DEVELOPMENT OF NOVEL PROTEOMIC STRATEGIES TO DISSECT PLANT PHOSPHOPROTEOMIC SIGNALING UNDER ENVIRONMENTAL STRESSES

by

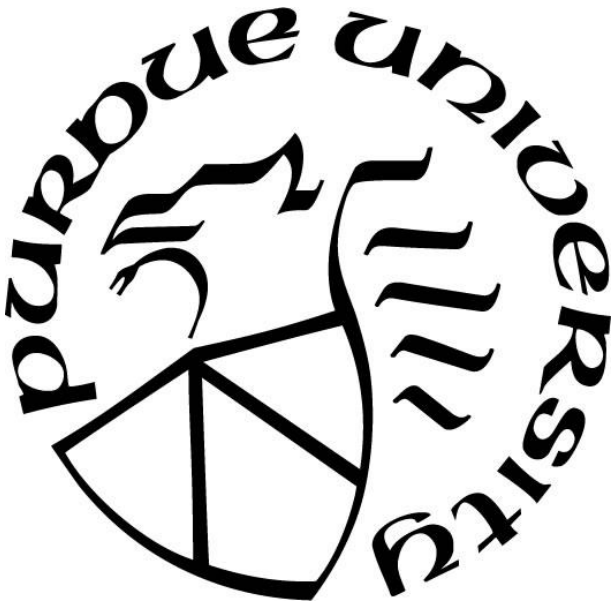
Chuan-Chih Hsu

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy



Department of Biochemistry

West Lafayette, Indiana

December 2017

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF DISSERTATION APPROVAL

Dr. W. Andy Tao, Chair
Department of Biochemistry

Dr. Chun-Ju Chang
Department of Basic Medical Sciences

Dr. Mark Hall
Department of Biochemistry

Dr. Jian-Kang Zhu
Department of Horticulture and Landscape Architecture

Approved by:

Dr. Christine A. Hrycyna
Head of the Departmental Graduate Program

To my father, who taught me to be independent as a person; my mother, who supported me without a word of complaint; and my sister, who has been my best friend since the beginning

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Dr. W. Andy Tao, for his guidance and support during four years. He made me enjoy doing research because his passion for research and development. I acknowledge the help of all the members of thesis committee, Dr. Chun-Ju Chang, Dr. Mark Hall, and Dr. Jian-Kang Zhu for the suggestions of my projects. I need to appreciate all my collaborators for not only providing plant materials but also inspiring and teaching me the importance of biological concepts. Specifically, I thank Dr. Jian-Kang Zhu, Dr. Pengcheng Wang, Dr. Yingfang Zhu, Dr. Chunzhao Zhao, Dr. Jun Yan, Dr. Cheng-Guo Duan, Dr. Yang Zhao, Dr. Yueh-Ju Hou, Dr. Mingguang Lai, Dr. Zhaobo Lang, for supporting me the plant projects.

I would like to acknowledge all the Tao lab mates that I have worked during the four years. I appreciate the mentorship of Dr. Liang Xue, Dr. Anton Iliuk, the many supports of Dr. Justine Arrington, J. Sebastian, Peipei Zhu, and the lasting friendship of I-Hsuan Chen since 2009. I will never forget all the fun and enjoyable memories I had with you in this lab.

Finally, I would like to thank my loving family; my parents Mu-Ju Hsu and Wen-Chih Lo, my sister Chuan-Hui Hsu, and my grandparents have been behind my whole life. Last but not least, I would like to thank my fiancée Yun Chien for your unconditional love, support, encouragement during the long-distance relationship of five years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CAA	2-chloroacetamide
CIPK	Calcineurin B-like protein-interacting protein kinases
CKII	Casein kinase 2
CKL2	Casein kinase 1 like protein 2
CPK	Calcium dependent protein kinase
FASP	Filter aided sample preparation
FDR	False discovery rate
GdnHCl	Guanidine hydrochloride
GO	Gene ontology
IMAC	Immobilized metal ion affinity chromatography
KALIP	Kinase assay linked phosphoproteomics
LC-MS	Liquid chromatography-mass spectrometry
MAPK	Mitogen activated protein kinase
OXI1	Oxidative signal inducible 1
PolyMAC	Polymer-based metal-ion affinity capture
SDC	Sodium deoxycholate
siKALIP	Stable isotope labeling kinase assay linked phosphoproteomics
SLS	Sodium lauroyl sarcosinate
SnRK2	Sucrose non-fermenting 1-related protein kinase 2
SOS2	Salt overly sensitive 2
TCEP	Tris(2-carboxyethyl)phosphine hydrochloride
TSAP	Temperature-sensitive alkaline phosphatase

ABSTRACT

Author: Hsu, Chuan-Chih. Ph.D.
Institution: Purdue University
Degree Received: December 2017
Title: Development of Novel Proteomic Strategies to Dissect Plant Phosphoproteomic Signaling under Environmental Stresses.
Major Professor: Weiguo Andy Tao

Protein phosphorylation is one of the important signaling mechanisms in plants which transduces environmental stimuli such as salinity, microbes, and hormones into intracellular signals and activates plant defense mechanisms. Thus, understanding the correlation between environmental stresses and alteration of plant phosphorylation requires system-wide phosphoproteomic analysis, which includes identification of kinase-substrate complexes and measurement of phosphorylation-mediated signaling changes. However, identification and quantification of plant phosphoproteome remains challenging due to the highly dynamic nature of plant proteome, interferences of cell wall, pigments, and secondary metabolites. Recently, mass spectrometry (MS) has been integrated with phosphopeptide enrichment approaches for identifying thousands of phosphorylation sites and for quantifying phosphoprotein stoichiometry. Although MS-based phosphoproteomics has revealed the global phosphorylation changes related to different physiological states of plants, many kinase-substrate networks involved in essential signaling pathways, such as the ABA-induced SNF-1-related protein kinase 2 (SnRK2) pathway and the mitogen-activated protein kinases (MAPKs) cascades, are still not completely understood. This dissertation discusses strategies for improving plant sample preparation and for identifying the direct substrates of the plant kinases. Chapter one highlights the low phosphopeptide identification rate by mass spectrometry. Chapter two details the development of a sample preparation protocol for the plant phosphoproteome analysis, and the application of the protocol for the study of tomato cold-induced phosphoproteomic changes. Chapter three shows the development of a novel approach for identification of the direct substrates of the plant kinases, whose activation regulates the signaling transductions of plant stress defense mechanisms.

CHAPTER 1. ESTIMATION OF THE EFFICIENCY OF PHOSPHOPEPTIDE IDENTIFICATION BY TANDEM MASS SPECTROMETRY

1.1 Summary

Mass spectrometry has played a significant role in the identification of unknown phosphoproteins and sites of phosphorylation in biological samples. Analyses of protein phosphorylation, particularly large scale phosphoproteomic experiments, have recently been enhanced by efficient enrichment, fast and accurate instrumentation, and better software, but challenges remain because of the low stoichiometry of phosphorylation and poor phosphopeptide ionization efficiency and fragmentation due to neutral loss. Phosphoproteomics has become an important dimension in systems biology studies, and it is essential to have efficient analytical tools to cover a broad range of signaling events. To evaluate current mass spectrometric performance, we present here a novel method to estimate the efficiency of phosphopeptide identification by tandem mass spectrometry. Phosphopeptides were directly isolated from whole plant cell extracts, dephosphorylated, and then incubated with one of three purified kinases—Casein Kinase II, Mitogen-activated protein kinase 6, and SNF-related protein kinase 2.6—along with $^{16}O_4$- and $^{18}O_4$-ATP separately for in vitro kinase reactions. Phosphopeptides were enriched and analyzed by LC-MS. The phosphopeptide identification rate was estimated by comparing phosphopeptides identified by tandem mass spectrometry with phosphopeptide pairs generated by stable isotope labeled kinase reactions. Overall, we found that current high speed and high accuracy mass spectrometers can only identify 20-40% of total phosphopeptides primarily due to relatively poor fragmentation, additional modifications, and low abundance, highlighting the urgent need for continuous efforts to improve phosphopeptide identification efficiency.

1.2 Introduction

It is estimated that there are over 500 protein kinases in mammals and over 1000 in plants, and an estimated half or more of all proteins are phosphorylated at certain points during their lifespan [1-2]. Reversible phosphorylation of proteins is involved in the regulation of most, if not all, cellular processes [3]. Abnormal phosphorylation has been implicated in a number of diseases, most notably cancer [4]. The accurate determination of sites of phosphorylation and dynamics of this modification in response to extracellular stimulation is important for elucidating complex disease mechanisms and global regulatory networks. Development of methods for analyzing phosphorylated proteins, therefore, has been an active field of research in the signaling, mass spectrometry, and proteomics communities.

Advances in mass spectrometry (MS)-based proteomics have driven increasing efforts to identify reliable approaches for the large scale analysis of phosphoproteins (phosphoproteomics) that include both the identification of protein phosphorylation sites and the quantification of changes in phosphorylation at individual sites [5]. Phosphorylation is often a low stoichiometric event [6]. To identify specific sites of phosphorylation, it is essential to have an efficient strategy for the selective enrichment of actual phosphopeptides. Current approaches include immobilized metal ion or metal oxide affinity chromatography (IMAC and MOAC) [7-10] and polymer-based metal ion affinity capture (PolyMAC) for general phosphopeptide enrichment [11] and anti-phosphotyrosine antibodies for the isolation of phosphotyrosine-containing peptides [12]. High throughput analysis of phosphorylation using directed enrichment methods followed by mass spectrometry (MS) has become a standard approach for phosphoprotein detection.

While phosphoproteomics has increasingly become an important, informative, dimension in omics studies, its challenges have persisted [13]. The primary challenge in examining protein phosphorylation is its low stoichiometry. Phosphorylated proteins, especially those involved in signaling, are often expressed in relatively low amounts in a cell, and few of these proteins exist in a phosphorylated form at any one time. Furthermore, phosphopeptides have low ionization efficiency due to their negatively charged phosphate groups [14], and they can exhibit poor fragmentation in tandem mass spectra due to neutral

loss of the phosphate groups [15]. Finally, informatics approaches for processing the results of mass spectrometry data for phosphopeptides are not yet mature [16].

There have been a number of attempts to improve phosphopeptide identification efficiency, particularly by alternative activation methods [17]. Faster and more accurate LC-MS systems have also made a significant contribution toward enhancing the coverage of the phosphoproteome, but it is not still clear what percentage of the phosphopeptide population is routinely identified by tandem mass spectrometry (MS/MS) and whether current phosphoproteomic strategies provide true representations of the phosphoproteome for systems biology analyses. Due to the dynamic nature of phosphorylation, low coverage of the phosphoproteome at certain cellular states might lead to biased or even incorrect conclusions. Here, we present a novel strategy to estimate the efficiency of phosphopeptide identification by tandem mass spectrometry. Instead of using a large pool of synthetic phosphopeptides which is costly to generate [18] and still incomprehensive, we created a phosphopeptide pool directly from whole cell extracts. To generate phosphopeptides with distinctively recognizable features in the mass spectra, we introduced *in vitro* kinase reactions with $^{16}O_4$- and $^{18}O_4$-ATP to generate phosphopeptide pairs with similar intensity that are separated by 6 Da on mass spectra. Previous literature and our own data indicate that the ^{18}O atoms on the -phosphoryl group do not exchange with water during kinase reactions [19]. Thus, the efficiency of phosphopeptide identification can be estimated by comparing the phosphopeptides identified by MS/MS with the total number of phosphopeptide pairs that demonstrate the distinctive mass shift.

1.3 Experimental Procedure

Unless otherwise noted, reagents were purchased from Sigma (St. Louis, MO), and buffers were prepared with water purified to a resistivity of 18.2 MΩ cm-1 using a Barnstead Nanopure water system (Thermo Fisher).

1.3.1 Plant Materials and Growth

The seeds of Col-0 wild type Arabidopsis were geminated on half-strength Murashige and Skoog (MS) medium (1 % sucrose with 0.6% phytogel). Five days after germination, seedlings were transferred into 40 ml half-strength MS liquid medium with 1% sucrose at

22 °C in continuous light on a rotary shaker set at 100 rpm. For osmotic stress treatment, twelve-day-old seedlings were transferred into fresh medium containing 800 mM Mannitol for 30 min. In parallel, the seedlings transferred into fresh medium were used as the control.

1.3.2 Protein Extraction and Digestion

Plant tissues were ground with mortar and pestle in liquid nitrogen, and the ground tissues were lysed in 6 M guanidine hydrochloride containing 100 mM Tris-HCl (pH = 8.5) with EDTA-free protease inhibitor cocktail (Roche, Madison, WI) and phosphatase inhibitor cocktail (Sigma-Aldrich, St. Louis, MO). Proteins were reduced and alkylated with 10 mM tris-(2-carboxyethyl)phosphine (TECP) and 40 mM chloroacetamide (CAA) at 95 °C for 5 min. Alkylated proteins were subjected to methanol-chloroform precipitation, and precipitated protein pellets were solubilized in 8 M urea containing 50 mM triethylammonium bicarbonate (TEAB). Protein amount was quantified by BCA assay (Thermo Fisher Scientific, Rockford, IL). Protein extracts were diluted to 4 M urea and digested with Lys-C (Wako, Japan) in a 1:100 (w/w) enzyme-to-protein ratio overnight at 37 °C. Digests were acidified with 10% trifluoroacetic acid (TFA) to a pH ~2 and desalted using a 100 mg Sep-Pak C18 column (Waters, Milford, MA).

1.3.3 Stable Isotope Labeled *In Vitro* Kinase Reaction

The *in vitro* kinase reaction was performed based on previous reports [20-21] with some modifications. The Lys-C digested peptides (200 µg) were treated with a thermosensitive alkaline phosphatase (TSAP) (Roche, Madison, WI) in a 1:100 (w/w) enzyme-to-peptide ratio at 37 °C overnight for dephosphorylation, and the dephosphorylated peptides were desalted using Sep-Pak C18 column. The desalted peptides were re-suspended in kinase reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, and 1 mM DTT, pH 7.5) with either 1 mM ^{16}O-ATP or γ-[$^{18}O_4$]-ATP (Cambridge Isotope Laboratories, MA). The suspended peptides were incubated with the recombinant SNF-related protein kinase 2.6 (SnRK2.6), Mitogen-activated protein kinase 6 (MPK6), or Casein Kinase II (CK2) (500 ng) at 30 °C overnight. The kinase reaction was quenched by acidifying with 10% TFA to a final concentration of 1%, and the peptides were desalted by Sep-Pak C18 column. The light

and heavy phosphopeptides were mixed and further digested by trypsin at 37 °C for 6 h. Tryptic phosphopeptides were desalted by Sep-Pak column, and then were enriched by PolyMAC-Ti reagent, and the eluates were dried in a SpeedVac for LC-MS/MS analysis.

1.3.4 PolyMAC Enrichment

Phosphopeptide enrichment was performed according to the reported PolyMAC-Ti protocol [11] with some modifications. Tryptic peptides (200 µg) were re-suspended in 100 µL of loading buffer (80% acetonitrile (ACN) with 1% TFA) and incubated with 25 µL of the PolyMAC-Ti reagent for 20 min. A magnetic rack was used to collect the magnetic beads to the sides of the tubes, and the flow-through was discarded. The magnetic beads were washed with 200 µL of washing buffer 1 (80% ACN, 0.2% TFA with 25 mM glycolic acid) for 5 min and washing buffer 2 (80% ACN in water) for 30 seconds. Phosphopeptides were then eluted with 200 µL of 400 mM NH_4OH with 50% ACN and dried in a SpeedVac.

1.3.5 LC-MS/MS Analysis

The phosphopeptides were dissolved in 5 µL of 0.3% formic acid (FA) with 3% ACN and injected into an Easy-nLC 1000 (Thermo Fisher Scientific). Peptides were separated on a 45 cm in-house packed column (360 µm OD × 75 µm ID) containing C18 resin (2.2 µm, 100Å, Michrom Bioresources) with a 30 cm column heater (Analytical Sales and Services) set at 50 °C. The mobile phase buffer consisted of 0.1% FA in ultra-pure water (buffer A) with an eluting buffer of 0.1% FA in 80% ACN (buffer B) run over a linear 60 min gradient of 5%-30% buffer B at a flow rate of 250 nL/min. The Easy-nLC 1000 was coupled online with a Velos Pro LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in the data-dependent mode in which a full MS scan (from m/z 350-1500 with the resolution of 30,000 at m/z 400) was followed by the 5 most intense ions being subjected to collision-induced dissociation (CID) fragmentation. CID fragmentation was performed and acquired in the linear ion trap (normalized collision energy (NCE) 30%, AGC 3e4, max injection time 100 ms, isolation window 3 m/z, and dynamic exclusion 60 s).

1.3.6 Data Processing

The raw files were searched directly against the *Arabidopsis thaliana* database (TAIR10) with no redundant entries using MaxQuant software (version 1.5.4.1) [22] with the Andromeda search engine. Initial precursor mass tolerance was set at 20 ppm, the final tolerance was set at 6 ppm, and the ITMS MS/MS tolerance was set at 0.6 Da. Search criteria included a static carbamidomethylation of cysteines (+57.0214 Da) and variable modifications of (1) oxidation (+15.9949 Da) on methionine residues, (2) acetylation (+42.011 Da) at the N-terminus of proteins, (3) phosphorylation (+79.996 Da), and (4) heavy phosphorylation (+85.979 Da) on serine, threonine or tyrosine residues. The match between runs function was enabled with 1.0 min match time window. The searches were performed with trypsin digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. The false discovery rates for proteins, peptides and phosphosites were set at 0.01. The minimum peptide length was six amino acids, and a minimum Andromeda score cut-off was set at 40 for modified peptides. A site localization probability of 0.75 was used as the cut-off for localization of phosphorylation sites. The MS/MS spectra can be viewed through the MaxQuant viewer. For the ProteomeDiscoverer searches, the raw files were searched directly against the same Arabidopsis thaliana database (TAIR10) with no redundant entries using the SEQUEST HT algorithm in Proteome Discoverer version 2.1 (Thermo Fisher Scientific). Peptide precursor mass tolerance was set at 10 ppm, and MS/MS tolerance was set at 0.6 Da. Search criteria included a static carbamidomethylation of cysteines (+57.0214 Da) and variable modifications of (1) oxidation (+15.9949 Da) on methionine residues, (2) acetylation (+42.011 Da) on protein N-termini, (3) phosphorylation (+79.996 Da), and (4) heavy phosphorylation (+85.979 Da) on serine, threonine or tyrosine residues. Searches were performed with full tryptic digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. Relaxed and strict false discovery rates (FDR) were set to 0.05 and 0.01, respectively. All localized phosphorylation sites were submitted to Motif-X [23] to determine kinase phosphorylation motifs with the TAIR10 database as the background. The significance was set at 0.000001, the width was set at 13, and the number of occurrences was set at 20. The light and heavy phosphopeptide and peak pairs were identified through the LAXIC algorithm [24]. All of the light and heavy

phosphopeptide and peak pairs are listed in the supplementary tables, and the raw data and analysis files for the proteomic analyses have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the jPOST partner repository (http://jpost.org) [25] with the data set identifier PXD005079.

1.4 Results and Discussion

The strategy to estimate the efficiency of phosphopeptide identification was devised based on our previous phosphoproteomic studies on kinase substrates [20-21]. The general strategy to estimate the efficiency of phosphopeptide identification by MS/MS is illustrated in Figure 1.1. To generate a comprehensive pool of phosphopeptides, proteins were extracted from whole cell lysates such as whole cell extracts from plants. After digestion with Lys-C to generate peptides, the peptides were incubated with a thermosensitive alkaline phosphatase overnight to remove phosphate groups on the peptides and to generate a pool of peptide candidates for the *in vitro* kinase reactions. We chose three kinases, casein kinase 2 (CK2), mitogen-activated protein kinase 6 (MPK6) and SNF-related protein kinase 2.6 (SnRK2.6), for their known high specificity toward acidic, proline-directed, and basic motifs, respectively. These three kinases also have high enzymatic activity *in vitro* and can potentially phosphorylate hundreds of substrates. The most important feature of this strategy is the generation of a large number of phosphopeptides that are invisible in MS/MS but have distinctive characteristics that can be unambiguously recognized even if their sequences are unknown. In doing so, we devised *in vitro* kinase reactions with γ-$^{16}O_4$- and γ-$^{18}O_4$-ATP in parallel. The kinase reaction transfers one or more γ-phosphate groups from ATP to substrate peptides, thus generating light- and heavy- phosphorylated peptides with similar intensities, assuming the same kinase has similar reactivity with γ-$^{16}O_4$- or γ-$^{18}O_4$-ATP. After the kinase reactions, samples were pooled together, and phosphopeptides were enriched with PolyMAC before LC-MS analyses. Data were searched against the appropriate protein database for sequence information. In-house LAXIC software was used to pick and quantify peptide pairs with the required characteristic features [24]. Finally, the efficiency of phosphopeptide identification can be estimated by comparing the phosphopeptides identified by MS/MS with the total number of phosphopeptide pairs.

We generated whole cell lysates from *Arabidopsis* seedlings in this study. The plant has over 1000 encoded kinases, and whole cell lysates likely contain tens of thousands of phosphorylation sites[2, 26]. Two plant kinases, MAPK6 and SnRK2.6, were recombinantly expressed and purified. Along with human recombinant CK2, the three kinases were incubated with *Arabidopsis* lysate and γ-$^{16}O_4$- or γ-$^{18}O_4$-ATP separately. Each kinase reaction can generate hundreds of phosphopeptides, which is sufficient for this study but still well within the capacity of a typical high resolution LC-MS system today. This strategy also minimizes the instrumentation factor. While it is conceivable that the speed of mass spectrometers affects the identification rate of phosphopeptides, this factor would be minor with the current approach.

As anticipated, we observed multiple peak doublets in the mass spectra. With a high speed mass spectrometer such as the Orbitrap Velos, most of these precursor ions were selected for MS/MS. Figure 1.2 to 1.4 illustrates examples of three peptides phosphorylated by CK2, MPK6, or SnRK2.6, respectively. The extracted ion chromatogram (XIC) and MS/MS spectra of the paired light/heavy NUP50 phosphopeptide ApSDIEEGDEVDSK are shown in Figure 1.2. The peptide was phosphorylated by CK2, and the doubly-charged, heavy phosphate-labeled phosphopeptide (red line) has a 3.00 m/z shift from the doubly-charged, light phosphate-labeled phosphopetide (blue line). No significant retention time shift was observed as a result of heavy phosphate labeling. The MS/MS spectrum shows the identification of paired light/heavy phosphopeptide with the expected acidic phosphorylation motif. Similarly, the XIC and MS/MS spectra of the paired light/heavy CAD5 phosphopeptide DPSGILpSPYTYTLR phosphorylated by MPK6 are shown in Figure 1.3. The phosphopeptide sequence has the characteristic proline-directed phosphorylation motif; Figure 1.4 shows the XIC and MS/MS spectra of the paired light/heavy AT5G05600 phosphopeptide VQpSLAESNLSSLPDR phosphorylated by SnRK2.6. The MS/MS spectra shows the identification of paired light/heavy phosphopeptide with the basic phosphorylation motif [-I-x-R-x-x-pS-]. Although the examples provided in Figure 2 show that the light- and heavy-labeled phosphopeptide pairs have similar intensity and were sequenced by MS/MS, not all phosphopeptide pairs have similar intensity. In many cases, the light-labeled phosphopeptide has higher intensity than its heavy-labeled counterpart. The exact cause is not clear. While we expected similar

kinase reactivity with light or heavy ATP, it is possible that γ-$^{18}O_4$-ATP has a bigger size which might prevent it from fitting inside the ATP binding pocket perfectly. We will investigate this phenomenon in a separate study. All doublet peaks with appropriate mass difference and similar intensity were deconvoluted and counted as phosphopeptides, no matter whether they were sequenced by MS/MS or not.

Six *in vitro* kinase reactions, the result of three kinases with light and heavy ATP separately, generated thousands of phosphopeptides *in vitro*. We searched MS/MS spectra against the *Arabidopsis* proteome database. In total, 1498, 1472, and 1837 doublet phosphopeptides were identified by CK2, MPK6, and SnRK2.6, respectively. The data indicates high, specific *in vitro* kinase activity for all three kinases. Motif analyses of the identified phosphopeptides resulted in the acidic motif for the CK2 kinase reaction, [-(pS/pT)-(D/E)-x-(D/E)-], the proline-directed phosphorylation motif for MPK6, [-(pS/pT)-P-], and the basic phosphorylation motif for SnRK2.6, [-(I/L)-x-R-x-x-(pS/pT)-] (Figure 1.5). The results from these motif analyses are highly consistent with previous literature reports and known substrate specificity of the three kinases [27-29].

The advancement of high speed and high accuracy mass spectrometers, along with ultra high performance liquid chromatography (UHPLC), has greatly improved the coverage of phosphoproteomes. However, considering the high dynamics of protein phosphorylation, it is not clear whether current LC-MS technology can provide sufficient coverage of most phosphoroteomes. In our phosphopeptide samples prepared before LC-MS analyses, virtually all phosphopeptides were generated from *in vitro* kinase reactions. Assuming similar reactivity with light- or heavy-ATP, we expected to observe all phosphopeptides in doublets with similar intensities. We applied our in-house software LAXIC [24] to identify all peaks pairs that were separated by $^{16}O_4$ and $^{18}O_4$ phosphoryl groups with similar intensities, and we calculated the successful phosphopeptide identification rate through three steps. First, the light and heavy peak pairs were identified from MS scans through two criteria: (1) the mass difference of 6 Da between the two peaks, and (2) the peaks were detected in the same full MS scan. Next, the light/heavy phosphopeptide pairs were selected from light and heavy phosphopeptides identified through MS/MS which fit the two criteria we mentioned above. Finally, the successful phosphopeptide identification rate was acquired via the ratio of the identified light/heavy phosphopeptide pairs over the

identified light/heavy peak pairs. In total, we identified 4752, 4053, and 6749 pairs in samples related to the kinase reactions of CK2, MPK6, and SnRK2.6, respectively, that meet the criteria. Accordingly, we calculated the efficiency of phosphopeptide identification (Table 1.1 and Figure 1.6) by comparing the number of phosphopeptides identified by MS^2 against the number of phosphopeptide peak pairs in the MS spectra. The percentages of successful phosphopeptide identification are 31%, 36%, and 27% for CK2, MPK6, and SnRK2.6, respectively. On average, only about 30% of phosphopeptides were identified by our current LC-MS instrument.

We also examined the effect of different search algorithms. Besides MaxQuant, we also searched the MS/MS spectra using Proteome Discoverer 2.1. MaxQuant is based on Andromeda while Proteome Discoverer uses Sequest HT as the search engine. Overall, with the same FDR cutoff value (FDR <1%), there are some obvious difference in the number of phosphopeptides identified by Andromeda (MaxQuant) or Sequest HT (Proteome Discoverer), but the efficiency of phosphopeptide identification is within a similar range (Table 1 and Figure 1.6).

There are multiple factors that may contribute to the relatively low efficiency of phosphopeptide identification (~30%). One obvious possibility is poor fragmentation of phosphopeptides in MS^2 spectra. We generated two plots to show the proportion and number of phosphopeptide doublets versus cut-off value (Figure 1.7A and 1.7B). The plots are quite informative. The maximum proportion of phosphopeptide doublets subjected to MS/MS is around 60%, indicating that 40% of the phosphopeptide doublets were not selected for MS/MS in our study. These phosphopeptide doublets that were not selected for MS/MS are likely of low abundance. When phosphopeptide abundance is low enough, the isotope pattern cannot be identified, and monoisotopic peaks cannot be recognized for MS/MS. Among the 60% of phosphopeptide pairs that were selected for MS/MS, the phosphopeptides in MPK6's samples have the highest identification efficiency. This is consistent with previous data that indicates proline-containing peptides have a high degree of peptide backbone fragment [30], which may facilitate identification, and the fact that most of MPK6's substrate peptides have the [(-pS/pT)-P-] motif. Moreover, other reasons such as additional modifications on the phosphopeptides [31] or variant isoforms not listed in the database may contribute to the high percentage of unassigned spectra.

1.5 Conclusion

Large scale analysis of protein phosphorylation, or phosphoproteomics, has become an important component of systems biology studies. While the advances of mass spectrometers in speed and accuracy, along with the introduction of ultra high performance liquid chromatograph, have greatly improved phosphoproteome coverage, it is critical to evaluate whether the phosphoproteomic data is comprehensive, especially considering that protein phosphorylation is highly dynamic. We have presented a novel method to estimate the efficiency of phosphopeptide identification by generating a large pool of phosphopeptides through direct isolation from cell lysates and *in vitro* kinase reactions. These phosphopeptides can be recognized according to specific features, though they may or may not be isolated for MS/MS. Examination of MS/MS data and MS features indicates that, on average, 30% of phosphopeptides were identified by MS/MS. Poor fragmentation and low abundance contribute to 70% of the phosphopeptides not being identified by MS/MS. This study highlights the need for additional efforts to increase the yield of phosphopeptides for MS analyses, possibly through better sample preparation, phosphopeptide enrichment, and LC resolution, and to further improve phosphopeptide fragmentation through alternative methods.

1.6 References

1. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S., The protein kinase complement of the human genome. *Science* **2002,** *298* (5600), 1912-+.

2. Schulze, W. X., Proteomics approaches to understand protein phosphorylation in pathway modulation. *Curr Opin Plant Biol* **2010,** *13* (3), 280-87.

3. Cohen, P., The origins of protein phosphorylation. *Nat Cell Biol* **2002,** *4* (5), E127-E130.

4. Hunter, T., Signaling - 2000 and beyond. *Cell* **2000,** *100* (1), 113-127.

5. Dephoure, N.; Zhou, C.; Villen, J.; Beausoleil, S. A.; Bakalarski, C. E.; Elledge, S. J.; Gygi, S. P., A quantitative atlas of mitotic phosphorylation. *P Natl Acad Sci USA* **2008,** *105* (31), 10762-10767.

6. Mann, M.; Ong, S. E.; Gronborg, M.; Steen, H.; Jensen, O. N.; Pandey, A., Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome. *Trends Biotechnol* **2002,** *20* (6), 261-268.

7. Stensballe, A.; Andersen, S.; Jensen, O. N., Characterization of phosphoproteins from electrophoretic gels by nanoscale Fe(III) affinity chromatography with off-line mass spectrometry analysis. *Proteomics* **2001,** *1* (2), 207-222.

8. Tsai, C. F.; Hsu, C. C.; Hung, J. N.; Wang, Y. T.; Choong, W. K.; Zeng, M. Y.; Lin, P. Y.; Hong, R. W.; Sung, T. Y.; Chen, Y. J., Sequential Phosphoproteomic Enrichment through Complementary Metal-Directed Immobilized Metal Ion Affinity Chromatography. *Anal Chem* **2014,** *86* (1), 685-693.

9. Larsen, M. R.; Thingholm, T. E.; Jensen, O. N.; Roepstorff, P.; Jorgensen, T. J. D., Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular & Cellular Proteomics* **2005,** *4* (7), 873-886.

10. Sugiyama, N.; Masuda, T.; Shinoda, K.; Nakamura, A.; Tomita, M.; Ishihama, Y., Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications. *Molecular & Cellular Proteomics* **2007,** *6* (6), 1103-1109.

11. 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), 2162-72.

12. Rush, J.; Moritz, A.; Lee, K. A.; Guo, A.; Goss, V. L.; Spek, E. J.; Zhang, H.; Zha, X. M.; Polakiewicz, R. D.; Comb, M. J., Immunoaffinity profiling of tyrosine phosphorylation in cancer cells. *Nature Biotechnology* **2005,** *23* (1), 94-101.

13. Engholm-Keller, K.; Larsen, M. R., Technologies and challenges in large-scale phosphoproteomics. *Proteomics* **2013,** *13* (6), 910-31.

14. Winter, D.; Seidler, J.; Ziv, Y.; Shiloh, Y.; Lehmann, W. D., Citrate boosts the performance of phosphopeptide analysis by UPLC-ESI-MS/MS. *J Proteome Res* **2009,** *8* (1), 418-24.

15. Leitner, A.; Foettinger, A.; Lindner, W., Improving fragmentation of poorly fragmenting peptides and phosphopeptides during collision-induced dissociation by

malondialdehyde modification of arginine residues. *Journal of mass spectrometry : JMS* **2007,** *42* (7), 950-9.

16. Iliuk, A. B.; Arrington, J. V.; Tao, W. A., Analytical challenges translating mass spectrometry-based phosphoproteomics from discovery to clinical applications. *Electrophoresis* **2014,** *35* (24), 3430-40.

17. Boersema, P. J.; Mohammed, S.; Heck, A. J., Phosphopeptide fragmentation and analysis by mass spectrometry. *J Mass Spectrom* **2009,** *44* (6), 861-78.

18. Marx, H.; Lemeer, S.; Schliep, J. E.; Matheron, L.; Mohammed, S.; Cox, J.; Mann, M.; Heck, A. J.; Kuster, B., A large synthetic peptide and phosphopeptide reference library for mass spectrometry-based proteomics. *Nature biotechnology* **2013,** *31* (6), 557-64.

19. Zhou, M.; Meng, Z.; Jobson, A. G.; Pommier, Y.; Veenstra, T. D., Detection of in vitro kinase generated protein phosphorylation sites using gamma[18O4]-ATP and mass spectrometry. *Anal Chem* **2007,** *79* (20), 7603-10.

20. Xue, L.; Wang, W. H.; Iliuk, A.; Hu, L. H.; Galan, J. A.; Yu, S.; Hans, M.; Geahlen, R. L.; Tao, W. A., Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. *P Natl Acad Sci USA* **2012,** *109* (15), 5615-5620.

21. Xue, L.; Wang, P.; Cao, P.; Zhu, J. K.; Tao, W. A., Identification of extracellular signal-regulated kinase 1 (ERK1) direct substrates using stable isotope labeled kinase assay-linked phosphoproteomics. *Mol Cell Proteomics* **2014,** *13* (11), 3199-210.

22. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **2008,** *26* (12), 1367-1372.

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

24. Xue, L.; Wang, P.; Wang, L.; Renzi, E.; Radivojac, P.; Tang, H.; Arnold, R.; Zhu, J. K.; Tao, W. A., Quantitative measurement of phosphoproteome response to osmotic stress in arabidopsis based on Library-Assisted eXtracted Ion Chromatogram (LAXIC). *Mol Cell Proteomics* **2013,** *12* (8), 2354-69.

25. Okuda, S.; Watanabe, Y.; Moriya, Y.; Kawano, S.; Yamamoto, T.; Matsumoto, M.; Takami, T.; Kobayashi, D.; Araki, N.; Yoshizawa, A. C.; Tabata, T.; Sugiyama, N.; Goto, S.; Ishihama, Y., jPOSTrepo: an international standard data repository for proteomes. *Nucleic Acids Res* **2016**.

26. Marx, H.; Minogue, C. E.; Jayaraman, D.; Richards, A. L.; Kwiecien, N. W.; Siahpirani, A. F.; Rajasekar, S.; Maeda, J.; Garcia, K.; Del Valle-Echevarria, A. R.; Volkening, J. D.; Westphall, M. S.; Roy, S.; Sussman, M. R.; Ane, J. M.; Coon, J. J., A proteomic atlas of the legume Medicago truncatula and its nitrogen-fixing endosymbiont Sinorhizobium meliloti. *Nat Biotechnol* **2016,** *34* (11), 1198-1205.

27. Tsai, C. F.; Wang, Y. T.; Yen, H. Y.; Tsou, C. C.; Ku, W. C.; Lin, P. Y.; Chen, H. Y.; Nesvizhskii, A. I.; Ishihama, Y.; Chen, Y. J., Large-scale determination of absolute phosphorylation stoichiometries in human cells by motif-targeting quantitative proteomics. *Nat Commun* **2015,** *6*.

28. Umezawa, T.; Sugiyama, N.; Takahashi, F.; Anderson, J. C.; Ishihama, Y.; Peck, S. C.; Shinozaki, K., Genetics and Phosphoproteomics Reveal a Protein Phosphorylation Network in the Abscisic Acid Signaling Pathway in Arabidopsis thaliana. *Sci Signal* **2013,** *6* (270).

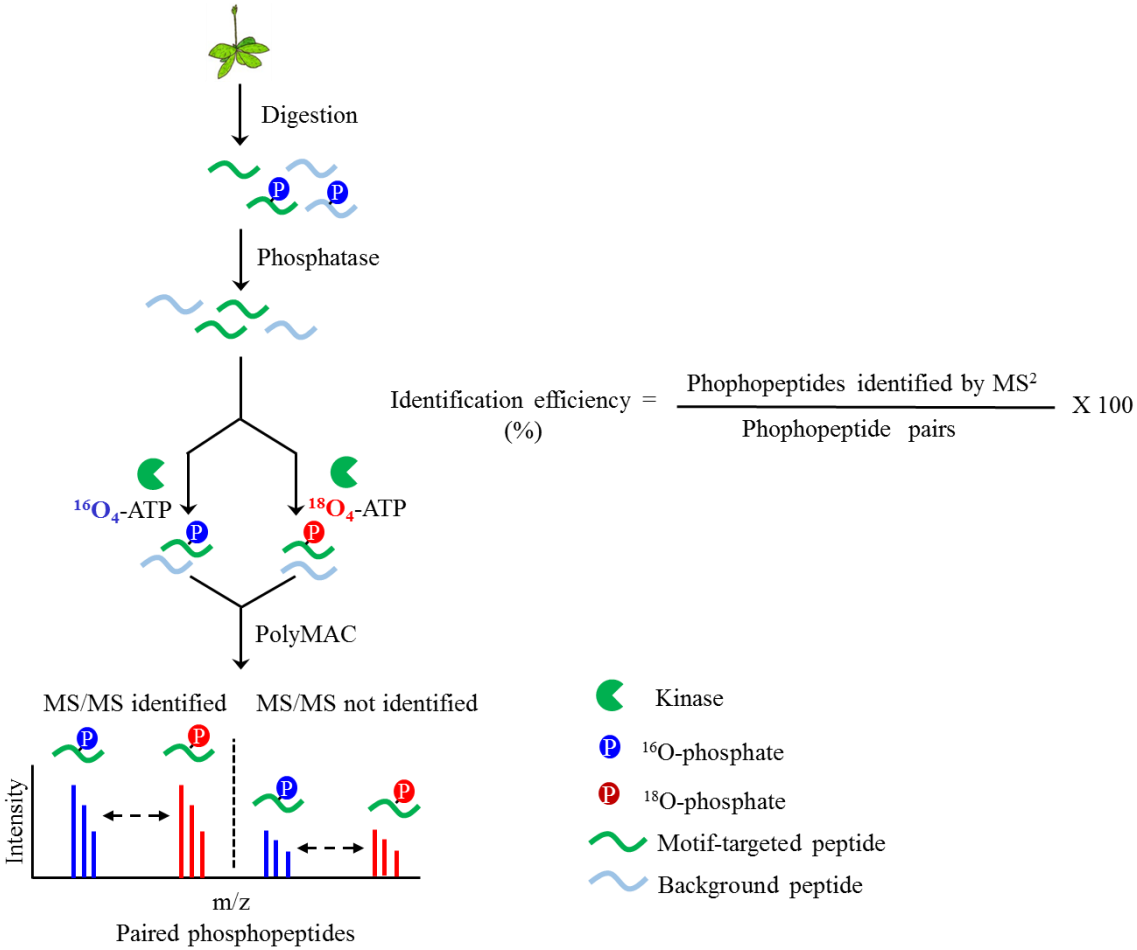
29. Wang, P. C.; Xue, L.; Batelli, G.; Lee, S.; Hou, Y. J.; Van Oosten, M. J.; Zhang, H. M.; Tao, W. A.; Zhu, J. K., Quantitative phosphoproteomics identifies SnRK2 protein kinase substrates and reveals the effectors of abscisic acid action. *P Natl Acad Sci USA* **2013,** *110* (27), 11205-11210.

30. Bleiholder, C.; Suhai, S.; Harrison, A. G.; Paizs, B., Towards understanding the tandem mass spectra of protonated oligopeptides. 2: The proline effect in collision-induced dissociation of protonated Ala-Ala-Xxx-Pro-Ala (Xxx = Ala, Ser, Leu, Val, Phe, and Trp). *J Am Soc Mass Spectrom* **2011,** *22* (6), 1032-9.

31. Chick, J. M.; Kolippakkam, D.; Nusinow, D. P.; Zhai, B.; Rad, R.; Huttlin, E. L.; Gygi, S. P., A mass-tolerant database search identifies a large proportion of unassigned spectra in shotgun proteomics as modified peptides. *Nat Biotechnol* **2015,** *33* (7), 743-9.

Table 1.1: Number of phosphopeptide pairs in MS spectra and phosphopeptides identified by MS^2. MQ is MaxQuant and PD is Proteome Discover.

Kinase	Doublet	Doublet-MQ	Doublet-PD
CK2	4752	1498	2213
MPK6	4053	1472	1816
SnRK2.6	6749	1837	1405



$$\text{Identification efficiency (\%)} = \frac{\text{Phophopeptides identified by MS}^2}{\text{Phophopeptide pairs}} \times 100$$

Figure 1.1: The workflow to estimate efficiency of phosphopeptide identification by tandem mass spectrometry. The ratio of phosphopeptides identified by MS/MS over total paired phosphopeptides identified by MS represents the phosphopeptide identification efficiency.

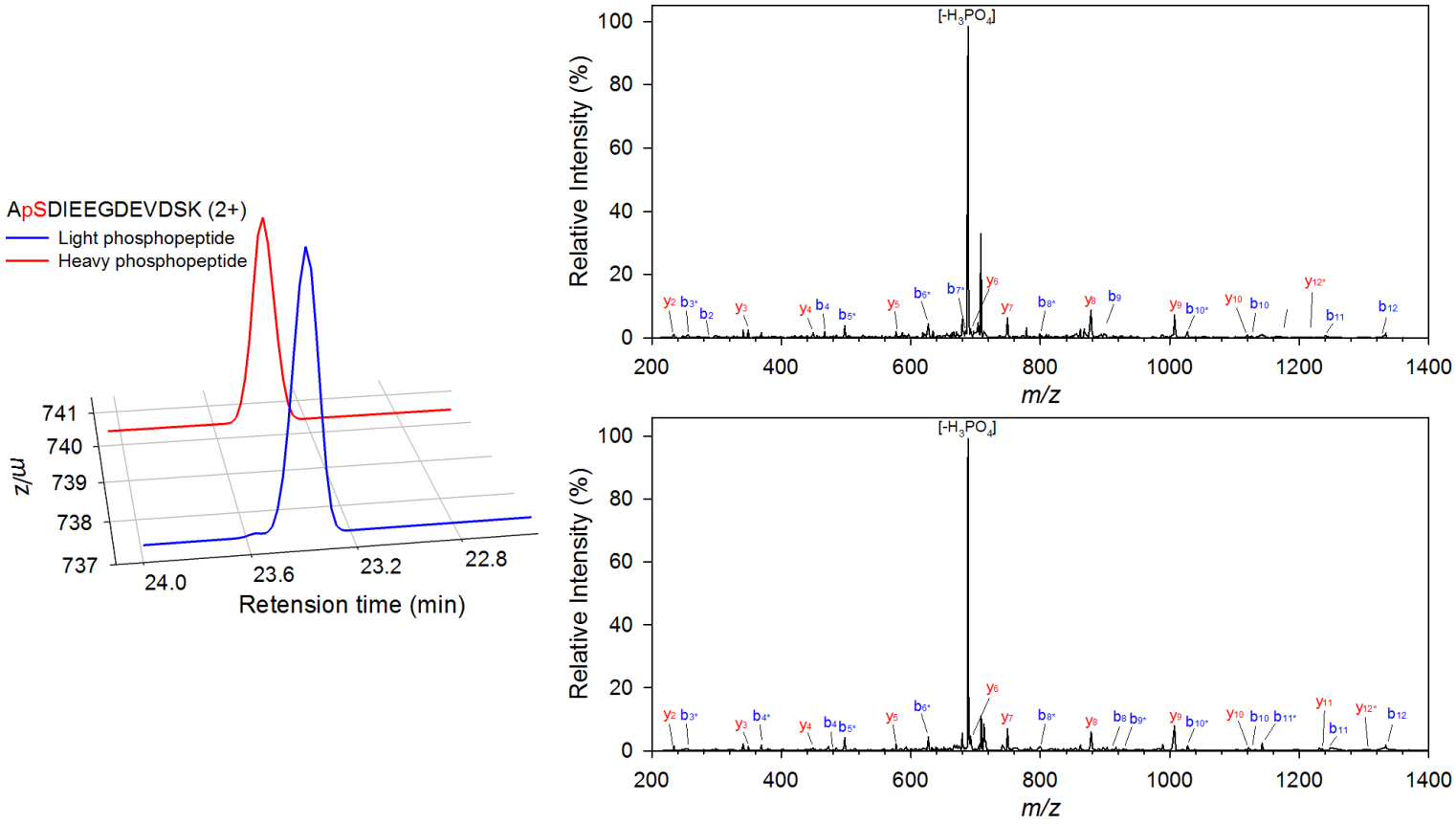


Figure 1.2: Selected examples of extracted ion chromatograms and MS/MS spectra of motif-targeted paired phosphopeptides from three *in vitro* kinase reactions. NUP50 phosphopeptide ApSDIEEGDEVDSK phosphorylated by CK2.

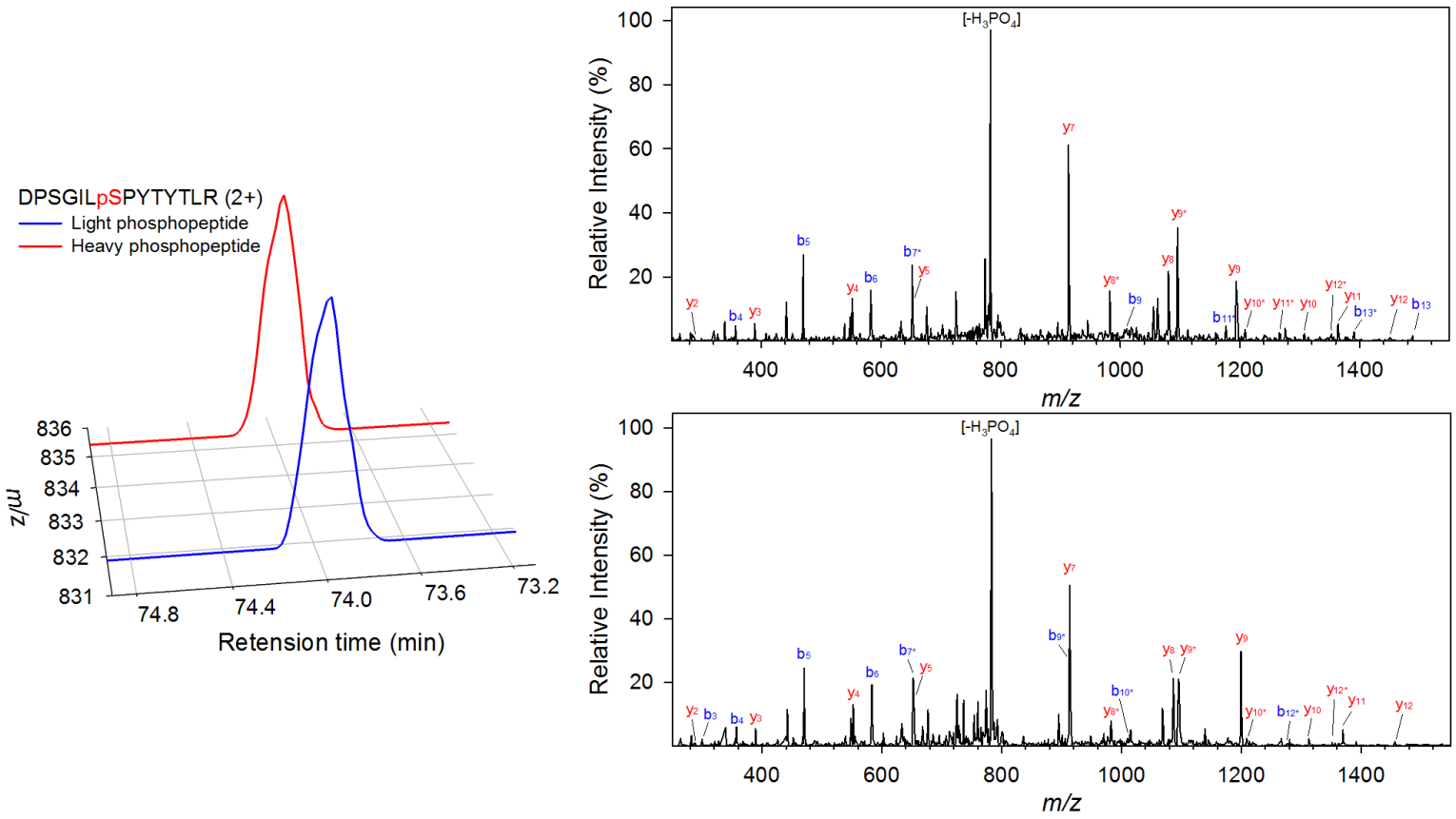


Figure 1.3: Selected examples of extracted ion chromatograms and MS/MS spectra of motif-targeted paired phosphopeptides from three *in vitro* kinase reactions. CAD5 phosphopeptide DPSGILpSPYTYTLR phosphorylated by MPK6.

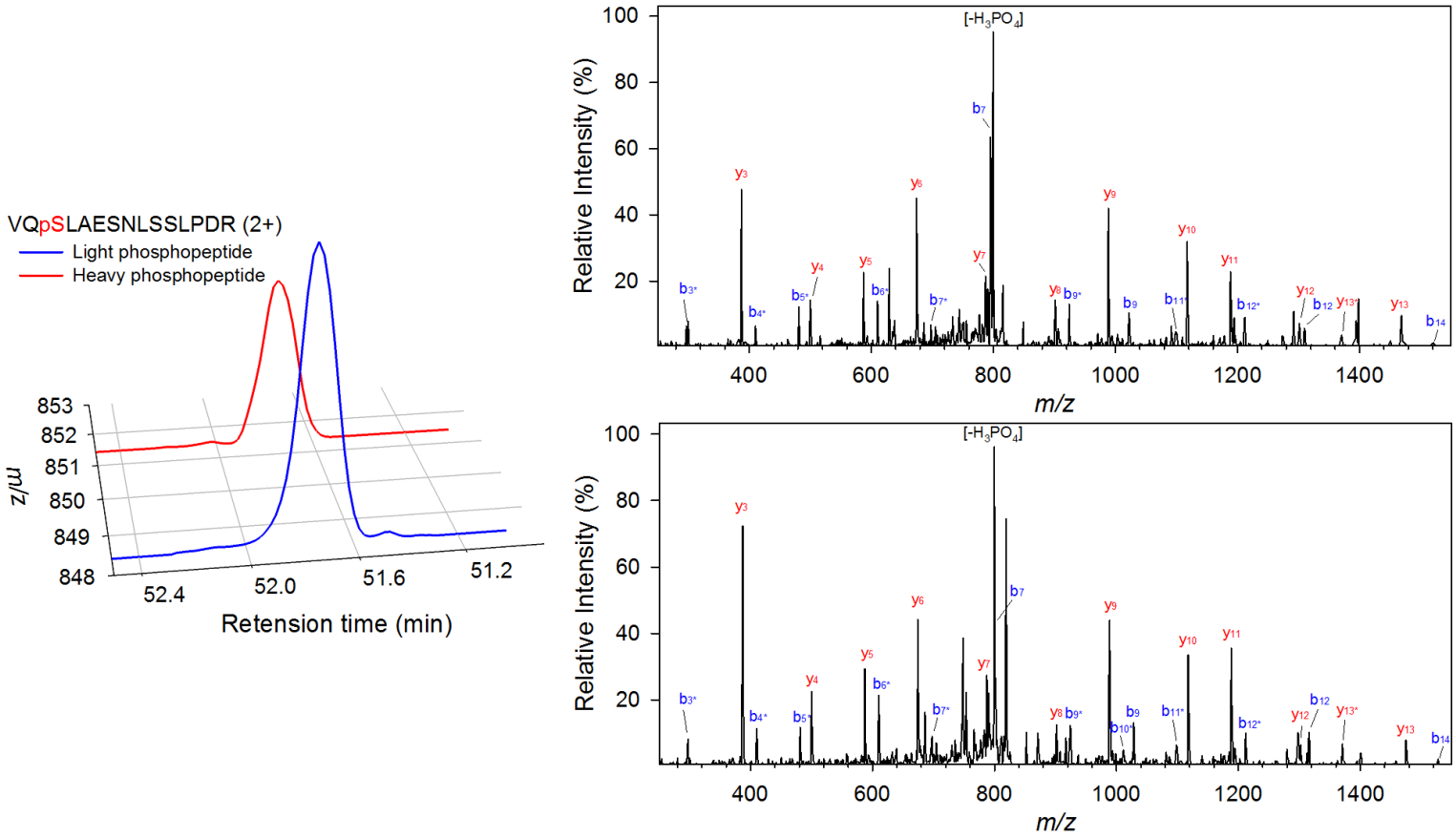


Figure 1.4: Selected examples of extracted ion chromatograms and MS/MS spectra of motif-targeted paired phosphopeptides from three *in vitro* kinase reactions. AT5G05600 phosphopeptide VQpSLAESNLSSLPDR phosphorylated by SnRK2.6.

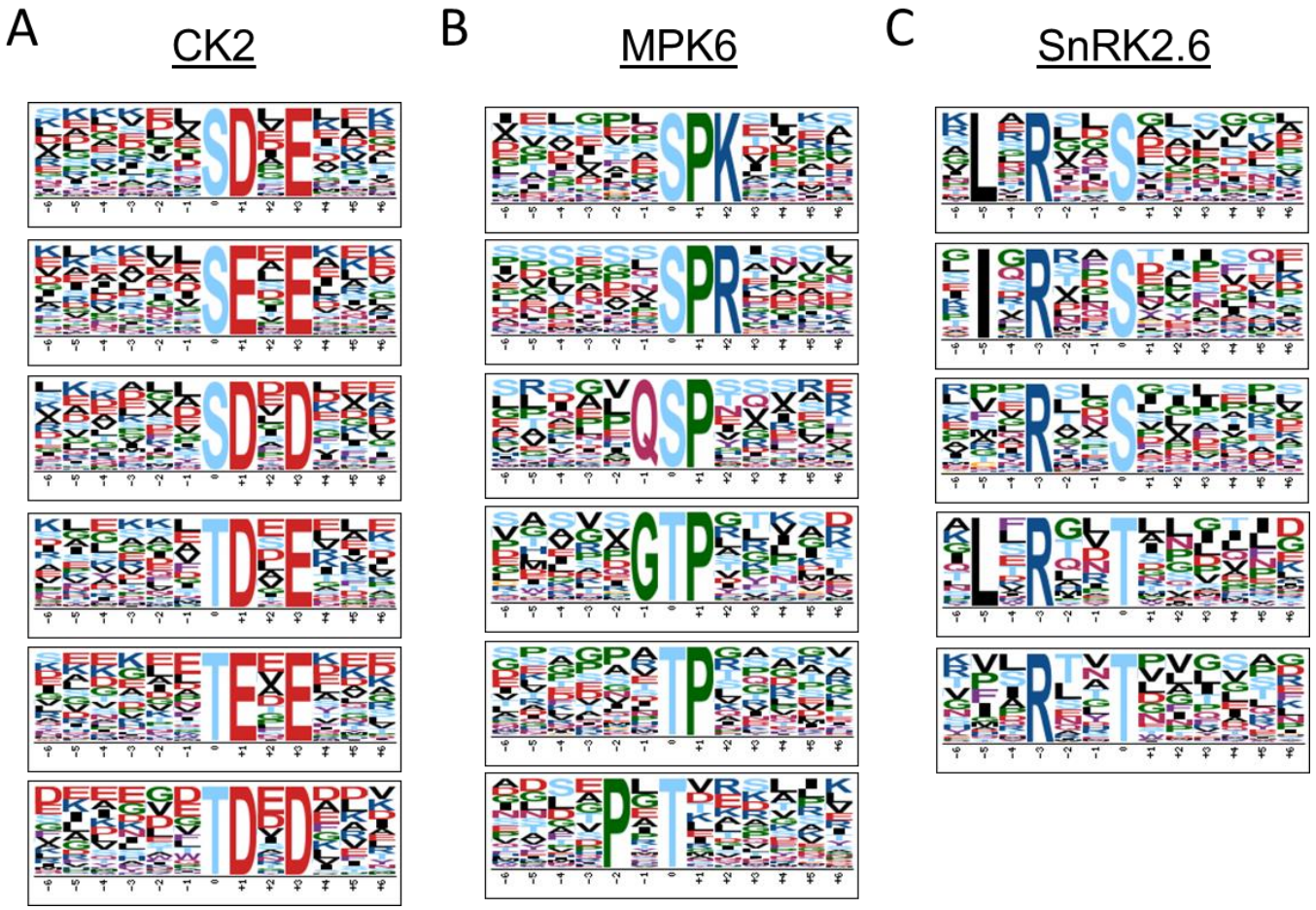


Figure 1.5: Motif analysis of identified light/heavy phosphopeptides. The phosphorylation motifs of identified light/heavy phosphopeptides were extracted by Motif-X from (A) CK2, (B) MPK6, and (C) SnRK2.6 kinase reactions.

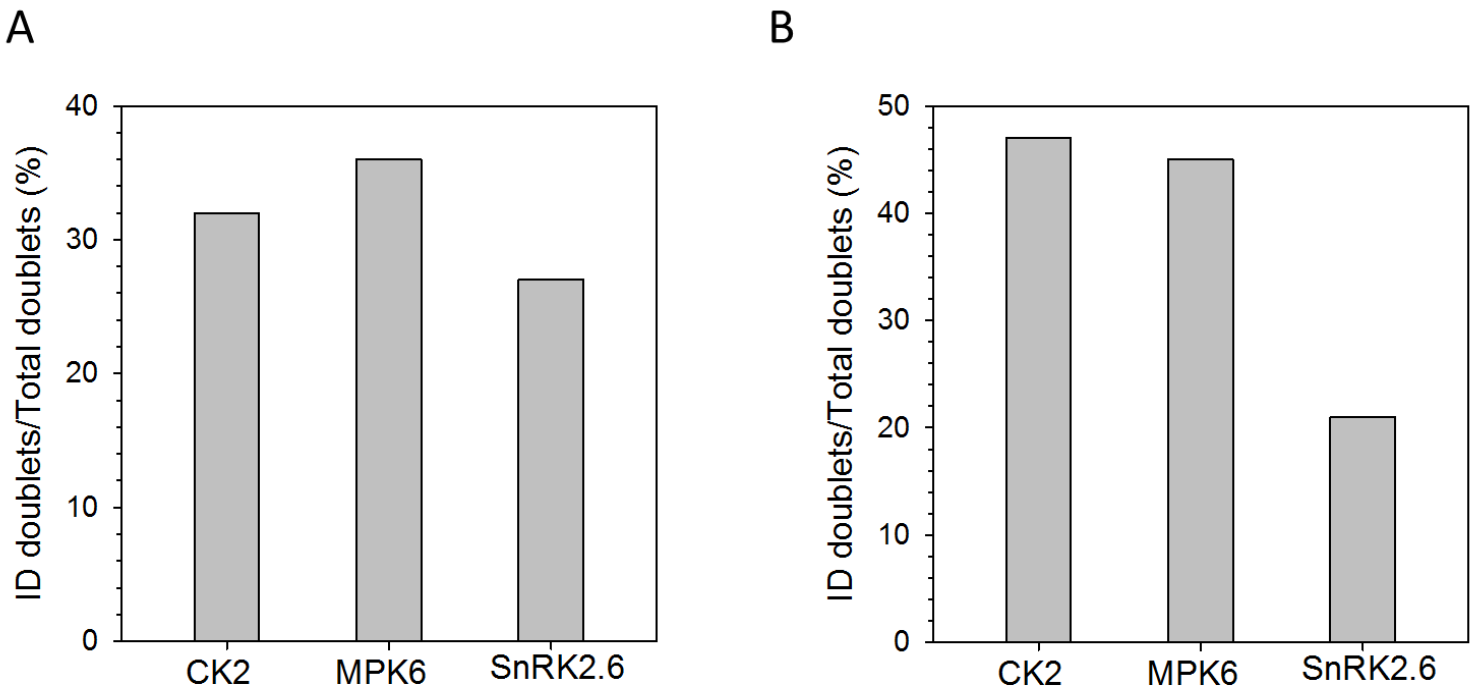


Figure 1.6: Comparison of phosphopeptide identification efficiencies between kinase reactions and search engines. The percentages of identification efficiency from (A) MaxQuant and (B) Sequest search engines were calculated by dividing phosphopeptides identified by MS/MS over total phosphopeptide pairs.

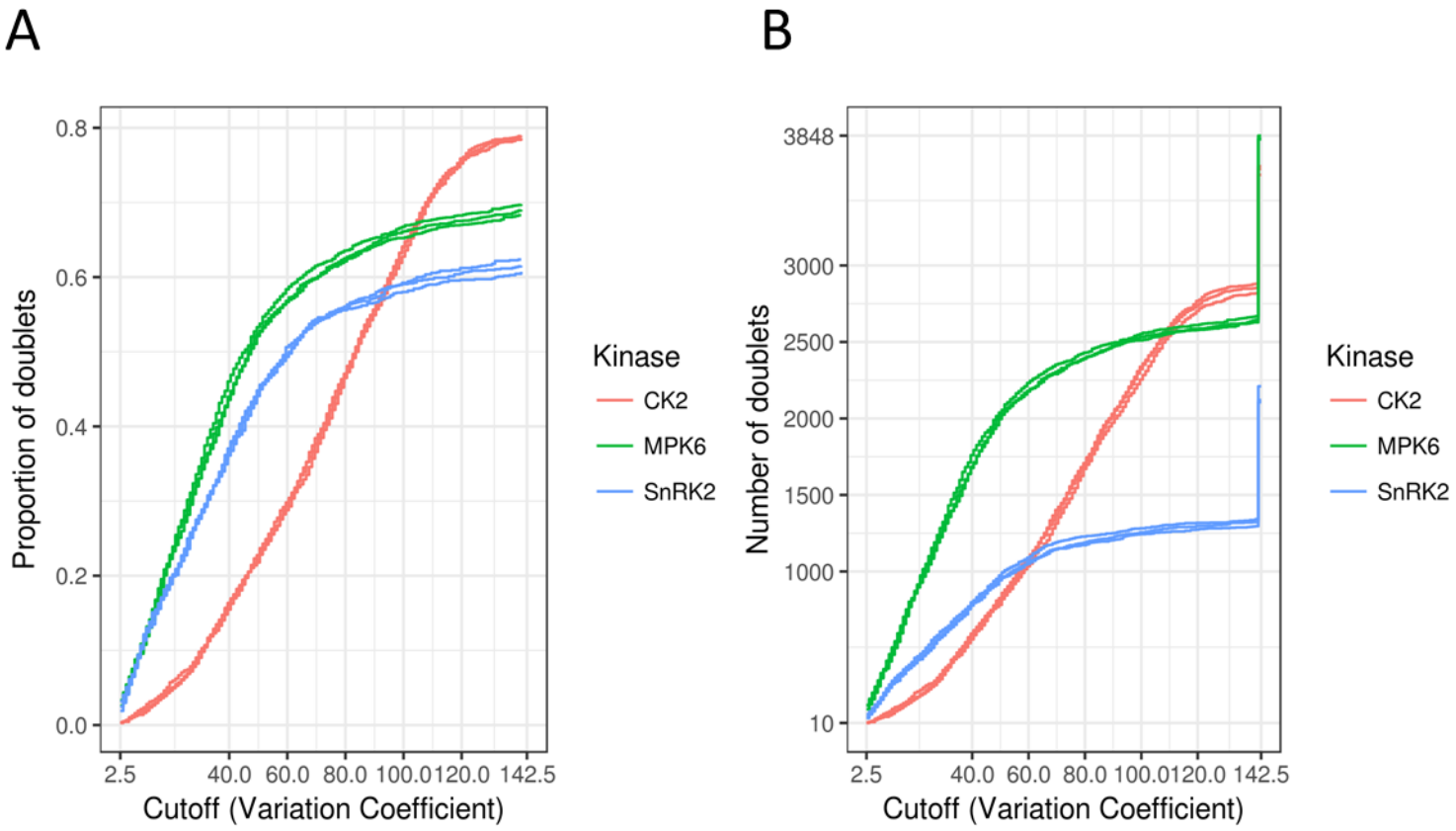


Figure 1.7: (A) Proportion of phosphopeptide doublets versus cut-off value; (B) Number of phosphopeptide doublets versus cut-off value.

CHAPTER 2. THE STREAMLINING PHOSPHOPROTEOMIC WORKFLOW TO STUDY PLANT SIGNALING AND ITS APPLICATION TO TOMATO PHOSPHOPROTEOMICS

2.1 Introduction

Protein phosphorylation is crucial for plant cells in perceiving and responding to environmental stimuli through transducing the signals from receptor kinases to targets[1-2]. Kinases have doubled in number and diversity in plants compared with that in human, implying the importance of plant kinome and phosphoproteome in regulating plant responses under abiotic and biotic stresses[3-4]. Therefore, profiling the phosphoproteomic changes in response to environmental stresses is an approach to understand and delineate the whole view of plant defense mechanisms. Mass spectrometry (MS) has been emerged as a power technology for identifying thousands of phosphorylation sites in a single shot from mammalian cells[5-6]. However, significantly analytical difficulty is encountered in plant phosphoproteomics due to the high dynamic range of proteome, toughness of plant cell walls, and the interference from chlorophyll and secondary metabolites[7-9]. These challenges hamper the sensitivity and reproducibility of detecting low abundant phosphorylation events upon outside stimuli.

Cold stress is a major environmental factor affects the growth, distribution of plants, and determines the yield of many important crops growing in tropical or subtropical area[10-12]. Under prolonged cold environment, plant cells alter thousands of genes expression to reach a cold acclimation status. Many important transcription factors are enhanced in regulating plant cold acclimation[10, 13]. For example, the increase of ICE1 expression plays a role in the modulation of cold-responsive genes (CORs) such as three *CBF* genes expression under cold stress in Arabidopsis[14]. The *ice1* mutant plant reduced the expression of *CBF* genes and displayed the cold sensitive phonotype[14]. Many kinases are activated and in the positive regulation of plant freezing tolerance at the posttranslational level. One of the important events is the activation of MEKK1-MKK2-MAPK4/6 cascades under short period of cold treatment, which has been linked to enhance freezing tolerance[15-16]. Another example is OST1, one of the core components of ABA pathway, which modulates

ICE1 protein turnover through phosphorylation at Ser278[17]. Phosphorylation of this site prevents the degradation of ICE1 protein under cold stress, which promotes cold tolerance in Arabidopsis. These examples suggest that the importance of plant phosphoproteome and kinome in the regulation of molecular events that trigger cold acclimation.

Tomato is the second most important horticulture crops, and recently we have started to utilize our proteomic approach to investigate the underlying mechanism of their cold tolerance. Considering the limited knowledge of molecular mechanism of tomato cold tolerance, tomatoes with distinct cold tolerances are suitable as the model plant to study the underlying mechanism of cold tolerance. Literatures have compared the cold tolerances of tomatoes in the view of transcriptome[18-19]. Distinct gene expression patterns were observed from different cold tolerant tomatoes, indicating the high complexity of cold-induce molecular mechanisms. However, there has been no large-scale study for characterizing the relation of cold tolerance with the regulation of tomato phosphoproteomics. We recently found that a cultivated tomato, *S. lycopersicum* (cultivar, N135 Green Gage), shows tolerance under chilling condition (4 ºC), while a wild tomato, *S. pimpinellifolium* (PIM, Atacames), are more sensitive to chilling stress.

To address the limits in accessing the whole view of plant phosphoproteomics, we develop an optimized sample preparation protocol specific for plant phosphoproteomics, which significantly increased the coverage of plant phosphoproteome. This protocol has been applied to study the phosphoproteomic perturbation of two tomato varieties under prolonged cold treatment. This in-depth phosphoproteomic resource reveals the phosphorylation sites implicated in kinase activation and cold-responsive gene expression. Coupling the data of in vitro kinase screening, we discovered a connection of SnRK2s activation and cold tolerance through phosphorylating their downstream kinases, which sheds lights on identification of critical tomato phosphoproteins conferring the cold tolerance.

2.2 Experimental Procedures

Unless otherwise noted, reagents were purchased from Sigma (St. Louis, MO), and buffers were prepared with water purified to a resistivity of 18.2 MΩ cm-1 using a Barnstead Nanopure water system (Thermo Fisher).

2.2.1 Plant Materials and Cold Treatment

Tomato were grown in green house at Purdue University. Tomato seeds were first germinated in soil and seedlings with identical growth were transferred to separate pots. Four-week-old tomato plants (about 30 cm height) were subjected to mock or cold treatment (4 ºC) for 5 days. Equal amount of leaf tissue from cold tolerance and sensitive tomato varieties were collected for phosphoproteomics analysis.

2.2.2 Electrolytic Leakage Measurement

The electrolyte leakage (EL) was measured as follows: at least three of fully expanded leaves from each tomato varieties were detached and immersed in 50 ml tubes with 25 ml distilled water. After gentle shaking overnight, The initial electrolyte conductivity (E1) was measured with conductivity meter. Then the samples were autoclaved. After cooling to room temperature, the conductivity (E2) was measured again. The relative electrolyte leakage was calculated as: E1/E2×100.

2.2.3 Protein Extraction

Ground tomato leaves were lysed in (1) Tris-HCl protocol: 100 mM Tris-HCl (pH 7.5), 250 mM NaCl, and 5 mM EDTA[20], (2) SDC-SLS protocol: 100 mM Tris-HCl (pH 8.5), 12 mM sodium deoxycholate (SDC), and 12 mM sodium lauroyl sarcosinate (SLS)[21], or (3) GdnHCl protocol: 100 mM Tris-HCl (pH 8.5) and 6 M guanidine chloride (GdnHCl) with EDTA-free protease and phosphatase inhibitor cocktail (Sigma)[22]. Proteins were reduced and alkylated with 10 mM tris-(2-carboxyethyl)phosphine (TECP) and 40 mM chloroacetamide (CAA) at 95 °C for 5 min. Protein amount was quantified by BCA assay (Thermo Fisher Scientific).

2.2.4 Filter Aided Sample Preparation (FASP)

The 200 µg of lysate was loaded into a Microcon 10k Da (Millipore) filter tube using the FASP protocol[23]. The GdnHCl buffer was replaced by 100 µl of 8 M urea buffer, and proteins were reduced by 5 mM DTT for 30 min and then alkylated by 15 mM IAA for 30

min in the dark. The urea buffer was replaced by 100 µl 50 mM ammonium bicarbonate (ABC). Lys-C was added into the filter with a 1:100 (w/w) ratio for 3 hours at 37 °C, and trypsin was subsequently added to a final 1:100 (w/w) ratio overnight. The digests were collected with two times of 100 µl 50 mM ABC buffer, and peptides were acidified with 10% trifluoroacetic acid (TFA) to a pH ~3, and desalted using a 50 mg of Sep-pak C18 column.

2.2.5 Methanol-Chloroform Precipitation

The protein precipitation was performed as previously described with some modifications[24]. The 200 µg of lysate was added four volumes of methanol, followed by an equal volume of chloroform with mixing. Three volumes of of ddH_2O was added to a tube with mixing. The solution was centrifuged at 16,000 g for 3 min. The upper aqueous layer was removed and was added four volumes of methanol and the tube was centrifuged again. The supernatant was discarded, and air dried the precipitated protein pellet.

2.2.6 Protein Digestion

Protein extracts (200 µg) were 5-fold diluted for the SDC-SLS protocol[25], 8-fold diluted for the urea protocol, or 10-fold diluted for the GdnHCl protocol using 50 mM TEAB and digested with Lys-C (Wako) in a 1:100 (w/w) enzyme-to-protein ratio for 3 hours at 37 °C, and trypsin (Sigma) was added to a final 1:100 (w/w) enzyme-to-protein ratio overnight. The SDC and SLS were removed by adding 100% acetyl acetate as previous reported. Digests were acidified with 10% trifluoroacetic acid (TFA) to a pH ~3, and desalted using a 50 mg of Sep-pak C18 column (Waters).

2.2.7 Dimethyl Labeling

Dimethyl-labeling experiment was performed as previous described[26]. The tryptic peptides were dissolved in 100 µl of 100 mM TEABC, and were mixed with 4 µl of 4%$^{13}CD_2O$ or $^{12}CH_2O$, and then 4 µl of freshly prepared 0.6 M sodium cyanoborohydride was immediately added. The mixture was agitated for 60 min at room temperature. The reaction was stopped by adding 16 µl of 1% ammonium hydroxide on ice and agitating the

mixture for 1 min. Then dimethyl labelling peptides were mixed with 20 µl of 10% formic acid, and the two labeled samples were mixed and desalted using the Sep-pak C18 column.

2.2.8 PolyMAC Enrichment

Phosphopeptides were enriched using PolyMAC-Ti kit (Tymora Analytical)[27]. Briefly, the digested peptides were resuspend with 20 µl of ultrapure water, and 200 µl of Loading buffer was added to the sample and vortex. Added 50 µl of the PolyMAC/Magnetic Capture beads to the sample and incubated for 20 min. The solvent was removed using the magnetic separator rack, and added 200 µl of Washing buffer 1 to the beads and incubated for 5 minutes. Incubated 200 µl of Washing buffer 2 with the beads for 5 minutes and removed the solvent. Incubated the beads twice with 100 µl of Elution buffer for 5 minutes, and dried the eluents completely in a SpeedVac.

2.2.9 Basic pH Reverse-phase Fractionation

Basic pH reverse-phase fractionation was performed with some modifications as previously described[28]. 2 mg of the C18-AQ beads (5 µm particles) were suspended in 100 µL of methanol and loaded into a 200 µL of StageTip with a 20 µm polypropylene frit. The C18 StageTips were activated with 100 µL of 40 mM NH_4HCO_2, pH 10, in 80% ACN, and equilibrated with 100 µL of 200 mM NH_4HCO_2, pH 10. The isolated phosphopeptides were suspend with 200 mM NH_4HCO_2, and the C18 StageTips were washed with 100 µL of 200 mM NH_4HCO_2, pH 10. The bound phosphopeptides were fractionated from the StageTip with 50 µL of 8 different ACN concentrations: 5%, 9%, 13%, 17%, 21%, and 80% of ACN in 200 mM NH_4HCO_2, pH 10. The eluted phosphopeptides were dried and stored at -20 °C.

2.2.10 In vitro Kinase Reaction

The in vitro kinase reaction was performed based on the siKALIP approach with some modifications[29]. The 200 µg of Lys-C digested peptides were dephosphorylated using thermosensitive alkaline phosphatase (TSAP) (Roche) in a 1:100 (w/w) enzyme-to-peptides ratio at 37 °C overnight. The reaction was terminated by heating the sample at 75 °C for 10 min. The dephosphorylated peptides were desalted using the Sep-pak column,

and suspended in kinase reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, and 1 mM γ-[$^{18}O_4$]-ATP, pH 7.5). The recombinant SnRK2E (1 µg) was incubated with the desalted peptides at 25 °C overnight. The kinase reaction was quenched by acidifying with 10% TFA to a final concentration of 1%, and the peptides were desalted by the Sep-pak C18 column. The heavy ^{18}O-phosphopeptides were further digested by trypsin at 37 °C for 6 h and enriched by polyMAC-Ti reagent mentioned above, and the eluates were dried in SpeedVac for LC-MS/MS analysis.

2.2.11 LC-MS/MS Analysis

The peptides were dissolved in 5 µL of 0.3% formic acid (FA) with 3% ACN and injected into an Easy-nLC 1000 (Thermo Fisher Scientific). Peptides were separated on a 45 cm in-house packed column (360 µm OD × 75 µm ID) containing C18 resin (2.2 µm, 100Å, Michrom Bioresources) with a 30 cm column heater (Analytical Sales and Services) set at 50 °C. The mobile phase buffer consisted of 0.1% FA in ultra-pure water (buffer A) with an eluting buffer of 0.1% FA in 80% ACN (buffer B) run over a linear 60 min (method comparisons) or 90 min (large-scale phosphoproteomics) gradient of 5%-30% buffer B at a flow rate of 250 nL/min. The Easy-nLC 1000 was coupled online with a Velos Pro LTQ-Orbitrap mass spectrometer (Thermo Fisher Scientific). The mass spectrometer was operated in the data-dependent mode in which a full MS scan (from m/z 350-1500 with the resolution of 30,000 at m/z 400) was followed by the 10 most intense ions being subjected to collision-induced dissociation (CID) fragmentation. CID fragmentation was performed and acquired in the linear ion trap (normalized collision energy (NCE) 30%, AGC 3e4, max injection time 100 ms, isolation window 3 m/z, and dynamic exclusion 60 s).

2.2.12 Data Processing

The raw files were searched directly against *Solanum lycopersicum* database from the Sol Genomics Network (http://solgenomics.net) version 3.0[30] with no redundant entries using MaxQuant software (version 1.5.5.1)[31] with 1% FDR cutoff at protein, peptide, and modification levels. The first peptide precursor mass tolerance was set at 20 ppm, and MS/MS tolerance was set at 0.6 Da. Search criteria included a static carbamidomethylation

of cysteines and variable modifications of (1) oxidation on methionine residues, (2) acetylation at N-terminus of protein, and (3) phosphorylation on serine, threonine or tyrosine residues were searched. For ^{18}O-phosphopeptides, heavy phosphorylation (+85.979 Da) was also set as a variable modification. Search was performed with full tryptic digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. Dimethyl-labeling quantitation was performed using MaxQuant by setting the multiplicity as 2, and the match between runs function was enabled. The false discovery rates of proteins, peptides and phosphosites were set at 0.01. The minimum peptide length was six amino acids, and a minimum Andromeda score was set at 40 for modified peptides. A site localization probability of 0.75 was used as the cut-off for localization of phosphorylation sites.

2.2.13 Data Analysis

All data were analyzed by using the Perseus software (version 1.5.6.0)[32]. The intensities of class I phosphorylation sites were extracted through MaxQuant, and the missing values of intensities were replaced from a normal distribution (width = 0.2 and down shift = 1.8) for statistical analysis. The significantly enriched phosphorylation sites were selected by the ANOVA test with a permutation-based FDR cut-off 0.05 with S0 was set at 0.2. Normalization was carried out by subtracting the medium of log2 transformed phosphopeptide intensities. All the localized and significantly changed phosphorylation sites were submitted to Motif-X[33] to determine the enriched phosphorylation motifs with ITAG 3.0 database as background. The significance was set at 0.000001, the width was set at 13, and the number of occurrences was set at 20.

2.3 Results

2.3.1 Comparison of Sample Preparation Protocols for Plant Phosphoproteome Analysis

Optimization of the sample preparation protocol for plant samples is important to enlarge the coverage of plant phosphoproteome and to dissect the signaling transductions

of plant cells. To improve the sensitivity of analyzing plant phosphoproteome, it is necessary to evaluate the factors during sample preparation that affect phosphopeptide identification such as the efficiency of protein extraction, removal of the interferences, and the activity of enzymatic digestion (Figure 2.1). One of the major challenges is the plant cell walls reduce the efficiency of plant protein extraction compared to mammalian cells. Many literatures have reported the use of Tris-HCl with salts[20], Tris-HCl with 30% sucrose[34], or SDS coupling with FASP[35] for extraction proteins from plant tissues. However, these results showed the varying identification efficiencies and yields. It is necessary to improve the lysis condition of mature plant tissues. Guanidine hydrochloride (GdnHCl) has been emerged as an efficient approach to extract RNA from plant tissues[36], so we hypothesized that GdnHCl is efficient to lyse the plant cell walls as well. To compare the efficiency of the lysis protocols for mature plant tissues, we have tested three lysis protocols: (1) Tris-HCl protocol, (2) SDC-SLS protocol, and (3) GdnHCl protocol (see methods) using matured tomato leaves (30 days). The 200 µg of proteins extracted from each protocol were diluted five folds and subsequently performed tryptic digestion. The phosphopeptides were enriched by PolyMAC and analyzed using 1 h gradient LC-MS/MS. Surprisingly, we observed that the GdnHCl protocol doubles the number of identified phosphopeptides compared with the other protocols, indicating that 6M GdnHCl provides better lysis efficiency for breaking the cell wall of plant cells (Figure 2.2A). By combining triplicate analyses, GdnHCl protocols identified more than 4,000 phosphopeptides relative to the other protocols (Figure 2.2B), which significantly enlarge the extent of tomato phosphoproteome. We further performed gene ontology (GO) tern analysis to access the significantly enriched cellular component categories of GdnHCl protocol from background using Fisher exact test (FDR < 0.05). The results show that the enriched cytosolic and organelle phosphoproteins using the GdnHCl protocol are overrepresented than that of the other protocols, indicating that using GdnHCl protocol extracted more intracellular phosphoproteins than the other protocols (Figure 2.3).

Another crucial factor that affects the phosphopeptide enrichment is that plant tissues generate many photosynthetic pigments and secondary metabolites. These interferences may increase the sample complexity by resulting in low phosphopeptide purification efficiency and specificity. Furthermore, it is difficult to separate these interferences from

digested peptides using C18 cartridge because most of them has similar hydrophobicity as digested peptides, suggesting the necessities to eliminate these interferences before enzymatic digestion. To alleviate the interfere of these compounds, we have compared the performance of Filter Aided Sample Prep (FASP) and methanol-chloroform precipitation to clean the unwanted interferences before tryptic digestion (see methods). We found that using 10k Da of FASP filter tube to filter out the interferences barely improves the number of identified phosphopeptides (Figure 2.4A), which may be attributed to the high sample loss during the numerous washing steps of FASP protocol. Performing methanol-chloroform precipitation increases the number of identified phosphopeptides by 20% compared with the direct digestion protocol (Figure 2.4A), demonstrating that it is necessary to clean the interferences of plant tissues prior to phosphopeptide enrichment.

The other step that influences phosphopeptide identification is the activity of enzymatic digestion. Masuda et al. reported that the SDC-SLS protocol significantly improves the phosphopeptide identification compared to the urea protocol using very small amount of cell lysate[21]. To evaluate the effect of digestion buffer on phosphopeptide identification, we compared two digestion protocols: one is the urea protocol and the other is the SDC-SLS protocol (see methods). The precipitated protein pellets were suspended in the two buffers with several pulses of sonication, and performed dilution for subsequently enzymatic digestions. The identification result shows that using SDC and SLS as surfactants improves the number of identified phosphopeptides compared to urea protocol with much lower percentage of missed cleavage, meaning that the enzymatic activities are enhanced in the SDC-SLS buffer compared to the urea digestion buffer (Figure 2.4B-C). Overall, using this optimized protocol enlarges the depth of tomato phosphoprteome more than two-fold, and around 8,100 unique phosphopeptides were identified from triplicate single-shot analyses of the matured tomato leaves.

2.3.2 Profiling of Tomato Phosphoproteomic Changes under Cold Stress

After optimization of the protocol, we applied this protocol to study the tomato global phosphoproteomic changes in response to cold stress and the underrepresented signaling mechanisms involved in cold tolerance. We compared two tomato varieties: one is Atacames (PIM, *Solanum pimpinellifolium*,), belonging to wild species, and it displayed

cold-sensitive phenotype. The other is the cultivated tomato N135 Green Gage (*Solanum lycopersicum*), which showed relatively cold-tolerant phenotype. To demonstrate their tolerance under prolonged cold environment, especially under chilling temperature, two tomato varieties were germinated in soil for about one month and then were transferred to 4 °C for five days or mock conditions (room temperature). It was shown that both tomato varieties grew well under room temperature. However, after 5 days of cold treatment (4 ºC), a few leaves in N135 Green Gage *(S. lycopersicum)* exhibited slight wilting symptom, while an aggravated wilting of leaves in Atacames *(S. pimpinellifolium*) was observed (Figure 2.5A). We further measured the degree of electrolytic leakage (EL) from the leaves of two tomato varieties, which represents the damage of the plasma membrane and the sensitivity of cold stress. In the room temperature, two tomato varieties had similar electrolytic leakage values (~ 6.6%). However, the EL value of *S. pimpinellifolium* was significantly increased to ~60% after 5 days of 4 °C treatment, which is about three-fold higher than that of *S. lycopersicum* (~16%) (Figure 2.5B). Consistent with the overserved cold phenotypes, the tomato *CBF3* expression is expressed at higher levels in N135 Green Gage than that in Atacames after cold treatments (Figure 2.5C). These findings suggest that Atacames (*S. pimpinellifolium*) was more sensitive to cold stress when compared to N135 Green Gage (*S. lycopersicum*). These data prompted us to compare the phosphoproteomic changes between those two tomato varieties to uncover the underrepresented signal transductions in surviving under prolonged cold stress.

To understand the underlying signals involved in cold stress tolerance, we performed a mass-spectrometry-based experiment to study the phosphoproteomic changes of the leaves of S. lycopersicum and S. pimpinellifolium treated with mock or 4 °C for five days (Figure 2.6). Each experiment was performed in biological triplicates. We used the optimized protocol for protein extraction, precipitation, and tryptic digestion. The digested peptides (from 1 mg protein) were dimethyl-labeled and all labeled peptides were combined. The pooled phosphopeptides were enriched using PolyMAC, and the purified phosphopeptides were fractionated into six fractions by basic pH reverse-phase StageTips and analyzed by a 2 h LC-MS/MS on a Velos Pro mass spectrometer. Using this optimized workflow, we identified ~ 5,500 phosphoproteins and ~30,500 unique phosphopeptides corresponding to ~23,000 unique phosphorylation sites (~14,500 class I sites, Figure 2.7A

and Table 2.1). Among all the class I phosphorylation sites, ~ 85% is localized at serine residues, and ~1.5% is at tyrosine residues (Figure 2.7B). To our knowledge, this is the first large-scale phosphoproteomic study which achieves the widest coverage of tomato phosphoproteome.

To evaluate the qualities of biological replicates of each sample, the principal component analysis (PCA) shows that all biological replicates are tightly together and each sample is well-separated, meaning the samples cluster by the nature of sample but not the batch (Figure 2.8A). The PCA analysis also indicates that two tomato varieties have distinct phosphoproteomic profiles (50%, PC1) and cold stress rewired the signal transductions of two tomato varieties (~30%, PC2). To gain insight the phosphoproteomic changes in response to cold stress in the two tomato varieties, we performed ANOVA test with the control of FDR < 0.05 to access the significantly regulated phosphorylation sites among four samples. Unsupervised hierarchical clustering further confirmed the PCA analysis that two major clusters are composed of the replicates of two tomatoes, respectively (Figure 2.8B). The heat map reveals five clusters of 1,159 most changed phosphorylation sites (Figure 2.9). Cluster 1 contains the phosphorylation sites which are relatively activated in *S. pimpinellifolium* (PIM, Atacames) under cold stress. The phosphorylation motif analysis ($p < 0.000001$) shows that [-pS-P-] motif is dominant in this cluster (Figure 2.10), indicating the cold-induced kinases recognize [-pS-P-] motif are crucial in the signaling of *S. pimpinellifolium* (PIM, Atacames). Cluster 2 is the induced phosphorylation sites under cold stress in both tomato varieties, and the [-R-X-X-pS-], [-pS-P-], [-L-X-X-X-X-pS], and [-G-pS-] motifs are enriched in this cluster (Figure 2.10). Cluster 3 and 4 are the up-regulated phosphorylation sites in *S. lycopersicum* (cultivar, N135 Green Gage) under cold condition. We found one acidic ([-pS-D-]) and a basic phosphorylation motifs ([-R-X-X-pS-]) are enriched in cluster 3 and 4 which are not identified from the cluster 1, revealing these proteins might be involved in cold-tolerant mechanisms in *S. lycopersicum* (Figure 2.10). Cluster 5 shows the repressed phosphorylation sites in two tomatoes under cold environment. Since many phosphorylation sites were significantly regulated in response to cold stress, implying the activities of kinases and phosphatases were perturbated by cold stress as well.

2.3.3 Kinase-Mediated Molecular Events under Cold Stress

Considering the phosphorylation on the kinases plays an important role in mediating kinase activities, we quantified the expression levels of 292 kinases corresponding to 669 phosphorylation sites in our large-scale phosphoproteomic data (Figure 2.11). Interestingly, we found that two tomato varieties have unique phosphorylation expression pattern in terms of kinome. For example, in *S. pimpinellifolium* (PIM, Atacames), many phosphorylation sites in the activation loop of mitogen-activated protein kinase (MAPK) cascades such as MAPK 1/2 which is the homolog of AtMAPK3/6 (pT221 and pY223), MAPK 9 which is the homolog of AtMAPK1 (pT191 and pY193), MAPK 14 (pY182), and MAPK (Solyc06g068990.3.1) (pY296) are highly activated, confirming that MAPKs are continuedly activated in *S. pimpinellifolium* under prolonged cold stress. The activation of MAPKs elucidates that the reason of the [-pS-P-] motif is more enriched in the cluster 1. It is therefore possible that *S. pimpinellifolium* have the same signal mechanisms as Arabidopsis in which cold stress is known to activate the activation loop of MKK4/5-MAPK3/6 and MKK1/2-MAPK4/6 to mediate cold tolerance[15]. In our data, we also identified two sites (pS436 and pS440) with [-pS-P-] motif on bHLH transcription factor 87 (Solyc06g068870.3.1) were highly phosphorylated after 5 days of cold treatment (Supplementary figure 3A). bHLH transcription factor 87 is the homolog protein of AtICE1 in Arabidopsis and this transcription factor is one of the substrates of MAPK3/6 in cold-treated Arabidopsis, indicating the importance of MAPKs signaling in *S. pimpinellifolium* under cold environment.

To identify the kinases responsible for the enriched acidic and basic phosphorylation motifs in cold-treated *S. lycopersicum* (cluster 3 and 4), we searched the phosphorylation sites on the kinases up-regulated in *S. lycopersicum* in response to cold treatment. We found that several phosphorylation sites on the casein kinases such as pS355 on the casein kinase family protein (Solyc07g040720.3.1) which is the homolog of AtCKL2, pS310 on the casein kinase (Solyc09g066440.3.1) which is the homolog of AtCKL6, and pS347 on the casein kinase family protein (Solyc12g007280.2.1) which is the homolog of AtCKL4 were significantly increased (> two-fold) in cold-treated *S. lycopersicum* (Figure 2.11).

Interestingly, we observed that cold stress highly induced two phosphorylation sites on two kinase family proteins (pS154 of Solyc01g103940.3.1 and pS173 of

Solyc01g108280.3.1) in *S. lycopersicum* which are the homologs of SnRK2A and SnRK2E in Arabidopsis, respectively (Figure 2.11). It has been reported that SRK2s autophosphorylate themselves in the activation loop to activate their kinase function and to phosphorylate many of their downstream substrates in response to ABA responses and osmotic stresses in Arabidopsis. Furthermore, we also observed two phosphorylation sites on the reported SRK direct substrates in Arabidopsis were only identified in *S. lycopersicum*. One is pS58 of abscisic acid responsive elements binding factor 2 (ABF2), which is an important transcription factor involving in activating ABA responsive gene expression in Arabidopsis. The other is pS878 of enhancer of mRNA decapping protein 4 (VCS) which has the [-R-X-X-pS-] phosphorylation motif, and VCS is one of the components of mRNA decapping complex[37]. It has been reported that VCS is a substrate of SnRK2s in Arabidopsis (Nature Plants 3, Article number: 16204 (2017)). These findings are consistent with the motif analysis that the [-R-X-X-pS-] which is the AtSRK recognition motif was enriched in the cluster 4. SRK2s were more activated in *S. lycopersicum* under cold environment, implying cold stress induces the activation of SRKs, and activated SRK2s are crucial in regulating cold responses in *S. lycopersicum*.

Besides the high phosphorylation of CKLs and SRK2s in *S. lycopersicum*, we also observed several phosphorylation sites (pS303, pS602, pS609, pS613, pS1387, and pS1390) on the serine/threonine-protein kinase ATM (Solyc04g049930.3.1) were significantly up-regulated in cold-treated *S. lycopersicum* compared with that of *S. pimpinellifolium*. Activation of ATM through phosphorylation is a key step that cells response to DNA damage under oxidative stresses. Plant cells tend to accumulate reactive oxygen species (ROS) in response to cold stress, and ROS act as signal transducer to activate downstream gene expression and kinase activity to achieve cold acclimation. It has been reported that the phosphorylated histone variant H2AX, one of the substrates of ATM, is one of the components of repairing complexes at the DNA damage sites. We observed two phosphorylation sites (pS121 and pS124) in cluster 4 at the C-terminal of histone H2A (Solyc05g014835.1.1) were only phosphorylated in *S. lycopersicum*, implying cold stress induces the DNA repairing signals in *S. lycopersicum*. It is possible that ATM kinases play a role in sustaining the genome integrity for cold acclimation in *S. lycopersicum*.

2.3.4 Identification of SRK2E Substrates in Tomato by siKALIP Strategy

These findings prompted us to systematically identify the direct substrates of SRK2E in response to cold stress in *S. lycopersicum* to delineate the downstream signal transductions of SRK2E under cold stress. To this end, we utilized recently developed stable isotope labeled kinase assay linked phosphoproteomics (siKALIP) approach which relies on coupling the in vitro kinase assay screening with in vivo kinase-dependent phosphoproteomic perturbation (Figure 2.12). For in vitro SRK2E screening, the leaves of *S. lycopersicum* were lysed and digested according the optimized protocol. The 200 µg of digested proteins were first dephosphorylated using temperature-sensitive alkaline phosphatase, and the reaction was terminated by heating the samples at 75°C for 5 min. The dephosphorylated peptides were incubated with recombinant SRK2E (Solyc01g108280.3.1) with γ-$^{18}O_4$-ATP, and the heavy ^{18}O-phosphopeptides phosphorylated by SRK2E were enriched through polyMAC and analyzed by 1 h LC-MS/MS. A total of 1,770 class I heavy ^{18}O-phosphorylation sites was identified from in vitro kinase screening, which provides more candidate substrates than our previous reports (Supplementary table S10). This result suggests that our optimized protocol not only enlarges the coverage of in vivo phosphoproteome but also improves the sensitivity of in vitro kinase screening. We sought to investigate whether the in vitro SRK2E kinase screening can identify the reported SRK2s phosphorylation sites on homologs of ABF2 and VCS. The result shows that the two known phosphorylation sites on the ABF2 (pS58) and VSC (p878) were both phosphorylated by SRK2E in vitro, confirming the sensitivity and specificity of the in vitro SRK2E kinase screening (Supplementary table S10). Next, we overlapped the identified heavy phosphorylation sites with the in vivo phosphorylation sites up-regulated (> 2-fold) after cold treatment in *S. lycopersicum*. A total of 90 phosphorylation sites corresponding to 84 phosphoproteins were selected as candidate SRK2E substrates (Figure 5B). Interestingly, we found several kinases, enzymes, and transcription factors that may involve in cold adaption in *S. lycopersicum*. For example, we observed SRK2E directly phosphorylated two previous mentioned kinases which are the pS355 of CKL2 and the pS609 of ATM in *S. lycopersicum* (Figure 5C). These findings suggest that cold stress may activate cold tolerance responses through SRK-CKL2 and SRK-ATM axis in *S. lycopersicum*.

2.4 Discussion

The high complexity of plant tissues is always a major challenge in studying plant phosphoproteomics. Our work achieved ab efficiently identification of the plant phosphoproteome by enhancing the lysis efficiency, removing the interferences, and optimizing the digestion protocols, which identified many low abundant phosphorylation sites from cytosol and organelles. This optimized protocol that we have developed is applicable to study the global phosphoproteomic remodeling of any given plant species after environmental perturbation. We have demonstrated that more than 14,000 class I phosphorylation sites can be identified upon two tomato varieties, which provides an informative resource to the community.

More importantly, this work not only is the deepest quantitative analysis of tomato cold-induced phosphoproteome to date, but also partially reveals the underlying mechanisms conferring the cold tolerance in tomato at the post-translational level involved in tomato cold tolerance. Through phosphorylation motif analysis, we represented known molecular mechanisms involved in cold-treated *S. pimpinellifolium* and further found the new kinase-substrate relationships that regulate cold tolerance in *S. lycopersicum*. This analysis shows that two tomato varieties have distinct kinome remodeling during prolonged cold stress. For example, several the MAPKs are activated in *S. pimpinellifolium* under cold environment, while two SRK2 kinases are more activated in *S. lycopersicum*. The data is supported from the literature that they observed the concentration of ABA is significantly increased in *S. lycopersicum* after 12 h and keeps increasing until 72 h under exposing cold temperature (10 °C)[38]. Since ABA is the upstream activator of the kinase activity of SRE2E, we propose that the induction of SRK2E activity is attributed to high production of ABA under prolonged cold stress.

Considering that casein kinases are known to recognize acidic phosphorylation motif, it is possible that the casein kinases are activated in S. lycopersicum to phosphorylate their substrates with the acidic motif [-pS-D-] in the cluster 3 and 4. It has been reported that AtCKL2 is induced during water loss and ABA treatment, and AtCKL2 participates the stomatal closure through phosphorylating and stabilizing actin depolymerizing factor 4 (ADF4) expression. Although we did not identify the phosphorylation sites of ADF4, casein kinases are likely involved in regulating cold tolerance in S. lycopersicum.The

combination of our phosphoproteomic data with in vitro SRK2E kinase-substrate screening enabled us to further link the cold-induced acidic ([-pS-D-]) motif with the activation of SRK2E in *S. lycopersicum*. Our results suggest that SRK2E phosphorylates CKL2 under cold stress in *S. lycopersicum*, and CKL2 further phosphorylates many downstream substrates with acidic motif that regulate cold tolerance. We also observed that several phosphorylation sites of ATM and histone H2A are increased in *S. lycopersicum*, implying the *S. lycopersicum* is likely to modulate DNA repairing through SRK-ATM-H2A axis. Further experiments are necessary to study the roles of CKL2 and ATM involve in cold tolerance and to identify the downstream substrates of the two kinases to delineate the underlying pathways of cold tolerance in tomato.

In sum, we have developed a novel phosphoproteomic pipeline for profiling the signal transductions of tomato exposed to low temperature. More importantly, we characterized the induction of distinct kinases in two tomato varieties which show distinct cold phenotypes. Our analysis has resulted in the identification of novel tomato phosphoproteins that changes in response to cold treatment. Together, we expect that this large-scale quantitative phosphoproteomic data will serve as a useful database to identify of key proteins that involve in cold tolerance in tomato.

2.5 References

1. Zhu, J. K., Abiotic Stress Signaling and Responses in Plants. *Cell* **2016,** *167* (2), 313-324.

2. Schulze, W. X., Proteomics approaches to understand protein phosphorylation in pathway modulation. *Curr Opin Plant Biol* **2010,** *13* (3), 280-87.

3. Silva-Sanchez, C.; Li, H.; Chen, S., Recent advances and challenges in plant phosphoproteomics. *Proteomics* **2015,** *15* (5-6), 1127-41.

4. Singh, D. K.; Calvino, M.; Brauer, E. K.; Fernandez-Pozo, N.; Strickler, S.; Yalamanchili, R.; Suzuki, H.; Aoki, K.; Shibata, D.; Stratmann, J. W.; Popescu, G. V.; Mueller, L. A.; Popescu, S. C., The tomato kinome and the tomato kinase library ORFeome: novel resources for the study of kinases and signal transduction in tomato and solanaceae species. *Mol Plant Microbe Interact* **2014,** *27* (1), 7-17.

5. Imamura, H.; Sugiyama, N.; Wakabayashi, M.; Ishihama, Y., Large-scale identification of phosphorylation sites for profiling protein kinase selectivity. *J Proteome Res* **2014,** *13* (7), 3410-9.

6. Tsai, C. F.; Hsu, C. C.; Hung, J. N.; Wang, Y. T.; Choong, W. K.; Zeng, M. Y.; Lin, P. Y.; Hong, R. W.; Sung, T. Y.; Chen, Y. J., Sequential Phosphoproteomic Enrichment through Complementary Metal-Directed Immobilized Metal Ion Affinity Chromatography. *Anal Chem* **2014,** *86* (1), 685-693.

7. Isaacson, T.; Damasceno, C. M.; Saravanan, R. S.; He, Y.; Catala, C.; Saladie, M.; Rose, J. K., Sample extraction techniques for enhanced proteomic analysis of plant tissues. *Nat Protoc* **2006,** *1* (2), 769-74.

8. Wang, W.; Tai, F.; Chen, S., Optimizing protein extraction from plant tissues for enhanced proteomics analysis. *J Sep Sci* **2008,** *31* (11), 2032-9.

9. Kim, Y. J.; Lee, H. M.; Wang, Y.; Wu, J.; Kim, S. G.; Kang, K. Y.; Park, K. H.; Kim, Y. C.; Choi, I. S.; Agrawal, G. K.; Rakwal, R.; Kim, S. T., Depletion of abundant plant RuBisCO protein using the protamine sulfate precipitation method. *Proteomics* **2013,** *13* (14), 2176-9.

10. Chinnusamy, V.; Zhu, J.; Zhu, J. K., Cold stress regulation of gene expression in plants. *Trends Plant Sci* **2007,** *12* (10), 444-51.

11. Miura, K.; Furumoto, T., Cold signaling and cold response in plants. *Int J Mol Sci* **2013,** *14* (3), 5312-37.

12. Thomashow, M. F., PLANT COLD ACCLIMATION: Freezing Tolerance Genes and Regulatory Mechanisms. *Annu Rev Plant Physiol Plant Mol Biol* **1999,** *50*, 571-599.

13. Medina, J.; Catala, R.; Salinas, J., The CBFs: three arabidopsis transcription factors to cold acclimate. *Plant Sci* **2011,** *180* (1), 3-11.

14. Chinnusamy, V.; Ohta, M.; Kanrar, S.; Lee, B. H.; Hong, X.; Agarwal, M.; Zhu, J. K., ICE1: a regulator of cold-induced transcriptome and freezing tolerance in Arabidopsis. *Genes Dev* **2003,** *17* (8), 1043-54.

15. Teige, M.; Scheikl, E.; Eulgem, T.; Doczi, R.; Ichimura, K.; Shinozaki, K.; Dangl, J. L.; Hirt, H., The MKK2 pathway mediates cold and salt stress signaling in Arabidopsis. *Mol Cell* **2004,** *15* (1), 141-52.

16. Furuya, T.; Matsuoka, D.; Nanmori, T., Phosphorylation of Arabidopsis thaliana MEKK1 via Ca(2+) signaling as a part of the cold stress response. *J Plant Res* **2013,** *126* (6), 833-40.

17. Ding, Y.; Li, H.; Zhang, X.; Xie, Q.; Gong, Z.; Yang, S., OST1 kinase modulates freezing tolerance by enhancing ICE1 stability in Arabidopsis. *Dev Cell* **2015,** *32* (3), 278-89.

18. Chen, H.; Chen, X.; Chen, D.; Li, J.; Zhang, Y.; Wang, A., A comparison of the low temperature transcriptomes of two tomato genotypes that differ in freezing tolerance: Solanum lycopersicum and Solanum habrochaites. *BMC Plant Biol* **2015,** *15*, 132.

19. Patade, V. Y.; Khatri, D.; Kumari, M.; Grover, A.; Mohan Gupta, S.; Ahmed, Z., Cold tolerance in Osmotin transgenic tomato (Solanum lycopersicum L.) is associated with modulation in transcript abundance of stress responsive genes. *Springerplus* **2013,** *2* (1), 117.

20. Wang, P. C.; Xue, L.; Batelli, G.; Lee, S.; Hou, Y. J.; Van Oosten, M. J.; Zhang, H. M.; Tao, W. A.; Zhu, J. K., Quantitative phosphoproteomics identifies SnRK2 protein kinase substrates and reveals the effectors of abscisic acid action. *P Natl Acad Sci USA* **2013,** *110* (27), 11205-11210.

21. Masuda, T.; Sugiyama, N.; Tomita, M.; Ishihama, Y., Microscale phosphoproteome analysis of 10,000 cells from human cancer cell lines. *Anal Chem* **2011,** *83* (20), 7698-703.

22. Humphrey, S. J.; Azimifar, S. B.; Mann, M., High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics. *Nat Biotechnol* **2015,** *33* (9), 990-5.

23. Wisniewski, J. R.; Zougman, A.; Nagaraj, N.; Mann, M., Universal sample preparation method for proteome analysis. *Nat Methods* **2009,** *6* (5), 359-62.

24. Wessel, D.; Flugge, U. I., A method for the quantitative recovery of protein in dilute solution in the presence of detergents and lipids. *Anal Biochem* **1984,** *138* (1), 141-3.

25. Masuda, T.; Tomita, M.; Ishihama, Y., Phase transfer surfactant-aided trypsin digestion for membrane proteome analysis. *J Proteome Res* **2008,** *7* (2), 731-40.

26. Boersema, P. J.; Raijmakers, R.; Lemeer, S.; Mohammed, S.; Heck, A. J., Multiplex peptide stable isotope dimethyl labeling for quantitative proteomics. *Nat Protoc* **2009,** *4* (4), 484-94.

27. 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), 2162-72.

28. Dimayacyac-Esleta, B. R.; Tsai, C. F.; Kitata, R. B.; Lin, P. Y.; Choong, W. K.; Lin, T. D.; Wang, Y. T.; Weng, S. H.; Yang, P. C.; Arco, S. D.; Sung, T. Y.; Chen, Y. J., Rapid High-pH Reverse Phase StageTip for Sensitive Small-Scale Membrane Proteomic Profiling. *Anal Chem* **2015,** *87* (24), 12016-23.

29. Xue, L.; Wang, P.; Cao, P.; Zhu, J. K.; Tao, W. A., Identification of extracellular signal-regulated kinase 1 (ERK1) direct substrates using stable isotope labeled kinase assay-linked phosphoproteomics. *Mol Cell Proteomics* **2014,** *13* (11), 3199-210.

30. Fernandez-Pozo, N.; Menda, N.; Edwards, J. D.; Saha, S.; Tecle, I. Y.; Strickler, S. R.; Bombarely, A.; Fisher-York, T.; Pujar, A.; Foerster, H.; Yan, A.; Mueller, L. A., The Sol Genomics Network (SGN)--from genotype to phenotype to breeding. *Nucleic Acids Res* **2015,** *43* (Database issue), D1036-41.

31. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **2008,** *26* (12), 1367-1372.

32. 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.

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

34. Vu, L. D.; Stes, E.; Van Bel, M.; Nelissen, H.; Maddelein, D.; Inze, D.; Coppens, F.; Martens, L.; Gevaert, K.; De Smet, I., Up-to-Date Workflow for Plant (Phospho)proteomics Identifies Differential Drought-Responsive Phosphorylation Events in Maize Leaves. *J Proteome Res* **2016,** *15* (12), 4304-4317.

35. Roitinger, E.; Hofer, M.; Kocher, T.; Pichler, P.; Novatchkova, M.; Yang, J.; Schlogelhofer, P.; Mechtler, K., Quantitative phosphoproteomics of the ataxia telangiectasia-mutated (ATM) and ataxia telangiectasia-mutated and rad3-related (ATR)

dependent DNA damage response in Arabidopsis thaliana. *Mol Cell Proteomics* **2015,** *14* (3), 556-71.

36. Logemann, J.; Schell, J.; Willmitzer, L., Improved method for the isolation of RNA from plant tissues. *Anal Biochem* **1987,** *163* (1), 16-20.

37. Soma, F.; Mogami, J.; Yoshida, T.; Abekura, M.; Takahashi, F.; Kidokoro, S.; Mizoi, J.; Shinozaki, K.; Yamaguchi-Shinozaki, K., ABA-unresponsive SnRK2 protein kinases regulate mRNA decay under osmotic stress in plants. *Nat Plants* **2017,** *3*, 16204.

38. Daie, J.; Campbell, W. F., Response of Tomato Plants to Stressful Temperatures : INCREASE IN ABSCISIC ACID CONCENTRATIONS. *Plant Physiol* **1981,** *67* (1), 26-9.

Table 2.1: Identification results in all biological replicates of two tomato varieties.

Sample	Amount	Phosphoproteins	Phosphopeptides	Class I sites
77-Replicate 1	1 mg	4,316	12,956	7,578
77-Replicate 2	1 mg	4,142	12,318	5,904
77-Replicate 3	1 mg	4,271	13,113	6,399
27-Replicate 1	1 mg	4,383	13,727	7,551
27-Replicate 2	1 mg	3,670	10,292	6,775
27-Replicate 3	1 mg	3,836	11,648	7,069
Total		5,471	30,924	14,489

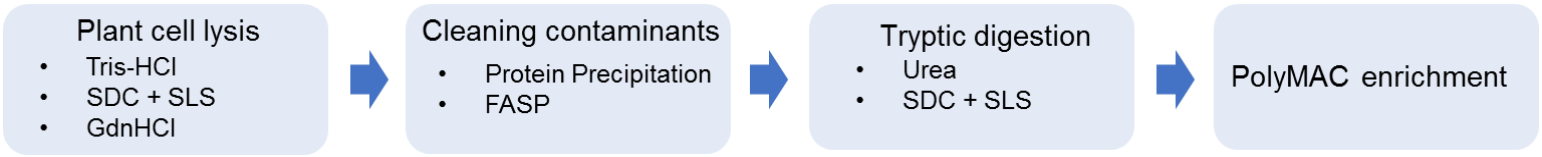


Figure 2.1: Optimization of the sample preparation protocols for plant phosphoproteomics. Schematic representation of the comparison of plant lysis buffers, cleaning contaminants, and tryptic digestion buffers to analyze tomato phosphoproteomics using 200 µg of tomato leaves protein.

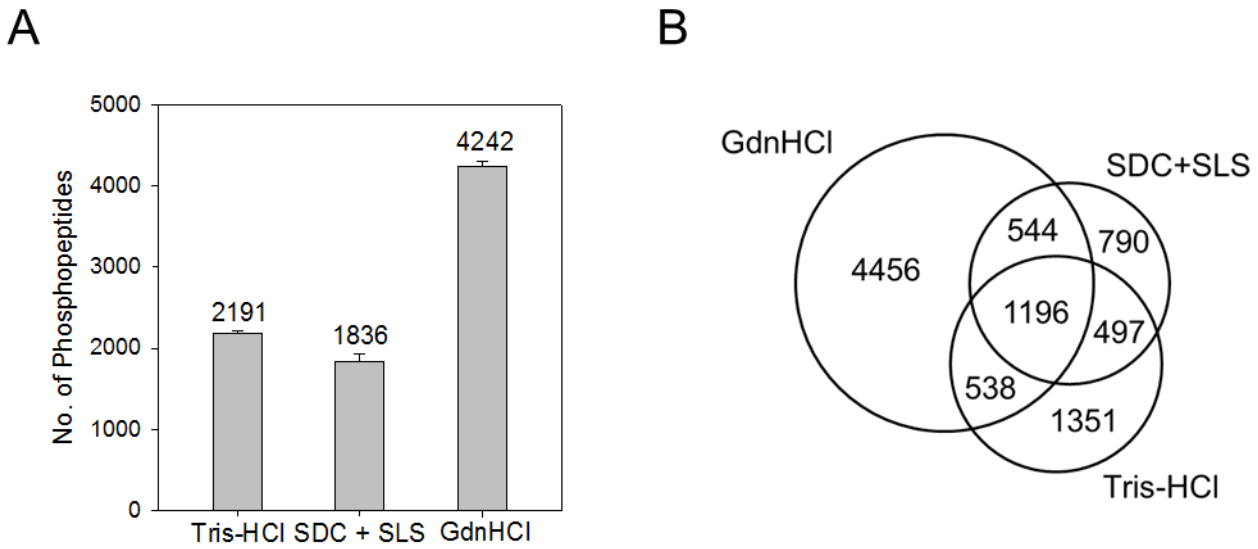


Figure 2.2: Optimization of the sample preparation protocols for plant phosphoproteomics. (A) Comparison of the plant tissue lysis efficiency using Tris-HCl, SDC and SLS mixture, and the GdnHCl lysis protocols. (B) Venn diagram shows the overlap of the number of identified phosphopeptides from triplicate analyses of the three lysis protocols.

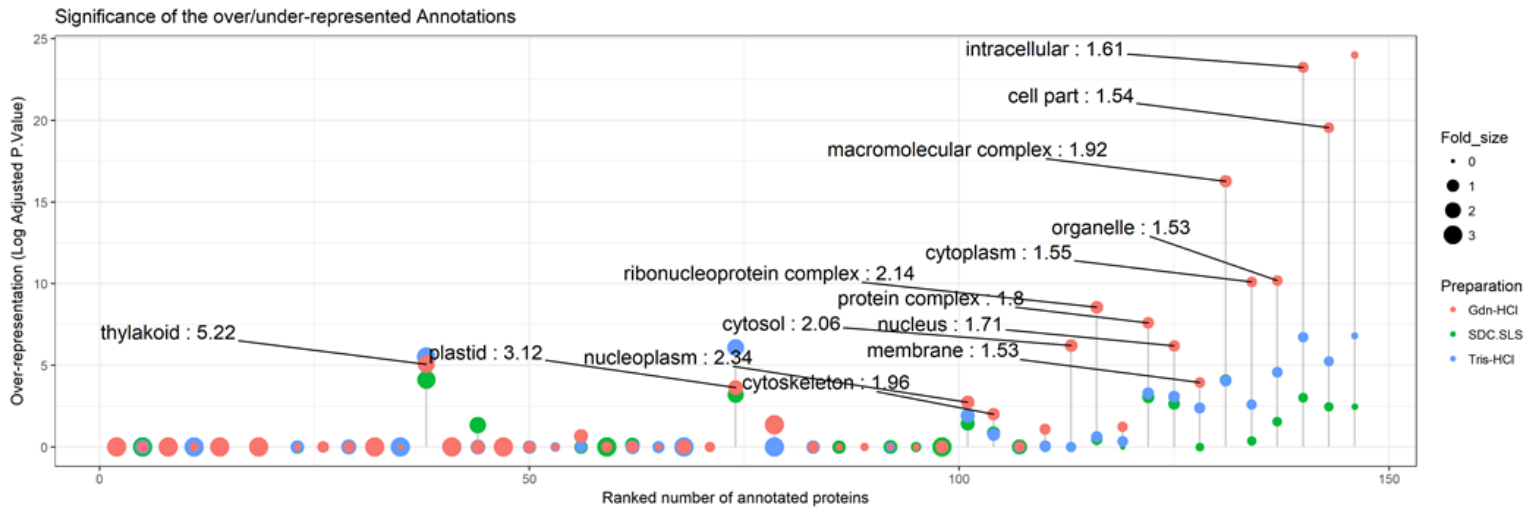


Figure 2.3: Comparison of the distribution of Gene ontology cellular components (GO-CCs) in the three protocols. The GO-CCs significantly overrepresented (p-value < 0.01 and fold change ≥ two-fold) in the three protocols compared to the background distribution were highlighted in the bar graph.

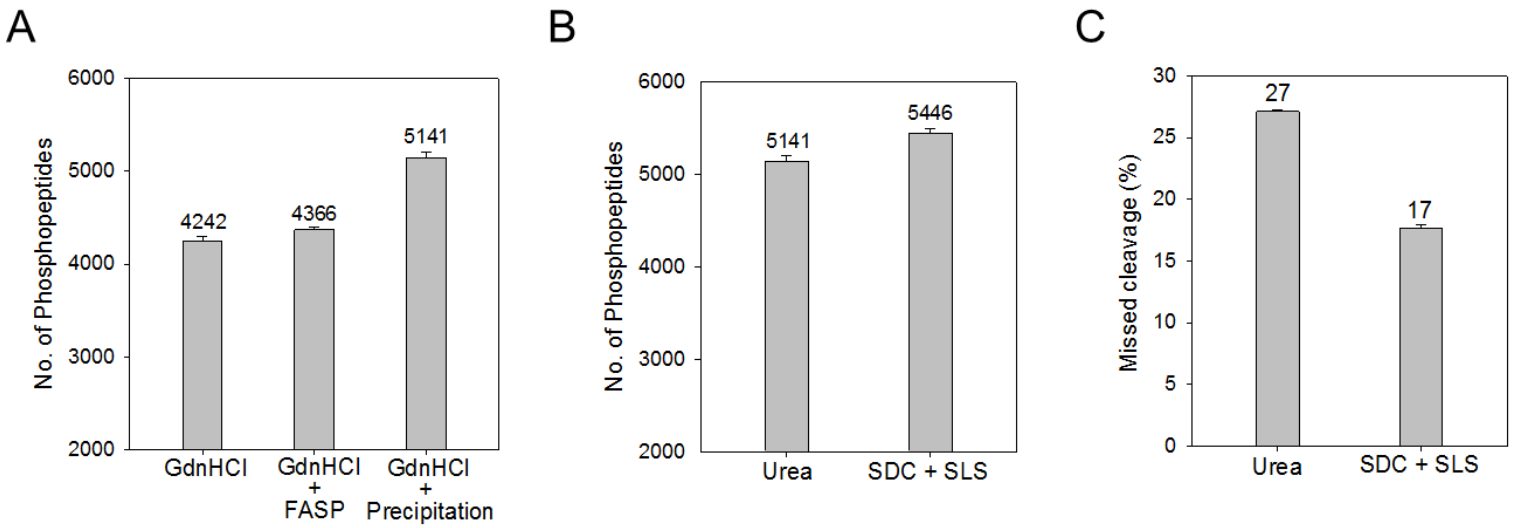


Figure 2.4: Optimization of the sample preparation protocols for plant phosphoproteomics. (A) Evaluation of the performance of FASP and methanol-chloroform precipitation for improving phosphopeptide identification. (B) Comparison of the phosphoproteomic coverage between the urea and SDC plus SLS tryptic digestion protocols. (C) Comparison of the trypsin activities between the urea and SDC plus SLS tryptic digestion protocols.

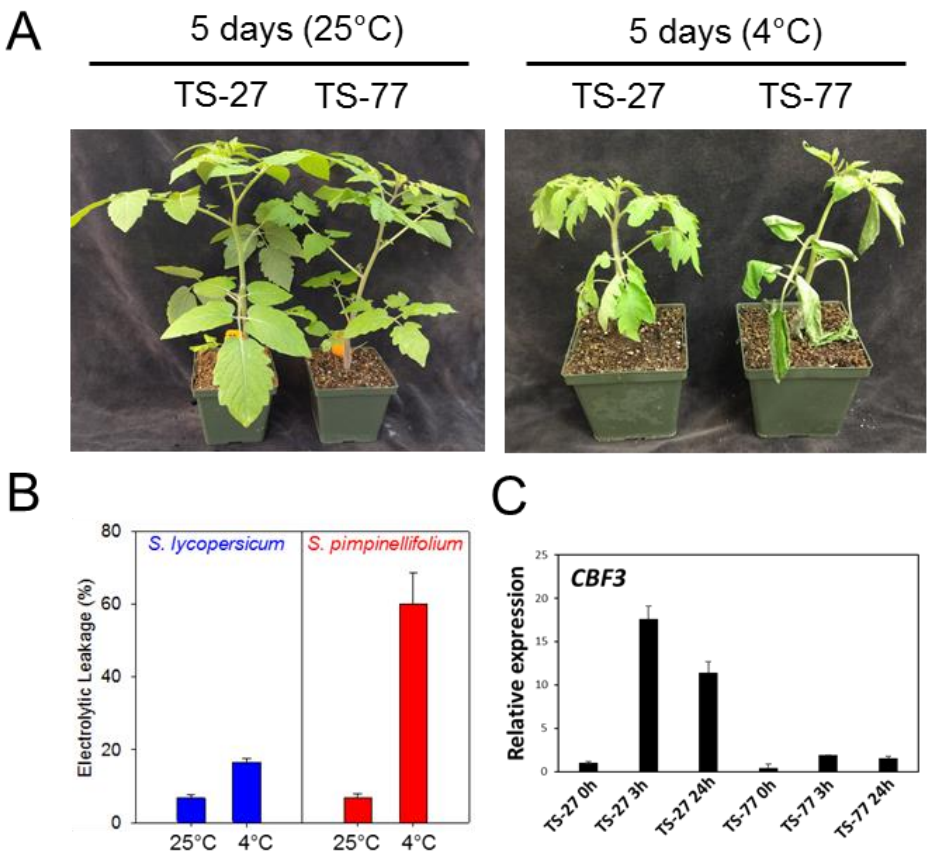


Figure 2.5: Morphological and physiological changes of two tomato species in response to prolonged cold treatment. (A) Morphological changes of *S. lycopersicum* and *S. pimpinellifolium* under 5d cold treatment were photographed. (B) Evaluation of the electrolytic leakage extent of the tomatoes between room temperature and cold environment. (C) The changes of *CBF3* expression level of two tomatoes after cold treatment are presented in the bar graph.

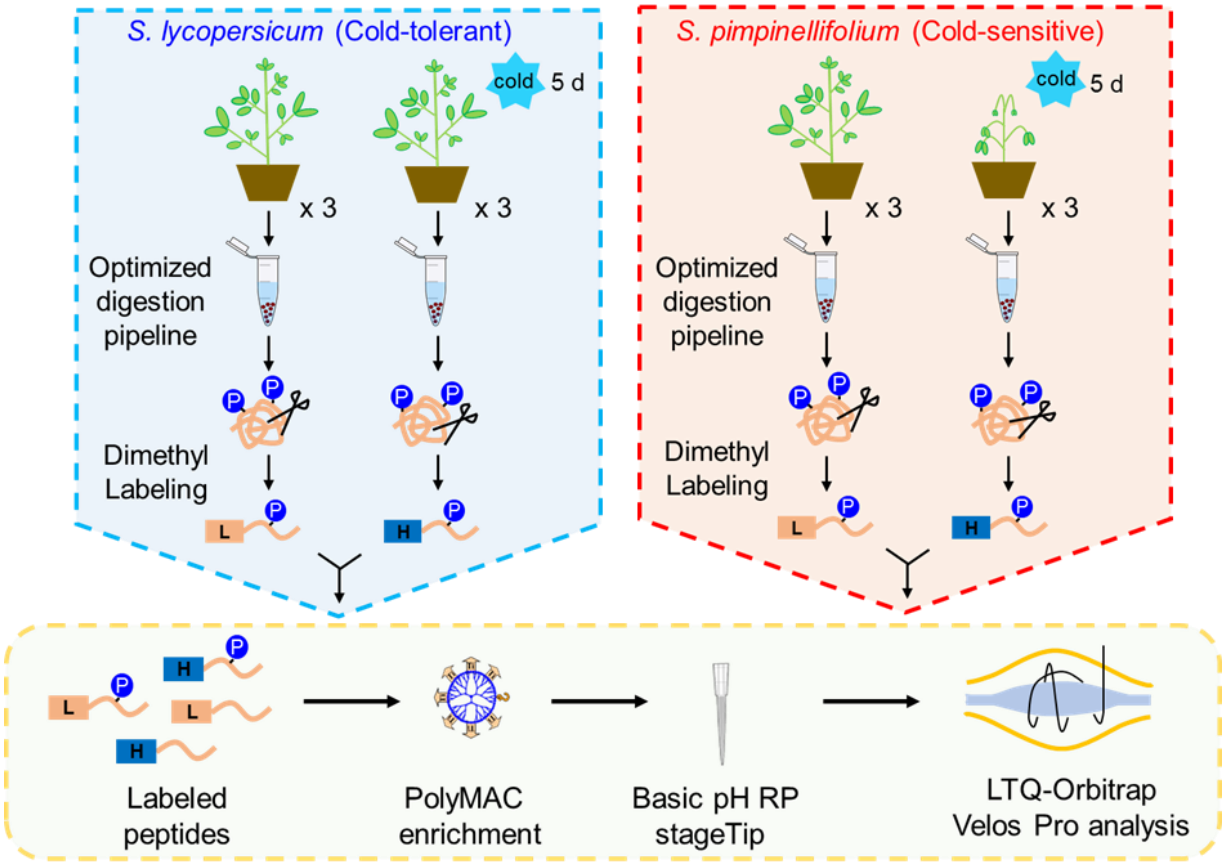


Figure 2.6: The workflow of profiling the phosphoproteome of tomato leaves in response to cold stress. The tomatoes leaves were grown in three biological replicates under five days room or cold temperature. Proteins were extracted, precipitated, and digested using the optimized protocol. The digested phosphopeptides were dimethyl-labeled, combined, and enriched using PolyMAC. To enlarge the phosphoproteome coverage, the isolated phosphopeptides were fractionated using basic pH reverse-phase StageTips prior to LC-MS/MS analysis.

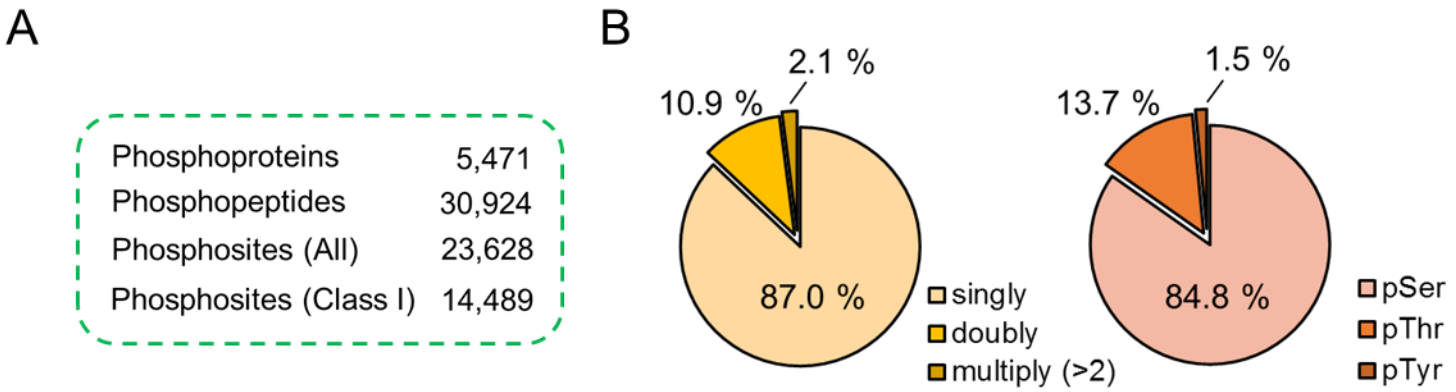


Figure 2.7: Profiling the phosphoproteome of tomato leaves in response to cold stress. (A) The total number of identified phosphoproteins, phosphopeptides, and phosphorylation sites in the phosphoproteome of tomato leaves. (B) Distribution of the number of phosphates and the amino acid residues for all detected phosphopeptides.

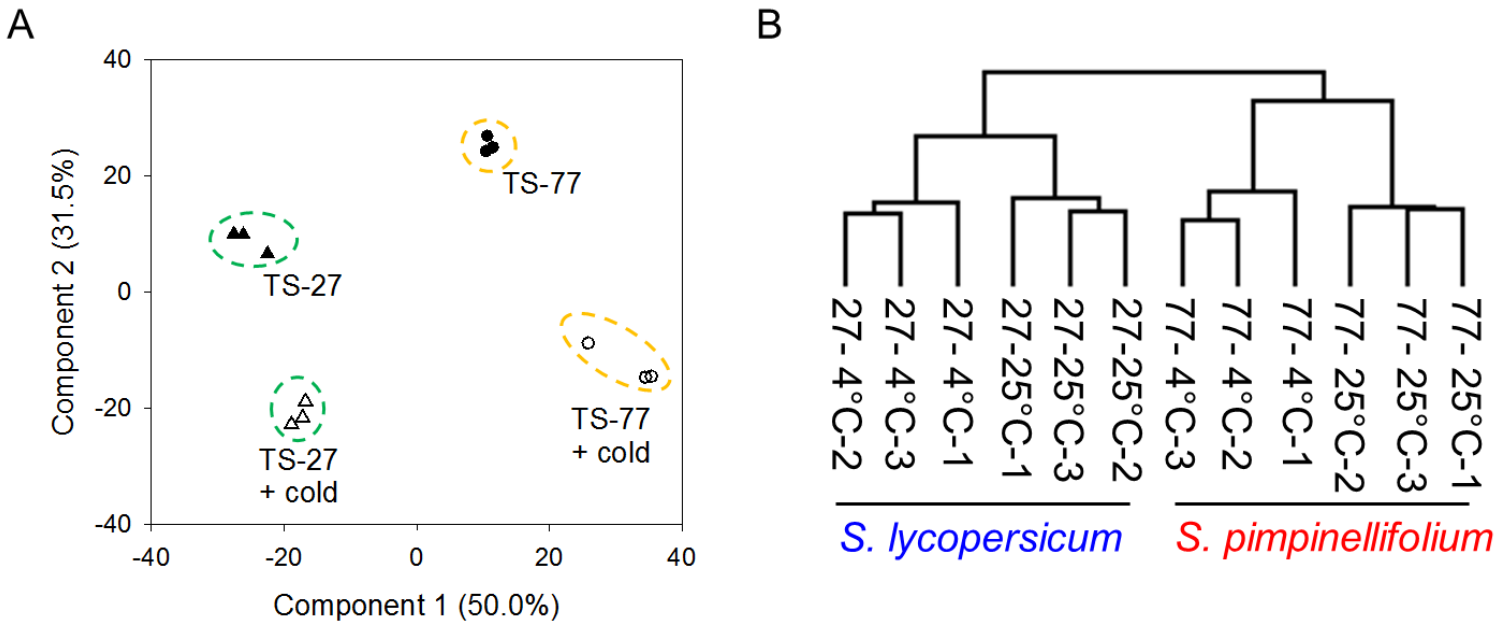


Figure 2.8: Evaluation of the quality and reproducibility of biological replicates in two tomatoes. (A) Principle component analysis (PCA) of the localized phosphorylation sites across all four samples representing distinct phosphoproteomic types for different treatments. (B) The dendrogram summarizes the hierarchical clustering analysis of all the biological replicates. Two main clusters were obtained based on the tomato varieties.

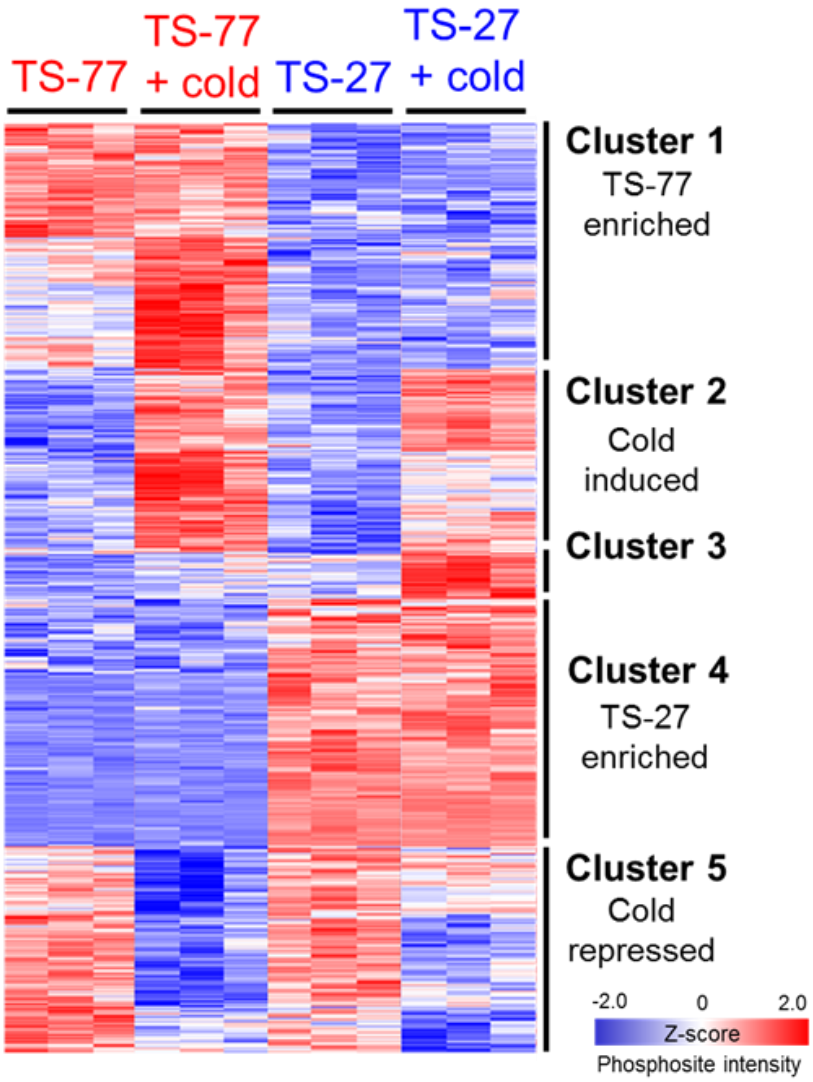


Figure 2.9 The phosphoproteomic perturbation of cold responses in tomatoes. Unsupervised hierarchical clustering of significantly changed phosphorylation sites of all biological replicates under cold stress (ANOVA, permutation-based FDR < 0.05). The color code of each phosphorylation site (row) at all samples (columns) indicates the low (blue) and high (red) Z-score normalized intensities.

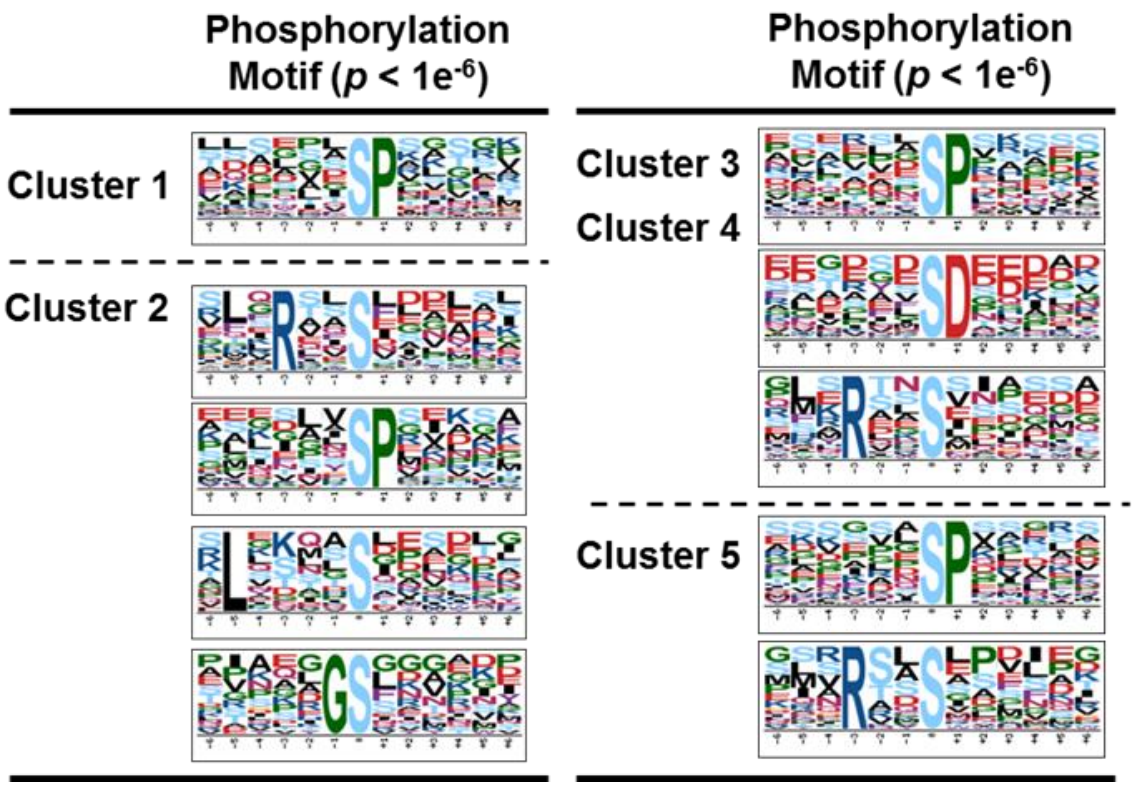


Figure 2.10: Phosphorylation motif analysis enriched the overrepresented phosphorylation motifs of each clusters in two tomatoes after cold stress treatment.

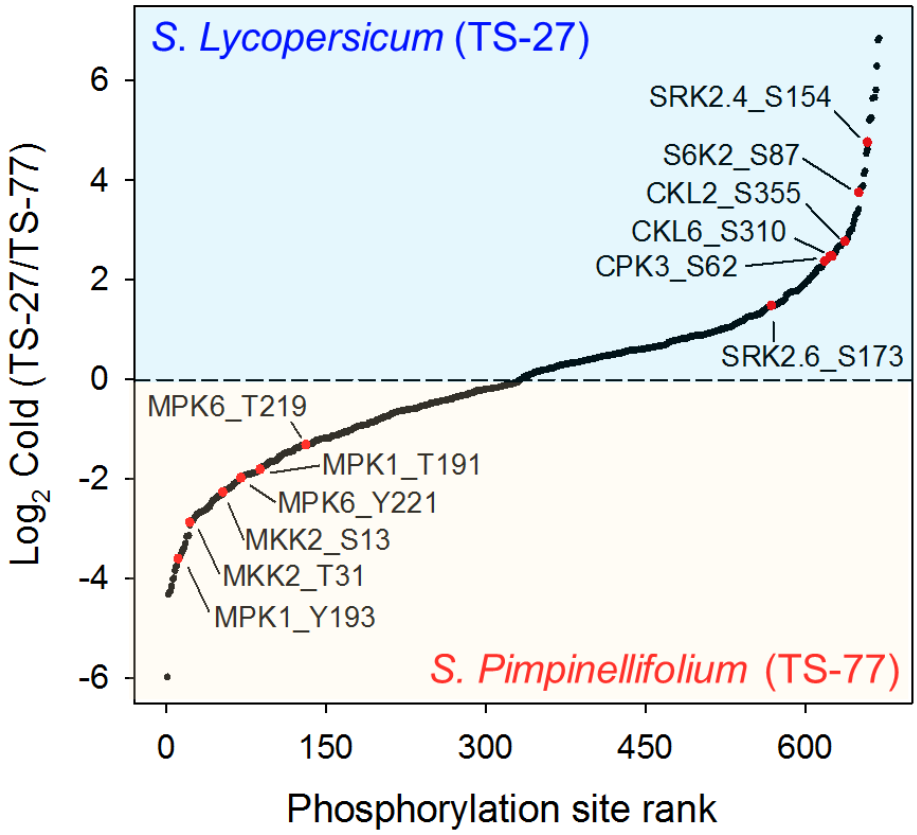


Figure 2.11 Kinome reprogramming of cold responses in two tomatoes. Ranked list of phosphorylation expression of kinases in response to cold treatment. Dots show the mean fold change of three biological replicates.

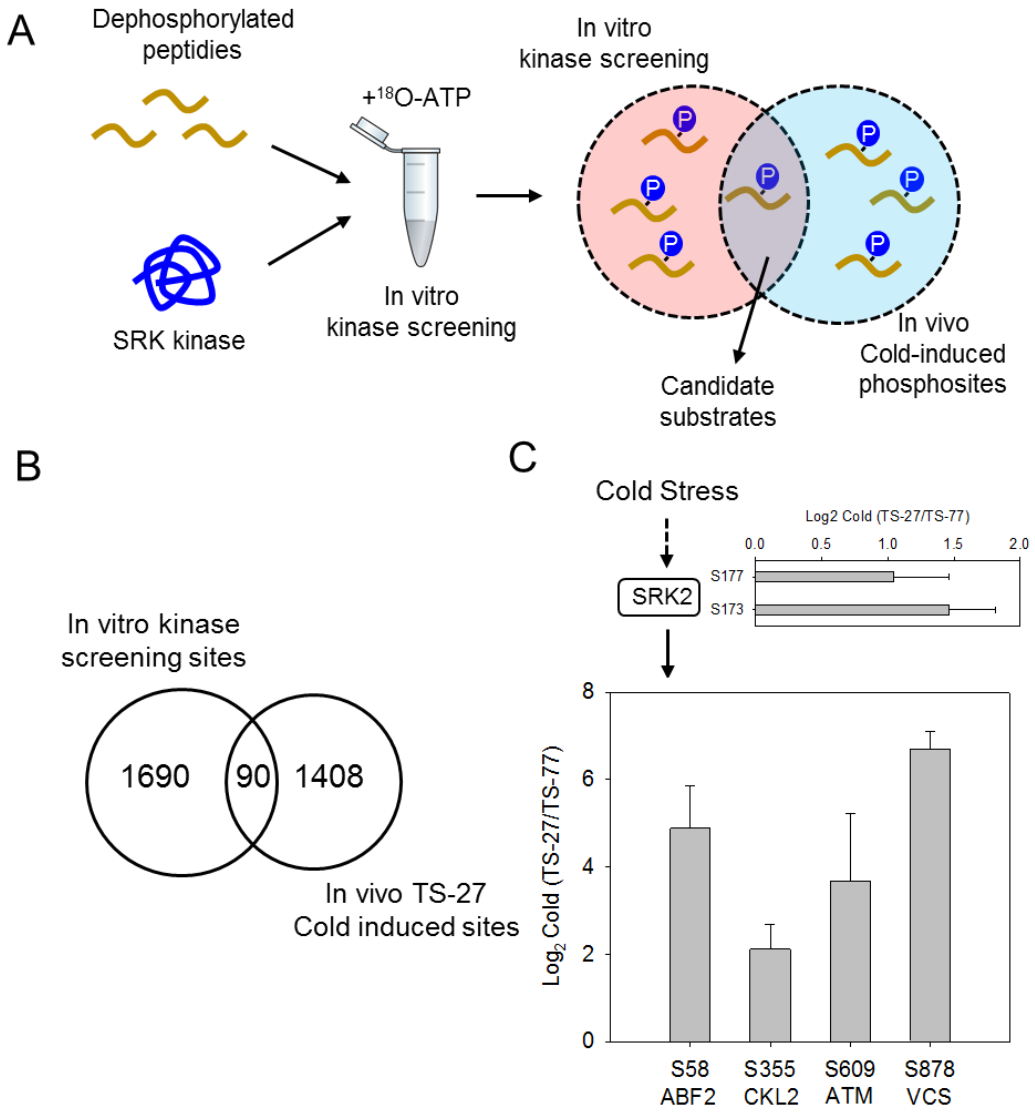


Figure 2.12: Identifying the direct substrates of SnRK2E involved in cold responses in *S. lycopersicum*. (A) The KALIP workflow applied to identify the direct SnRK2E substrates in *S. lycopersicum* with high confidence. (B) The overlap of in vitro kinase screening and in vivo phosphoproteomic result to acquire the candidate substrates of SnRK2E. (C) Schematic representation of the proposed cold-induced mechanisms through activation of SnRK2E in *S. lycopersicum*. The bar graph shows the cold-induced changes in phosphorylation levels of proteins between two tomatoes.

CHAPTER 3. IDENTIFICATION OF THE DIRECT SUBSTRATES OF THE PLANT KINASES IN RESPONSE TO ENVIRONMENTAL STRESSES

3.1 Summary

Protein kinases are major regulatory components in almost all cellular processes in eukaryotic cells. By adding phosphate groups, protein kinases regulate the activity, localization, protein-protein interactions and other profiles of their target proteins. It is known that protein kinases such as SNF1 related protein kinase 2 and 3 (SnRK2 and SnRK3), calcium dependent protein kinases (CDPKs), and mitogen activated protein kinases (MAPKs) are central components in plant response to environmental stresses such as drought, high salinity, cold and pathogen attack. Although the phosphorylation-dependent signaling plays critical roles in plant stress biology, only a few targets of these protein kinases have been identified. Moreover, how these protein kinases regulate the downstream biological processes and mediate the stress responses is still largely unknown. In this study, we refined the mass spectrometry-based strategy termed Kinase Assay LInked Phosphoproteomics (KALIP) with enriched phosphoproteome as kinase-substrates pool to identify the putative substrates of 9 protein kinases that function in plant abiotic and biotic responses. As a result, we identified more than 4000 putative target sites of osmotic stress-activated SnRK2.4 and SnRK2.6, ABA activated protein kinases SnRK2.6 and Casein-Kinase Like 2 (CKL2), elicitor-activated protein kinase CDPK11 and MPK6, cold-activated protein kinase MPK6 and Calcium/calmodulin-regulated Receptor-Like Kinase 1, H2O2-activated protein kinase OXI1 and MPK6, salt-induced protein kinase SOS1, as well as the low-potassium-activated protein kinase CIPK23. We also analyzed the conserved motif and functional ontology of these putative kinase targets. These results provide comprehensive information on the role of these protein kinases in the control of cellular activities and advance our understanding of plant response to environmental stresses.

3.2 Introduction

Eukaryotic protein kinases are major regulatory components in all cellular functions[1]. By adding phosphate groups to substrate proteins, protein kinases direct the activity, localization, association with other proteins, and overall function of many proteins and serve to orchestrate the activity of almost all cellular processes[2]. Compare to other eukaryotes, flowering plants have more genes encoding protein kinases in their genomes[3-4]. In the model plant Arabidopsis, more than 1000 protein kinases and 20,000 phosphosites have been identified, suggested the critical functions of protein kinase and their substrates in plants[5].

It has been well known that some protein kinases are central components in plant response to environmental changes[6]. Sucrose non-fermenting (SNF1)-related protein kinase 2s (SnRK2s) are central components in plant response to drought and osmotic stresses[7]. There are ten members in the SnRK2 family in Arabidopsis (SnRK2.1 to -2.10) and rice (SAPK1 to –SAPK10). Three of them in Arabidopsis, SnRK2.2, SnRK2.3, and SnRK2.6, are activated by ABA, and play critical roles in ABA responses. The *snrk2.2/3/6* triple-knockout mutant has a dramatic ABA-insensitive phenotype in seed germination, root growth inhibition, stomatal opening, and gene regulation[8-9]. All SnRK2s except SnRK2.9 can be activated by osmotic stress caused by NaCl or mannitol treatment in Arabidopsis[10]. The *snrk2.1/2/3/4/5/6/7/8/9/10* decuple mutant is hypersensitive to PEG-induced growth inhibition, and has a dramatically reduced shoot and root growth even under moderate osmotic stress treatment[11]. In the decuple mutant, osmotic stress-induced gene expression and metabolic changes (e.g., IP_3 and proline accumulation) were also impaired[11]. Moreover, the decuple mutant has a strong dwarf phenotype and cannot produce seeds even under high-humidity conditions, suggesting the importance of SnRK2s in controlling normal growth and development.

In addition to SnRK2s, calcium dependent protein kinases (CDPKs/CPKs) are also required for stomatal movement in response to ABA[12]. Mutant plants defective in four redundant calcium-dependent protein kinases, CPK5, CPK6, CPK11, and CPK23, lost the capability of stomata closing in response to ABA[13]. CPKs phosphorylate effectors like SLAC1 to trigger the stomata closure [13-14]. Interesting, another calcium signaling activated protein kinase CIPK23 also phosphorylate SLAC1 and contribute to ABA induced stomata closing[15]. CIPK26 also phosphorylates RbohF and control ABA induced reactive oxygen

species (ROS) generation[16]. Casein kinase like 2 (CKL2) regulates actin filament stability and stomatal closure by phosphorylation of actin depolymerizing factor ADF4[17].

SALT OVERLY SENSITIVE 2 (SOS2), also known as CIPK24, is a key regulator in salinity response in plant[18]. High salinity increases the cytosolic calcium content, which is sensed by calcium sensor SOS3 (CBL4) in root and SCaBP8/CBL10 in shoot[19]. SOS3 and CBL10 activates SOS2. SOS2 then phosphorylates and activates plasma membrane Na^+/H^+ antiporter SOS1 to extrude Na^+ into soil solution or load Na^+ to xylem[20]. The SOS1 is only known substrate of SOS2 so far, and how SOS3-SOS2 regulates salinity response is still large unknown.

Mitogen activated protein kinases (MAPKs) cascades play pivotal roles in regulation of plant cell to both environmental changes and pathogen attacks. Activation of MAPKs is one of the earliest signaling events after plant sensing of pathogen/microbe-associated molecular patterns (PAMPs/MAMPs) and pathogen effectors[21-22]. As the most well-studied protein kinase in plant biology, several protein substrates of MPK3 and MPK6 have been identified[23-24]. These substrates mediated the biosynthesis/signaling of plant stress/defense hormones, reactive oxygen species (ROS) generation, stomatal closure, defense gene activation, phytoalexin biosynthesis, cell wall strengthening, and hypersensitive response (HR) cell death. In addition to MPK3/6, CPK4, CPK5, CPK6, and CPK11 also were reported that critical for transcriptional reprogramming in plant innate immune signaling[25]. Interestingly, rice CPK18 could phosphorylate MPK5 and active MAPK signaling in dependent of upstream MAP kinase kinase, suggested the crosstalk between calcium and MPK signaling[26]. CPK5 also phosphorylates respiratory burst oxidase homolog D (RBOHD), a plant NADPH oxidase, to promote the synthesis of ROS locally and facilitate rapid signal propagation as a prerequisite for defense response activation at distal sites within the plant[27].

To systematically identify the direct substrates of interested kinase with high confidence, we have developed the kinase assay linked phosphoproteomics (KALIP) approach for identifying the direct substrates of target kinase by coupling the *in vitro* kinase-assay based substrate screening with *in vivo* kinase-dependent phosphoproteomic analysis[28]. To further reduce the interference of background phosphorylation, stable isotope labeled-ATP has been integrated with KALIP approach termed stable isotope

labeled KALIP (siKALIP) to qualitatively identify the heavy ^{18}O-phosphopeptides as the candidate substrates *in vitro*[29]. However, the overlapped number of *in vitro* kinase-based substrate screening and *in vivo* kinase-dependent quantitative phosphoproteomics still leaves room for improvement, which stems from the different physiological states of kinase reaction between *in vitro* and *in vivo* and the complexity of proteome.

To further improve the sensitivity and accuracy of *in vitro* kinase screening, we have developed KALIP 2.0 approach by using the phosphoproteome as substrates pool for kinase reaction, which significantly reduces the complexity of plant proteome prior to large-scale *in vitro* kinase screening. The rationale is that bona fide kinase substrates will already be phosphorylated in vivo in the kinase-dependent biological status, and using these phosphopeptides in a subsequent kinase reaction helps retain the physiological relevance that is usually lost with an in vitro reaction. Next, quantitative analysis is performed to select the heavy isotope labeled substrates which are significantly expressed in the kinase-activated condition. By coupling the two steps, this strategy offers the candidate substrates of targeted kinase with high sensitivity, reliability, and simplicity. To demonstrate the feasibility of the KALIP 2.0, we used an Arabidopsis model in which the interested kinases are significantly induced after stress treatment. This system fulfills the requirement of KALIP 2.0 that the kinase of interest is considerably altered in comparison to a mock-treated sample in which the interested kinase only retains basal activity.

3.3 Experimental Procedures

Unless otherwise noted, reagents were purchased from Sigma (St. Louis, MO), and buffers were prepared with water purified to a resistivity of 18.2 MΩ cm^{-1} using a Barnstead Nanopure water system (Thermo Fisher).

3.3.1 Plant Materials and Growth

The seeds of Col-0 wild type Arabidopsis were geminated on half-strength Murashige and Skoog (MS) medium (1 % sucrose with 0.6% phytogel). Five days after germination, seedlings were transferred into 40 ml half-strength MS liquid medium with 1% sucrose at 22 °C in continuous light on a rotary shaker set at 100 rpm. For osmotic stress treatment,

twelve-day-old seedlings were transferred into fresh medium containing 800 mM Mannitol for 30 min. In parallel, the seedlings transferred into fresh medium were used as the control.

3.3.2 Protein Extraction and Digestion

Plant tissues were ground with liquid nitrogen in mortar, and the ground tissues were lysed in 6 M guanidine hydrochloride containing 100 mM Tris-HCl (pH = 8.5) with EDTA-free protease inhibitor cocktail (Roche, Madison, WI) and phosphatase inhibitor cocktail (Sigma-Aldrich, St. Louis, MO). Proteins were reduced and alkylated with 10 mM tris-(2-carboxyethyl)phosphine (TECP) and 40 mM chloroacetamide (CAA) at 95 °C for 5 min. Alkylated proteins were subject to methanol-chloroform precipitation, and precipitated protein pellets were solubilized in 8 M urea containing 50 mM triethylammonium bicarbonate (TEAB). Protein amount was quantified by BCA assay (Thermo Fisher Scientific, Rockford, IL). Protein extracts were diluted to 4 M urea and digested with Lys-C (Wako, Japan) in a 1:100 (w/w) enzyme-to-protein ratio for 3 hours at 37 °C, and further diluted to a 1 M urea concentration. Trypsin (Promega, Madison, WI) was added to a final 1:100 (w/w) enzyme-to-protein ratio overnight. Digests were acidified with 10% trifluoroacetic acid (TFA) to a pH ~2, and desalted using a 100 mg of Sep-pak C18 column (Waters, Milford, MA).

3.3.3 Stable Isotope Labeled *In Vitro* Kinase Reaction

The tryptic peptides (400 µg) from mock-treated and ABA-treated seedlings were incubated with polyMAC-Ti reagent[30] (Tymora Analytical, West Lafayette, IN) to enrich phosphopeptides, respectively. The enriched phosphopeptides were treated with thermosensitive alkaline phosphatase (TSAP) (Roche, Madison, WI) in a 1:100 (w/w) enzyme-to-peptides ratio at 37 °C overnight for dephosphorylation, and the dephosphorylated peptides were desalted using SDB-XC StageTips. The desalted peptides were re-suspended in kinase reaction buffer (50 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM DTT, and 1 mM γ-[$^{18}O_4$]-ATP, pH 7.5). The recombinant SnRK2.6 (1 µg) was incubated with the desalted peptides at 30 °C overnight. The kinase reaction was quenched by acidifying with 10% TFA to a final concentration of 1%, and the peptides were desalted by SDB-XC

StageTip. The heavy ^{18}O-phosphopeptides were further digested by trypsin at 37 °C for 6 h and enriched by second polyMAC-Ti reagent, and the eluates were dried in SpeedVac for LC-MS/MS analysis.

3.3.4 Phosphopeptide Enrichment

Phosphopeptide enrichment was performed according to the reported polyMAC-Ti protocol with some modifications[30]. Tryptic peptides (200 µg) were re-suspended in 100 µL of loading buffer (80% acetonitrile (ACN) with 1% TFA) and incubated with 25 µL of the PolyMAC-Ti reagent for 20 min. Use a magnetic rack to collect the magnetic beads to the sides of the tubes and discard the flow-through. The magnetic beads were washed with 200 µL of washing buffer 1 (80% ACN, 0.2% TFA with 25 mM glycolic acid) for 5 min and washing buffer 2 (80% ACN in water) for 30 seconds, respectively. Phosphopeptides were eluted with 200 µL of 400 mM NH_4OH with 50% ACN and dried in SpeedVac.

3.3.5 LC-MS/MS Analysis

Desalted phosphopeptides were dissolved in 12 µL of 0.3% formic acid (FA) with 3% ACN and injected in triplicate into an Easy nLC 1000 system coupled online to a LTQ-Orbitrap Velos Pro mass spectrometer (Thermo Scientific). Phosphopeptides were separated on an analytical column (360 µm OD, 75 µm ID, 45 cm length) packed in-house with 2.2 µm/120 Å ProntoPEARL C18 AQ resin (Bischoff Chromatography) that was maintained at 50°C with a column heater (Analytical Sales and Services, Flanders, NJ). The reverse phase separation took place at 250 nL/min using mobile phases of 0.1% formic acid in water (A) and 0.1% formic acid in 80% ACN (B) over the course of 60 min. The gradient was set to increase linearly from 5% B to 30% B over 40 min; this was followed by an increase to 50% B over 10 min, a 5 min wash at 95% B, and a return to equilibrium. The mass spectrometer was operated in the data dependent mode where one survey scan in the Orbitrap from 350-1700 m/z was followed by selection of the top 10 most abundant ions for MS/MS. Precursor ions were fragmented with CID (NCE = 30%) and analyzed in the linear ion trap. The AGC target was set to 3e4 with a maximum injection time of 100

ms. Charge states of +1 were rejected for MS/MS, and the dynamic exclusion time was set to 60 s.

3.3.6 Data Processing

The raw files were searched directly against *Arabidopsis Thaliana* database (TAIR10) with no redundant entries using MaxQuant software (version 1.5.4.1)[31] with Andromeda search engine. Initial precursor mass tolerance was set at 20 p.p.m. and the final tolerance was set at 6 p.p.m., and ITMS MS/MS tolerance was set at 0.6 Da. Search criteria included a static carbamidomethylation of cysteines (+57.0214 Da) and variable modifications of (1) oxidation (+15.9949 Da) on methionine residues, (2) acetylation (+42.011 Da) at N-terminus of protein, and (3) phosphorylation (+79.996 Da) on serine, threonine or tyrosine residues were searched. For identification of ^{18}O-phosphopeptides, heavy phosphorylation (+85.979 Da) was added as a variable modification and the match between run function was enabled. Search was performed with Trypsin/P digestion and allowed a maximum of two missed cleavages on the peptides analyzed from the sequence database. The false discovery rates of proteins, peptides and phosphosites were set at 0.01. The minimum peptide length was six amino acids, and a minimum Andromeda score was set at 40 for modified peptides. A site localization probability of 0.75 was used as the cut-off for localization of phosphorylation sites. All the peptide spectral matches and MS/MS spectra can be viewed through MaxQuant viewer. All the localized and up-regulated ^{18}O-phosphorylation sites under ABA stimulation were submitted to Motif-X[32] with TAIR10 database as background. The significance was set at 0.000001, the width was set at 13, and the number of occurrences was set at 20. The candidate substrate-substrate interaction networks were predicted in STRING version 10.0[33] with the interaction score $\geq$ 0.7 and maximum 10 interactor nodes, and the signal networks were visualized using Cytoscape version 3.4.0[34] with MCODE plugin version 1.4.2. DAVID (Database for Annotation, Visualization and Integrated Discovery)[35] was utilized to identify over-represented GO terms of MCODE-enriched clusters with *p*-value less than 0.05 and TAIR10 database was set as the background.

3.3.7 Data Analysis

All data were analyzed by using the Perseus software (version 1.6.0.2)[36]. For quantification of *in vitro* kinase screening, the intensities of heavy ^{18}O-phosphopeptides were extracted through MaxQuant, and the missing values of intensities were replaced by a constant value for statistical analysis. The significantly enriched ^{18}O-phosphopeptides were identified by the p-value is significant from a two sample t-test with a permutation-based FDR cut-off 0.05. For *in vivo* phosphoproteomic analysis, the significantly enriched phosphopeptides were identified by performing a two sample t-test with a permutation-based FDR cut-off 0.05 with S0 was set at 0.2. Normalization was carried out by subtracting the medium of log2 transformed phosphopeptide intensities.

3.4 Results

In our previous study, we developed a phosphoproteomic method to identify the putative substrates of ABA activated SnRK2s in Arabidopsis[37]. By this method, we identified 58 putative SnRK2 substrates by comparison the ABA responsive phosphoproteomics between wild type and *snrk2.2/2.3/2.6* triple mutant. Due to the limitation of MS detection, lots of low abundance substrates were not detected. In addition, this method depends on the genetic mutant which lacking the three major ABA activated SnRK2s and couldn't distinguish the direct SnRK2 targets or indirect substrates that phosphorylated by the kinase(s) downstream of SnRK2s. We also used Kinase assay-linked phosphoproteomics (KALIP) to identify the direct substrates of protein kinase. By this method, a whole cell extract was dephosphorylated and subjected to incubate with the chosen kinase and the ^{18}O-ATP as the phosphate donor. We identified 38 direct substrates of extracellular signal-regulated kinase 1 (ERK1)[29]. Though KALIP shows good potential to identify the kinase substrate in dependence of the genetic mutant, the major challenge for this system-wide kinase-substrate identification is that the low coverage between *in vitro* kinase reaction and *in vivo* phosphoproteomics, indicating the complexity of proteome and the difference of physiological statuses still limit the accuracy of *in vitro* kinase-substrates identification.

To overcome defects of pervious methods, we refined an advanced KALIP approach based approach termed KALIP 2.0 using the enriched phosphopeptides as the substrate pool, which significantly reduce the complexity of proteome. The workflow of this advanced approach integrating stable isotope-labeled *in vitro* kinase screening with quantitative measurement is shown in Figure 3.1. The extracted proteins are firstly digested by Lys-C to generate longer peptides and to retain the phosphorylation motif of in the peptides especially the kinase recognizes the basic motifs (e.g. [-R-x-x-(pS/pT)-]). The Lys-C digested phosphopeptides from either mock or kinase-activated Arabidopsis seedlings are enriched by polyMAC. The enriched phosphopeptides are treated with alkaline phosphatase to remove the background phosphorylation. The dephosphorylated peptides are subject to the targeted kinase reaction with heavy isotope ATP as phosphate donor. The heavy phosphate-labeled peptides are further digested by trypsin and enriched by second polyMAC and analyzed by LC-MS/MS. The candidate substrates of targeted kinase are determined through selecting the significantly up-regulated heavy ^{18}O-phosphopeptides in the kinase-activated seedlings, and the kinase phosphorylation sites on substrates are determined through the highest site localization probabilities within the ^{18}O-phosphopeptides.

We utilized this approach to identify the direct substrates of SnRK2 activated under ABA stimulus. The seedlings were either with or without 4 μM of ABA solution treatment for 30 min. The mock and ABA-treated (SnRK2-activated) seedlings were lysed and aliquoted into two parts: one for *in vitro* kinase screening, and the other for *in vivo* ABA-activated phosphoproteomic perturbation comparison. The recombinant SnRK2.6 was utilized as reaction kinase to phosphorylate the dephosphorylated peptides of mock and ABA-activated phosphoproteome, and the results of in vitro SnRK2.6 reaction were compared with *in vivo* phosphoproteomic identification to demonstrate the validity of this approach. For *in vitro* SnRK2 reaction, a total of 1,185 and 2,124 unique ^{18}O-phosphopeptide sequences were identified from mock and ABA-treated seedlings, respectively (Figure 2.2). In comparison with mock seedlings, two-fold number of ^{18}O-phosphopeptides were identified from ABA-treated seedlings, demonstrating that many SnRK2 substrates were activated by SnRK2 under ABA stimulation. Moreover, up to 965 (45%) identified ^{18}O-phosphopeptide sequences match to the 8,390 phosphopeptidess

identified from *in vivo* ABA-activated phosphoproteome, indicating the advanced siKAIP strategy offers sensitivity and specificity for identifying the ABA responsive SnRK2 substrates. The reproducibility of SnRK2 reaction was evaluated by comparing the intensities of identified ^{18}O-phosphopeptides in duplicate biological experiments, and the pearson coefficient is r = 0.894 for mock seedlings, and r = 0.898 for ABA-treated seedlings (Figure 2.3). The results indicate that the validity of *in vitro* kinase screening for selecting the candidate SnRK2 substrates through quantitative measurement.

To access the ABA-induced SnRK2 substrates with high confidence, we applied quantitative analysis to select the highly up-regulated ^{18}O-phosphopeptides in ABA-treated seedlings. Of the 2,124 identified ^{18}O-phosphopeptides, 1,389 ^{18}O-phosphopeptides that exist in all 3 replicates in either mock- or ABA-treated seedlings were used for further quantitative analysis. A permutation-based two sample t-test was performed to assess the significantly differences between the mock- and ABA-treated seedlings (Figure 2.4). In all, 778 ^{18}O-phosphopeptides corresponding to 614 phosphoproteins were significantly enriched (FDR < 0.01, and s0 > 0.35) based on the t-test results (Figure 2.5), which are the candidate substrates of SnRK2 in response to ABA responses.

Next, the motif analysis was performed to extract the SnRK2 recognition sequences from the significantly enriched ^{18}O-phosphopeptides in ABA-treated seedlings (Figure 2.5). Five high confidence motifs were detected: [-L-x-R-x-x-pS-] (Motif 1), [-R-x-x-(pS/pT)-] (Motif 2), [-L-x-x-x-x-pS] (Motif 3), and [-G-pS-] (Motif 4). Enrichment of Motif 1 to Motif 3 indicates that SnRK2 prefers the motifs of Leucine at -5 and Arginine at -3 position. (Figure 2.5). The Motif 1 and Motif 2 are recognized as well-known SnRK2 phosphorylation motifs, indicating that the good recognition specificity of *in vitro* SnRK2 screening at phosphopeptide level.

Interestingly, we found that a new SnRK2 motif, Motif 4 ([-G-pS-]), from *in vitro* results in which the glycine residue at -1 position has not been reported. Furthermore, several previous confirmed SnRK2 substrates such as AREB1/ABF2, AREB3, MSL9, and BTR1L are categorized as Motif 4 in the *in vitro* results, which suggests the Motif 4 is one of the SnRK2 phosphorylation motifs (Table 3.1). Several reported putative SnRK2 substrates from large-scale phosphoproteomic analysis were also identified including CIP7, VCS[38], COP1-interacting protein-related, and SAP-domain containing protein with Motif

1 and Motif 2, indicating the wide coverage of confirmed SnRK2 substrates by this approach.

Despite hundreds of mammalian MPK substrates regulating specific physiological functions have been identified, the direct substrates of plant MPK3/6 activated in response to these environmental stresses are remained elusive. To identify the direct substrates of MAPK6 in the regulation of cold, oxidative, and pathogen stress, the seedlings were treated with 4 °C, 10 μM of H_2O_2 solution, and 10 μM of flg22 solution treatment for 30 min, respectively, and compared with mock seedlings. The phosphopeptides from mock and stress-treated seedlings were purified by polyMAC and dephosphorylated by alkaline phosphatase before *in vitro* kinase reaction. A total of 3,699, 4,484, 4,517, and 4,642 heavy ^{18}O-phosphopeptides were identified from mock, cold, oxidative, and pathogen treated seedlings, respectively. Quantitative analyses were then performed to select the ^{18}O-phosphopeptides highly expressed in stress-stimulated seedlings with proline-directed phosphorylation motif, which are candidate the direct substrates of MAPK3/6 involving in stress responses. A total of 1,998, 2,047, and 2,134 heavy ^{18}O-phosphopeptides were significantly up-regulated (FDR $\leq$ 0.01, s0 is 0.5) upon cold, oxidative, and pathogen stimulation, respectively (Figure 3.6). The Pearson correlation between duplicate kinase reaction experiments is r = 0.89 for duplicate cold stress experiments, and r = 0.91 for duplicate pathogen stress experiments, indicating the in vitro kinase reaction has high reliability for selecting of the direct substrates through quantifying the ^{18}O-phosphopeptides amount difference between mock and stress-treated samples (Figure 3.7). The significantly expressed ^{18}O-phosphorylation sites were analyzed by Motif-X to extract the overrepresented phosphorylation motifs, and the results show that [-(pS/pT)-P-] motifs are dominant in motif analysis, which is the reported MAPK phosphorylation motif (Figure 3.8).

Since KALIP 2.0 approach identified several the direct substrates of SnRK2.6 and MPK6, we further applied this approach to identify the ABA-activated Casein-Kinase Like 2 (CKL2), elicitor-activated protein kinase CDPK11, H2O2-activated protein kinase OXI1, salt-induced protein kinase SOS1, as well as the low-potassium-activated protein kinase CIPK23. The identification and phosphorylation motif results are shown in Table 3.2.

3.5 Discussion

In this study, we describe the KALIP 2.0 approach which uses activated phosphoproteome as kinase-substrate pool in combination with quantitative measurement to select the direct substrates of interested kinase under kinase perturbation. This approach effectively reduces the complexity of the plant proteome through incubating the targeted kinase with stress-activated phosphoproteome. Higher overlap of *in vitro* identification with *in vivo* phosphoproteomic result by this approach indicates better specificity for identifying the direct substrates compared to the use of the whole proteome as reaction pool. The reproducible *in vitro* quantification results enable comparison the kinase perturbation in small kinase-scale phosphoproteome for determining the candidate substrates with high confidence.

The practicality of this approach was shown in identifying the direct SnRK2 substrates in response to ABA stimulation. Identification of several known SnRK2 substrates and classic SnRK2 phosphorylation motifs such as Motif 1 and Motif 2 from our results demonstrates the sensitivity and specificity of this method. Our results suggest that SnRK2 recognizes a new phosphorylation motif [-G-pS-]. We found that the phosphorylation at Ser^{45} of AREB1 and at Ser^{43} of AREB3 which are two known SnRK2 substrates are in the context of phosphorylation Motif 4 (Table 3.1), and two studies also detected the amount of these two phosphorylation sites regulated by SnRK2 perturbation[7, 37]. The glycine residue adjacent to the phosphorylation site is also conserved in the AREB family, indicating the glycine residue is important for SnRK2 phosphorylation on AREB family.

MAPK cascades are also key signaling pathways for initiating physiological responses under environmental stresses. Studies demonstrated that MAPK cascades such as MPK1/2 and MPK 3/6 are activated under ABA stimulation[7], and the potential up-stream activator of MAPK cascades is generation of reactive oxygen species (ROS) by ABA[39-40]. These studies suggested that SnRK2 somehow regulates ABA-dependent MAPK activation. However, the connection between SnRK2 and MAPK cascades remains elusive. Interestingly, we found two candidate SnRK2 substrates involved in MAPK cascades activation. First, a ^{18}O-phosphopeptide, ^{23}FLTQpSGpTFK^{31}, of MKK1 was phosphorylated by SnRK2.6 *in vitro*, and the Thr^{29} phosphorylation site is belong to Motif

4, [-G-pS-]. This phosphopeptide is significantly up-regulated in ABA-activated in vivo phosphoproteomic data as well, indicating that MKK1 is a candidate SnRK2 substrates (Figure 3.9). Second, SnRK2 phosphorylated MAPKKK7 at Ser^{338} with Motif 3 [L-x-x-x-x-pS] *in vitro*, and the amount of corresponding phosphopeptide, ^{331}LPLVGVSpSFR^{340}, is also increased after ABA stimulation *in vivo* (Figure 3.10). These results suggest that SnRK2 may act as alternative regulator to activate MAPK cascades under ABA signaling.

Calcium ion-dependent protein kinases (CDPKs) are another important kinase family involved in ABA signaling. It has been reported that CDPK family is activated as ABA regulator in guard cells and PP2C mediates the activation of SnRK2 and CDPKs[12]. Furthermore, some SnRK2 substrates, AREB/ABF and SLAC1 are reported to be phosphorylated by CDPKs as well[14]. These results indicate that there may be regulatory connections between SnRK2 and CDPK. To our knowledge, the upstream CDPK regulators activating CDPK under ABA stimulation are still not understood. In our *in vitro* kinase reaction dataset, two ^{18}O-phosphopeptides at C-terminus of CDPKs were detected: ^{521}LMREGpSLQLEGEN^{533} of CDPK19 which has Motif 1 and Motif 4 and ^{470}TESSLQPEGELLPIIN^{485} of CDPK27. The same phosphopeptide of CDPK27 was also detected *in vivo* and up-regulated in response to ABA stimulation (Figure 3.11), and the result is consistent with the quantitative result reported by Wang et al.[37]. These results suggest that SnRK2 may be one of the upstream kinases that modulates CDPKs activation under ABA stimulation.

In our list, we also found an interesting protein, PAT1, which is one of the components of mRNA decapping complex or processing bodies. The decapping enzyme DCP1/2 removes the 5' cap of mRNA, and then exoribonuclease degrades the monophosphorylated mRNA. The function of PAT1 in this complex is a mRNA decapping enhancer. Our data show that PAT1 is phosphorylated by SnRK2 *in vitro*, and the phosphorylation amount is increased in response to ABA treatment *in vivo* (Figure 3.12). We generated a T-DNA insertion *pat1*-knockdown mutant plant, and the mutant plants show smaller leaves compared to wild type Arabidopsis. Under ABA treatment, the mutant plants showed poorer growth relative to wild-type plants such as even smaller leaves and shorter primary roots under ABA condition (Figure 3.13). This observation indicates that PAT1 is important in regulating plant growth under osmotic stress conditions.

This method offers deeper extent of identifying underrepresented substrates and delineates the role of kinases in connecting regulatory pathways such as MAPK and CDPK cascades under ABA signaling. We believe that this approach can be applied to study different biological systems and be integrated with other techniques such as genetic mutant for system-wide studying of kinase-substrate networks.

3.6 References

1. Hunter, T., Signaling - 2000 and beyond. *Cell* **2000,** *100* (1), 113-127.

2. Manning, G.; Whyte, D. B.; Martinez, R.; Hunter, T.; Sudarsanam, S., The protein kinase complement of the human genome. *Science* **2002,** *298* (5600), 1912-34.

3. Singh, D. K.; Calvino, M.; Brauer, E. K.; Fernandez-Pozo, N.; Strickler, S.; Yalamanchili, R.; Suzuki, H.; Aoki, K.; Shibata, D.; Stratmann, J. W.; Popescu, G. V.; Mueller, L. A.; Popescu, S. C., The tomato kinome and the tomato kinase library ORFeome: novel resources for the study of kinases and signal transduction in tomato and solanaceae species. *Mol Plant Microbe Interact* **2014,** *27* (1), 7-17.

4. Silva-Sanchez, C.; Li, H.; Chen, S., Recent advances and challenges in plant phosphoproteomics. *Proteomics* **2015,** *15* (5-6), 1127-41.

5. Marx, H.; Minogue, C. E.; Jayaraman, D.; Richards, A. L.; Kwiecien, N. W.; Siahpirani, A. F.; Rajasekar, S.; Maeda, J.; Garcia, K.; Del Valle-Echevarria, A. R.; Volkening, J. D.; Westphall, M. S.; Roy, S.; Sussman, M. R.; Ane, J. M.; Coon, J. J., A proteomic atlas of the legume Medicago truncatula and its nitrogen-fixing endosymbiont Sinorhizobium meliloti. *Nat Biotechnol* **2016,** *34* (11), 1198-1205.

6. Zhu, J.-K., Abiotic Stress Signaling and Responses in Plants. *Cell* **2016,** *167* (2), 313-324.

7. Umezawa, T.; Sugiyama, N.; Takahashi, F.; Anderson, J. C.; Ishihama, Y.; Peck, S. C.; Shinozaki, K., Genetics and Phosphoproteomics Reveal a Protein Phosphorylation Network in the Abscisic Acid Signaling Pathway in Arabidopsis thaliana. *Sci Signal* **2013,** *6* (270).

8. Fujii, H.; Zhu, J. K., Arabidopsis mutant deficient in 3 abscisic acid-activated protein kinases reveals critical roles in growth, reproduction, and stress. *Proc Natl Acad Sci U S A* **2009,** *106* (20), 8380-5.

9. Umezawa, T.; Sugiyama, N.; Mizoguchi, M.; Hayashi, S.; Myouga, F.; Yamaguchi-Shinozaki, K.; Ishihama, Y.; Hirayama, T.; Shinozaki, K., Type 2C protein phosphatases directly regulate abscisic acid-activated protein kinases in Arabidopsis. *P Natl Acad Sci USA* **2009,** *106* (41), 17588-17593.

10. Boudsocq, M.; Barbier-Brygoo, H.; Lauriere, C., Identification of nine sucrose nonfermenting 1-related protein kinases 2 activated by hyperosmotic and saline stresses in Arabidopsis thaliana. *J Biol Chem* **2004,** *279* (40), 41758-66.

11. Fujii, H.; Verslues, P. E.; Zhu, J. K., Arabidopsis decuple mutant reveals the importance of SnRK2 kinases in osmotic stress responses in vivo. *Proc Natl Acad Sci U S A* **2011,** *108* (4), 1717-22.

12. Asano, T.; Hayashi, N.; Kikuchi, S.; Ohsugi, R., CDPK-mediated abiotic stress signaling. *Plant Signal Behav* **2012,** *7* (7), 817-21.

13. Brandt, B.; Munemasa, S.; Wang, C.; Nguyen, D.; Yong, T.; Yang, P. G.; Poretsky, E.; Belknap, T. F.; Waadt, R.; Aleman, F.; Schroeder, J. I., Calcium specificity signaling mechanisms in abscisic acid signal transduction in Arabidopsis guard cells. *Elife* **2015,** *4*.

14. Geiger, D.; Scherzer, S.; Mumm, P.; Marten, I.; Ache, P.; Matschi, S.; Liese, A.; Wellmann, C.; Al-Rasheid, K. A.; Grill, E.; Romeis, T.; Hedrich, R., Guard cell anion channel SLAC1 is regulated by CDPK protein kinases with distinct Ca2+ affinities. *Proc Natl Acad Sci U S A* **2010,** *107* (17), 8023-8.

15. Maierhofer, T.; Diekmann, M.; Offenborn, J. N.; Lind, C.; Bauer, H.; Hashimoto, K.; KA, S. A.-R.; Luan, S.; Kudla, J.; Geiger, D.; Hedrich, R., Site- and kinase-specific phosphorylation-mediated activation of SLAC1, a guard cell anion channel stimulated by abscisic acid. *Sci Signal* **2014,** *7* (342), ra86.

16. Drerup, M. M.; Schlucking, K.; Hashimoto, K.; Manishankar, P.; Steinhorst, L.; Kuchitsu, K.; Kudla, J., The Calcineurin B-like calcium sensors CBL1 and CBL9 together with their interacting protein kinase CIPK26 regulate the Arabidopsis NADPH oxidase RBOHF. *Mol Plant* **2013,** *6* (2), 559-69.

17. Zhao, S.; Jiang, Y.; Zhao, Y.; Huang, S.; Yuan, M.; Zhao, Y.; Guo, Y., CASEIN KINASE1-LIKE PROTEIN2 Regulates Actin Filament Stability and Stomatal Closure via Phosphorylation of Actin Depolymerizing Factor. *Plant Cell* **2016,** *28* (6), 1422-39.

18. Zhu, J. K., Salt and drought stress signal transduction in plants. *Annu Rev Plant Biol* **2002,** *53*, 247-73.

19. Choi, W. G.; Toyota, M.; Kim, S. H.; Hilleary, R.; Gilroy, S., Salt stress-induced Ca2+ waves are associated with rapid, long-distance root-to-shoot signaling in plants. *Proc Natl Acad Sci U S A* **2014,** *111* (17), 6497-502.

20. Shi, H.; Ishitani, M.; Kim, C.; Zhu, J. K., The Arabidopsis thaliana salt tolerance gene SOS1 encodes a putative Na+/H+ antiporter. *Proc Natl Acad Sci U S A* **2000,** *97* (12), 6896-901.

21. Tena, G.; Asai, T.; Chiu, W. L.; Sheen, J., Plant mitogen-activated protein kinase signaling cascades. *Curr Opin Plant Biol* **2001,** *4* (5), 392-400.

22. Meng, X.; Zhang, S., MAPK cascades in plant disease resistance signaling. *Annu Rev Phytopathol* **2013,** *51*, 245-66.

23. Feilner, T.; Hultschig, C.; Lee, J.; Meyer, S.; Immink, R. G.; Koenig, A.; Possling, A.; Seitz, H.; Beveridge, A.; Scheel, D.; Cahill, D. J.; Lehrach, H.; Kreutzberger, J.; Kersten, B., High throughput identification of potential Arabidopsis mitogen-activated protein kinases substrates. *Mol Cell Proteomics* **2005,** *4* (10), 1558-68.

24. Hoehenwarter, W.; Thomas, M.; Nukarinen, E.; Egelhofer, V.; Rohrig, H.; Weckwerth, W.; Conrath, U.; Beckers, G. J., Identification of novel in vivo MAP kinase substrates in Arabidopsis thaliana through use of tandem metal oxide affinity chromatography. *Mol Cell Proteomics* **2013,** *12* (2), 369-80.

25. Boudsocq, M.; Willmann, M. R.; McCormack, M.; Lee, H.; Shan, L.; He, P.; Bush, J.; Cheng, S. H.; Sheen, J., Differential innate immune signalling via Ca(2+) sensor protein kinases. *Nature* **2010,** *464* (7287), 418-22.

26. Ding, Y.; Li, H.; Zhang, X.; Xie, Q.; Gong, Z.; Yang, S., OST1 kinase modulates freezing tolerance by enhancing ICE1 stability in Arabidopsis. *Dev Cell* **2015,** *32* (3), 278-89.

27. Dubiella, U.; Seybold, H.; Durian, G.; Komander, E.; Lassig, R.; Witte, C. P.; Schulze, W. X.; Romeis, T., Calcium-dependent protein kinase/NADPH oxidase activation circuit is required for rapid defense signal propagation. *Proc Natl Acad Sci U S A* **2013,** *110* (21), 8744-9.

28. Xue, L.; Wang, W. H.; Iliuk, A.; Hu, L. H.; Galan, J. A.; Yu, S.; Hans, M.; Geahlen, R. L.; Tao, W. A., Sensitive kinase assay linked with phosphoproteomics for identifying direct kinase substrates. *P Natl Acad Sci USA* **2012,** *109* (15), 5615-5620.

29. Xue, L.; Wang, P.; Cao, P.; Zhu, J. K.; Tao, W. A., Identification of extracellular signal-regulated kinase 1 (ERK1) direct substrates using stable isotope labeled kinase assay-linked phosphoproteomics. *Mol Cell Proteomics* **2014,** *13* (11), 3199-210.

30. 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), 2162-72.

31. Cox, J.; Mann, M., MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nature Biotechnology* **2008,** *26* (12), 1367-1372.

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

33. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K. P.; Kuhn, M.; Bork, P.; Jensen, L. J.; von Mering, C., STRING v10: protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res* **2015,** *43* (D1), D447-D452.

34. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N. S.; Wang, J. T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T., Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res* **2003,** *13* (11), 2498-2504.

35. Huang, D. W.; Sherman, B. T.; Lempicki, R. A., Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **2009,** *4* (1), 44-57.

36. 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.

37. Wang, P. C.; Xue, L.; Batelli, G.; Lee, S.; Hou, Y. J.; Van Oosten, M. J.; Zhang, H. M.; Tao, W. A.; Zhu, J. K., Quantitative phosphoproteomics identifies SnRK2 protein

kinase substrates and reveals the effectors of abscisic acid action. *P Natl Acad Sci USA* **2013,** *110* (27), 11205-11210.

38. Soma, F.; Mogami, J.; Yoshida, T.; Abekura, M.; Takahashi, F.; Kidokoro, S.; Mizoi, J.; Shinozaki, K.; Yamaguchi-Shinozaki, K., ABA-unresponsive SnRK2 protein kinases regulate mRNA decay under osmotic stress in plants. *Nat Plants* **2017,** *3*, 16204.

39. Xing, Y.; Jia, W.; Zhang, J., AtMKK1 mediates ABA-induced CAT1 expression and H2O2 production via AtMPK6-coupled signaling in Arabidopsis. *Plant J* **2008,** *54* (3), 440-51.

40. Wang, X. J.; Zhu, S. Y.; Lu, Y. F.; Zhao, R.; Xin, Q.; Wang, X. F.; Zhang, D. P., Two coupled components of the mitogen-activated protein kinase cascade MdMPK1 and MdMKK1 from apple function in ABA signal transduction. *Plant Cell Physiol* **2010,** *51* (5), 754-66.

Table 3.1: Selected examples of confirmed and putative SnRK2 substrates identified in our list.

Protein	^{18}O-Phosphopeptides	Motif	Status
SnRK2.2	^{168}SSVLHSQPKpSTVGTPAYIAPEILLR^{192}	-	confirmed
AREB1	^{42}DFGpSMNMDELLK^{53}	[-G-pS-]	confirmed
AREB3	^{40}ALGpSMNLDELLK^{51}	[-G-pS-]	confirmed
MSL9	^{115}SLREQFGAGpSFAR^{127}	[-G-pS-]	confirmed
BTR1L	^{39}FLVpSNAAAGpSVIGK^{52}	[-G-pS-]	confirmed
FSD2	^{253}LVSWEpTVSTRLESAIAR^{269}	[-Lx-x-x-x-pS-]	confirmed
CIP7	^{485}ISpSTALNMASEVVRK^{499}	[-Lx-R-x-x-pS-]	putative
VCS	^{568}PSIIVNRSEpSANK^{580}	[-R-x-x-pS-]	confirmed
AT1G72410	^{541}NLpSELRFSDDSK^{552}	[-R-x-x-pS-]	putative
AT4G39680	^{429}RDFSRSDpSpSVSEDGPK^{444}	[-R-x-x-pS-]	putative

Table 3.2: The results of five *in vitro* kinase reactions.

Kinase	Treatment	Up-regulated ^{18}O-phosphosites	Up-regulated Phosphoproteins	Major Motif
SOS2	NaCl	345	301	-L-x-R-x-x-(pS/T)-
CKL2	ABA	281	227	-E-x-(pS/T)-E-
OXI1	H2O2	74	71	-(pS/T)-F-
CIPK23	Low K	65	62	-R-x-x-(pS/T)-
CPK11	Flg22	123	109	-R-x-x-(pS/T)-

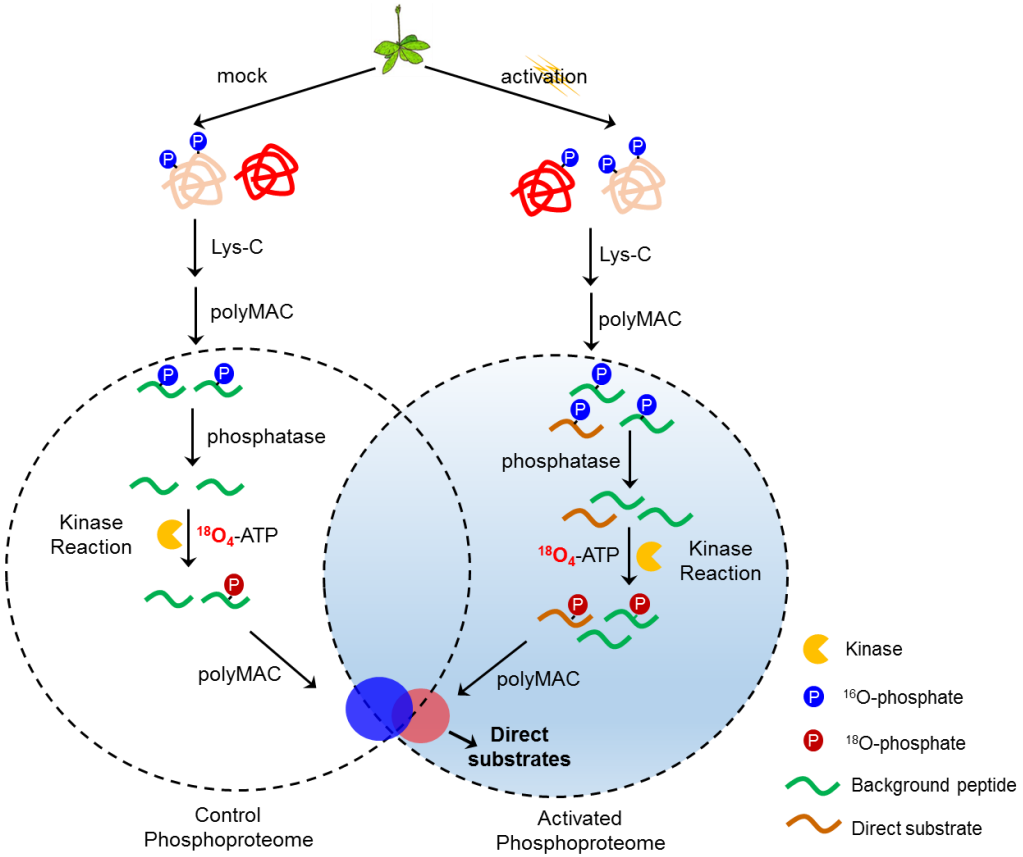


Figure 3.1: The KALIP 2.0 workflow for kinase-substrate identification through quantitative *in vitro* kinase screening**.** The mock and stress-activated seedlings were digested with Lys-C, and the resulting long phosphopeptides (blue) were enriched by polyMAC. The phosphopeptides were dephosphorylated and incubated with recombinant kinase with γ-$^{18}O_4$-ATP, and ^{18}O-phosphate (red) labeled-peptides were enriched again through polyMAC and analyzed by LC-MS/MS. The direct substrates of targeted kinase are selected from stress-activated seedlings in which the ^{18}O-phosphopeptides (red) were significantly enriched than those (green) were unchanged in both seedlings.

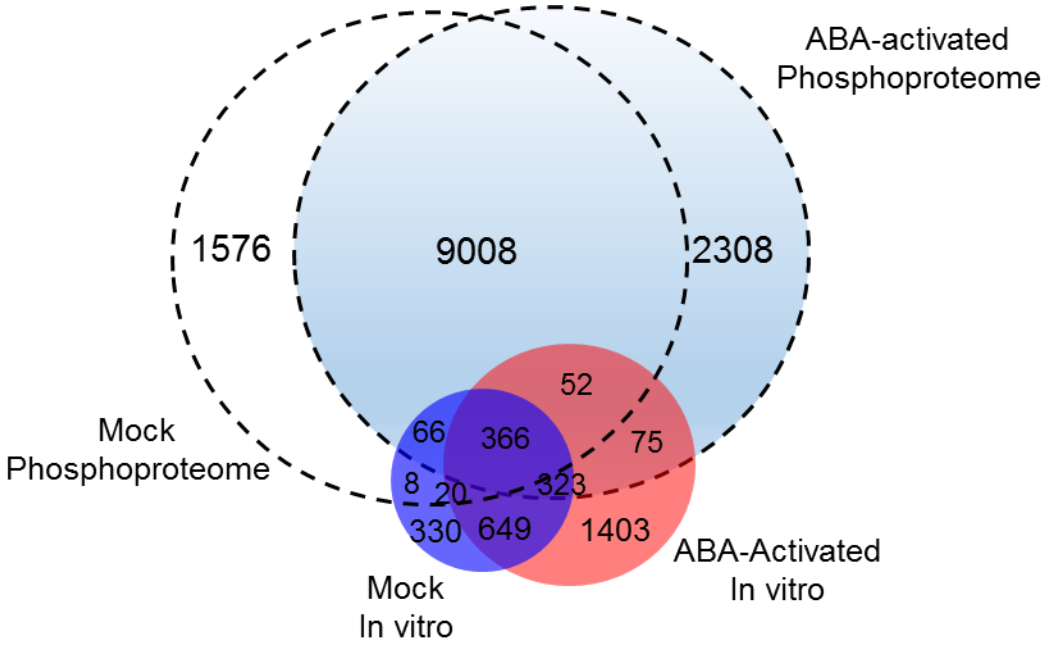


Figure 3.2: The sensitivity of the KAILP 2.0 approach. Venn diagram shows the overlap of identified in vivo phosphopeptides and in vitro ^{18}O-phosphopeptides. Up to 45% of identified *in vitro* ^{18}O-phosphopeptides from ABA-treated seedlings are matched to *in vivo* ABA-activated phosphoproteomics, indicating the sensitivity and specificity of the KALIP 2.0 approach.

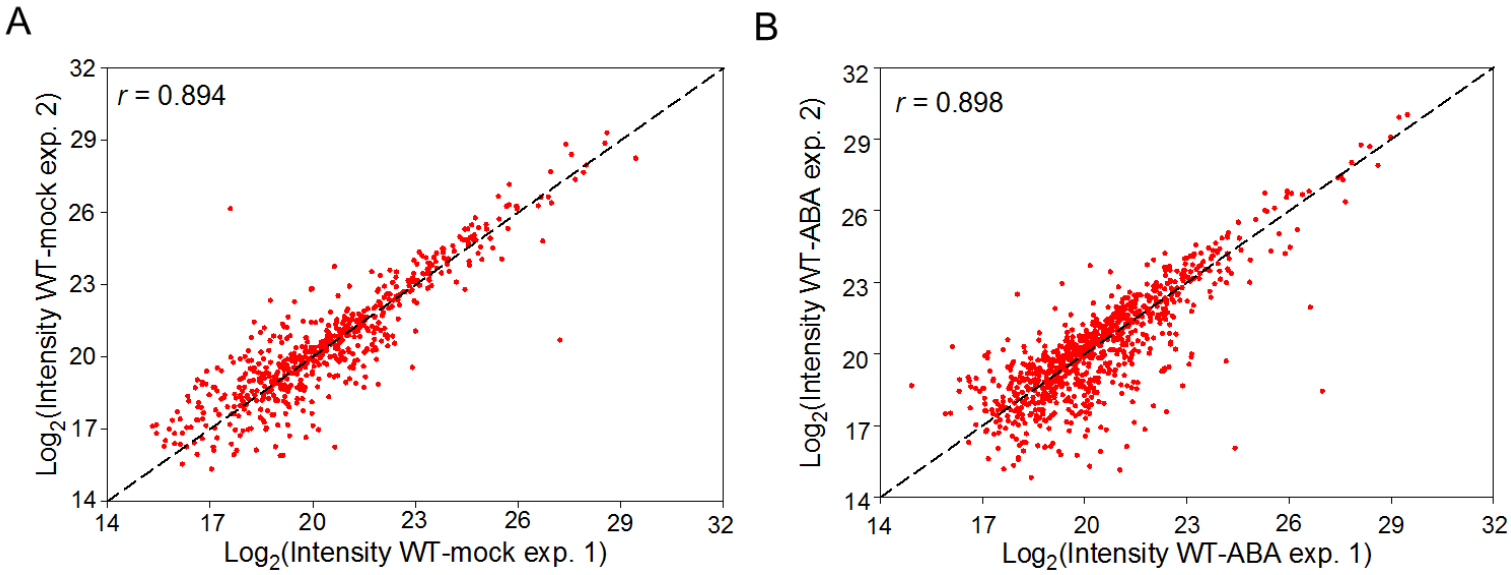


Figure 3.3: The reproducibility of the KAILP 2.0 approach. The scatter plots show the $\log_2$ transformed extracted ion chromatogram (XIC) of identified ^{18}O-phosphopeptides from duplicate experiments of (A) mock and (B) ABA-treated seedlings. The Pearson coefficient between replicate experiments is 0.894 for mock and 0.898 for ABA-treated *in vitro* kinase screening respectively, indicating the good reproducibility of ^{18}O-phosphopeptide quantification for identifying the direct substrates of SnRK2.6.

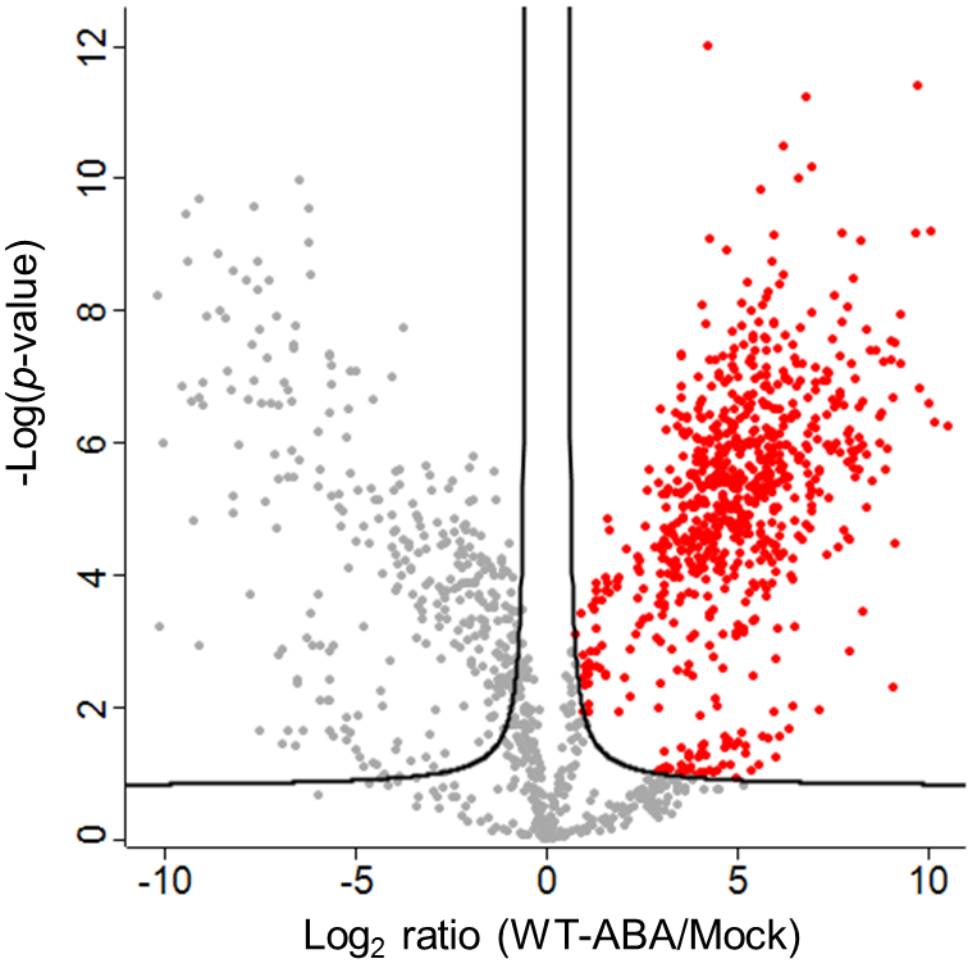


Figure 3.4: Filtering of ABA-induced ^{18}O-phosphopeptides from SnRK2.6 *in vitro* kinase screening. The volcano plot represents the comparison of ABA-stimulated versus mock sample. ABA responsive ^{18}O-phosphopeptides that were identified through a permutation-based FDR t-test (FDR = 0.01, and S0 = 0.35). The ^{18}O-phosphopeptides enriched in response to ABA stimulation are colored in red.

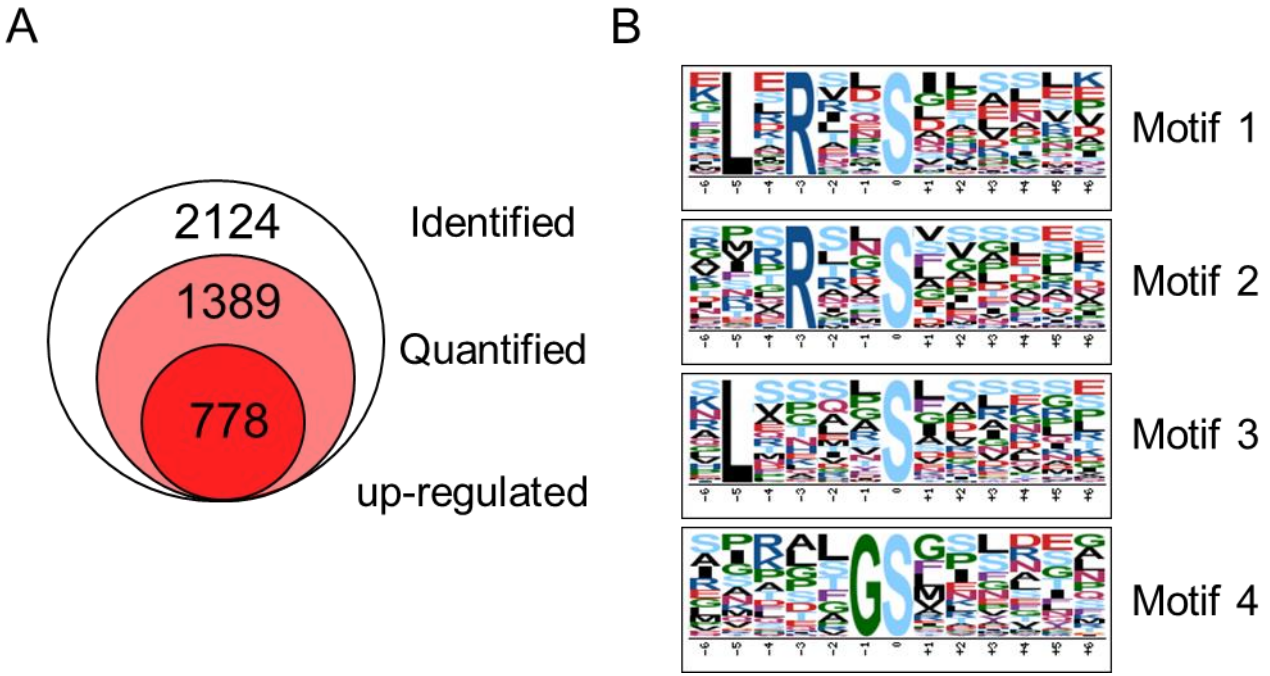


Figure 3.5: Analysis of ABA-induced ^{18}O-phosphopeptides from SnRK2.6 *in vitro* kinase screening. (A) The number of identified (class 1), quantified (class 2), and significantly up-regulated ^{18}O-phosphopeptides (class 3) after kinase reaction. (B) SnRK2.6 phosphorylation motifs were enriched from the sequence windows of significantly enriched ^{18}O-phosphorlyation sites in ABA-treated seedlings using Motif-X algorithm.

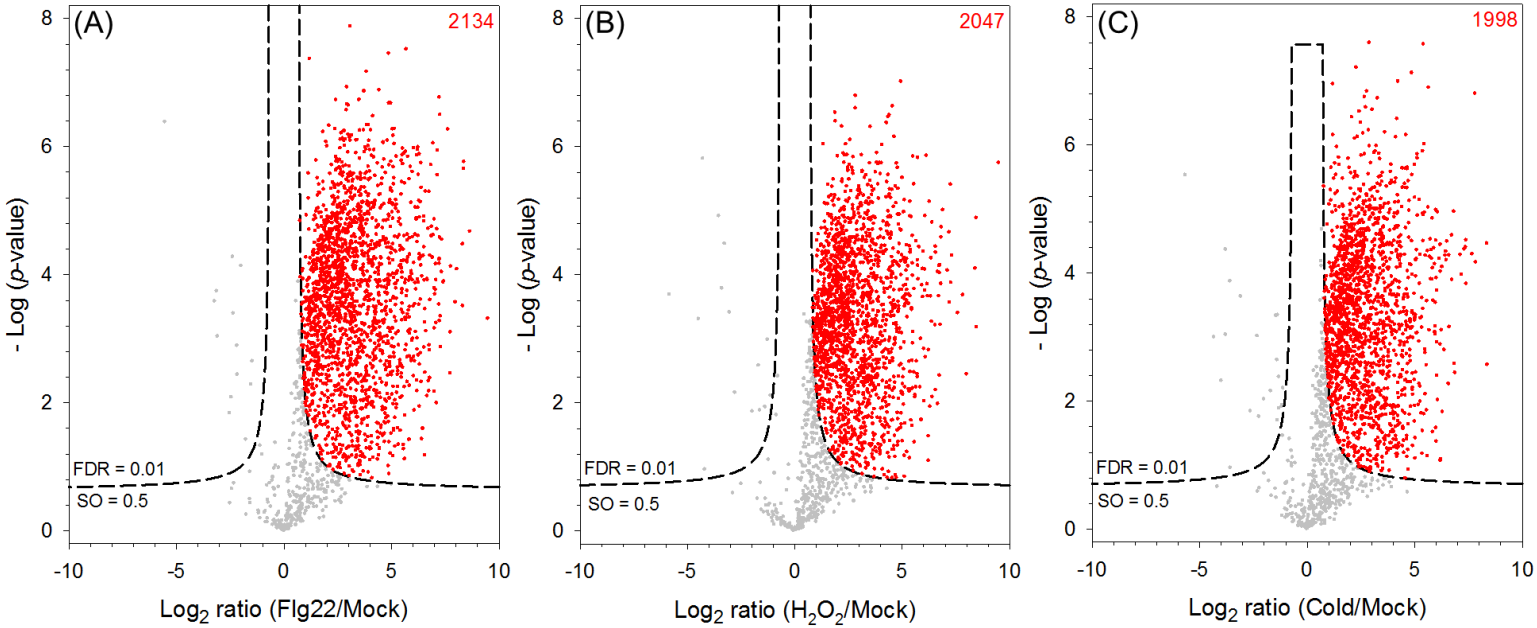


Figure 3.6: Selecting of stress-inducible phosphopeptides from in vitro MAPK6 kinase reaction. The volcano plots representing the comparison of (A) pathogen, (B) oxidative, and (C) cold stress-stimulated versus control seedlings. Stress-inducible phosphopeptides were identified through a permutation-based FDR t test (FDR ≤ 0.01, SO = 0.5). Phosphorylation sites up-regulated in response to stress are colored in red. The number of significantly up-regulated phosphopeptides after stress treatment is shown in up right side of the figure.

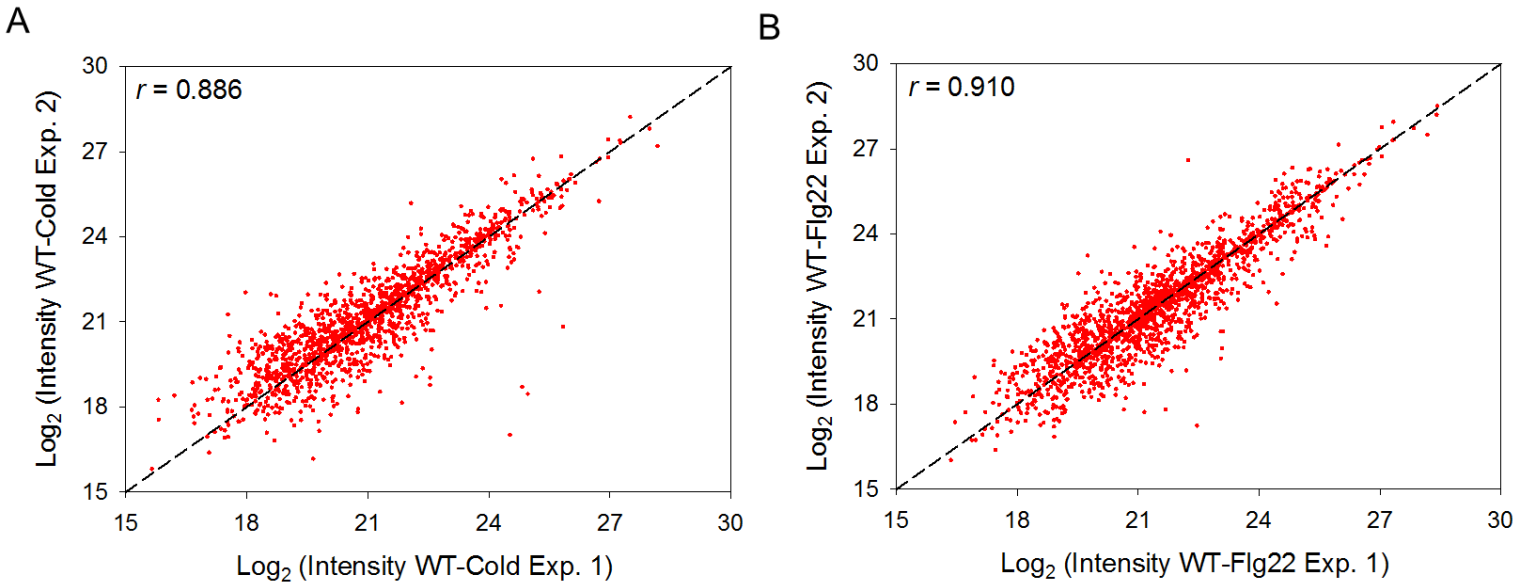


Figure 3.7: The reproducibility of the in vitro MAPK6 kinase reactions. The scatter plots show the log_2 transformed extracted ion chromatogram (XIC) of identified ^{18}O-phosphopeptides from duplicate experiments of cold and flg22-treated seedlings. The Pearson coefficient between replicate experiments is 0.886 for cold-treated and 0.910 for flg22-treated *in vitro* kinase screening respectively, indicating the good reproducibility of ^{18}O-phosphopeptide quantification for identifying the direct substrates of MAPK6.

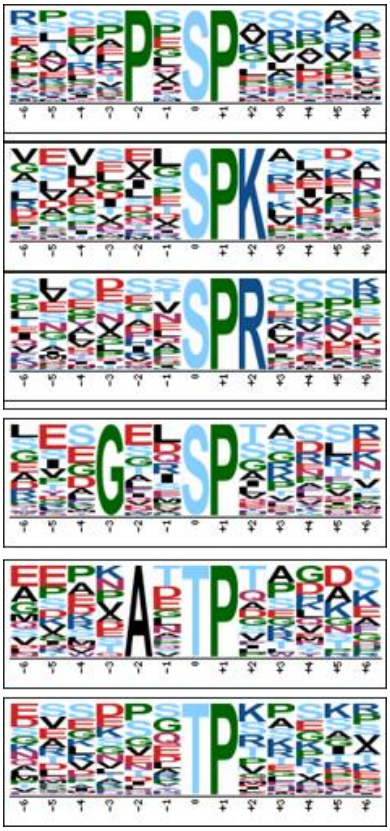


Figure 3.8: Motif analysis of Candidate MAPK 6 substrate by Motif-X, and the results indicate that proline-directed motifs are dominant in candidate MAPK 6 substrates.

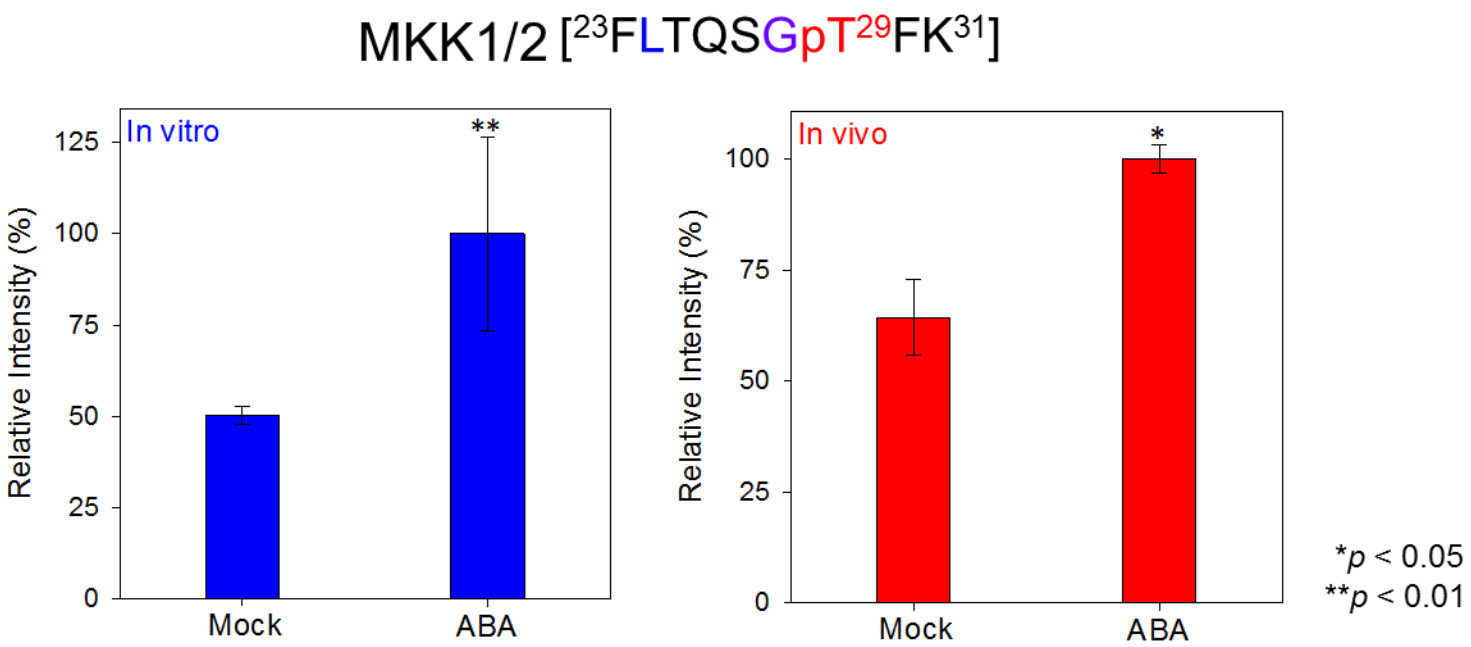


Figure 3.9: MKK1/2 are candidate SnRK2.6 direct substrates. The label-free quantitation of the Thr[29] [18]O-phosphopeptide of MKK1/2 in *in vitro* and *in vivo* mock-treated and ABA-treated seedlings is represented in two bar chart.

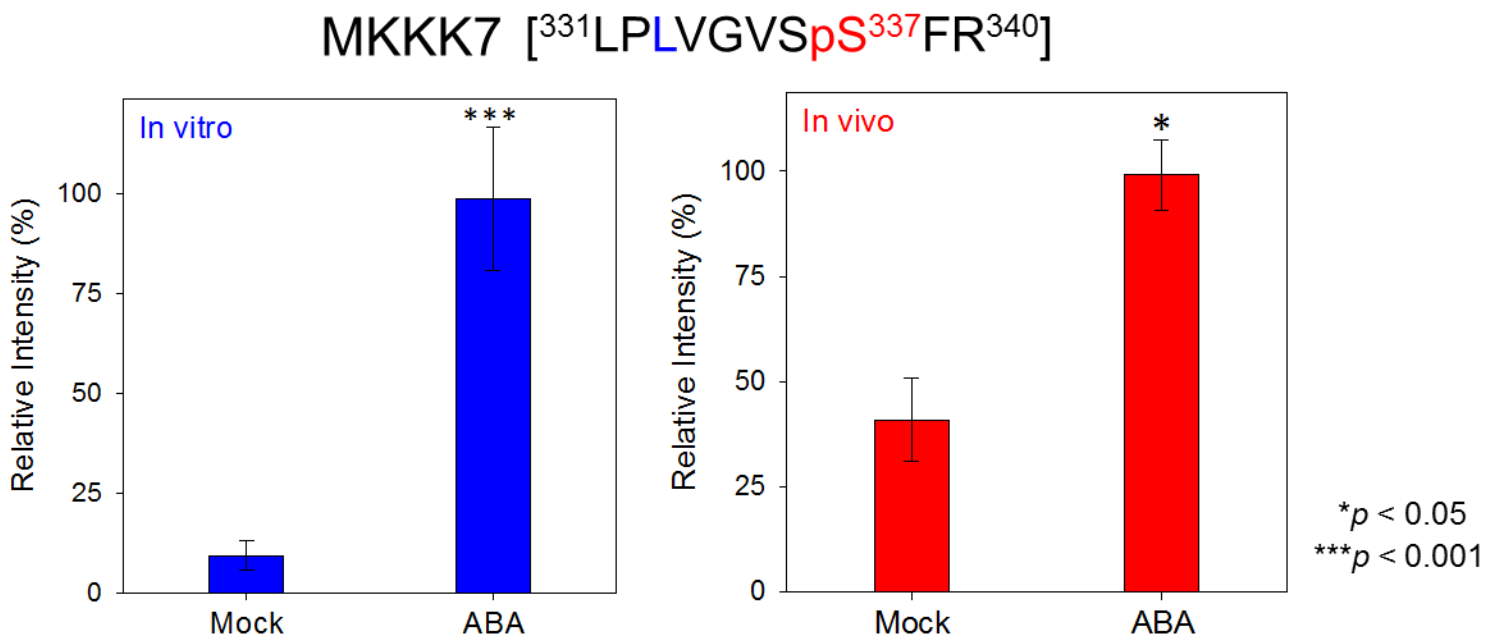


Figure 3.10: MAPKKK7 are candidate SnRK2.6 direct substrates. The label-free quantitation of the Ser^{337} ^{18}O-phosphopeptide of MAPK kinase kinase (MAPKKK) in *in vitro* and *in vivo* mock-treated and ABA-treated seedlings is represented in two bar chart.

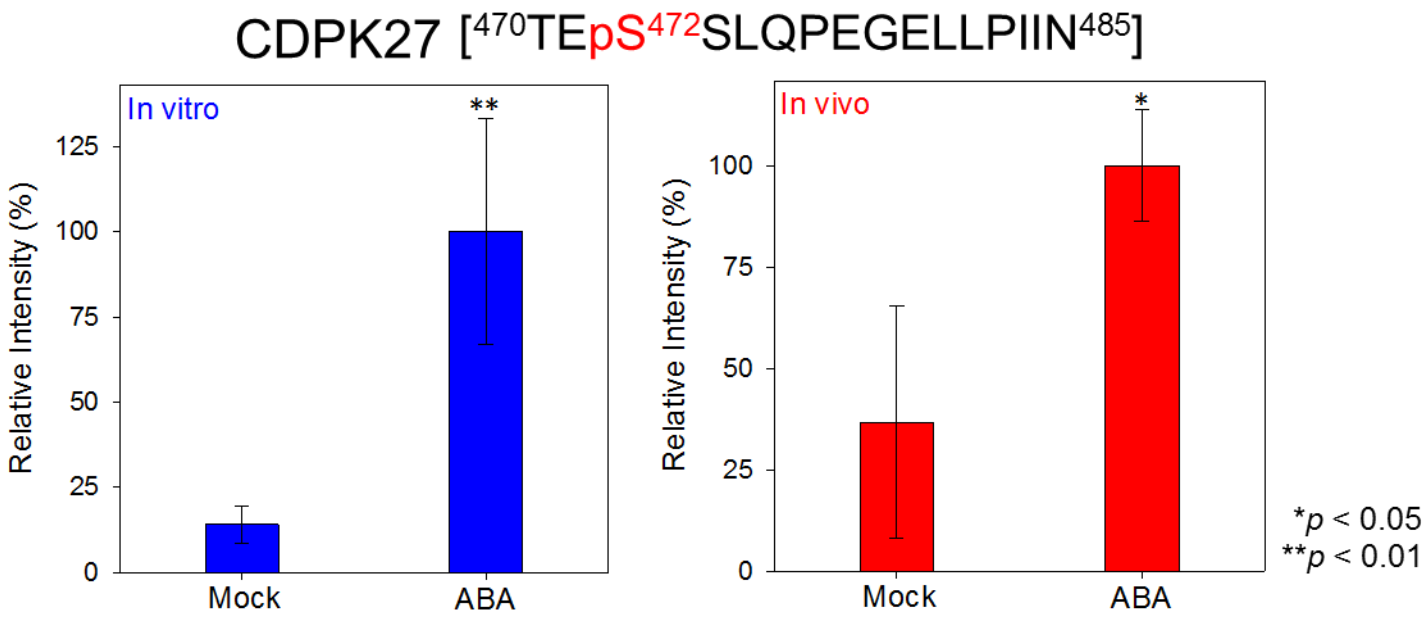


Figure 3.11: CDPK27 are candidate SnRK2.6 direct substrates. The extracted ion chromatogram (XIC) of the Ser^{472} ^{18}O-phosphopeptide of Ca^{2+}-dependent protein kinase 27 (CDPK27) in *in vitro* and *in vivo* mock-treated and ABA-treated seedlings is represented in two bar chart.

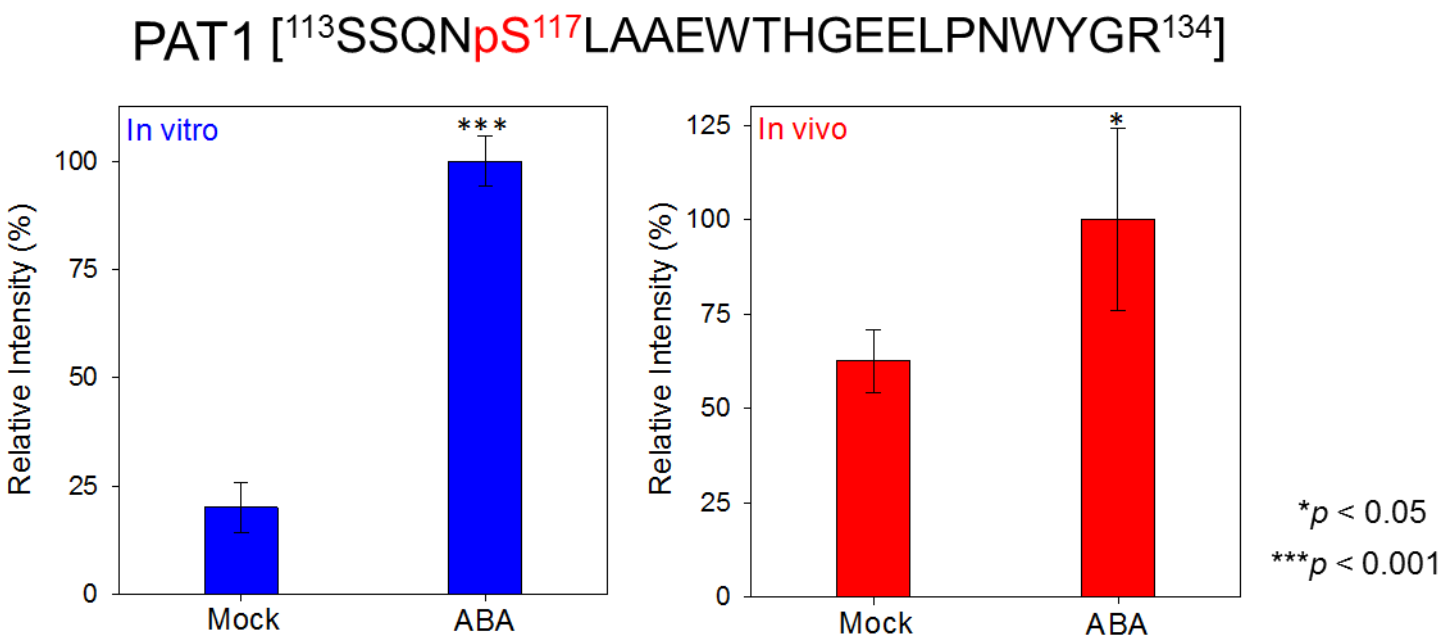


Figure 3.12: PAT1 are candidate SnRK2.6 direct substrates. The extracted ion chromatogram (XIC) of the Ser^{117} ^{18}O-phosphopeptide of PAT1 in *in vitro* and *in vivo* mock-treated and ABA-treated seedlings is represented in two bar chart.

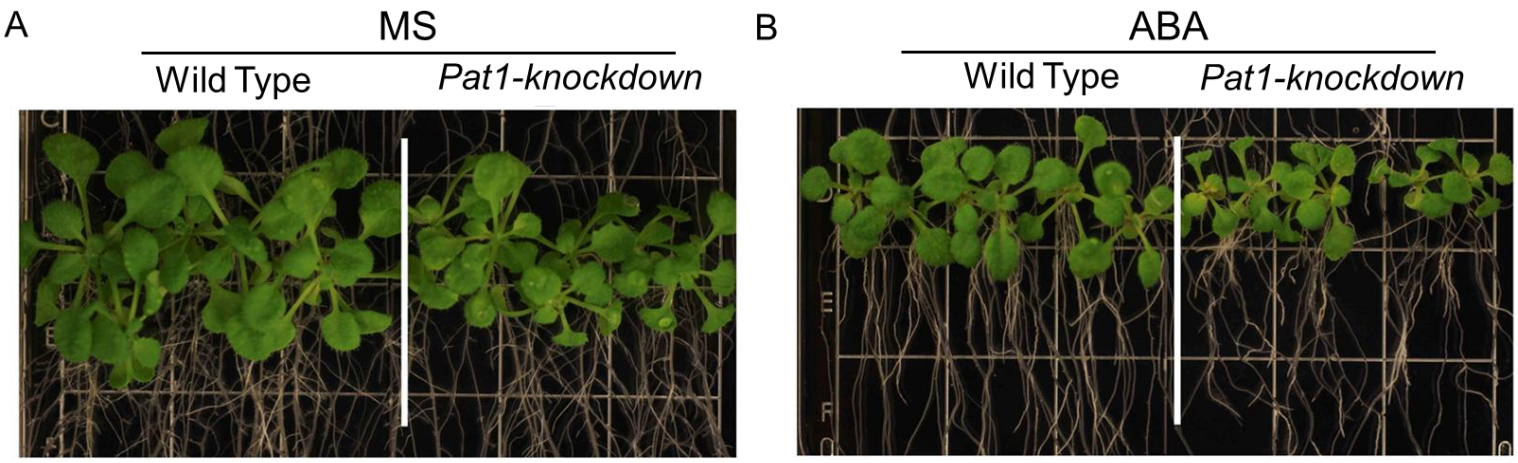


Figure 3.13: Effect of ABA on wild type and *pat1*-knockdown mutant seedlings. Treatment of ABA inhibited the growth of *pat1*-knockdown seedlings, indicating PAT1 regulates the growth of plant under ABA condition.

VITA

Chuan-Chih Hsu is the son of Mu-Yu Hsu and Wen-Chih Lo. Originally from Taichung City, Taiwan, Chuan-Chih graduated from Taichung First Senior High School in 2003. He then went on to obtain a BS in Agricultural Chemistry from National Taiwan University in Taipei City, Taiwan in 2007. Chuan-Chih studied in the lab of Dr. Ren-Shih Chung at Department of Agricultural Chemistry during his final year as an undergraduate student. Chuan-Chih then decided to study analytical chemistry so he spent two years to complete a MS in Chemistry at National Taiwan University with Dr. Yu-Ju Chen as his advisor in 2009. The title of his MS thesis is Developed a Nanoprobe-based Immobilized Metal Affinity Chromatography (NB-IMAC) Approach for Multiply Phosphorylated Peptides Enrichment. Chuan-Chih liked his research so much that he decided to continually work in the lab of Dr. Yu-Ju Chen for the research and development substitute service until June, 2012.

Chuan-Chih then moved to West Lafayette, Indiana, to pursue a PhD in Biochemistry at Purdue University in the fall of 2012. After one year lab rotation, he joined the lab of Dr. Weiguo Andy Tao and spent the rest of his time at Purdue designing novel (phospho)proteomic strategies toward comprehensive phosphoproteomic analysis and investigating plant phosphorylation signaling pathways in response to environmental stresses.

PUBLICATIONS

1. Zhao, C.; Wang, P.; Si, T.; Hsu, C.-C.; Wang, L.; Zayed, O.; Yu, Z.; Zhu, Y.; Dong, J.; Tao, W. A.; Zhu, J.-K. Antagonistic MAP kinase cascades regulate cold response by modulating ICE1 protein stability. Dev. Cell [under revision]
2. Wang, P.*; Zhao, Y.*; Li, Z.*; Hsu, C.-C.*; Liu, X.; Fu, L.; Hou, Y.-J.; Xie, S.; Zhang, C.; Gao, J.; Cao, M.; Huang, X.; Zhu, Y.; Tang, K.; Tao, W. A.; Xiong, Y.; Zhu, J.-K. Reciprocal regulation of the TOR kinase and ABA receptor balances plant growth and defense. Mol. Cell [under revision] (*: equal contribution)
3. Zhu, Y.; Wang, B.; Xie, S.; Hsu, C.-C.; Du, H.; Tang, K.; Yang, Y.; Tao, W. A.; Zhu, J.-K. An Arabidopsis Nucleoporin NUP85 is essential for plant responses to ABA and salt stress. PLoS Genet. [under revision]
4. Sen. A.; Hsieh, W.-C.; Hanna, C. B.; Hsu, C.-C.; Pearson, M.; Tao, W. A.; Aguilar, R. C. Coincidence Detection of Ubiquitination and Phosphorylation is Required for Internalization of an Epsin-Specific Cargo in S. cerevisiae. Proc. Natl. Acad. Sci. U.S.A. [under revision]
5. Duan, C.; Wang, X.; Zhang, L.; Xiong, X.; Zhang, Z.; Tang, K.; Pan, L.; Hsu, C.-C.; Xu, H.; Tao, W. A.; Zhang, H.; Zhu, J.-K. A protein complex regulates RNA processing of intronic heterochromatin-containing genes in Arabidopsis. Proc. Natl. Acad. Sci. U.S.A. 2017 [accepted]
6. Hsu, C.-C.; Arrington, J. V.; Xue, L.; Tao, W. A. Identification of Plant Kinases Substrates using Stable Isotope Labeled Kinase Assay Linked Phosphoproteomics Kinase Signaling Networks, 2017, 327-335.
7. Chen, I.-H.; Xue, L.; Hsu, C.-C.; Paez, J. S. P.; Pan, L.; Andaluz, H.; Went, M. K.; Iliuk, A. B.; Zhu, J.-K.; Tao, W. A. Phosphoproteins in extracellular vesicles as candidate markers for breast cancer. Proc. Natl. Acad. Sci. U.S.A. 2017, 114, 3175-3180.
8. Yan, J.; Wang, P.; Wang, B.; Hsu, C.-C.; Zhang, H.; Hou, Y.-J.; Zhou, Y.; Tang, K.; Wang, Q.; Zhao, C.; Zhu, X.; Tao, W. A.; Li, J.; Zhu, J.-K. The SnRK2 kinases modulate miRNA accumulation in Arabidopsis. PLoS Genet. 2017, 13 (4), e1006753.
9. Hsu, C.-C.; Xue, L.; Arrington, J. V.; Wang, P.; Paez, J. S. P.; Zhou, Y.; Zhu, J.-K.; Tao, W. A. Estimate phosphopeptide identification rate by tandem mass spectrometry. J. Am. Soc. Mass Spectrom. 2017, 28, 1127-1135.
10. Arrington, J. V.; Hsu, C.-C.; Tao, W. A. Kinase assay-linked phosphoproteomics: Discovery of direct kinase substrates. Methods Enzymol. 2017, 586, 453-471.
11. Duan, C.; Wang, X.; Xie, S.; Pan, L.; Miki, D.; Tang, K.; Hsu, C.-C.; Lei, M.; Zhong, Y.; Hou, Y.-J.; Wang, Z.; Zhang, Z.; Mangrauthia, S.; Xu, H.; Zhang, H.; Dilkes, B.; Tao, W. A., Zhu, J.-K. A Pair of Transposon-derived Proteins Function in A Histone Acetyltransferase Complex for Active DNA Demethylation. Cell Res. 2017, 27, 226

12. Duan, C.; Wang, X.; Tang, K.; Zhang, H.; Mangrauthia, S. K.; Lei, M.; Hsu, C.-C.; Hou, Y.-J.; Wang, C.; Li, Y.; Tao, W. A.; Zhu, J.-K. MET18 Connects the Cytosolic Iron-Sulfur Cluster Assembly Pathway to Active DNA Demethylation in Arabidopsis. PLoS Genet. 2015, 11, e1005559.
13. Pan, L.; Wang, L.; Hsu, C.-C.; Zhang, J.; Iliuk, A.; Tao, W. A. A sensitive assay to measure total protein phosphorylation level in complex protein samples. Analyst 2015, 140, 3390-3396.
14. Lang, Z.; Lei, M.; Wang, X.; Tang, K.; Miki, D.; Zhang, H.; Mangrauthia, S. K.; Liu, W.; Nie, W.; Ma, G.; Yan J.; Duan, C.-G.; Hsu, C.-C.; Wang, C.; Tao, W. A.; Gong, Z.; Zhu J.-K. The Methyl-CpG-Binding Protein MBD7 Facilitates Active DNA Demethylation to Limit DNA Hyper-Methylation and Transcriptional Gene Silencing. Mol. Cell 2015, 57, 971-983.
15. Wang, P.; Du, Y.; Ho, J.-J.; Zhao, Y.; Hsu, C.-C.; Yuan, F.; Zhu, X.; Tao, W. A.; Song, C.-P.; Zhu, J.-K. Nitric oxide negatively regulates abscisic acid signaling in guard cells by S-nitrosylation of OST1. Proc. Natl. Acad. Sci. U.S.A. 2015, 112, 613-618.