mg C18 cartridge (Waters). Each cartridge was activated with 1 mL methanol, equilibrated with 1 mL 80% ACN/0.1% TFA (Buffer B), and then washed with 1 mL of 5% ACN/0.1% TFA (Buffer A). Following sample loading, the cartridge was washed with 1mL of 5% Buffer A, and samples were eluted with 1 mL Buffer B. Desalted samples were dried to completion in a vacuum concentrator. Dried peptide samples were stored at -80°C for further use.

For trypsin digestion after PTP1B phosphatase reaction. PTP1B dephosphorylation reactions were quenched by adding 100 µL 24 mM SDC/SLC, diluted 4X 50 mM TEAB and further digested with trypsin (Sigma) (1 µg/100 µg sample for additional of 6 hours at 37 ℃). TFA was added to a final concentration of 1%, and detergents were removed by ethyl acetate. The digested peptides were desalted using C18 Stage tips (Glygen). Each SDB-XC StageTip was

equilibrated with 20 µL Buffer B, and washed with the same volume of Buffer A. Following sample loading, the tip was washed with 20 µL Buffer A, and samples were eluted with 20 µL Buffer B. Desalted samples were dried to completion in a vacuum concentrator. Dried peptide samples were stored at -80°C for further use.

3.3.3 PT-66 enrichment

15 µL anti-phosphotyrosine antibodies (clone PT-66) conjugated to agarose beads (Sigma) were washed trice with 100 mM of Tris-HCl buffer, pH 7.5. 200 µg dried peptides were dissolved in 100 mM of Tris-HCl buffer, pH 7.5, and added to washed PT-66 beads, incubated for 12-16 h at 4 °C with rotation. 200 µg dried peptides were used for TMT labeling, and 300 µg were used for dimethyl labeling. On the next day, the bead-bound peptides were washed three times with 100 mM Tris-HCl buffer (pH 7.5) in a 10 µm filter tube (Mo Bi Tec). Phosphopeptides were eluted at room temperature with 3X 100 µL of 0.1% TFA for 10 min each. All eluates were combined and dried in a vacuum concentrator.

3.3.4 PTP1B phosphatase assay

The PT-66 enriched peptides were resuspended in PTP1B phosphatase reaction buffer (50 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 1 mM DTT, and 100 nM DADEpYLIPQQG peptides (Santa Cruz)). Saved 100 µL of sample in a new tube and labeled it as 0 min reaction (Ctrl) sample. The reactions were performed by adding 0.1 nM active PTP1B enzyme at 37 °C, and aliquots (100 µL) from reactions were removed at different time points (*TMT labeling samples:* 1 min, 5 min, 10 min, 15 min, and 30 min. *Dimethyl labeling samples:* 1 min and 10 min). Each sample was quenched by adding 100 µL of 24 mM SDC/SLS, and further digested with trypsin.

3.3.5 Dimethyl labeling

The desalted peptides dissolved in 100 µL of 100 mM TEAB were mixed with formaldehyde (4% in water) and 0.6 M freshly prepared sodium cyanoborohydride added immediately. The mixture was vortexed for 1 hr at room temperature, quenched with 16 µL of ammonium hydroxide for 1 min, and added 20 µL of 10% FA. The sample undergone 10 min PTP1B reaction was labeled

using light formaldehyde (CH_2O), whereas deuterated formaldehyde (CD_2O) was used to label 1 min PTP1B reaction sample, and the control sample used heavy formaldehyde ($^{13}CD_2O$). The same amount of three samples were combined and desalted using C18 StageTip (Glygen).

3.3.6 TMT labeling

Desalted peptides were labeled with TMT (6-plex) reagents (ThermoFisher Scientific). Peptides in different time points (0 min, 1 min, 5 min, 10 min, 15 min, and 30 min) were resuspended in 100 µL of 50 mM TEAB, and each sample was labeled with 50 µg of TMT reagent. Samples were incubated at room temperature for 1 hr with gentle agitation. TMT reaction was quenched with 2 µL of 5% hydroxylamine for 15 min. TMT labeled samples were combined, dried to completion, reconstituted in 200 µL of 0.1% TFA, and desalted on C18 StageTips (Glygen).

3.3.7 LC-MS data acquisition

The labeled peptides were resuspended in 10.8 µL of MS-grade 0.05% TFA/2% ACN and analyzed by online nanoflow liquid chromatography-tandem mass spectrometry (LC-MS/MS) using Q Exactive HF-X (QE HF-X) coupled online to an Ultimate_Nano HPLC (ThermoFisher Scientific). 10 µL of the sample was loaded at onto a trap column (75 µm X 2 cm C18, 3 µm 100 Å) (ThermoFisher Scientific) that was connected to a house packed analytical column (~30 cm with ReproSil Pur 120 C18 AQ 1.9 µm beads (Dr.Maisch Gmbh) and heated to 50 ℃. Mobile phase flow rate was 300 nL/min, comprises 0.1% FA (Solvent A) and 80% ACN/0.1%FA (Solvent B). The 90-min LC-MS/MS method used to following gradient profile for dimethyl labeling sample: (min:%B) 0:5; 8:5; 68:35; 69:90; 70:5; 71:90; 72:5; 73:90; 74:5; 75:90; 79:90; 80:5; 90:5. TMT labeled sample also used 90-min method with following gradient profile: (min:%B) 0:5; 5:5; 8:10; 68:45; 69:90; 70:5; 71:90; 72:5; 73:90; 74:5; 75:90; 79:90; 80:5; 90:5. The QE HF-X was operated in the data-dependent mode for dimethyl labeling sample, acquiring MS2 scans (resolution=30,000) after each MS1 scan (resolution=60,000, scan range of 375-1500 m/z) on the top 15 most abundant ions using an MS1 AGC of 1e6, and an MS2 AGC of 2e4. The maximum ion time utilized for MS2 scans was 200 ms; normalization collision energy was set to 28; the dynamic exclusion on time was set to 60 s; isolation window of 1.6 m/z. The QE HF-X was operated in TMT mode for TMT sample acquiring HCD MS2 scans (resolution=30,000) after each

MS1 scan (resolution=60,000) on the top 10 most abundant ions using an MS1 target of 3e6 and an MS2 target of 2e5. The maximum ion time utilized for MS2 scans was 100 ms; isolation window of 0.7 m/z; the HCD-normalization collision energy was set to 32.

3.3.8 Data search and analysis

The raw files were searched against the *Homo sapiens* database with no redundant entries using the SEQUEST search engine built into Proteome Discoverer (version 2.3). Peptide precursor mass tolerance for the main search was set to 10 ppm, and MS/MS tolerance was set to 0.02 Da. Searches were performed with full tryptic digestion, and peptides were allowed a maximum of two missed cleavages. Search criteria included a static modification of cysteine residues of +57.021 Da to account for alkylation and a variable modification of +15.995 Da for potential oxidation of methionine residues and acetyl of +42.011 Da at the N-terminus of protein. A variable modification of +79.996 Da was also set on serine, threonine, and tyrosine residues to identify phosphorylation sites. The false discovery rate (FDR) for PSMs, proteins, and phosphorylation sites was set to 1% for each analysis using a reverse decoy database. Proteins matching the reverse decoy database, or the common laboratory contaminant database were discarded.

For TMT modifications. The reporter ions quantifier node was enabled, and the integration tolerance was set to 20 ppm with HCD activation type. A variable modification of +229.163 Da was set on N-terminus peptide and a static modification of +229.163 Da at lysine residue.

For dimethyl labeling. Static modifications of 1) light dimethyl of +28.031 Da; 2) intermediate dimethyl of +32.056 Da; 3) heavy dimethyl of +34.063 Da were set at lysine and any N-term peptide.

3.3.9 Western Blotting

40 µg of protein lysate were mixed with 1X LDS loading buffer and 10 mM DTT, boiled for 5 min. Loaded to an SDS-PAGE gel with 4-12% gradient (ThermoFisher Scientific) and ran using 1X MOPS buffer. The sample on the gel was transferred to polyvinylidene difluoride membrane using 1X transferring buffer (25 mM bicine, 25 mM Bis-Tris, 1 mM EDTA, 13% MeOH) for 75 min. The membrane was then incubated in blocking buffer (1% BSA diluted in TBS-T) with gentle shaking for 1 hr at room temperature. The membrane was incubated in 1:1,000

anti-PTP1B (Cell Signaling), or anti-pY100 (Cell Signaling), or anti-FLAG (Thermo) antibody overnight at 4 ℃ with gentle shaking. Next, the membrane was rinsed three times in TBS-T for 5 min each with 0.1% Tween-20. Subsequently, the membrane was incubated for 1 hr at room temperature in 1: 5,000 dilutions of HRP-conjugated secondary antibody: anti-mouse (Thermo), anti-rabbit (Thermo). The membrane was washed three times quickly in TBS-T with 0.1% Tween-20 for 5 min each. Detection was performed using ECL buffer (Thermo), and blots were developed in G153 developing buffer (HealthCare) in a dark room.

3.4 Results

3.4.1 *In vivo* global phosphoproteome for identification of insulin pathway substrates

Diabetes mellitus is a metabolic and endocrine disorder characterized by chronic hyperglycemia that results in biochemical changes as well as tissue damage. The number of adults with diabetes is expected to rise from 22.3 million in 2014 to 39.7 million in 2030 and 60.6 million in 2060[19]. Substantial evidence indicates that PTP1B plays a significant part in the insulin signaling system, and PTP1B inhibitor was considered a promising therapeutic target for treating type 2 diabetes[12,20–25]. PTP1B dephosphorylates insulin receptor substrate 1 (IRS-1) in the insulin signaling pathway, reducing insulin sensitivity, or turning off signaling. A significant variety of PTP1B inhibitors have been generated and tested for their capacity to enhance insulin signaling, whether they are bioactive compounds derived from natural products or synthetic small molecules[21,26–30].

We optimized our new protocol to profile pTyr in PTP1B inhibitor compound III[26]-treated cells in the insulin signaling pathway. As seen in Fig. 3-1 our workflow is straightforward, and cells were collected after PTP1B inhibitor and insulin incubation for subsequent pTyr peptide enrichment and LC-MS/MS analysis. We identified 332 and 297 total peptides in the PTP1B inhibitor treated and DMSO sample (Fig. 3-2 A), respectively. 271 pTyr sites were found in the PTP1B inhibitor-treated sample and 224 pTyr sites in DMSO (Fig. 3-2 B). A volcano plot was created to depict the quantitative data graphically, and it showed 114 pTyr sites upregulated in PTP1B inhibitor-treated sample. Several known PTP1B substrates were highlighted in blue (CTTN[31], EGFR[23,32], FER[33], IRS1[34,35], PTPN11[32], SRC[36], TYK2[37]) (Fig. 3-2 C). Cellular compartment and biological process classification of the 114 pTyr proteins were analyzed using

DAVID 6.8 to compare to known PTP1B activities[38]. Most proteins are membrane-bound and cytoplasmic proteins, but there are also large numbers of nucleus proteins (Fig. 3-2 D). In terms of biological processes, most proteins are involved in insulin pathways, as expected, RPTK, EGFR, PI3K, and phosphatidylinositol-mediated signaling, and mRNA and RNA splicing (Fig. 3-2 E). Hence, the analysis shows that many hit proteins in the pTyr protein list are consistent with known PTP1B activity. Next, we will overlap this pTyr protein list with a dimethyl labeled list to eliminate the false positive hits to increase the confidence of PTP1B substrate identification.

3.4.2 Profiling of *in vitro* PTP1B phosphatase substrate via kinetic profile

Label-free quantification relies on comparing MS1 peptide signal intensities between MS runs, so it is easy to incorporate into our protocol. As seen in Fig. 3-3 that PTP1B-substrate complexes were isolated using FLAG-IP then followed by in-solution digestion. The resulted peptides were dephosphorylated by active PTP1B for 15 and 45 min. However, each time points of samples were measured individually in label-free quantitation, and as a result, we had "missing quantitation values" between runs (data not shown). To overcome this, TMT tags were used to label different time points, and our new protocol was refined, as shown in Fig. 3-4.

TMT labeled multiplexed phosphoproteome resulted in more phosphopeptides identification[39–41] and accurate quantitation. Hence, our new protocol (Fig. 3-4) TMT is used to tag six different time points to profile PTP1B kinetics. Few parameters need to be considered to monitor enzyme activity *in vitro* over a given time. First, to ensure a similar substrate concentration in this protocol, we enriched 200 µg lysate with PT-66 antibody for generation of at least 1500 pTyr peptides. Second, we optimized PTP1B enzyme assay timing and concentration. The standard curve of PTP1B activity was generated by a straightforward time-course experiment combined with different enzyme concentrations. As seen in Fig. 1-5 A, D-F, too much enzyme (1nM, 10 nM, 20 nM, and 50 nM) led to the completion of the reaction too quickly, and no signal was detected after 1 min. Whereas Fig. 1-5 C shows too little enzyme, the readout is not strong enough to measure relative background signal reliably. As a result, 0.1 nM PTP1B gave a standard curve of the enzyme activity (Fig. 1-5 B). We used this concentration for the rest of the labeled experiments.

To access the quality for TMT labeling, we performed a correlation analysis of reporter intensities for all channels and visualized in Fig. 1-6 A, with the Pearson value of >0.9. A total of 1056 peptides were identified in this experiment, with 1038 peptides were labeled with TMT tags,

representing 98% of TMT labeling efficiency, and among 736 were pTyr peptides (Fig. 1-6 B). The intensity of proteins or peptides in different time points was illustrated in the heatmap in which PTP1B enzymes were active enough to dephosphorylate most proteins and peptides after 1 min of reaction (Fig. 1-6 C & D). Next, we analyzed the kinetics of PTP1B enzyme. We took only a single phosphorylation site of pTyr peptides into consideration in calculating the first-order reaction. The rate and the substrate concentration were directly proportional to each other. A total of 158 pTyr sites were passed the criteria in the kinetics analysis. The distribution of these 158 pTyr peptides in PTP1B enzyme rates in the reaction was shown in Fig. 3-7 A, with an average Kcat/Km rate of E07 $s^{-1}M^{-1}$ was also consistent with prior work (Fig. 3-7 B). Motif analysis of phosphosites did not reveal enrichment for specific motif patterns for these 158 pTyr peptides, but it preferred higher acidic residues around the pTyr site (Fig. 3-7 C). The kinetics analysis of some known PTP1B substrates did not show the consistency of dephosphorylation (Fig. 3-7 D), and this put a question to our method analysis. Is it because of inactive PTP1B enzyme during the reaction or ratio suppression in TMT labeling that affects our kinetic analysis? To answer this question, we modified the protocol from TMT labeling to dimethyl labeling, which uses MS1 in quantitation and eliminates the ratio suppression issue. Dimethyl labeling was utilized in place of TMT tags. Lastly, we overlapped this pTyr protein list with a dimethyl labeled list to eliminate the false positive hits to increase the confidence of PTP1B substrate identification.

As seen in Fig. 3-8, the workflow for dimethyl labeling is straightforward, for a total of 3 time point. The first time point sample was labeled with light formaldehyde (CH_2O), and the second time point sample was labeled with deuterated formaldehyde (CD_2O), and the last time point sample with heavy carbon-13 ($^{13}CD_2O$). Overall, in four replicates, we identified 4786 total dimethyl labeled peptides, in which 2777 were phosphopeptides, and 2345 were pTyr peptides (Fig. 3-9 A). In each labeling, we identified more than 800 phosphopeptides with more than 700 pTyr. However, the quantifiable pTyr were limited to 348 in light labeled samples. 348 pTyr peptides were further analyzed based on their dephosphorylation profile in 10- and 30-min PTP1B phosphatase assay, 158 pTyr sites were the final list that showed dephosphorylation profile that meet the criteria (Fig. 3-9 B). To further evaluate 158 phosphoproteins, we overlapped it with literature reported PTP1B substrates and nine proteins from our list showed consistency (Fig. 3-10 A). Finally, the list from dimethyl labeling and insulin experiments were further overlapped with each other and with literature reported PTP1B substrate list (Fig 3-10 B). Four of proteins

were found to be common in our list and nine were potential PTP1B substrates. Proteins from the list needs to validate further to check the validity of our experiment.

3.4.3 Detection of *in vitro* PTP1B protein substrates

We also attempted to work on the protein level to isolate PTP1B-substrate using the FLAG tag and eluted substrates from the complex by sodium orthovanadate (Fig. 3-11). *PTP1BD181A* plasmid was transfected into the HEK293 cells using PEI cationic polymer. PEI cationic polymer is cost-effective, but it has greater toxicity to the cells. To evaluate whether PEI is suitable for HEK293 cells and the duration of PEI incubation, we transfected a control GFP plasmid into HEK293. To access the expression of GFP, we visualized the cells under the microscope (Fig. 3-12 A). The optimum PEI incubation time is around 48h when saw >85% GFP expression with few cells' death. Hence, we used the same condition for our *PTP1B* plasmid transfection using PEI.

Whether PTP1B is expressed in the mammalian cells or not, we evaluated it in Western blot. As seen in Fig. 3-12 B that PTP1B proteins were successfully isolated using FLAG immunoprecipitation (MW ~55 kDa). We also evaluated whether pervanadate worked or not. The first line of input showed smearing in the left Western blot indicated that phosphorylation of tyrosine residues was protected from pervanadate stimulation. Next, we used sodium orthovanadate to elute the PTP1B substrate from the complex. The substrates pool was divided into two equal portions, one treated with PTP1B active enzyme and another incubated with PTP1B mutated version. Consequently, both samples were digested with Lys-C and trypsin. Phosphopeptides were enriched using PolyMAC, and the resulted peptides were subjected to LC-MS/MS analysis. We identified a total of 192 pTyr sites in the control sample (PTP1B inactive enzymes were used in the PTP1B phosphatase reaction) and 21 pTyr peptides in the WT sample (Fig. 3-13 A). 176 out of 192 pTyr in the control sample were classified as Class I pTyr peptides, and within 172 pTyr sites showed at least a 1.5-fold change in phosphorylation intensity (Fig. 3-13 B). These 172 pTyr sites correspond to 111 pTyr proteins that might be PTP1B substrate candidates. Among the hits, few known PTP1B substrates were identified, such as CTTN[31], EGFR[23,42], CAV1[43], CTNND1[44], INSR[45,46], IGF1R[24,25,44], GAB1[32], BCAR1[47,48], and ARHGAP12[44,49]. Next, we performed cellular compartment and biological process analysis of the 111 pTyr proteins using DAVID 6.8[38] to compare to known PTP1B activities (Fig. 3-13 C & D).

PTP1B is well known to be located mainly at the ER and accesses substrates via endocytosis, ER network movement and biosynthesis[50,51]. Hence, it is expected to see proteins that are either membrane-bound or cytosolic present in our list. Consistent with prior work, most of our proteins in the list are either membrane-bound or cytoplasmic (Fig. 3-13 C). PTP1B and substrates are involved in cell-cell communication, insulin signaling pathway[24,25,46,51–53]. As expected, our proteins in the list are involved in cell migration, tyrosine phosphorylation, cell adhesion, and insulin-related signaling pathways like EGFR, MAP, and GTPase activity cascades (Fig 3.13 D).

However, for trusted MS-based data, it requires either biological or technical replicates or another method to validate the results. Failed to replicate the protein level experiment due to either active PTP1B enzyme or lost selectivity of PTP1B in the lysate. PTP1B is a ubiquitously expressed phosphatase in cells, and its activity sensitivity is maintained in a subcellular context. However, in our method, it lost its sensitiveness. It not only dephosphorylates pTyr peptides also pSer and pThr. We attempted to optimize the PTP1B enzyme condition to suit our method (Fig. 3-14) by lowering PTP1B amount (Fig 3-14 A), reducing reaction temperature (Fig. 3-14 B), lowering PTP1B reaction concentration (Fig 3-14 C), or varying incubation time in the enzyme reaction. By changing each condition, it is clear that PTP1B dephosphorylates pTyr and pSer and pThr. Besides, we also noticed that not all experiments carried out the same amount of pTyr at the starting point and wondered what step led to this. We blotted the PTP1B substrates pool eluted by sodium orthovanadate, and it showed that this competitive inhibitor was not consistent in different experiments (Fig 3-15). Therefore, we optimized the PTP1B-substrate elution conditions (Fig. 3-15).

3.5 References

(1) Galic, S.; Hauser, C.; Kahn, B. B.; Haj, F. G.; Neel, B. G.; Tonks, N. K.; Tiganis, T. Coordinated Regulation of Insulin Signaling by the Protein Tyrosine Phosphatases PTP1B and TCPTP. *Mol. Cell. Biol.* **2005**, *25* (2), 819.

(2) Asante-Appiah, E.; Kennedy, B. P. Protein Tyrosine Phosphatases: The Quest for Negative Regulators of Insulin Action. *Am. Physiol. Soc.* **2003**, *284* (4), 663–670.

(3) Bence, K. K.; Delibegovic, M.; Xue, B.; Gorgun, C. Z.; Hotamisligil, G. S.; Neel, B. G.; Kahn, B. B. Neuronal PTP1B Regulates Body Weight, Adiposity and Leptin Action. *Nat. Med.* **2006**, *12* (8), 917–924.

(4) Abdelsalam, S. S.; Korashy, H. M.; Zeidan, A.; Agouni, A. The Role of Protein Tyrosine Phosphatase (PTP)-1B in Cardiovascular Disease and Its Interplay with Insulin Resistance. *Biomolecules* **2019**, *9* (7), 286.

(5) L, L.; M, S.; ML, T. The Two Faces of PTP1B in Cancer. *Biochim. Biophys. Acta* **2010**, *1804* (3), 613–619.

(6) Hoekstra, E.; Das, A. M.; Swets, M.; Cao, W.; Woude, C. J. van der; Bruno, M. J.; Peppelenbosch, M. P.; Kuppen, P. J. K.; Hagen, T. L. M. ten; Fuhler, G. M.; et al. Increased PTP1B Expression and Phosphatase Activity in Colorectal Cancer Results in a More Invasive Phenotype and Worse Patient Outcome. *Oncotarget* **2016**, *7* (16), 21922–21938.

(7) Sommer, S. Le; Morrice, N.; Pesaresi, M.; Thompson, D.; Vickers, M. A.; Murray, G. I.; Mody, N.; Neel, B. G.; Bence, K. K.; Wilson, H. M.; et al. Deficiency in Protein Tyrosine Phosphatase PTP1B Shortens Lifespan and Leads to Development of Acute Leukemia. *Cancer Res.* **2018**, *78* (1), 75–87.

(8) Xu, Q.; Wu, N.; Li, X.; Guo, C.; Li, C.; Jiang, B.; Wang, H.; Shi, D. Inhibition of PTP1B Blocks Pancreatic Cancer Progression by Targeting the PKM2/AMPK/MTOC1 Pathway. *Cell Death Dis. 2019 1012* **2019**, *10* (12), 1–15.

(9) Yu, M.; Liu, Z.; Liu, Y.; Zhou, X.; Sun, F.; Liu, Y.; Li, L.; Hua, S.; Zhao, Y.; Gao, H.; et al. PTP1B Markedly Promotes Breast Cancer Progression and Is Regulated by MiR-193a-3p. *FEBS J.* **2019**, *286* (6), 1136–1153.

(10) Ricke, K. M.; Cruz, S. A.; Qin, Z.; Farrokhi, K.; Sharmin, F.; Zhang, L.; Zasloff, M. A.; Stewart, A. F. R.; Chen, H.-H. Neuronal Protein Tyrosine Phosphatase 1B Hastens Amyloid β-Associated Alzheimer's Disease in Mice. *J. Neurosci.* **2020**, *40* (7), 1581–1593.

(11) Lantz, K. A.; Hart, S. G. E.; Planey, S. L.; Roitman, M. F.; Ruiz-White, I. A.; Wolfe, H. R.; McLane, M. P. Inhibition of PTP1B by Trodusquemine (MSI-1436) Causes Fat-Specific Weight Loss in Diet-Induced Obese Mice. *Obesity* **2010**, *18* (8), 1516–1523.

(12) Krishnan, N.; Konidaris, K. F.; Gasser, G.; Tonks, N. K. A Potent, Selective, and Orally Bioavailable Inhibitor of the Protein-Tyrosine Phosphatase PTP1B Improves Insulin and Leptin Signaling in Animal Models. *J. Biol. Chem.* **2018**, *293* (5), 1517–1525.

(13) Swarbrick, M. M.; Havel, P. J.; Levin, A. A.; Bremer, A. A.; Stanhope, K. L.; Butler, M.; Booten, S. L.; Graham, J. L.; McKay, R. A.; Murray, S. F.; et al. Inhibition of Protein Tyrosine Phosphatase-1B with Antisense Oligonucleotides Improves Insulin Sensitivity and Increases Adiponectin Concentrations in Monkeys. *Endocrinology* **2009**, *150* (4), 1670.

(14) Vainonen, J. P.; Momeny, M.; Westermarck, J. Druggable Cancer Phosphatases. *Sci. Transl. Med.* **2021**, *13* (588), 1–14.

(15) Zhu, S.; Bjorge, J. D.; Fujita, D. J. PTP1B Contributes to the Oncogenic Properties of Colon Cancer Cells through Src Activation. *Cancer Res.* **2007**, *67* (21), 10129–10137.

(16) Cortesio, C. L.; Chan, K. T.; Perrin, B. J.; Burton, N. O.; Zhang, S.; Zhang, Z.-Y.; Huttenlocher, A. Calpain 2 and PTP1B Function in a Novel Pathway with Src to Regulate Invadopodia Dynamics and Breast Cancer Cell Invasion. *J. Cell Biol.* **2008**, *180* (5), 957–971.

(17) Lessard, L.; Labbé, D. P.; Deblois, G.; Bégin, L. R.; Hardy, S.; Mes-Masson, A.-M.; Saad, F.; Trotman, L. C.; Giguère, V.; Tremblay, M. L. PTP1B Is an Androgen Receptor–Regulated Phosphatase That Promotes the Progression of Prostate Cancer. *Cancer Res.* **2012**, *72* (6), 1529–1537.

(18) Longo, P. A.; Kavran, J. M.; Kim, M.-S.; Leahy, D. J. Transient Mammalian Cell Transfection with Polyethylenimine (PEI). *Methods Enzymol.* **2013**, *529*, 227–240.

(19) Lin, J.; Thompson, T. J.; Cheng, Y. J.; Zhuo, X.; Zhang, P.; Gregg, E.; Rolka, D. B. Projection of the Future Diabetes Burden in the United States through 2060. *Popul. Heal. Metrics 2018 161* **2018**, *16* (1), 1–9.

(20) ZY, Z.; SY, L. PTP1B Inhibitors as Potential Therapeutics in the Treatment of Type 2 Diabetes and Obesity. *Expert Opin. Investig. Drugs* **2003**, *12* (2), 223–233.

(21) H, H.; IR, G.; G, A.; SM, A.; W, H.; I, A. Protein Tyrosine Phosphatase 1B (PTP1B) Inhibitors as Potential Anti-Diabetes Agents: Patent Review (2015-2018). *Expert Opin. Ther. Pat.* **2019**, *29* (9), 689–702.

(22) Bakke, J.; Haj, F. G. Protein-Tyrosine Phosphatase 1B Substrates and Metabolic Regulation. *Semin. Cell Dev. Biol.* **2015**, *0*, 58.

(23) Salmeen, A.; Andersen, J. N.; Myers, M. P.; Tonks, N. K.; Barford, D. Molecular Basis for the Dephosphorylation of the Activation Segment of the Insulin Receptor by Protein Tyrosine Phosphatase 1B Quired for Maximal Activation (White et Al., 1988). Follow-Ing Activation, Additional Autophosphorylation Reactions Occur in The. *Mol. Cell* **2000**, *6*, 1401–1412.

(24) Blanquart, C.; Boute, N.; Lacasa, D.; Issad, T. Monitoring the Activation State of the Insulin-Like Growth Factor-1 Receptor and Its Interaction with Protein Tyrosine Phosphatase 1B Using Bioluminescence Resonance Energy Transfer. *Mol. Pharmacol.* **2005**, *68* (3), 885–894.

(25) Bandyopadhyay, D.; Kusari, A.; Kenner, K. A.; Liu, F.; Chernoff, J.; Gustafson, T. A.; Kusari, J. Protein-Tyrosine Phosphatase 1B Complexes with the Insulin Receptor in Vivo and Is Tyrosine-Phosphorylated in the Presence of Insulin *. *J. Biol. Chem.* **1997**, *272* (3), 1639–1645.

(26) He, R.; Zeng, L.-F.; He, Y.; Zhang, S.; Zhang, Z.-Y. Small Molecule Tools for Functional Interrogation of Protein Tyrosine Phosphatases. *FEBS J.* **2013**, *280* (2), 731.

(27) Zhao, B. T.; Nguyen, D. H.; Le, D. D.; Choi, J. S.; Min, B. S.; Woo, M. H. Protein Tyrosine Phosphatase 1B Inhibitors from Natural Sources. *Arch. Pharmacal Res. 2017 412* **2017**, *41* (2), 130–161.

(28) Wang, L.-J.; Jiang, B.; Wu, N.; Wang, S.-Y.; Shi, D.-Y. Natural and Semisynthetic Protein Tyrosine Phosphatase 1B (PTP1B) Inhibitors as Anti-Diabetic Agents. *RSC Adv.* **2015**, *5* (60), 48822–48834.

(29) Yang, Y.; Tian, J. Y.; Ye, F.; Xiao, Z. Identification of Natural Products as Selective PTP1B Inhibitors via Virtual Screening. *Bioorg. Chem.* **2020**, *98*, 103706.

(30) Jiang, C.; Liang, L.; Guo, Y. Natural Products Possessing Protein Tyrosine Phosphatase 1B (PTP1B) Inhibitory Activity Found in the Last Decades. *Acta Pharmacol. Sin.* **2012**, *33* (10), 1217.

(31) M, S.; N, D.; ML, T. PTP1B Regulates Cortactin Tyrosine Phosphorylation by Targeting Tyr446. *J. Biol. Chem.* **2008**, *283* (23), 15740–15746.

(32) Ferrari, E.; Tinti, M.; Costa, S.; Corallino, S.; Nardozza, A. P.; Chatraryamontri, A.; Ceol, A.; Cesareni, G.; Castagnoli, L. Identification of New Substrates of the Protein-Tyrosine Phosphatase PTP1B by Bayesian Integration of Proteome Evidence. *J. Biol. Chem.* **2011**, *286* (6), 4173.

(33) Zhang, J.; Wang, Z.; Zhang, S.; Chen, Y.; Xiong, X.; Li, X.; Tonks, N. K.; Fan, G. Spatial Regulation of Signaling by the Coordinated Action of the Protein Tyrosine Kinases MET and FER. *Cell. Signal.* **2018**, *50*, 100.

(34) Goldstein, B. J.; Bittner-Kowalczyk, A.; White, M. F.; Harbeck, M. Tyrosine Dephosphorylation and Deactivation of Insulin Receptor Substrate-1 by Protein-Tyrosine Phosphatase 1B: POSSIBLE FACILITATION BY THE FORMATION OF A TERNARY COMPLEX WITH THE GRB2 ADAPTOR PROTEIN *. *J. Biol. Chem.* **2000**, *275* (6), 4283–4289.

(35) A, G.-R.; JA, M. G.; S, S.-G.; M, R.; DJ, B.; AM, V. Inhibition of PTP1B Restores IRS1-Mediated Hepatic Insulin Signaling in IRS2-Deficient Mice. *Diabetes* **2010**, *59* (3), 588–599.

(36) EG, A.-S.; F, H.; C, D.; B, M.; A, K.-F.; BC, F.; B, F.; BG, N.; SJ, S. PTP-1B Is an Essential Positive Regulator of Platelet Integrin Signaling. *J. Cell Biol.* **2005**, *170* (5), 837–845.

(37) MP, M.; JN, A.; A, C.; ML, T.; CM, H.; JP, P.; A, S.; D, B.; NK, T. TYK2 and JAK2 Are Substrates of Protein-Tyrosine Phosphatase 1B. *J. Biol. Chem.* **2001**, *276* (51), 47771–47774.

(38) Dennis, G.; Sherman, B. T.; Hosack, D. A.; Yang, J.; Gao, W.; Lane, H. C.; Lempicki, R. A. DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biol. 2003 49* **2003**, *4* (9), 1–11.

(39) X, J.; R, B.; J, B.; DL, D.; AM, R.; R, V.; AR, H. Sensitive and Accurate Quantitation of Phosphopeptides Using TMT Isobaric Labeling Technique. *J. Proteome Res.* **2017**, *16* (11), 4244–4252.

(40) Hogrebe, A.; von Stechow, L.; Bekker-Jensen, D. B.; Weinert, B. T.; Kelstrup, C. D.; Olsen, J. V. Benchmarking Common Quantification Strategies for Large-Scale Phosphoproteomics. *Nat. Commun. 2018 91* **2018**, *9* (1), 1–13.

(41) Zecha, J.; Satpathy, S.; Kanashova, T.; Avanessian, S. C.; Kane, M. H.; Clauser, K. R.; Mertins, P.; Carr, S. A.; Kuster, B. TMT Labeling for the Masses: A Robust and Cost-Efficient, In-Solution Labeling Approach *[S]. *Mol. Cell. Proteomics* **2019**, *18* (7), 1468–1478.

(42) Flint, A. J.; Tiganis, T.; Barford, D.; Tonks, N. K. Development of "Substrate-Trapping" Mutants to Identify Physiological Substrates of Protein Tyrosine Phosphatases. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (5), 1680–1685.

(43) Lee, H.; Xie, L.; Luo, Y.; Lee, S.-Y.; Lawrence, D. S.; Wang, X. B.; Sotgia, F.; Michael P. Lisanti, A.; Zhang, Z.-Y. Identification of Phosphocaveolin-1 as a Novel Protein Tyrosine Phosphatase 1B Substrate. *Biochemistry* **2005**, *45* (1), 234–240.

(44) P, M.; HC, E.; J, R.; JV, O.; ML, T.; M, M.; A, U.; H, D. Investigation of Protein-Tyrosine Phosphatase 1B Function by Quantitative Proteomics. *Mol. Cell. Proteomics* **2008**, *7* (9), 1763–1777.

(45) Dadke, S.; Chernoff, J. Interaction of Protein Tyrosine Phosphatase (PTP) 1B with Its Substrates Is Influenced by Two Distinct Binding Domains. *Biochem. J.* **2002**, *364* (Pt 2), 377.

(46) Boute, N.; Boubekeur, S.; Lacasa, D.; Issad, T. Dynamics of the Interaction between the Insulin Receptor and Protein Tyrosine-Phosphatase 1B in Living Cells. *EMBO Rep.* **2003**, *4* (3), 313.

(47) F, L.; MA, S.; J, C. Transformation Suppression by Protein Tyrosine Phosphatase 1B Requires a Functional SH3 Ligand. *Mol. Cell. Biol.* **1998**, *18* (1), 250–259.

(48) F, L.; DE, H.; J, C. Direct Binding of the Proline-Rich Region of Protein Tyrosine Phosphatase 1B to the Src Homology 3 Domain of P130(Cas). *J. Biol. Chem.* **1996**, *271* (49), 31290–31295.

(49) Banh, R. S.; Iorio, C.; Marcotte, R.; Xu, Y.; Cojocari, D.; Rahman, A. A.; Pawling, J.; Zhang, W.; Sinha, A.; Rose, C. M.; et al. PTP1B Regulates Non-Mitochondrial Oxygen Consumption via RNF213 to Promote Tumour Survival during Hypoxia. *Nat. Cell Biol.* **2016**, *18* (7), 803.

(50) FG, H.; PJ, V.; A, S.; BG, N.; PI, B. Imaging Sites of Receptor Dephosphorylation by PTP1B on the Surface of the Endoplasmic Reticulum. *Science* **2002**, *295* (5560), 1708–1711.

(51) FG, H.; O, S.; A, K.; S, W.-K.; V, R.; HM, H.; M, G.; M, B.; C, A.; BG, N.; et al. Regulation of Signaling at Regions of Cell-Cell Contact by Endoplasmic Reticulum-Bound Protein-Tyrosine Phosphatase 1B. *PLoS One* **2012**, *7* (5).

(52) JE, B.; Á, G.; CO, A. PTP1B Promotes Focal Complex Maturation, Lamellar Persistence and Directional Migration. *J. Cell Sci.* **2013**, *126* (Pt 8), 1820–1831.

(53) Arregui, C. O.; González, Á.; Burdisso, J. E.; Wusener, A. E. G. Protein Tyrosine Phosphatase PTP1B in Cell Adhesion and Migration. *Cell Adh. Migr.* **2013**, *7* (5), 418.

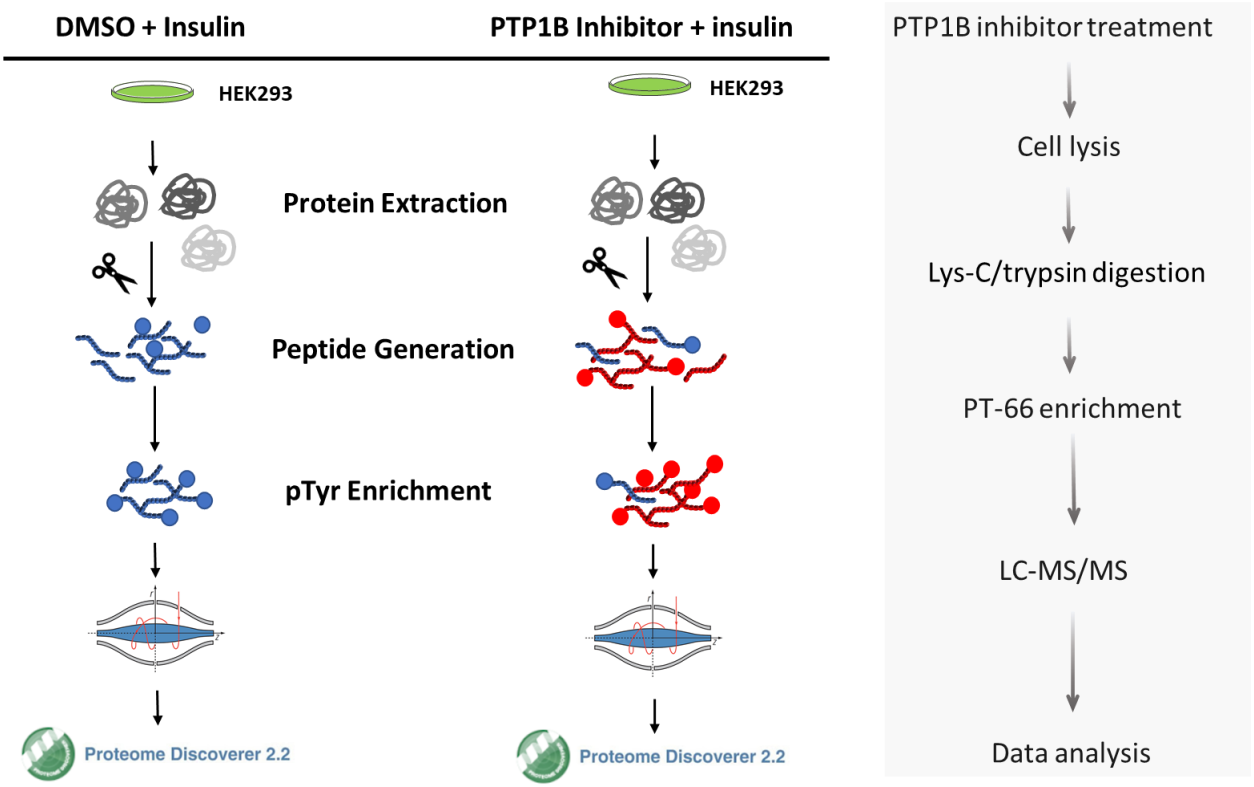
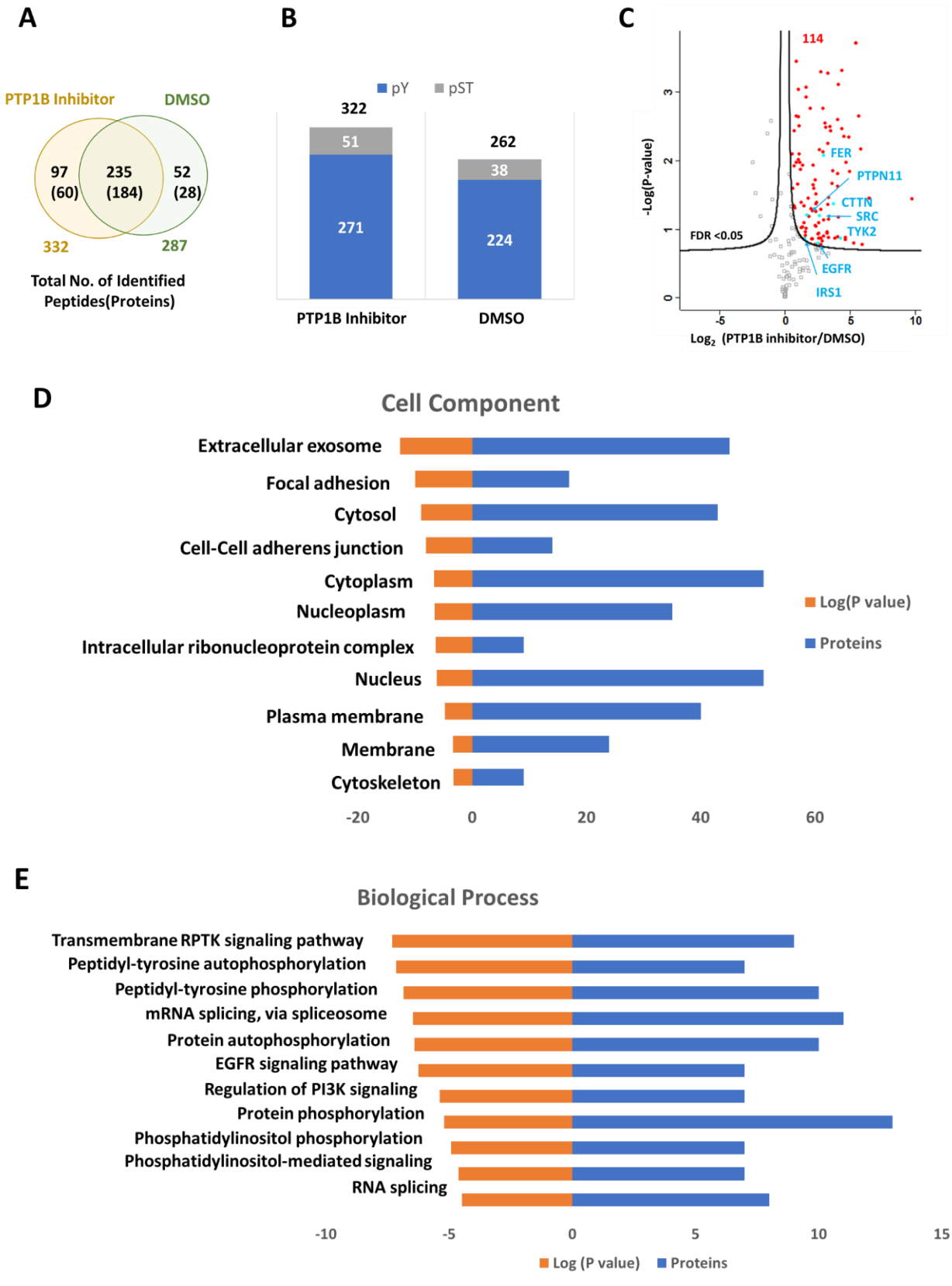


Figure 3-1: Workflow of PTP1B-dependent activation on the insulin signaling pathway.

Figure 3-2: A) Number of peptides/proteins identified in PTP1B inhibitor and DMSO treated cells. B) Total number of phosphopeptides identified in PTP1B inhibitor and DMSO treated samples. C) Volcano plot illustrates significantly upregulated pTyr sites found in PTP1B inhibitor treated sample; known PTP1B substrates indicated in blue. D) Cell component analysis of 114 pTyr proteins using David version 6.8. E) Biological process analysis of 114 pTyr proteins using David version 6.8.


A
PTP1B Inhibitor
DMSO
97
(60)
235
(184)
52
(28)
332
287
Total No. of Identified
Peptides(Proteins)
B
pY
pST
322
51
271
262
38
224
PTP1B Inhibitor
DMSO
C
114
FER
PTPN11
CTTN
SRC
TYK2
EGFR
IRS1
FDR <0.05
-Log(P-value)
Log2 (PTP1B inhibitor/DMSO)
D
Cell Component
Extracellular exosome
Focal adhesion
Cytosol
Cell-Cell adherens junction
Cytoplasm
Nucleoplasm
Intracellular ribonucleoprotein complex
Nucleus
Plasma membrane
Membrane
Cytoskeleton
Log(P value)
Proteins
E
Biological Process
Transmembrane RPTK signaling pathway
Peptidyl-tyrosine autophosphorylation
Peptidyl-tyrosine phosphorylation
mRNA splicing, via spliceosome
Protein autophosphorylation
EGFR signaling pathway
Regulation of PI3K signaling
Protein phosphorylation
Phosphatidylinositol phosphorylation
Phosphatidylinositol-mediated signaling
RNA splicing
Log (P value)
Proteins

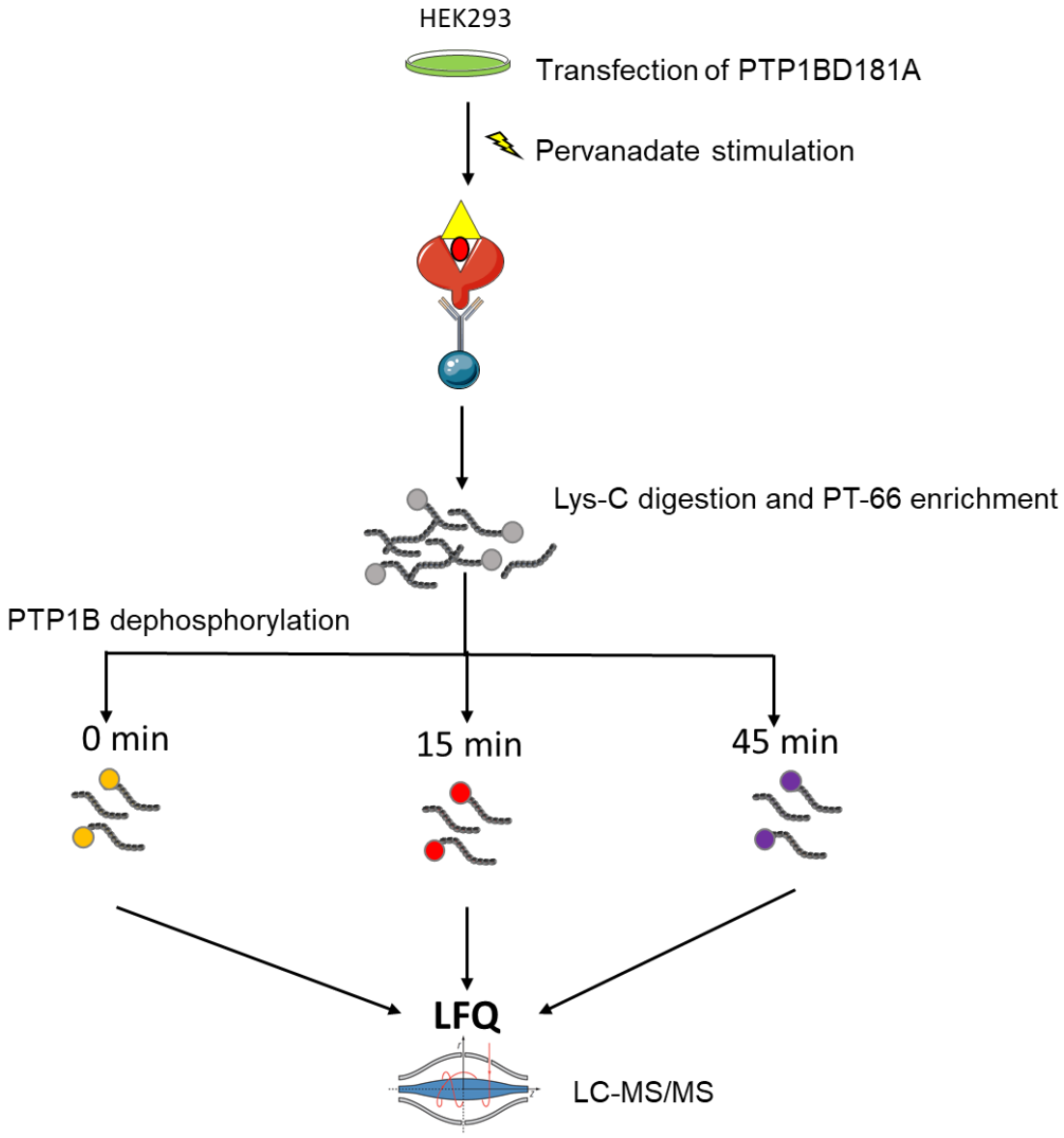


Figure 3-3: Workflow of label-free quantitation of pTyr in PTP1B phosphatase reaction.

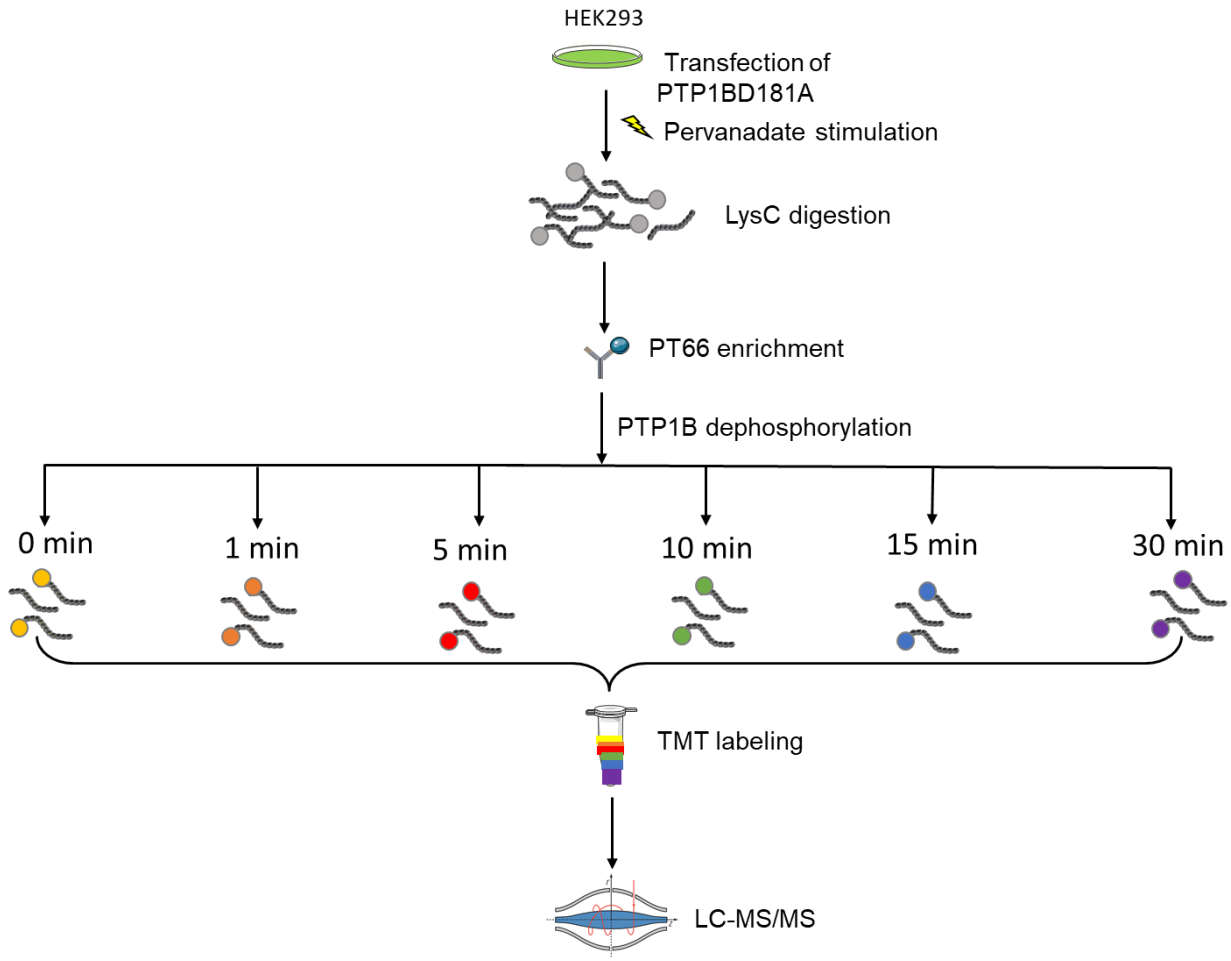


Figure 3-4: Workflow of TMT labeled kinetic assay of PTP1B enzyme.

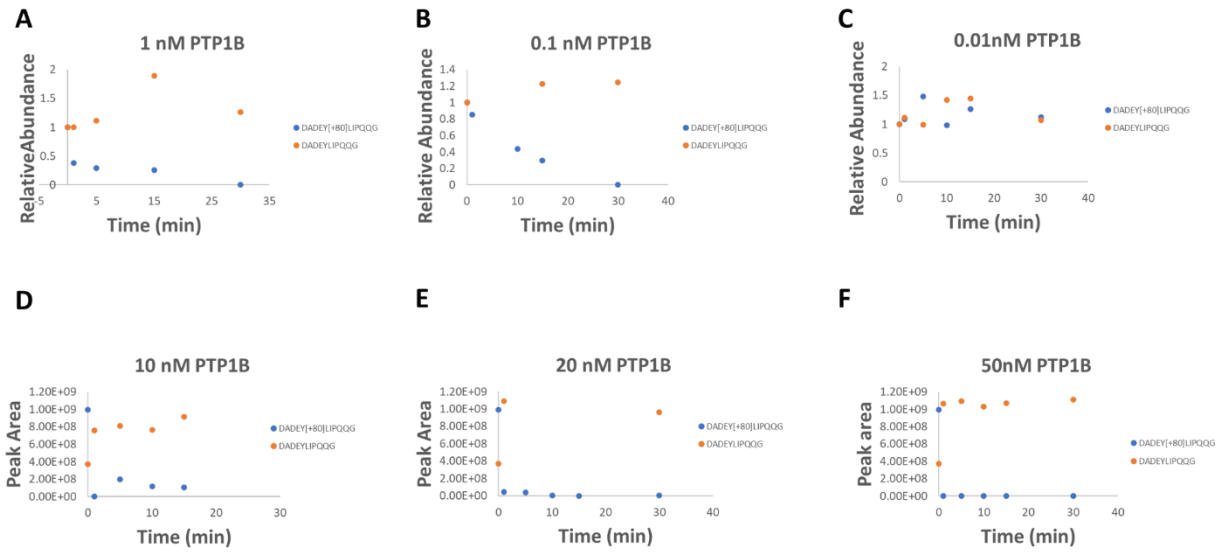


Figure 3-5: Optimization of PTP1B concentration using known phosphorylated PTP1B substrate.

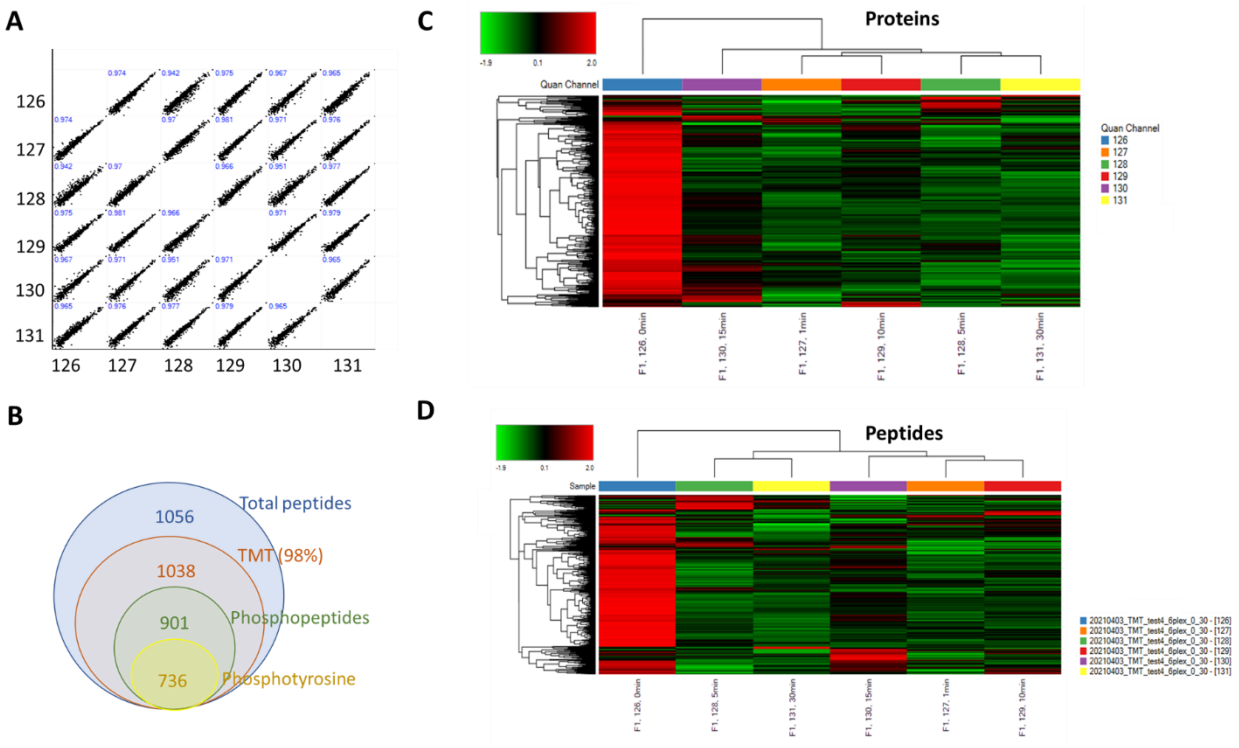
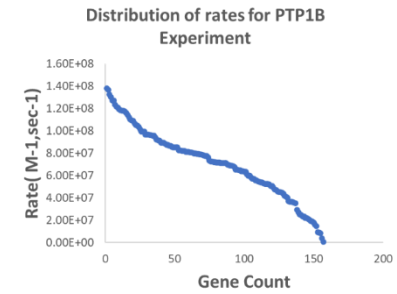


Figure 3-6: A) Correlation of reporter intensities for all channels. B) Number of total peptides, TMT labeled peptides, phosphopeptides, and pTyr peptides identified. C) Heat map of proteins intensity found in all channels. D) Heat map of peptides intensity found in all channels.

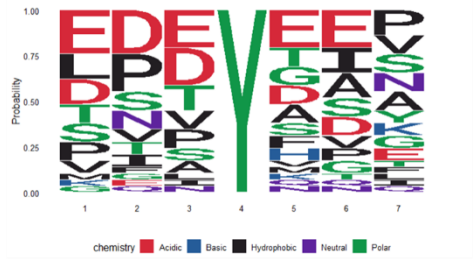
A



B

Publication	Peptide	Kcat/Km (s-1 M-1)
Zhang et al. *PNAS* 1993	DADEpYLIPQQG	2.9E+07
Sarmiento et al. *JBC* 1998		1.1E+07
Davis et al. *Bioconjug. Chem* 2012		2.4E+06
Hon et al. *Biol Inorg Chem* 2017		1.1E+07
Our data		6.5E+07

C



D

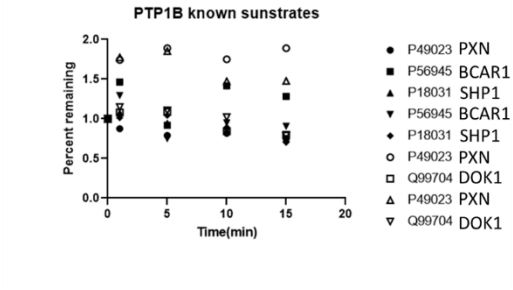


Figure 3-7: Distribution of rates for PTP1B phosphatase reaction. B) Kcat/Km of DADEpYLIPQQG peptide in PTP1B phosphatase reaction in comparison with reported values. C) Sequence logos (WebLogo) showing amino acid preferences at position flanking phosphorylated tyrosine residue in PTP1B phosphatase reaction. D) The percentage of PTP1B known substrates remaining in PTP1B phosphatase reaction.

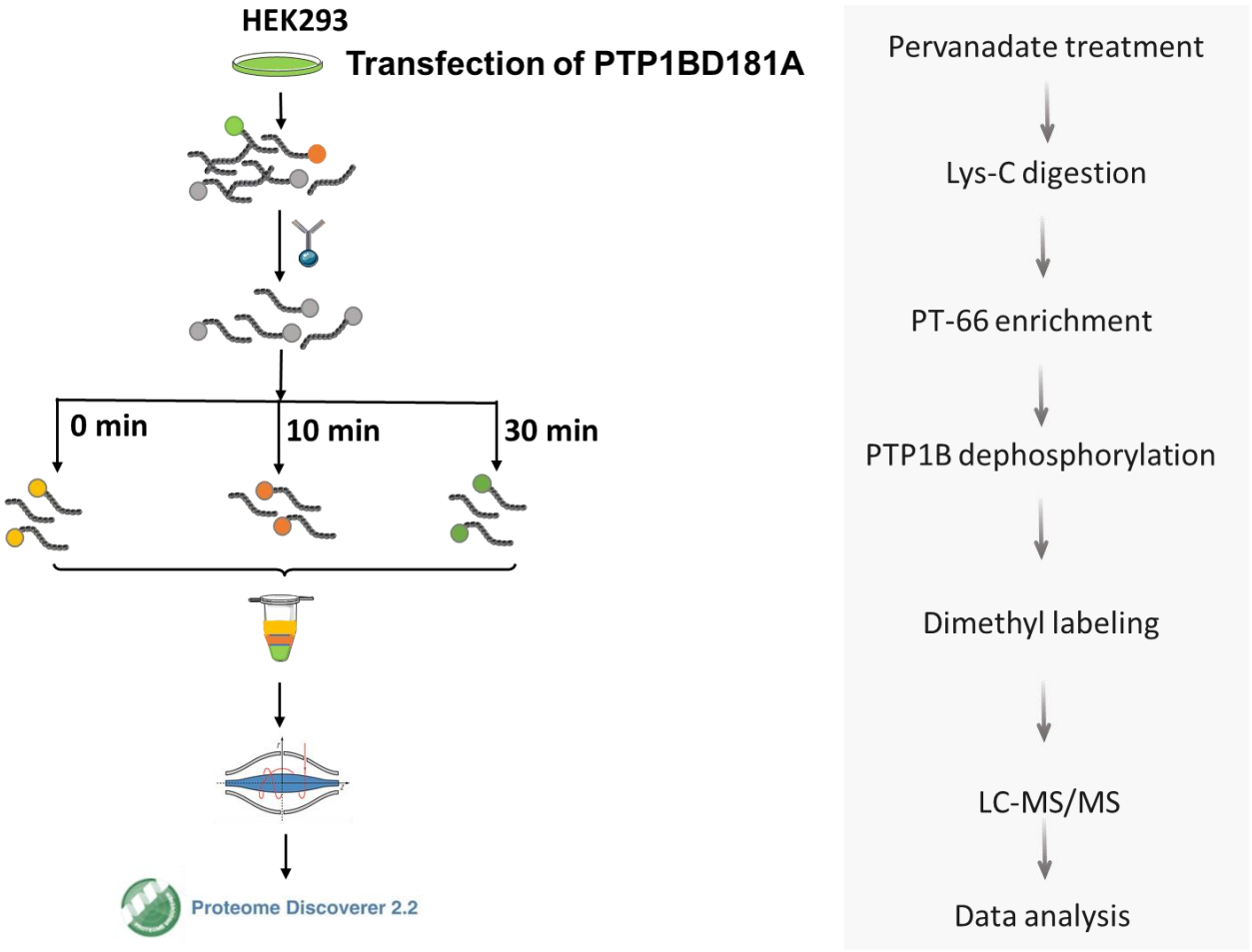


Figure 3-8: Workflow of *dimethyl labeling* kinetic assay of PTP1B.

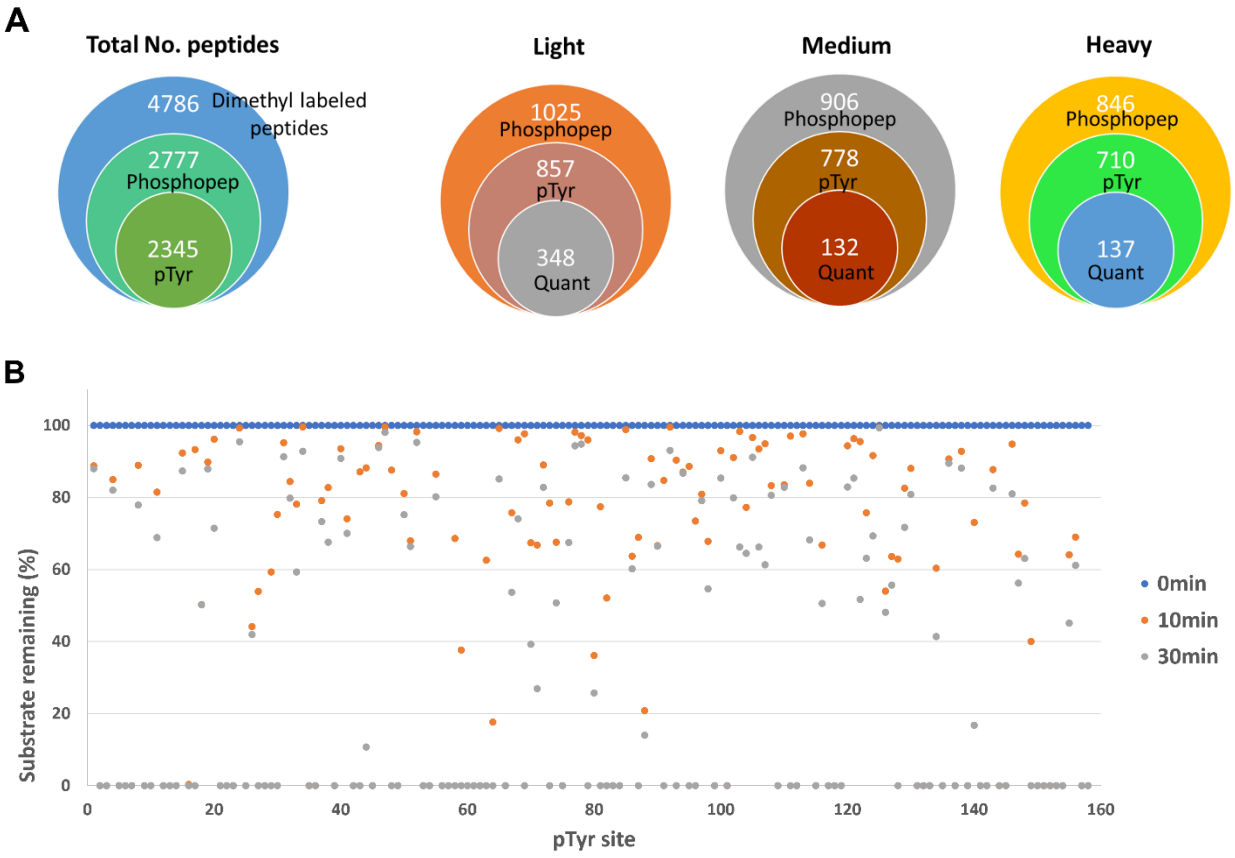
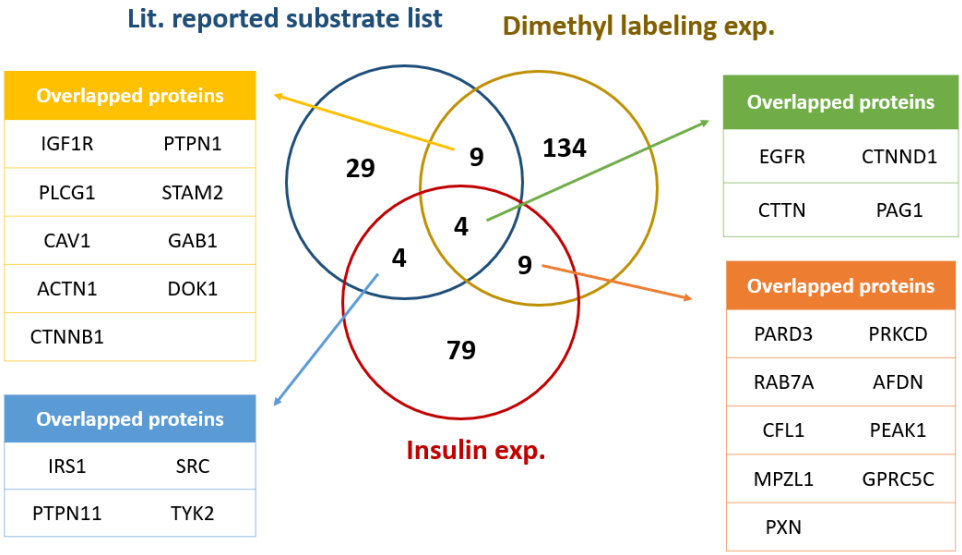


Figure 3-9: Result from dimethyl labeling experiment. A) Number of identified dimethyl labeled peptides, phosphopeptides, and phosphotyrosine found in total, light, medium and heavy labeled dimethyl experiment. B) Scatter plot of pTyr site that showed dephosphorylation after 10- and 30-min incubation with PTP1B phosphatase with p value < 0.05.

A

Overlapped proteins		
EGFR	IGF1R	PLCG1
ACTN1	CTTN	CAV1
PAG1	CTNND1	DOK1

B



Overlapped proteins	
IGF1R	PTPN1
PLCG1	STAM2
CAV1	GAB1
ACTN1	DOK1
CTNNB1	

Overlapped proteins	
IRS1	SRC
PTPN11	TYK2

Overlapped proteins	
EGFR	CTNND1
CTTN	PAG1

Overlapped proteins	
PARD3	PRKCD
RAB7A	AFDN
CFL1	PEAK1
MPZL1	GPRC5C
PXN	

Figure 3-10: A) Overlap of the list from dimethyl labeling experiment with literature reported PTP1B substrate list. B) Overlap of the lists from dimethyl labeling and insulin experiments with literature reported PTP1B substrate list.

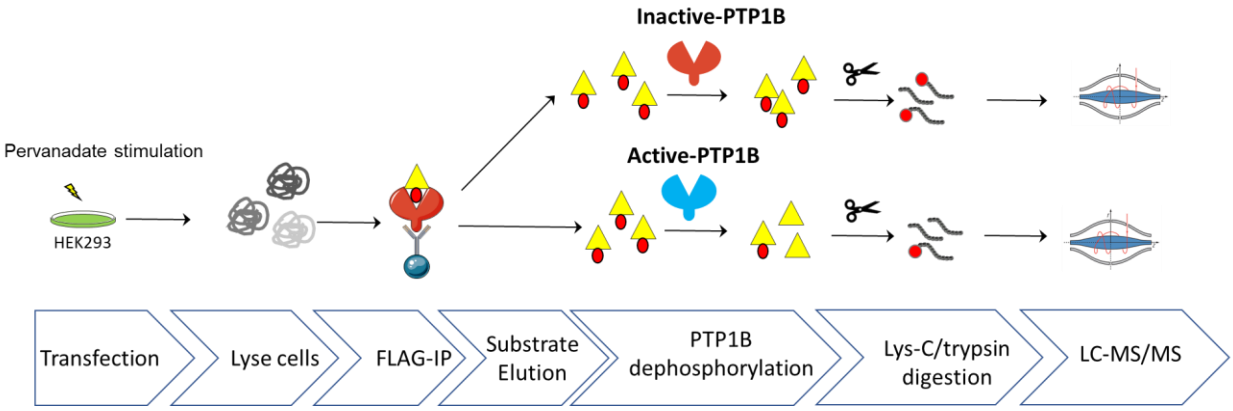


Figure 3-11: Workflow on the protein level to isolate PTP1B substrate using FLAG tag beads

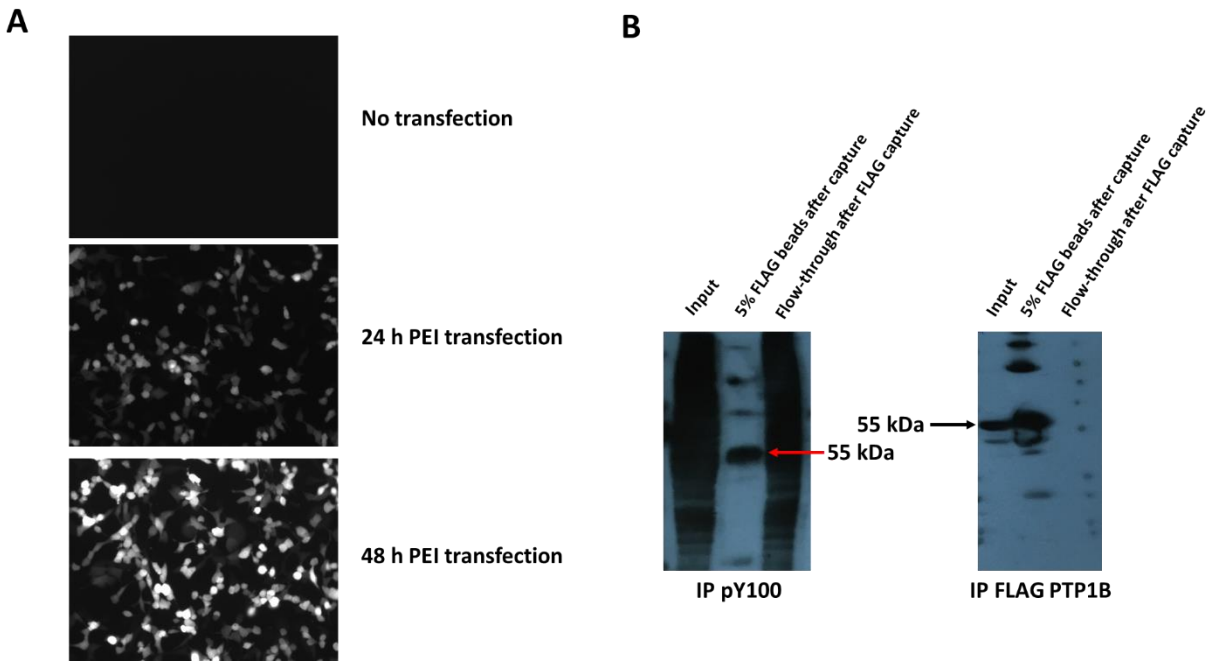


Figure 3-12: A) PTP1BD181A transfection into HEK293 cells using PEI system at time 0, 24h, and 48h. B) Left: pervanadate stimulation increased pTyr level in the cells; Right: FLAG-IP was efficient to isolate PTP1B in the cell lysate.

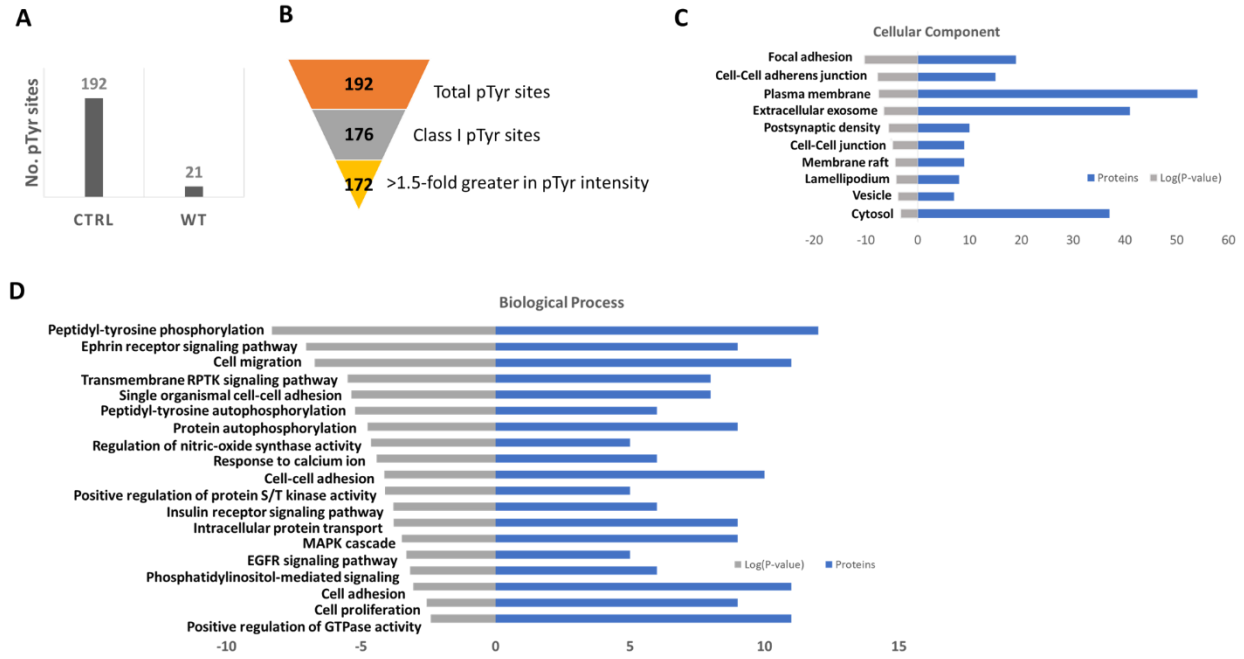


Figure 3-13: A) Number of identified pTyr site in Ctrl and WT samples. B) Number of pTyr found in PTP1B phosphatase assay incubated with inactive PTP1B enzyme. C) Functional classification of 111 hit proteins classified in cellular component. D) Functional classification of 111 hit proteins classified in biological processes. Functional classification used DAVID 6.8, with both P value (<0.05) and at least five proteins shown.

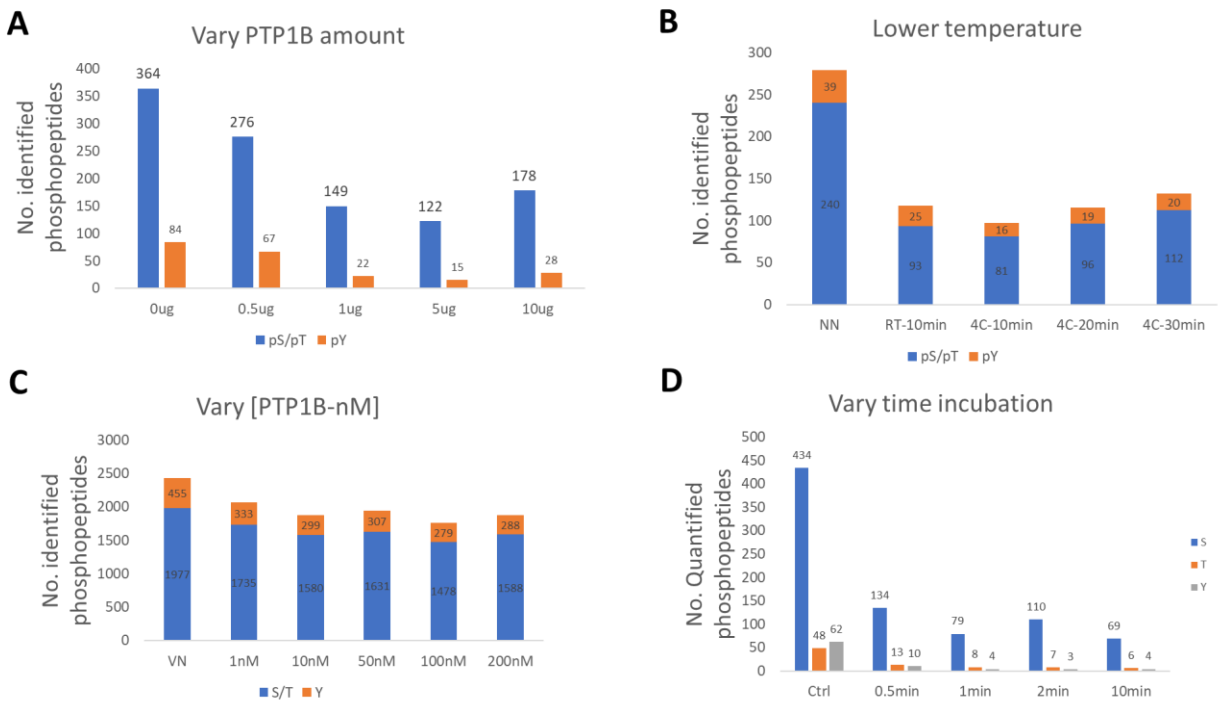
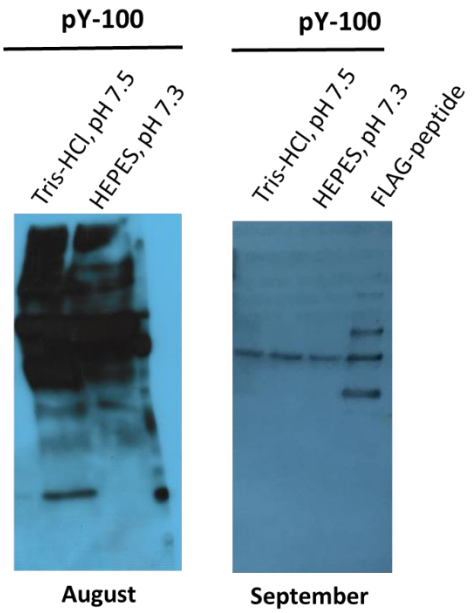


Figure 3-14: Optimization of PTP1B enzyme reaction, in terms of enzyme concentrations, incubation temperatures, and incubation times.



	August	September
1	25mM Tris-HCl, pH 7.5/10 mM sodium orthovanadate	25mM Tris-HCl, pH 7.5/10 mM sodium orthovanadate
2	25mM HEPES, pH 7.3/10 mM sodium orthovanadate	25mM HEPES, pH 7.3/10 mM sodium orthovanadate
3		100ug/ml FLAG-peptide

Figure 3-15: Optimization of sodium orthovanadate elution conditions in two different dates.

VITA

Peipei Zhu is originally from Zhejiang, China, but migrated to Albania, Europe, when she was 12. She graduated from ACS Athens high school in Athens, Greece, with an International Baccalaureate diploma. In 2013, Peipei earned a bachelor's degree in Chemistry from Rochester Institute of Technology in Rochester, New York. Peipei wanted to pursue a career in biochemistry, so she spent two years earning an MS in Chemistry under the guidance of Dr. Suzanne O'Handley. The title of her MS thesis is An ApnAase/mRNA decapping Nudix hydrolase from *Mycobacterium leprae.* Peipei was so intrigued with her research that she chose to start a Ph.D. program at Purdue University in the fall of 2015. She joined the lab of Dr. Weiguo Andy Tao. She spent the rest of her time at Purdue developing novel phosphoproteomics methods to identify phosphatase substrates and investigating phosphorylation signaling pathways in EGF and insulin activation.

LIST OF PUBLICATIONS

(1) **Zhu, P**; Zhang, R; Hsu, CC; Zhang, ZY; Tao, W. A; (**2021**). Systematic Identification of SHP2 Phosphatase Substrates. (Manuscript preparation)

(2) DeMarco, A. G.; Milholland, K. L.; Pendleton, A. L.; Whitney, J. J.; **Zhu, P**.; Wesenberg, D. T.; Nambiar, M.; Pepe, A.; Paula, S.; Chmielewski, J.; et al. Conservation of Cdc14 Phosphatase Specificity in Plant Fungal Pathogens: Implications for Antifungal Development. *Sci. Reports 2020 101* **2020**, *10* (1), 1–14.

(2) Zhang, Y.; **Zhu, P**.; Wu, X.; Yuan, T.; Su, Z.; Chen, S.; Zhou, Y.; Tao, W. A. Microcystin-LR Induces NLRP3 Inflammasome Activation via FOXO1 Phosphorylation, Resulting in Interleukin-1β Secretion and Pyroptosis in Hepatocytes. *Toxicol. Sci.* **2021**, *179* (1), 53–69.

(3) Wang, P.; Hsu, C.-C.; Du, Y.; **Zhu, P**.; Zhao, C.; Fu, X.; Zhang, C.; Paez, J. S.; Macho, A. P.; Tao, W. A.; et al. Mapping Proteome-Wide Targets of Protein Kinases in Plant Stress Responses. *Proc. Natl. Acad. Sci.* **2020**, *117* (6), 3270–3280.

(4) Zhao, C.; Zayed, O.; Zeng, F.; Liu, C.; Zhang, L.; **Zhu, P**.; Hsu, C.-C.; Tuncil, Y. E.; Tao, W. A.; Carpita, N. C.; et al. Arabinose Biosynthesis Is Critical for Salt Stress Tolerance in Arabidopsis. *New Phytol.* **2019**, *224* (1), 274–290.

(5) Zhao, C.; Zayed, O.; Yu, Z.; Jiang, W.; **Zhu, P**.; Hsu, C.-C.; Zhang, L.; Tao, W. A.; Lozano-Durán, R.; Zhu, J.-K. Leucine-Rich Repeat Extensin Proteins Regulate Plant Salt Tolerance in Arabidopsis. *Proc. Natl. Acad. Sci.* **2018**, *115* (51), 13123–13128.

(6) Zhang, C.; Du, X.; Tang, K.; Yang, Z.; Pan, L.; **Zhu, P**.; Luo, J.; Jiang, Y.; Zhang, H.; Wan, H.; et al. Arabidopsis AGDP1 Links H3K9me2 to DNA Methylation in Heterochromatin. *Nat. Commun.* **2018**, *9* (1).

(7) Hsu, C.-C.; Zhu, Y.; Arrington, J. V; Paez, J. S.; Wang, P.; **Zhu, P**.; Chen, I.-H.; Zhu, J.-K.; Tao, W. A. Universal Plant Phosphoproteomics Workflow and Its Application to Tomato Signaling in Response to Cold Stress. *Mol. Cell. Proteomics* **2018**, *17* (10), 2068–2080.