Technologies for Phosphoproteomic Interrogation of T Cell Signaling

By

Judson McKnight Belmont

B.S., University of Rochester, Rochester, NY, 2010

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in the Department of Molecular Biology, Cell Biology, and Biochemistry in the Division of Biology and Medicine at Brown University

Providence, Rhode Island

May 2018

This dissertation by Judson McKnight Belmont is accepted in its present form by the Division of Biology and Medicine as satisfying the dissertation requirement for the degree of Doctor of Philosophy.

Date _____ _____

Arthur R. Salomon, Ph.D., Advisor

Recommended to the Graduate Council

Date _____ _____

Phillip A. Gruppuso, M.D. Reader

Date _____ _____

Nicola Neretti, Ph.D., Reader

Date _____ _____

Christopher de Graffenreid, Ph.D., Reader

Date _____ _____

Scott Gerber, Ph.D., Outside Reader
Dartmouth College

Approved by the Graduate Council

Date _____ _____

Andrew G. Campbell, Ph.D.
Dean of the Graduate School

Curriculum Vitae

Education

2012-2017 Ph.D., Molecular Biology, Cell Biology, and Biochemistry Graduate Program
Concentration in Functional and Computational Genomics
Brown University, Providence, RI
Dr. Arthur R. Salomon, Advisor

2006-2010 Bachelor of Science, Molecular Genetics
Bachelor of Arts, Philosophy
University of Rochester, Rochester, NY
Cum laude, Distinction in the Field

Professional Research Experience

2012-2017 Graduate Student, Brown University, Providence, RI
Molecular Biology, Cell Biology, and Biochemistry (Dr. Arthur Salomon)
Research: Development of computational tools and strategies for the analysis of phosphoproteomic data, and application of these methods to investigate signal transduction mechanisms in T cell receptor signaling.

2010-2012 SUNY Upstate Medical University, Syracuse, NY
Department of Neuroscience (Dr. Eric Olson)
Research: Physiological characterization of the roles of focal adhesion proteins in early cortical development.

2009-2010 University of Rochester, Rochester, NY
Department of Biology (Dr. Daniel Garrigan)
Research: Investigation of the global distribution of chromosomal inversions in human populations.

Publications

Belmont J, Gu T, Mudd A, Salomon AR (2017) A PLC-γ1 Feedback Pathway Regulates Lck Substrate Phosphorylation at the T Cell Receptor and SLP-76 Complex. *Journal of Proteome Research* PMID: 2864403

Ahsan N, Belmont J, Chen Z, Clifton J, Salomon AR (2017) Highly Reproducible Improved Label-Free Quantitative Analysis of Cellular Phosphoproteome by Optimization of LC-MS/MS Gradient and Analytical Column Construction. *Journal of Proteomics* PMID: 28634120

Soemedi R, Vega H, Belmont JM, Ramachandran S, Fairbrother WG (2014) Genetic Variation and RNA Binding Proteins: Tools and Techniques to Detect Functional Polymorphisms. *Systems Biology of RNA Binding Proteins. Advances in Experimental Medicine and Biology* Volume 825, 2014, pp 227-266. PMID:25201108

Publications in Preparation

Belmont J, Ji Q, Salomon AR (2017) Integrated Tools for Quantitative Post-Acquisition Analysis of Proteomic Data in a Modern LC-MS Data Management and Visualization Platform (In preparation)
Rashid M, Belmont J, Carpenter D, Turner CE, Olson EC (2017) Conditional Knockout of Paxillin and Hic-5 Focal Adhesion Adaptor Proteins Slows Migration Speed and Delays Cortical Layer Formation (In review)

Conference Abstracts and Presentations

Belmont J, Ahsan N, Salomon AR. Development of a Novel Software Platform for Streamlining Modern LC-MS Workflows. *Rhode Island EPSCoR Research Symposium* (2016) Kingston, RI.
Belmont J, Ahsan N, Ramratnam B, Salomon AR. HTAPP: High-Throughput Autonomous Proteomic Pipeline for Automated Acquisition and Insightful Analysis of MS/MS Data. *US HUPO Annual Conference* (2016) Boston, MA.
Belmont J, Salomon AR. High Throughput Autonomous Proteomic Pipeline (HTAPP): A Software Platform Supporting Modern Proteomic Workflows. Brown University Molecular Biology, Cell Biology, and Biochemistry Seminar Series – Data Club (2010) Providence, RI.
Belmont J, Salomon AR. Adaptation of Gene Set Enrichment Analysis for Phosphoproteomics Data. Brown University Molecular Biology, Cell Biology, and Biochemistry Seminar Series - 1st Year Rotation Talks (2010) Providence, RI.

Techniques and Expertise

Programming Languages: Java, R, Python, Perl, SQL, XML, PHP

Bioinformatics: Mascot, Thermo XCalibur, Galaxy

Other Computational: Git, Apache, IIS, UNIX shell, Filemaker

Proteomics: HPLC, liquid chromatography-tandem mass spectrometry (LC-MS/MS), labeled (SILAC and TMT) and label-free quantitation

Biochemistry: co-immunoprecipitation, C18 desalt and cleanup

Molecular/Cellular Biology: Western blot, immunohistochemistry, PCR, organotypic culture, mammalian cell culture, confocal microscopy

Animal work: *in utero* procedures, colony maintenance, husbandry

Teaching Experience

Graduate Teaching Assistantship, Brown University
Quantitative Approaches to Biology, Spring 2014

Acknowledgements

For his mentorship and inspiration, I would like to thank my advisor, Prof. Art Salomon. This work would not be possible without his expertise, and I am grateful for the time and effort he has invested in my growth as a scientist.

For their guidance, encouragement, and intellectual contributions to my research, I would also like to recognize my thesis committee, Profs. Philip Gruppuso, Nicola Neretti, and Christopher de Graffenreid. I am grateful for their dedication to ensuring and improving the quality of this dissertation.

For their love and unyielding support throughout all my often-changing academic whims, I would like to thank my parents, Judson R. and Janice Belmont, as well as my siblings, Jon, Erica, and Cheryl.

For his technical expertise and the assistance that he has provided me, I would like to thank Prof. Nagib Ahsan, as well as the various CCRD staff members I've had the pleasure of interacting with.

For their camaraderie and intellectual input, I would like to recognize the past and present members of the Salomon lab, including but not limited to Ynes Helou, Qinqin Ji, Tao Gu, Zhuo Chen, Xien Yu Chua, Nick Mroz, and Tim Aballo, as well as the many talented rotation students and undergraduates I've had the good fortune to work alongside.

Table of Contents

List of Figures

List of Supplemental Figures

List of Supplemental Tables

List of Abbreviations

APC	Antigen presenting cell
API	Application programming interface
AP-1	Activator protein 1
CARMA1	Caspase recruitment domain-containing protein 11
CD	Cluster of differentiation
CTL	Cytotoxic T lymphocyte
DAG	Diacylglycerol
ERK	Extracellular signal-regulated kinase
ESI	Electrospray ionization
FDR	False detection rate
GEF	Guanine nucleotide exchange factor
GSEA	Gene set enrichment analysis
GUI	Graphical user interface
HPLC	High-pressure liquid chromatography
HPRD	Human Protein Reference Database
HTAPP	High Throughput Autonomous Proteomics Pipeline
IL-2	Interleukin 2
IMAC	Immobilized metal affinity chromatography
IP_3	Inositol 1,4,5-trisphosphate
ITAM	Immunoreceptor tyrosine-based activation motif
Itk	IL2-inducible T-cell kinase
iTRAQ	Isobaric tags for relative and absolute quantitation
JDBC	Java Database Connectivity
LAT	Linker for activation of T-cells
LC-MS	Liquid chromatography coupled to mass spectrometry
Lck	Lymphocyte-specific protein tyrosine kinase
MALDI	Matrix-assisted laser desorption ionization
MEK	Mitogen-activated protein kinase kinase
MHC	Major histocompatibility complex
MOAC	Metal oxide affinity chromatography
m/z	Mass-to-charge ratio
NFAT	Nuclear factor of activated T cells
ODBC	Open Database Connectivity

PH	Pleckstrin homology
PIP_2	Phosphatidylinositol 4,5-bisphosphate
PKC	Protein Kinase C
PLC-γ1/2	Phospholipase C-γ1/2
PTM	Post-translational modification
RasGRP1	Ras guanyl release protein 1
RPLC	Reverse-phase liquid chromatography
SH2/3	Src homology 2/3
SHP-1	Src homology region 2 domain-containing phosphatase-1
SIC	Selected ion chromatogram
SLP-76	SH2 domain-containing leukocyte phosphoprotein of 76 kDa
SOS1	Son of sevenless homolog 1
SQL	Structured Query Language
TCR	T cell Receptor
T_H	Helper T cell
TMT	Tandem mass tags
TSAd	T cell specific adaptor
UHPLC	Ultra-high-pressure liquid chromatography
XML	Extensible markup language
xSEA	Extensible set enrichment analysis
Zap-70	Zeta-associated protein of 70 kD

Abstract

The ability to detect and adapt to changing conditions and circumstances is a requirement of life at both the organismal and cellular levels. Inside the cell, rapid and fine-tuned responses to environmental cues can be achieved through dynamic post-translational modifications (PTMs) of proteins. Significant insights into the activity and dependencies of signaling pathways can therefore be gleaned from global surveys of PTM events, and in particular protein phosphorylation, as it is absolutely essential in the signaling cascades mediating the downstream functions of receptor tyrosine kinase systems.

In T cells, coordinated phosphorylation of tyrosine residues is critical for transmission of signals from the activated T cell receptor. Investigations into the dynamic phosphoproteome of T cells have been greatly assisted by advances in mass spectrometry, which allows for the identification and quantitation of thousands of PTMs simultaneously without a requirement for site-specific antibodies. However, despite the increasing resolution and decreasing cost of mass spectrometry assays, the complex network of phosphorylation events governing T cell activation has been only incompletely characterized. In particular, while it is now appreciated that numerous feedback pathways within the T cell phosphoproteome work together to regulate the initiation of immune responses, when and where feedback loops occur in the pathway is poorly understood.

To better understand the role that critical T cell signaling components play in feedback regulation of the T cell receptor pathway, we conducted a phosphoproteomic study into the role of Phospholipase C-γ1 in mediating the activation state of the receptor-proximal signaling machinery. Additionally, we investigated the application of optimized chromatography configurations to T cell phosphoproteomics, and demonstrated the immense potential to achieve greater qualitative and quantitative insights into the T cell signaling machinery. Finally, to support the visualization, management, annotation, and analysis of this proteomic data, we

designed and implemented new tools to expedite discoveries from large LC-MS datasets. The subject of this work is therefore the insights we have gained into the mechanisms governing T cell activation and regulation as well as the technical and technological innovations that made those insights possible.

Chapter 1

Introduction

1.1. The Role of T Cells in Adaptive Immunity

1.1.1. Innate and Adaptive Immunity

The strategies by which a host identifies and eliminates foreign entities collectively make up the organism's immune system. In vertebrates, the varied mechanisms of immunity are broadly divided into two categories, the innate and adaptive immune systems. The innate immune system is comprised of passive mechanisms—such as skin and mucosal membranes–that prevent pathogens from entering or establishing themselves in the host as well as cytotoxic cells with germline-encoded specificities.[1] Traditionally, the hallmark of innate immunity mechanisms is that they do not possess immunologic memory: past encounters with a pathogen have no bearing on the host's future immune capabilities. In contrast, the adaptive immune

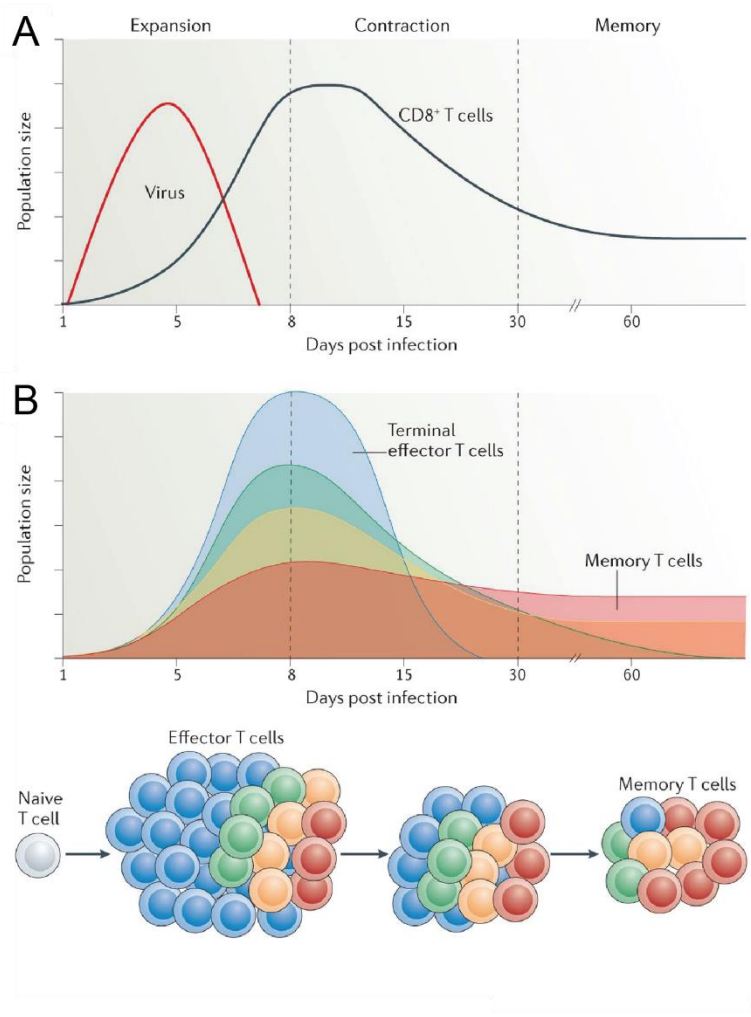


Figure 1.1: Time course of T cell population response. A) Detection of a foreign pathogen induces expansion and differentiation of naïve CD4+ and CD8+ T cell populations. In the days after the pathogen is cleared, the T cell count drops, but a proportion of CD8+, and a smaller proportion of CD4+, T cells remain behind. B) Phenotypic characterization of T cell populations during and post-immune response reveals that the spike in T cells is primarily due to the transient increase in effector T cells. The persisting population of T cells is comprised of memory T cells, which allow for more efficient immune responses upon re-exposure to the same pathogen. (Adapted from Kaech and Cui 2012)

system responds more efficiently to previously encountered antigens due to memory cells that are left behind from previous immune challenges (Figure 1.1). Adaptive immunity is mediated by the activities of T and B cells, which differ in their approaches to fighting off pathogens. Whereas B cells contribute to immunity through their capability to directly recognize free antigen and subsequently produce and secrete neutralizing antibodies, T cells interact with antigens that have already undergone processing by immune cells, and differentiate into effector T cells that either destroy infected cells or mediate the recruitment and regulation of other immune cell populations. Further details of the development and activation of T cells, as well as the relevance of PLC-γ1 to both processes, are discussed below.

1.1.2. Activation and Immune Responses of T Cells

Before T cells can initiate immune responses, they must be alerted to the presence of a foreign antigen. The nature of an invading species is initially defined by antigen-presenting cells (APCs), which scavenge for peptides in their environment before processing and conjugating them to MHC molecules for external display on their surface (Figure 1.2). Subsequently, the binding affinity between the peptide/MHC complexes on APCs and T cell surface receptors determine if an immune response will trigger as well as the strength of that response. Once activated by high-affinity binding to a cognate peptide/MHC, the T cell receptor (TCR) initiates an intracellular signaling cascade culminating in clonal expansion of T cell populations, migration into regions of high antigen density,[2] targeted removal of infected cells, and production of cytokines to provide cues and support to other immune cell populations. The latter two responses, targeted cell killing and cytokine production, are handled by specialized cytotoxic (CTL) and helper (T_H) T cells, respectively. These two populations can be differentiated by the co-receptor they express, and their consequent affinity for different MHC types: CTLs express CD8 on their surface, allowing them to interact with class I MHCs, while T_h express CD4, which binds class II MHCs exclusively.[3]

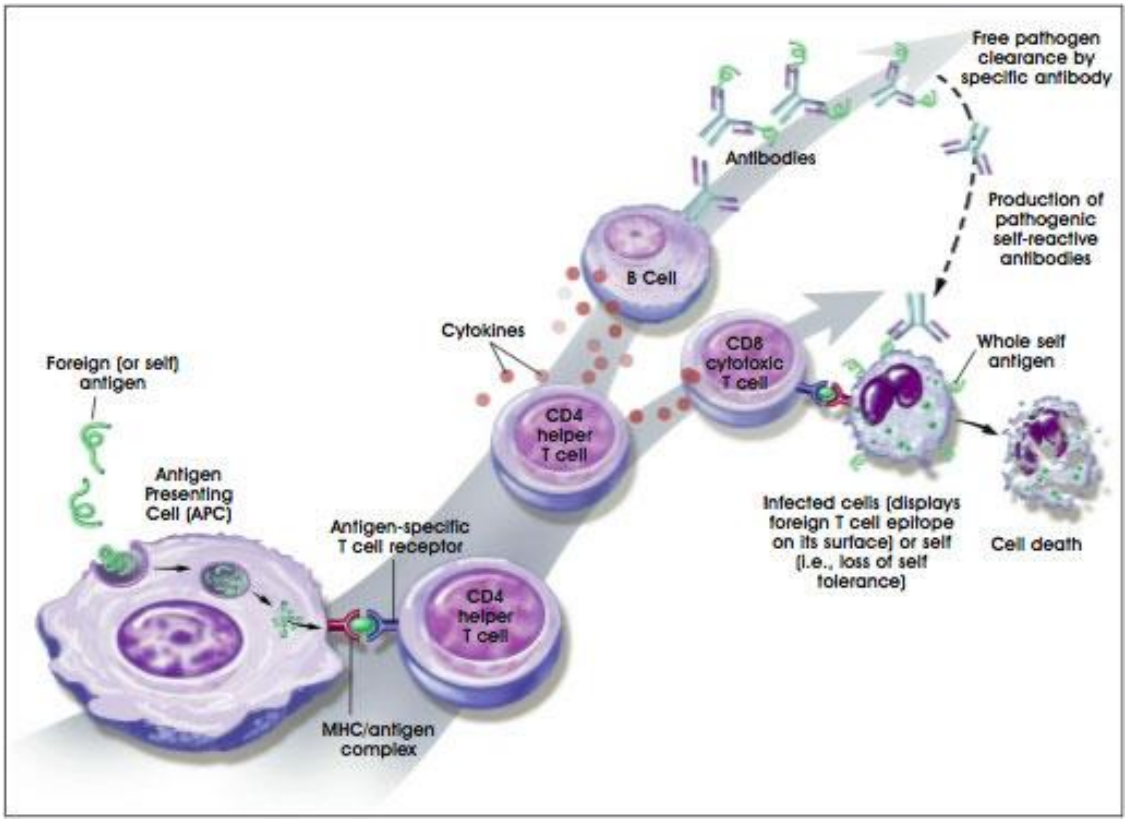


Figure 1.2: Stimulation of the TCR induces cell type-specific effector functions. Activation induces CD4+ T cells to produce cytokines promoting the expansion of immune cells and secretion of antibodies by B cells. Activated CD8+ T cells secrete cytoxic granules and express FasL on their surface to terminate infected cells.(From *http://stemcells.nih.gov/info/scireport/chapter6.asp*)

1.1.3. T Cell Receptor Form and Function

Both the structure of the TCR as well as the process by which that structure develops are essential for proper T cell function. The TCR is a transmembrane heterodimer comprised of two polypeptide chains. In total, four individual TCR chains—referred to as α, β, γ, and δ—exist, and pair off to form two distinct TCR dimers, α/β and γ/δ, with 95% of T cells exhibiting the α/β form.[3] As the structure of the TCR's extracellular domain determines its antigenic recognition capabilities, there is a clear immune advantage to expressing a structurally diverse repertoire of TCRs. To maximize the potential space of recognizable antigens, the unique structure of the variable region of each TCR chain is determined on a cell-by-cell basis by stochastic *V(D)J* recombination events.[4] In this way, an organism's T cell cohort achieves a diverse sampling of the estimated 1×10^{13} possible functional TCRs.[5]

As the generation of the unique variable regions of each TCR chain is not templated by antigenic peptides, a subsequent quality control step is required to ensure both the functionality and specificity of the combinatorically generated receptors. This is achieved during the sequential processes of positive and negative selection in T cell maturation.[6] During positive selection, cells expressing TCRs capable of binding to self peptide/MHC complexes are permitted to continue on to negative selection, while those exhibiting no reactivity are eliminated. Negative selection then further narrows down the pool of candidate T cells by eliminating those that bind too efficiently to self peptide/MHC complexes and represent potential auto-immune threats. Throughout the selection process, T cells are 'double positive'—they express both CD4 and CD8 and can bind class I and II MHCs. Subsequently, each cell ceases to express one of the co-receptors to differentiate into a mature, single positive CD4+ or CD8+ T lymphocyte. The exact mechanism by which CD4+ or CD8+ identity is determined is still unresolved, but current models maintain that fate is determined either stochastically or by instruction from the co-receptor with the predominating signal strength.[3]

1.1.4. The T Cell Receptor Complex

By itself, TCR α/β does not possess the capability to induce intracellular signaling cascades. Instead, TCR α/β noncovalently associates with three other dimeric signaling molecules, CD3ε/CD3δ, CD3ε/CD3γ, and the homodimeric ζ/ζ, to form the TCR complex.[7, 8] Like TCR α/β, CD3 and ζ chains are embedded in the cell membrane; however, unlike TCR α/β, these proteins extend into the cytoplasm far enough to allow for extensive interactions with receptor-proximal components of the signaling pathway.[7] The CD3 and ζ chains serve as the interface between TCR α/β and the intracellular signaling machinery via their multiple immunoreceptor tyrosine-based activation motif (ITAM) domains. Defined as a dual YxxL/I domain separated by a variable region 6-8 amino acids long (YxxL/I-x_{6-8}-YxxL/I), ITAMs' role in TCR signal transduction is mediated primarily via their dual tyrosine residues.[9-11] As

discussed in the next section, phosphorylation of the dual tyrosines in the TCR complex ITAMs is among the earliest events in the TCR signaling cascade.[12]

1.2. The TCR Signaling Pathway

1.2.1. Overview

TCR signaling is triggered by the stable binding of a cognate peptide/MHC complex to the extracellular portion of the TCR, followed shortly thereafter by extensive phosphorylation of the cytoplasmic ITAMs by apical T cell kinases. The exact mechanism by which TCR engagement is coupled to phosphorylation of TCR ITAMs is unclear: one current model posits that receptor binding induces a conformational change rendering the ITAMs accessible to upstream kinases,[13] while another proposes that engagement of the TCR may increase the local density of those kinases.[14] Still a third model proposes that the tight apposition of the APC to the T cell membrane creates an exclusion zone that is inaccessible to the phosphatase CD45 due to its large extracellular domain.[15, 16]

While the mechanistic details of TCR activation require further elucidation, what is known for certain is that it directly precedes the recruitment of essential T cell kinase Lck. Lck is the Src family kinase primarily responsible for the phosphorylation of ITAMs following receptor engagement.[17, 18] Following phosphorylation of the TCR ITAMs by Lck, the phosphorylated tyrosines serve as a docking site for the Syk family kinase, zeta-associated protein of 70 kD (Zap-70). There, Zap-70 undergoes both autophosphorylation and phosphorylation by Lck on regulatory sites to assume its catalytically active form.[19-21] Activated Zap-70 subsequently phosphorylates a host of downstream targets, including the molecular scaffold, SH2 domain-containing leukocyte phosphoprotein of 76 kDa (SLP-76). Phosphorylated SLP-76 nucleates a multiprotein signalosome complex coordinating many of the downstream effects of TCR signaling. In particular, interleukin-2-inducible T-cell kinase (Itk) binds phosphorylated Slp-76, and is catalytically active following its phosphorylation by Lck.[22, 23] The substrates of Itk

include phospholipase C γ1 (PLC-γ1), which associates with a region of Slp-76 adjacent to Itk's binding site.[24, 25] Activation of PLC-γ1 via Itk allows the phospholipase to carry out its primary function, which is cleaving phosphatidylinositol 4,5-bisphosphate (PIP_2). The products of this reaction, inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG), both serve crucial signaling roles as second messengers in the pathway. DAG is a cofactor for various protein kinase C (PKC) isoforms, recruiting the kinases to the cell membrane and potentiating their activity.[26] DAG also binds the Ras GEF RasGRP1, and DAG's recruitment of both PKCs and their substrate RasGRP1 to the membrane feeds into the Ras—Raf—MEK—ERK pathway culminating in activation of the AP-1 transcription factor.[27, 28] IP_3 stimulates the opening of ligand-gated Ca^{2+} channels in the ER to increase the concentration of cytoplasmic Ca^{2+}. Elevated Ca^{2+} levels mediate multiple processes in T cells, ranging from calmodulin- and calcineurin-dependent dephosphorylation and activation of the nuclear factor of activated T cells (NFAT) transcription factor to microtubule and actin polarization towards the engaged APC.[29] Additionally, Ca^{2+} and DAG together have been shown to synergistically activate PKCs, paving the way for PKCθ-mediated activation of the transcription factor NF-κβ.[26, 30] The transcriptional programs enacted by IP_3- and DAG-dependent factors ultimately induce cell division, mobilization, differentiation into effector subtypes, and secretion of paracrine signaling factors to potentiate immune responses.

1.2.2. Lck Structure and Regulation

As the preeminent kinase responsible for ITAM phosphorylation, Lck plays a pivotal role in the initiation of TCR signaling. In addition to its kinase domain, Lck contains both a Src homology 2 (SH2) and SH3 domain, both of which contribute to regulation of the kinase's activity. SH2 domains coordinate binding primarily to phosphotyrosine residues, but show some capability for interacting with phospholipids as well.[31] Lck's SH2 domain participates in an auto-inhibitory mechanism whereby it binds to the phosphorylated negative regulatory site on

Lck, Tyr^{505}, folding the protein back on itself and rendering its kinase domain inaccessible.[32-34] SH3 domains principally bind to PxxP motifs,[35] but like SH2 domains, exhibit affinity for phospholipid substrates in select instances.[36-38] Lck's SH3 domain has been proposed to play a role in targeting the kinase to non-receptor-proximal substrates,[39] but the mechanism through which this occurs is not clearly defined. A possible explanation for this phenomenon is discussed in Chapter 2.

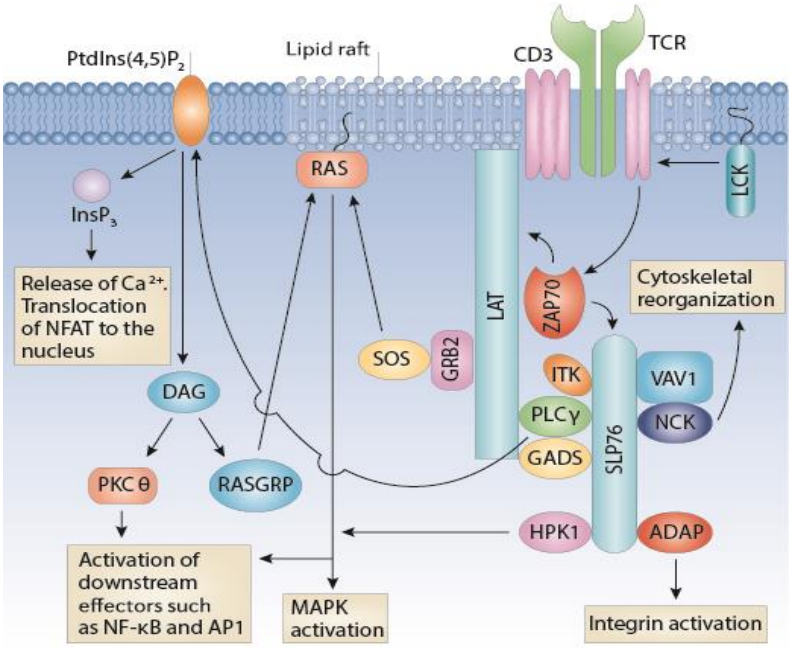


Figure 1.3: Overview of signaling events downstream of the T cell receptor. Stimulation of the TCR leads to phosphorylation of TCR ITAMs by Lck, and subsequent activation of Zap-70. In turn, activated Lck and Zap-70 phosphorylate the molecular scaffold Slp-76 to initiate a signalosome assembly including PLC-γ1 and its upstream kinase, Itk. Following activation of PLC-γ1 by Itk, PLC-γ1 catalyzes the formation of DAG and IP_3 from PIP_2. Both molecules serve critical roles as second messengers in the pathway, promoting activation of PKC and elevated levels of cytoplasmic Ca^{2+} (Adapted from Koretzky et al 2006)

1.2.3. Feedback Regulation of Lck

Fine-tuned control of the strength and duration of the TCR signaling pathway is achieved in part by multiple feedback networks allowing downstream events to influence receptor-proximal signaling machinery. Many of these feedback pathways converge on Lck, as Lck's kinase activity is crucial for the activation of downstream signaling modules.[17, 40] An elegant mechanism for discrimination between strong and weak TCR stimulation has previously been defined whereby Lck activity is negatively regulated by weak stimuli to prevent initiation of an immune response by non-pathogenic agents.[41] This mechanism functions by regulating the association of tyrosine phosphatase Src homology region 2 domain-containing phosphatase-1

(SHP-1) with Lck, which is induced by TCR stimulation and negatively regulates Lck activity.[42] This association is overridden by strong TCR stimulation leading to ERK activation, which in turn phosphorylates Lck on Ser^{59} and disrupts the stable association of Lck and SHP-1.[41] Another feedback pathway mitigating Lck's activity centers on the phosphorylation of Lck's Tyr^{192} by Itk. Phosphorylation of this residue has been shown to alter the binding affinity of Lck's SH2 domain, favoring a stable association of Lck with T cell-specific adaptor (TSAd), and correlates with decreased phosphorylation of TCR complex ITAM sites.[43] Together, these observations support a model wherein TSAd titrates Lck away from the receptor to modulate signal strength after a signal sufficient to activate Itk has been generated.

Mechanisms regulating the activity of Lck and other TCR signaling components can also act directly to modulate protein abundance. Negative regulation of Lck via Cbl-directed degradation has also been demonstrated as a viable regulatory mechanism active in TCR signaling at later timepoints post-stimulation. Cbl's association with Lck is induced by TCR stimulation, and depends on Lck kinase activity.[44] This mechanism may involve a direct interaction between Lck's SH3 domain and Cbl, based on regulatory associations that have been observed between Cbl and other Src family kinases.[44, 45]

1.2.4. PLC-γ1 Structure and Regulation

The PLC family of enzymes catalyze the conversion of PIP_2 to IP_3 and DAG, and is divided into subtypes β, γ, δ, η, ζ, and ε. The universal feature of all PLC enzymes is the X and Y domains, regions of high sequence homology that together form the enzyme's catalytic site (Figure 1.4). In PLC-γ1 and its structurally identical isoform, PLC-γ2, the X and Y domains lie on either side of a linker region, allowing for basal auto-inihibition of the enzyme via separation of the domains. PLC-γ1 is distinguished from the rest of the PLC family by the presence of tandem SH2 domains as well as a single SH3 domain, all of which are located in the X-Y linker. PLC-γ1's N-terminal SH2 domain mediates PLC-γ1's association with the transmembrane scaffold

protein, linker of activated T cells (LAT), while its C-terminal SH2 domain participates in an intramolecular interaction with the X domain of the enzyme's catalytic site. This forces PLC-γ1 into an auto-inhibited conformation that is relieved by phosphorylation of Tyr^{783} by Itk. The SH3 domain of PLC-γ1 mediates the enzyme's association with the proline-rich region of SLP-76.[24] Unlike the association of PLC-γ1 with LAT, which is induced by TCR stimulation and requires phosphorylation of LAT by Zap-70,[46, 47] PLC-γ1's SH3-dependent association with Slp-76 is observed in both active and resting T cells.[24] PLC-γ1 also contains two pleckstrin homology (PH) domains, one of which is conserved among nearly all PLC enzymes, and a second split PH domain divided into two halves located on either side of the enzyme's SH2 and SH3 domains within the X-Y linker. PLC-γ1's intact PH domain coordinates the binding of PLC-γ1 to PIP_2 and PIP_3, while the split domain appears to modulate PLC-γ1 activity through its association with the Rho family GTPase Rac.[48, 49]

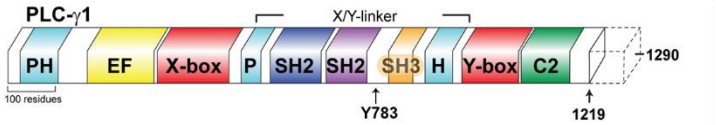


Figure 1.4: Domain structure of PLC-γ1. PLC-γ1 and the structurally similar PLC-γ2's core functional domains are conserved across nearly all PLC family enzymes, and include the N-terminal PH domain, the Ca^{2+}-binding EF hand, and the X and Y regions that together make up the phospholipase active site. In contrast to other PLC enzymes, PLC-γ1's X and Y regions are separated by a linker containing a split PH domain, two SH2 domains, and a single SH3 domain. The N-terminal and C-terminal SH2 domains participate in intramolecular and intermolecular interactions, respectively, with the latter serving an autoinhibitory function that is relieved by phosphorylation of Tyr^{783}. The SH3 domain is required for localization of PLC-γ1 to SLP-76, while the split PH domain binds to Rac, although this interaction may be specific to PLC-γ2 only. The ~70 amino acids at the C terminus comprise an intrinsically disordered region. (Adapted from Gresset et al. 2010)

1.2.5. PLC-γ1 in Development and Disease

PLC-γ1 is ubiquitously expressed, and plays a key developmental role in many tissues. This is due to the fact that PLC-γ1 is integral to the signaling capabilities of multiple growth factor receptors, including Trk,[50] EGFR,[51, 52] HGFR,[53] VEGFR,[54] FGFR,[55] and PDGFR.[56, 57] Accordingly, PLC-γ1 knockout mice are nonviable, and do not live past embryonic day 9 due to failed vasculogenesis resulting from impaired VEGF signaling,[58] while chimeric mice exhibit severe renal dysplasia and defects in hematopoiesis.[59]

In the thymus, PLC-γ1 is indispensable for selection and differentiation events during T cell maturation. In conditional knockout mice lacking PLC-γ1 in double-positive thymocytes, the inability to induce signaling in response to TCR stimulation results in a significant decrease in cell survival following positive selection.[60] Among the lymphocytes that do survive to maturity, TCR-induced proliferation and cytokine production are reduced. Despite the impaired signaling capabilities and reduced numbers of PLC-γ1-deficient lymphocytes, PLC-γ1 conditional knockout mice exhibit autoimmune symptoms due to decreases in regulatory T cell populations.

Discussions of PLC-γ1's clinical relevance primarily center on its role in metastasis and cell invasion. Activating mutations in PLC-γ1 have been identified as a common signature of angiosarcomas.[61] Likewise, abnormally high levels of PLC-γ1 are observed in breast tissue,[62] colorectal,[63] and squamous cell [64] carcinomas. PLC-γ1's modulation of tumor expansion and invasion is thought to occur through the relaxation of PIP_2-mediated inhibition of actin remodeling factors [65, 66] as well as mechanisms linking DAG and IP_3 to activation of Rac.[67, 68] However, activation of the Ras GEF SOS1 by PLC-γ1's SH3 domain has also been reported, highlighting a potential role for phospholipase activity-independent mitogenic signaling through PLC-γ1.[69, 70]

1.3. Mass Spectrometric Methods for Proteomic Analysis

1.3.1. Background

When the Human Genome Project announced the completion of its initial organizational goals in 2003,[71] it heralded an era of increasingly cheaper and readily available genome sequencing. While the complete sequence of the human genome has yielded considerable insights into genetic variants and their roles in disease, a more complete understanding of the molecular bases of human disease must consider not only genes themselves, but their functional products and the post-translational modifications thereof. Investigation of these topics

falls within the purview of proteomics, which both profits from and complements the genomic data delivered by high-throughput sequencing.

The goal of proteomics is the high-throughput identification, quantitation, and analysis of an organism's protein complement delineated by tissue and time of expression. Like genomics, proteomics was realized by technological innovations allowing biological molecules to be assayed on a previously unattainable scale. In the past, methods such as 2-dimensional gel electrophoresis and protein microarrays enjoyed widespread adoption for protein profiling *en masse*. However, advances in mass spectrometry (MS) delivering increasingly higher resolution and deeper sequencing depth have allowed MS to supplant these technologies as the preeminent high-throughput method for characterizing proteomic samples. In proteomics applications, the joint liquid chromatography-mass spectrometry (LC-MS) method is most commonly used, in which mass spectrometry is coupled with on-line liquid chromatography and an ion source. Liquid chromatography provides separation of complex mixtures of samples on the basis of chemical properties, and is immediately followed by ionization of analytes prior to their introduction into the MS. Ionization of samples is required for MS, as the instrument separates and detects proteins or peptides on the basis of mass-to-charge ratio, referred to hereafter as m/z. The generation of ions for MS analysis is mostly achieved by either matrix-assisted laser desorption ionization (MALDI) or electrospray ionization (ESI), as these 'soft' ionization methods induce only minimal fragmentation of peptides.

1.3.2. Bottom-up vs. Top-down Proteomics Methods

The application of LC-MS to study proteins has led to the emergence of two analytical paradigms, bottom-up and top-down proteomics. Bottom-up proteomics aims to characterize proteins and their modifications by analyzing their constituent peptides (Figure 1.5, bottom). In this approach, proteins mixtures are trypsinized prior to LC-MS to cleave the amino acid backbones at lysine and arginine residues, producing peptides of ~600-2000 Da for analysis. In

contrast, top-down proteomics is the characterization of proteins in their entirety, and involves separating and injecting intact proteins in the mass spectrometer (Figure 1.5, top). In both approaches, the starting analyte undergoes further fragmentation prior to MS/MS sequencing. The sequencing of large peptide fragments derived from intact proteins is useful for the study of PTMs, as co-occurrences of non-adjacent PTMs can be observed. Moreover, top-down analysis ensures that the entirety of a protein's amino acid sequence will be captured, whereas only a small subset of tryptic peptides derived from a single protein will typically be detected in a bottoms-up procedure. Currently, widespread adoption of top-down proteomics is hampered by a number of technical hurdles, such as the difficulty of reliably generating MS/MS fragments from proteins larger than ~50 kDa[72] and the challenges associated with separating complex and heterogenous protein mixtures. However, some of the benefits of top-down proteomics can be obtained by performing bottom-up proteomics with longer tryptic peptides, and this synthetic approach has been referred to as middle-down proteomics (Figure 1.6).

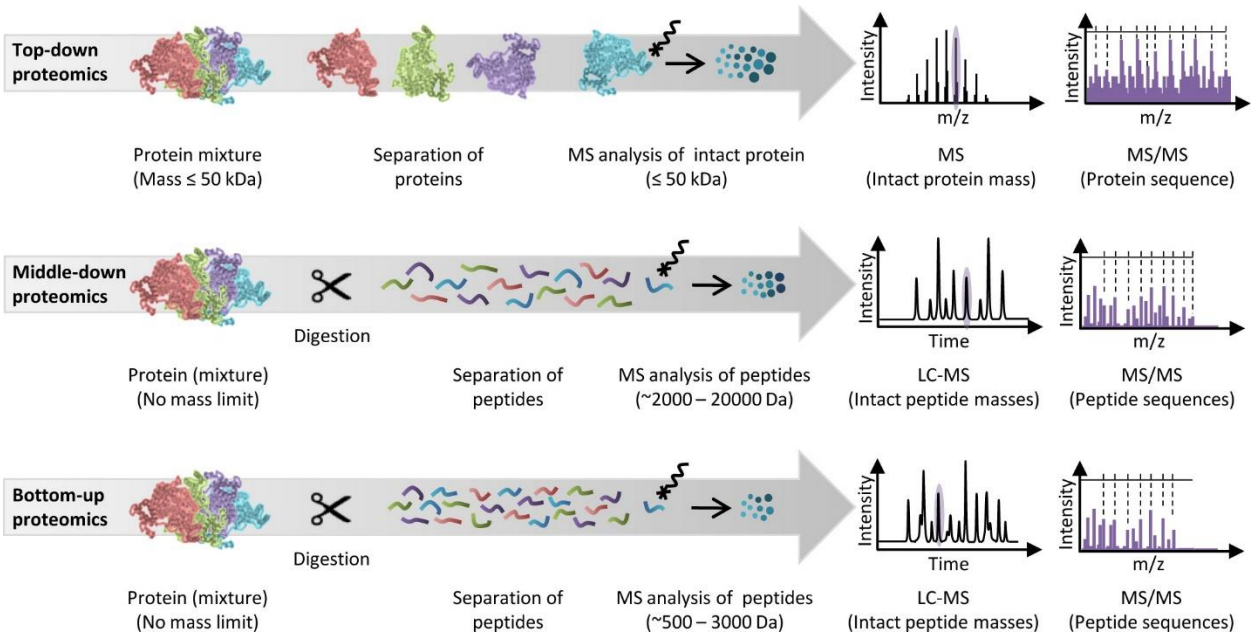


Figure 1.5: Top-down, middle-down, and bottom-up proteomics workflows. (From Switzar *et al*. 2013).

1.3.3. Reverse-Phase Liquid Chromatography

The resolving power of MS is greatly increased by an on-line liquid chromatography separation of analytes prior to injection. Among the liquid chromatography methods in use, reverse-phase liquid chromatography (RPLC) has been referred to as the “gold standard” for proteomics owing to the compatibility of its buffers with ESI as well as its applicability to mixtures containing a wide range of physicochemical properties.[73] RPLC operates on the principle that a mixture of molecules of varying polarities can be partitioned between a non-polar stationary phase and a polar mobile phase. The stationary phase is established by packing the chromatography column with particles conjugated to hydrophobic alkyl chains. Separation of complex mixtures in reverse-phase chromatography is achieved with the addition of an increasing gradient of pressurized organic solvent to titrate analytes out of the stationary phase.

The performance of a chromatographic column is expressed as the number of theoretical plates, *N*, it demonstrates for a given analyte. This metric describes column efficiency in terms of the theoretical number of equilibrium stages between the mobile and stationary phases it provides, with higher values indicating better performance. For a given peak, *i*, the theoretical plate number is expressed as

$$N = 5.545 \left(\frac{retention\ time_i}{full\ width\ at\ half\ height_i} \right)^2$$

The values of *N* within a specific retention time window are used to quantify the efficacy of chromatographic separation. Specifically, *N* is used to determine peak capacity (*n*), which refers to the “number of peaks that can be separated with a resolution of unity in a given time [interval] (t_1-t_n)”, [74] and is defined as

$$n = 1 + \int_{t_1}^{t_n} \frac{\sqrt{N}}{4} \frac{\mathrm{d}t}{t}$$

From this equation it can be seen that peak capacity is increased by maximizing *N*. In turn, *N* is optimized by longer elution gradients and reductions in peak width.

Methods to increase peak capacity are of significant interest for proteomics researchers utilizing LC-MS due to the fact that a linear relation between peak capacity and number of peptides identified has been previously demonstrated.[75] Ultra-high performance liquid chromatography (UHPLC) pumps operating at extremely high pressures allow for the use of smaller stationary phase particles, minimizing peak width and thus enhancing peak resolution compared to standard HPLC pumps. However, the higher price point of UHPLC systems and additional maintenance required due to the stress of high pressure operation have led many to pursue method development and optimization on HPLC pumps to achieve similar performance. To this end, optimizations including longer gradients,[76] longer analytical columns,[75] and reduced stationary phase bead size [77] have all been thoroughly tested on HPLC systems and shown to provide higher quality separations of complex peptide mixtures. In Chapter 3 we demonstrate how optimized chromatography conditions yield deeper insights into the dynamic phosphoproteome of activated T cells.

1.3.4. Peptide Sequencing by Mass Spectrometry

Identification of the proteins represented in a heterogenous mixture of peptides is one of the central goals in many proteomics investigations. Sequence data is not derived directly from the observed m/z of a species injected into the mass spectrometer (Figure 1.6A). Instead, after conducting a scan across the full range of monitored m/z's, the instrument selects candidate precursor ions from among the most abundant species recorded in the scan to undergo MS/MS (Figure 1.6B). First, precursor ions are fragmented through either bombardment with highly accelerated neutral molecules (collision-induced dissociation, or CID) or addition of electrons to the multiply positively charged molecular ion (electron transfer dissociation, or ETD). Then, the mass spectrum of the resulting fragments is recorded. With MS/MS spectra obtained at high

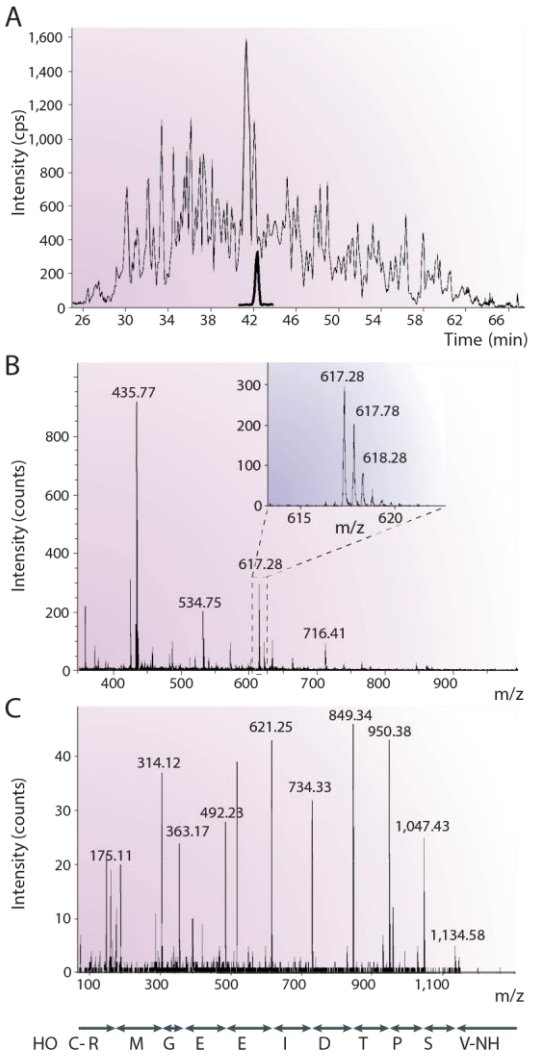


Figure 1.6: Representation of LC-MS data gathered in a single run. A) Total ion chromatogram depicting the intensities from all mass spectra collected during the course of the experiment as a function of retention time. The intensity obtained from a particular peptide eluting around 42 minutes is outlined in bold. The representation of the ion current from a single peptide as a function of time within a fixed mass window is referred to as a selected ion chromatogram. B) Mass spectrum obtained from the selected peptide outlined in bold in (A). The inset depicts the mass peaks obtained from peptides containing varying amount of the C^{13} isotope. C) MS/MS spectrum of the ion selected for sequencing. The full sequence of the peptide is written carboxy- to amino-terminal, and the mass of each y-ion is annotated on the plot. (Steen and Mann 2004).

mass accuracy and resolution in which sufficient coverage of mass fragments is obtained, it is possible to sequence the peptide *de novo* using the shifts in mass between sequential product ions (Figure 1.6C). More commonly, however, sequence assignment is approached as a database matching problem, in which observed MS/MS spectra are searched against databases of predicted fragments ions. While database searching is necessarily limited by the quality and coverage of the genome sequence for the species under investigation, its demands on computational resources and data quality are more reasonable than those required by *de novo* sequencing.[78]

1.3.5. Peptide Quantitation by Mass Spectrometry

Identification of the proteins present in a sample is often accompanied by quantitation of the observed peptides. Mass spectrometry is not quantitative in an absolute sense, as a peptide's unique physicochemical properties determine its ionization efficiency, thus confounding direct comparison of ionic abundances between different peptides. However, relative quantitation of peptide abundance is readily achieved using either labeled or label-free methods. As these methods of quantitation impose unique restrictions on experimental design and sample preparation, selection of a quantitation method is guided in part by the study's research objectives.

Labeled quantitation refers to a class of methods in which peptides originating from different samples are differentiated by the addition of a tag or label of variable mass (Figure 1.7, right). This allows for samples prepared separately to be combined prior to LC-MS, as the mass differential between peaks can be used during quantitation to determine sample identity. Stable-isotope labeling by amino acids in cell culture (SILAC) is one commonly used labeled method in which the mass labels are amino acids (typically lysine and arginine, due to the cutting specificity of trypsin) isotopically labeled with ^{14}C, ^{15}N, and/or ^{18}O. These labels are introduced metabolically by depleting the nutrient source of specific amino acids and reconstituting with their labeled counterparts. Following LC-MS, the relative abundance between two SILAC-labeled peptides can be obtained by integrating the area under the curve for both extracted ion chromatograms (Figure 1.6A), then calculating the ratio between them. As SILAC allows for the combination and co-processing of samples early in the workflow, it may offer the most accurate quantitation of all the labeled and label-free methods.[79] However, the limited number of combinations of isotopically distinct labels may preclude SILAC's use in studies involving more than 5 comparisons.[80]

Quantitation of SILAC data is achieved primarily through integration of extracted ion chromatograms, and is therefore considered MS-based quantitation. In contrast, MS/MS-based quantitation is used by isobaric labeling systems, such as tandem mass tags (TMT) and isobaric tags for relative and absolute quantitation (iTRAQ). In these approaches, different chemical tags of equal total mass are affixed to the N-terminal amines of trypsin-digested peptides. Each of these tags is comprised of two regions—a reporter fragment and a mass equalization fragment—separated by a labile linker that is cleaved during MS/MS fragmentation by very high energy collision dissociation. Cleavage of the linker during MS/MS fragmentation releases the reporter fragment so that its abundance and mass can be recorded in the MS/MS spectrum. As the reporter fragment mass of each tag is unique, the abundance of the reporter is a proxy for the peptide abundance in the corresponding sample, provided that the efficiency of the tagging procedure is 100%. While isobaric mass tagging does not allow samples to be combined as early as SILAC does, the approach allows for differential tagging and pooling of up to 10 distinct samples,[81] and can be used in situations where metabolic labeling is infeasible. Additionally, hybrid “hyperplexing” approaches, in which the isobaric tagging workflow is applied to metabolically labeled samples, has been successfully used to achieve as many as 18 distinct labels in a single LC-MS run.[82]

Label-free quantitation presents an alternative to labeled approaches that avoids the use of expensive labeling reagents entirely. In this approach, samples are not pooled, and must be processed in separate LC-MS runs. This comes at the cost of some additional inter-run variability; however, the payoff of label-free proteomics is superior sequence coverage as well as the absence of restraints imposed by a limited number of mass labels.[83, 84] Quantitation of label-free data can be achieved by MS-based quantitation of precursor ions, as discussed above, or via MS/MS-based spectral counting, in which the abundance of the peptide is measured by the number of MS/MS spectra confidently assigned to it. While this approach is

intuitive and straightforward to implement, it suffers from a number of deficiencies, including its incompatibility with dynamic exclusion and poor accuracy when quantifying peptides with low spectral counts.[85] Furthermore, spectral counting by definition excludes reproducible quantitative data from peptides that did not trigger an MS/MS, which are readily accessible using MS-based precursor ion quantitation.

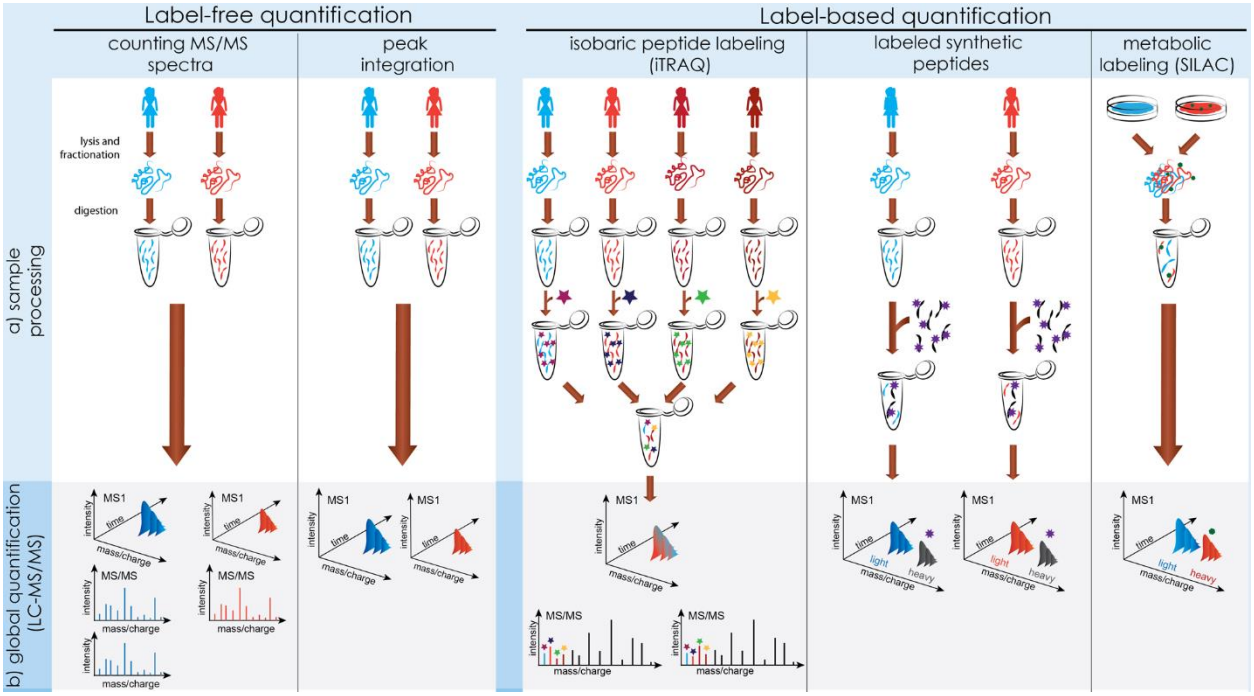


Figure 1.7: Summary of quantitation methods for LC-MS-based proteomics. (Left) Label-free workflows require separate processing and LC-MS analysis of each sample, as peptides are not labeled. Quantitation is achieved either at the MS level by integrating precursor peak areas or at the MS/MS level by summing the number of observations of each peptide (Right) Label-based quantitation allows peptides from distinct samples to be combined prior to LC-MS due to differential mass tags applied to each sample. Isobaric peptide labeling employs chemical labeling of peptides after digestion using a labile tag that is recovered and quantified at the MS/MS level. SILAC labeling is achieved metabolically by adding isotopically labeled amino acids to an organism's nutrient source, resulting in distinct mass shifts that can be used to determine a peak's sample of origin. Isotopically labeled synthetic peptides can also be spiked into label-free samples in known quantities to allow for absolute quantitation of peptide abundance. (Adapted from Käll and Vitek 2011).

1.3.6. Computational Platforms for Proteomics

Owing to the scale and complexity that can now be achieved in modern LC-MS datasets, bioinformatics plays a significant role in proteomics workflows at each step from the assignment and quantitation of spectra all the way to the collation, statistical analysis, and visualization of large datasets. For many labs, these steps are carried out by an assortment of specialized tools

that may be joined into custom pipelines or run manually. However, for workgroups where the principle focus is not computational method development, integrated solutions containing a comprehensive suite of data processing capabilities can greatly increase the throughput of data analysis. Freely available software platforms such as MaxQuant,[86] OpenMS,[87] and PeptideShaker [88] strive to provide feature-rich computing environments that enable complex analytical tasks to be carried out in a single program. Our lab's own software, Peptide Depot is another entry in this category,[89] and functions as part of a larger analysis pipeline that automates many rote pre-processing tasks including database searching, quantitation, and validation of proteomics data.[90] However, many useful post-acquisition analytical tasks concerning the interpretation of data are not yet implemented in these platforms, and require additional processing outside the software. In particular, while phosphoproteomic data analysis is supported by all these software packages, support for the annotation and pathway analysis of phosphoproteomes varies among them. Chapter 4 concerns the expansion of the Peptide Depot bioinformatics platform and the new modes of post-acquisition phosphoproteomics analysis enabled by it.

1.3.7. Phosphoproteomics

Reversible phosphorylation of hydroxyl amino acids is the most common experimentally observed post-translational modification.[91] As many as one- to two-thirds of all human proteins are believed to be phosphorylated in some capacity,[92, 93] and combinatorial modification patterns involving phosphorylation may give rise to distinct functions not mediated by phosphorylation alone.[94-96] Thus, a tremendous amount of functional diversity at the molecular level is achieved through dynamic regulation of the phosphoproteome.

The capability to precisely define the location and relative abundance of phosphorylated sites has made bottom-up proteomics an invaluable tool for investigating the phosphoproteome. Owing to the low stoichiometry of phosphorylation *in vivo*, enrichment of phosphorylated

peptides prior to LC-MS is required to achieve useful sequencing depth of the phosphoproteome. Global enrichment of phosphorylated peptides is typically performed using either immobilized metal affinity chromatography (IMAC) or metal oxide affinity chromatography (MOAC), with the latter showing some evidence of greater specificity for phosphorylated peptides.[97-99] However, global enrichment strategies are better suited to the analysis of serine- and threonine-phosphorylated peptides than they are tyrosine-phosphorylated peptides, as the latter are estimated to make up only 0.5% of the phosphoproteome,[100] and are thus sparsely recovered by global phosphorylation enrichment methods. The phosphoproteomics studies presented in this work primarily concern the phosphotyrosine portion of the T cell phosphoproteome, as the receptor-proximal signals passed between the TCR, Lck, Zap-70, Slp-76, Itk, and PLC-γ1 are transduced predominantly through tyrosine phosphorylation. Deep sequencing of the tyrosine phosphoproteome in our studies is achieved through immunoaffinity pulldowns via immobilized anti-phosphotyrosine antibodies.

1.3.8. Prediction of Pathways from Phosphoproteomics Data

The unbiased, bird's-eye view of signal transduction events provided by *omics data provides an insight into systems-level events that is not easily achieved by traditional low-throughput methods. To capitalize on the scale of this data, computational methods to detect perturbation of functionally related features in large datasets are an active area of research, where features may be observations of RNA, peptide, or phosphopeptide levels. Many of these methods proceed from systems of ontologies, which are categorizations of features by function, cellular location, or implication in disease such as those provided by GO [101] and KEGG.[102, 103] These ontologies serve as the basis for both over-representation analysis approaches, which consider only features with statistically significant changes in expression,[104, 105] as well as enrichment analysis approaches. The latter category of methods is exemplified by the gene set enrichment analysis (GSEA) method, which excels at the detection of subtle yet

reproducible changes of features that would otherwise be missed by examining only lists of statistically significant observations.[106] While GSEA was initially developed for use with microarray data, the approach has been successfully applied to proteomics data.[107, 108]

The increased availability of phosphosite-specific annotation databases such as Scansite [109] and HPRD [110] presents both new opportunities and challenges for the prediction of pathways from phosphoproteomic data. As phosphorylation events can exert either positive or negative regulatory effects, phosphorylation may provide more nuanced insight into protein activity than protein levels. Furthermore, substrate phosphorylation levels provide clues to the overall activity of their upstream kinase. However, the functional independence and varying phosphorylation levels of sites on the same protein mean that categorizing phosphopeptide observations into traditional gene- or protein-level ontologies is likely of little predictive value. At minimum, meaningful set enrichment analysis for phosphoproteomics data requires the use of new ontologies containing annotations of subsequence-level features (Figure 1.8). To date, a handful of approaches have attempted to adapt enrichment set analysis-like methods to phosphoproteomic data, including Kinase Enrichment Analysis [111] and Kinase Substrate Enrichment Analysis,[112] which are designed for detecting enrichments of kinase activity, as well as Phosphorylation Set Enrichment Analysis, which identifies potentially novel substrates of known kinases.[113] However, no single approach combining kinase-substrate ontologies as well as known and *in silico* predicted functional domains has emerged, let alone one integrated into a fully featured data management platform. Our efforts to address this deficiency are discussed in Chapter 4.

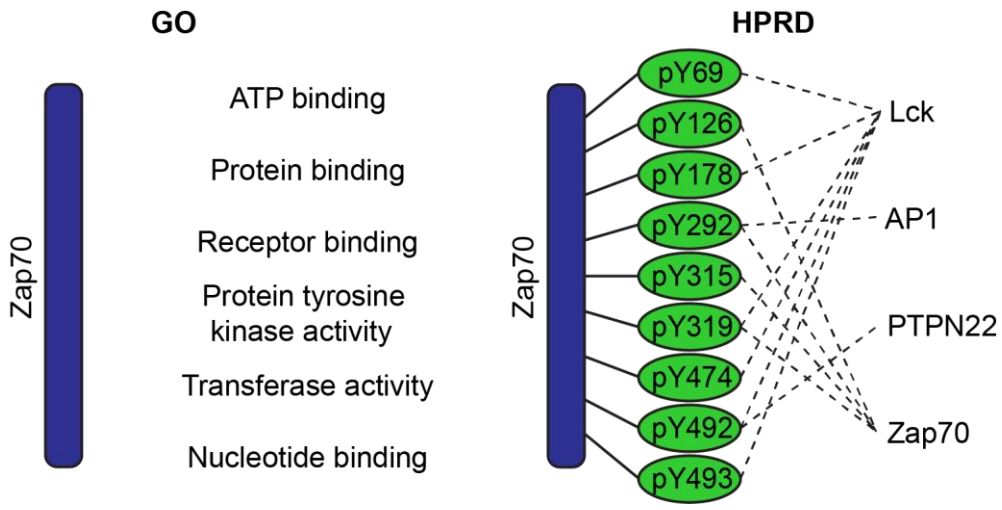


Figure 1.8: Differential resolution of protein-level and phosphosite-level ontologies as exemplified by GO and HPRD. (Left) Select GO annotations of TCR kinase Zap-70. These annotations describe function at the protein level, and thus cannot be assigned to any subsequence. (Right) Select HPRD annotations of known modulators of Zap-70 phosphorylation. Enrichment analyses using these ontologies provides qualitatively different, but more nuanced, information about signaling events.

1.4. Overview of Chapters in This Dissertation

It is now well understood that multiple feedback mechanisms compete and cooperate to establish the phosphorylation of receptor-proximal signaling proteins in the TCR signaling pathway, and that these phosphorylation events track with the activation state of the pathway. However, our understanding of when and where feedback loops occur to set basal and active phosphorylation levels of the TCR and its downstream signaling effectors is still incomplete. In particular, PLC-ɣ1's role in mediating the activation state of both its proximal and distal upstream kinases has not been previously established. We conducted a bottom-up phosphoproteomic screen to investigate how phosphorylation-mediated TCR signaling is perturbed in the absence of PLC-ɣ1 using the Jurkat-derived, PLC-ɣ1-deficient Jgamma1 cell line. To our knowledge, this work constitutes the first timecourse of the tyrosine phosphoproteome in PLC-ɣ1-deficient T cells, and reveals surprising dependencies of both Lck and TCR ITAM phosphorylation on PLC-ɣ1.

In Chapter 3, we optimized and assessed an LC-MS workflow for achieving better qualitative and quantitative characterizations of the T cell phosphoproteome. In this study, we

examined the benefits of enhanced chromatographic efficiency achieved by an RPLC procedure using sub-2μm beads coupled with longer chromatography columns, showing that it afforded greater sensitivity to detect dynamic phosphorylation of critical TCR signaling proteins. Importantly, the qualitative and quantitative gains we demonstrated here do not require the use of a UHPLC or a column heater. Thus, this work paves the way for future phosphoproteomic insights by showing that deeper and more reproducible analysis of the T cell phosphoproteome can be achieved through readily implemented chromatographic optimizations.

Chapter 4 details the development of a software platform supporting the management, visualization, and analysis of large LC-MS datasets. Nowadays, many labs' capabilities to generate massive proteomic and phosphoproteomic data sets at considerable sequencing depth far outpaces their ability to derive biological insight from this data. Oftentimes, the bottleneck in proteomics investigations is due to the gap in expertise between LC-MS experts and bioinformatics analysts, as facility with both domains is critical for efficient and insightful proteomics analysis. This work seeks to bridge that gap by providing a feature-rich software environment to assist researchers with post-acquisition analysis tasks. Additionally, this software features novel modes of analysis for phosphoproteomics that leverage existing bioinformatics resources to assist in the exploratory analysis phase of phosphoproteomics studies.

1.5. References

1. Dempsey, P.W., S.A. Vaidya, and G. Cheng, *The art of war: Innate and adaptive immune responses.* Cell Mol Life Sci, 2003. **60**(12): p. 2604-21.
2. Krummel, M.F., F. Bartumeus, and A. Gerard, *T cell migration, search strategies and mechanisms.* Nat Rev Immunol, 2016. **16**(3): p. 193-201.
3. Germain, R.N., *T-cell development and the CD4-CD8 lineage decision.* Nat Rev Immunol, 2002. **2**(5): p. 309-22.
4. Janeway, C.A.J.T., P.; Walport, M. et al, *Immunobiology: The Immune System in Health and Disease*. 5 ed. 2001: Garland Science.
5. Nikolich-Zugich, J., M.K. Slifka, and I. Messaoudi, *The many important facets of T-cell repertoire diversity.* Nat Rev Immunol, 2004. **4**(2): p. 123-32.

6. Robey, E. and B.J. Fowlkes, *Selective events in T cell development.* Annu Rev Immunol, 1994. **12**: p. 675-705.
7. Wucherpfennig, K.W., et al., *Structural biology of the T-cell receptor: insights into receptor assembly, ligand recognition, and initiation of signaling.* Cold Spring Harb Perspect Biol, 2010. **2**(4): p. a005140.
8. Kuhns, M.S. and H.B. Badgandi, *Piecing together the family portrait of TCR-CD3 complexes.* Immunol Rev, 2012. **250**(1): p. 120-43.
9. Love, P.E. and S.M. Hayes, *ITAM-mediated signaling by the T-cell antigen receptor.* Cold Spring Harb Perspect Biol, 2010. **2**(6): p. a002485.
10. Reth, M., *Antigen receptor tail clue.* Nature, 1989. **338**(6214): p. 383-4.
11. Koyasu, S., et al., *Delineation of a T-cell activation motif required for binding of protein tyrosine kinases containing tandem SH2 domains.* Proc Natl Acad Sci U S A, 1994. **91**(14): p. 6693-7.
12. Samelson, L.E., et al., *Antigen activation of murine T cells induces tyrosine phosphorylation of a polypeptide associated with the T cell antigen receptor.* Cell, 1986. **46**(7): p. 1083-90.
13. Xu, C., et al., *Regulation of T cell receptor activation by dynamic membrane binding of the CD3epsilon cytoplasmic tyrosine-based motif.* Cell, 2008. **135**(4): p. 702-13.
14. Li, Q.J., et al., *CD4 enhances T cell sensitivity to antigen by coordinating Lck accumulation at the immunological synapse.* Nat Immunol, 2004. **5**(8): p. 791-9.
15. Cordoba, S.P., et al., *The large ectodomains of CD45 and CD148 regulate their segregation from and inhibition of ligated T-cell receptor.* Blood, 2013. **121**(21): p. 4295-302.
16. James, J.R. and R.D. Vale, *Biophysical mechanism of T-cell receptor triggering in a reconstituted system.* Nature, 2012. **487**(7405): p. 64-9.
17. Straus, D.B. and A. Weiss, *Genetic evidence for the involvement of the lck tyrosine kinase in signal transduction through the T cell antigen receptor.* Cell, 1992. **70**(4): p. 585-93.
18. van Oers, N.S., N. Killeen, and A. Weiss, *Lck regulates the tyrosine phosphorylation of the T cell receptor subunits and ZAP-70 in murine thymocytes.* J Exp Med, 1996. **183**(3): p. 1053-62.
19. Neumeister, E.N., et al., *Binding of ZAP-70 to phosphorylated T-cell receptor zeta and eta enhances its autophosphorylation and generates specific binding sites for SH2 domain-containing proteins.* Mol Cell Biol, 1995. **15**(6): p. 3171-8.
20. Chan, A.C., et al., *Activation of ZAP-70 kinase activity by phosphorylation of tyrosine 493 is required for lymphocyte antigen receptor function.* EMBO J, 1995. **14**(11): p. 2499-508.
21. Kong, G., et al., *Distinct tyrosine phosphorylation sites in ZAP-70 mediate activation and negative regulation of antigen receptor function.* Mol Cell Biol, 1996. **16**(9): p. 5026-35.
22. Heyeck, S.D., et al., *Lck phosphorylates the activation loop tyrosine of the Itk kinase domain and activates Itk kinase activity.* J Biol Chem, 1997. **272**(40): p. 25401-8.
23. Berg, L.J., et al., *Tec family kinases in T lymphocyte development and function.* Annu Rev Immunol, 2005. **23**: p. 549-600.
24. Yablonski, D., T. Kadlecek, and A. Weiss, *Identification of a phospholipase C-gamma1 (PLC-gamma1) SH3 domain-binding site in SLP-76 required for T-cell receptor-mediated activation of PLC-gamma1 and NFAT.* Mol Cell Biol, 2001. **21**(13): p. 4208-18.
25. Bogin, Y., et al., *SLP-76 mediates and maintains activation of the Tec family kinase ITK via the T cell antigen receptor-induced association between SLP-76 and ITK.* Proc Natl Acad Sci U S A, 2007. **104**(16): p. 6638-43.
26. Tan, S.L. and P.J. Parker, *Emerging and diverse roles of protein kinase C in immune cell signalling.* Biochem J, 2003. **376**(Pt 3): p. 545-52.
27. Roose, J.P., et al., *A diacylglycerol-protein kinase C-RasGRP1 pathway directs Ras activation upon antigen receptor stimulation of T cells.* Mol Cell Biol, 2005. **25**(11): p. 4426-41.

28. Genot, E. and D.A. Cantrell, *Ras regulation and function in lymphocytes.* Curr Opin Immunol, 2000. **12**(3): p. 289-94.
29. Kuhne, M.R., et al., *Linker for activation of T cells, zeta-associated protein-70, and Src homology 2 domain-containing leukocyte protein-76 are required for TCR-induced microtubule-organizing center polarization.* J Immunol, 2003. **171**(2): p. 860-6.
30. Sun, Z., et al., *PKC-theta is required for TCR-induced NF-kappaB activation in mature but not immature T lymphocytes.* Nature, 2000. **404**(6776): p. 402-7.
31. Rameh, L.E., C.S. Chen, and L.C. Cantley, *Phosphatidylinositol (3,4,5)P3 interacts with SH2 domains and modulates PI 3-kinase association with tyrosine-phosphorylated proteins.* Cell, 1995. **83**(5): p. 821-30.
32. Xu, W., et al., *Crystal structures of c-Src reveal features of its autoinhibitory mechanism.* Mol Cell, 1999. **3**(5): p. 629-38.
33. Abraham, N. and A. Veillette, *Activation of p56lck through mutation of a regulatory carboxy-terminal tyrosine residue requires intact sites of autophosphorylation and myristylation.* Mol Cell Biol, 1990. **10**(10): p. 5197-206.
34. Bergman, M., et al., *The human p50csk tyrosine kinase phosphorylates p56lck at Tyr-505 and down regulates its catalytic activity.* EMBO J, 1992. **11**(8): p. 2919-24.
35. Yu, H., et al., *Structural basis for the binding of proline-rich peptides to SH3 domains.* Cell, 1994. **76**(5): p. 933-45.
36. Heuer, K., et al., *The helically extended SH3 domain of the T cell adaptor protein ADAP is a novel lipid interaction domain.* J Mol Biol, 2005. **348**(4): p. 1025-35.
37. Heuer, K., et al., *Lipid-binding hSH3 domains in immune cell adapter proteins.* J Mol Biol, 2006. **361**(1): p. 94-104.
38. Perez, Y., et al., *Lipid binding by the Unique and SH3 domains of c-Src suggests a new regulatory mechanism.* Scientific Reports, 2013. **3**.
39. Rudd, M.L., A. Tua-Smith, and D.B. Straus, *Lck SH3 domain function is required for T-cell receptor signals regulating thymocyte development.* Mol Cell Biol, 2006. **26**(21): p. 7892-900.
40. Hauck, F., et al., *Primary T-cell immunodeficiency with immunodysregulation caused by autosomal recessive LCK deficiency.* J Allergy Clin Immunol, 2012. **130**(5): p. 1144-1152 e11.
41. Stefanova, I., et al., *TCR ligand discrimination is enforced by competing ERK positive and SHP-1 negative feedback pathways.* Nat Immunol, 2003. **4**(3): p. 248-54.
42. Lorenz, U., et al., *Lck-dependent tyrosyl phosphorylation of the phosphotyrosine phosphatase SH-PTP1 in murine T cells.* Mol Cell Biol, 1994. **14**(3): p. 1824-34.
43. Granum, S., et al., *The kinase Itk and the adaptor TSAd change the specificity of the kinase Lck in T cells by promoting the phosphorylation of Tyr192.* Sci Signal, 2014. **7**(355): p. ra118.
44. Rao, N., et al., *Negative regulation of Lck by Cbl ubiquitin ligase.* Proc Natl Acad Sci U S A, 2002. **99**(6): p. 3794-9.
45. Panchamoorthy, G., et al., *p120cbl is a major substrate of tyrosine phosphorylation upon B cell antigen receptor stimulation and interacts in vivo with Fyn and Syk tyrosine kinases, Grb2 and Shc adaptors, and the p85 subunit of phosphatidylinositol 3-kinase.* J Biol Chem, 1996. **271**(6): p. 3187-94.
46. Finco, T.S., et al., *LAT is required for TCR-mediated activation of PLCgamma1 and the Ras pathway.* Immunity, 1998. **9**(5): p. 617-26.
47. Zhang, W., et al., *LAT: the ZAP-70 tyrosine kinase substrate that links T cell receptor to cellular activation.* Cell, 1998. **92**(1): p. 83-92.
48. Piechulek, T., et al., *Isozyme-specific stimulation of phospholipase C-gamma2 by Rac GTPases.* J Biol Chem, 2005. **280**(47): p. 38923-31.

49. Walliser, C., et al., *rac regulates its effector phospholipase Cgamma2 through interaction with a split pleckstrin homology domain.* J Biol Chem, 2008. **283**(44): p. 30351-62.
50. Medina, D.L., et al., *TrkB regulates neocortex formation through the Shc/PLCgamma-mediated control of neuronal migration.* EMBO J, 2004. **23**(19): p. 3803-14.
51. Wu, D., et al., *EGFR-PLCgamma1 signaling mediates high glucose-induced PKCbeta1-Akt activation and collagen I upregulation in mesangial cells.* Am J Physiol Renal Physiol, 2009. **297**(3): p. F822-34.
52. Thomas, S.M., et al., *Epidermal growth factor receptor-stimulated activation of phospholipase Cgamma-1 promotes invasion of head and neck squamous cell carcinoma.* Cancer Res, 2003. **63**(17): p. 5629-35.
53. Davies, G., et al., *Phospholipase-C gamma-1 (PLCgamma-1) is critical in hepatocyte growth factor induced in vitro invasion and migration without affecting the growth of prostate cancer cells.* Urol Oncol, 2008. **26**(4): p. 386-91.
54. Meyer, R.D., C. Latz, and N. Rahimi, *Recruitment and activation of phospholipase Cgamma1 by vascular endothelial growth factor receptor-2 are required for tubulogenesis and differentiation of endothelial cells.* J Biol Chem, 2003. **278**(18): p. 16347-55.
55. Peters, K.G., et al., *Point mutation of an FGF receptor abolishes phosphatidylinositol turnover and Ca2+ flux but not mitogenesis.* Nature, 1992. **358**(6388): p. 678-81.
56. Hess, J.A., et al., *Analysis of platelet-derived growth factor-induced phospholipase D activation in mouse embryo fibroblasts lacking phospholipase C-gamma1.* J Biol Chem, 1998. **273**(32): p. 20517-24.
57. Larose, L., et al., *Identification of residues in the beta platelet-derived growth factor receptor that confer specificity for binding to phospholipase C-gamma 1.* Oncogene, 1993. **8**(9): p. 2493-9.
58. Liao, H.J., et al., *Absence of erythrogenesis and vasculogenesis in Plcg1-deficient mice.* J Biol Chem, 2002. **277**(11): p. 9335-41.
59. Shirane, M., et al., *Deficiency of phospholipase C-gamma1 impairs renal development and hematopoiesis.* Development, 2001. **128**(24): p. 5173-80.
60. Wen, R., et al., *An important role of phospholipase Cgamma1 in pre-B-cell development and allelic exclusion.* EMBO J, 2004. **23**(20): p. 4007-17.
61. Behjati, S., et al., *Recurrent PTPRB and PLCG1 mutations in angiosarcoma.* Nat Genet, 2014. **46**(4): p. 376-9.
62. Arteaga, C.L., et al., *Elevated content of the tyrosine kinase substrate phospholipase C-gamma 1 in primary human breast carcinomas.* Proc Natl Acad Sci U S A, 1991. **88**(23): p. 10435-9.
63. Noh, D.Y., et al., *Elevated content of phospholipase C-gamma 1 in colorectal cancer tissues.* Cancer, 1994. **73**(1): p. 36-41.
64. Xie, Z., et al., *Phospholipase C-gamma1 is required for the epidermal growth factor receptor-induced squamous cell carcinoma cell mitogenesis.* Biochem Biophys Res Commun, 2010. **397**(2): p. 296-300.
65. Wang, W., R. Eddy, and J. Condeelis, *The cofilin pathway in breast cancer invasion and metastasis.* Nat Rev Cancer, 2007. **7**(6): p. 429-40.
66. Mouneimne, G., et al., *Phospholipase C and cofilin are required for carcinoma cell directionality in response to EGF stimulation.* J Cell Biol, 2004. **166**(5): p. 697-708.
67. Jones, N.P. and M. Katan, *Role of phospholipase Cgamma1 in cell spreading requires association with a beta-Pix/GIT1-containing complex, leading to activation of Cdc42 and Rac1.* Mol Cell Biol, 2007. **27**(16): p. 5790-805.
68. Price, L.S., et al., *Calcium signaling regulates translocation and activation of Rac.* J Biol Chem, 2003. **278**(41): p. 39413-21.

69. Kim, M.J., et al., *Direct interaction of SOS1 Ras exchange protein with the SH3 domain of phospholipase C-gamma1.* Biochemistry, 2000. **39**(29): p. 8674-82.
70. Xie, Z., et al., *The SH3 domain, but not the catalytic domain, is required for phospholipase C-gamma1 to mediate epidermal growth factor-induced mitogenesis.* Biochem Biophys Res Commun, 2010. **398**(4): p. 719-22.
71. *International Consortium Completes Human Genome Project*. 2003; Available from: https://www.genome.gov/11006929/.
72. Han, X., et al., *Extending top-down mass spectrometry to proteins with masses greater than 200 kilodaltons.* Science, 2006. **314**(5796): p. 109-12.
73. Yates, J.R., C.I. Ruse, and A. Nakorchevsky, *Proteomics by mass spectrometry: approaches, advances, and applications.* Annu Rev Biomed Eng, 2009. **11**: p. 49-79.
74. Shen, Y.L., ML, *General Equation for Peak Capacity in Column Chromatography.* Anal Chem, 1998. **70**: p. 3853-2856.
75. Kocher, T., R. Swart, and K. Mechtler, *Ultra-high-pressure RPLC hyphenated to an LTQ-Orbitrap Velos reveals a linear relation between peak capacity and number of identified peptides.* Anal Chem, 2011. **83**(7): p. 2699-704.
76. Thakur, S.S., et al., *Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation.* Mol Cell Proteomics, 2011. **10**(8): p. M110 003699.
77. Wilson, I.D., et al., *High resolution "ultra performance" liquid chromatography coupled to oa-TOF mass spectrometry as a tool for differential metabolic pathway profiling in functional genomic studies.* J Proteome Res, 2005. **4**(2): p. 591-8.
78. Steen, H. and M. Mann, *The ABC's (and XYZ's) of peptide sequencing.* Nat Rev Mol Cell Biol, 2004. **5**(9): p. 699-711.
79. Zhang, G., D. Fenyo, and T.A. Neubert, *Evaluation of the variation in sample preparation for comparative proteomics using stable isotope labeling by amino acids in cell culture.* J Proteome Res, 2009. **8**(3): p. 1285-92.
80. Chen, X., et al., *Quantitative proteomics using SILAC: Principles, applications, and developments.* Proteomics, 2015. **15**(18): p. 3175-92.
81. Rauniyar, N. and J.R. Yates, 3rd, *Isobaric labeling-based relative quantification in shotgun proteomics.* J Proteome Res, 2014. **13**(12): p. 5293-309.
82. Dephoure, N. and S.P. Gygi, *Hyperplexing: a method for higher-order multiplexed quantitative proteomics provides a map of the dynamic response to rapamycin in yeast.* Sci Signal, 2012. **5**(217): p. rs2.
83. Latosinska, A., et al., *Comparative Analysis of Label-Free and 8-Plex iTRAQ Approach for Quantitative Tissue Proteomic Analysis.* PLoS One, 2015. **10**(9): p. e0137048.
84. Bantscheff, M. and B. Kuster, *Quantitative mass spectrometry in proteomics.* Anal Bioanal Chem, 2012. **404**(4): p. 937-8.
85. Old, W.M., et al., *Comparison of label-free methods for quantifying human proteins by shotgun proteomics.* Mol Cell Proteomics, 2005. **4**(10): p. 1487-502.
86. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.
87. Rost, H.L., et al., *OpenMS: a flexible open-source software platform for mass spectrometry data analysis.* Nat Methods, 2016. **13**(9): p. 741-8.
88. Vaudel, M., et al., *PeptideShaker enables reanalysis of MS-derived proteomics data sets.* Nat Biotechnol, 2015. **33**(1): p. 22-4.

89. Yu, K. and A.R. Salomon, *PeptideDepot: flexible relational database for visual analysis of quantitative proteomic data and integration of existing protein information.* Proteomics, 2009. **9**(23): p. 5350-8.
90. Yu, K., et al., *Integrated platform for manual and high-throughput statistical validation of tandem mass spectra.* Proteomics, 2009. **9**(11): p. 3115-25.
91. Khoury, G.A., R.C. Baliban, and C.A. Floudas, *Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database.* Sci Rep, 2011. **1**.
92. Cohen, P., *The origins of protein phosphorylation.* Nat Cell Biol, 2002. **4**(5): p. E127-30.
93. Vlastaridis, P., et al., *Estimating the total number of phosphoproteins and phosphorylation sites in eukaryotic proteomes.* Gigascience, 2017. **6**(2): p. 1-11.
94. Busino, L., et al., *Cdc25A phosphatase: combinatorial phosphorylation, ubiquitylation and proteolysis.* Oncogene, 2004. **23**(11): p. 2050-6.
95. Paz, J.C., et al., *Combinatorial regulation of a signal-dependent activator by phosphorylation and acetylation.* Proc Natl Acad Sci U S A, 2014. **111**(48): p. 17116-21.
96. Sawicka, A. and C. Seiser, *Histone H3 phosphorylation - a versatile chromatin modification for different occasions.* Biochimie, 2012. **94**(11): p. 2193-201.
97. Larsen, M.R., et al., *Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns.* Mol Cell Proteomics, 2005. **4**(7): p. 873-86.
98. Gates, M.B., K.B. Tomer, and L.J. Deterding, *Comparison of metal and metal oxide media for phosphopeptide enrichment prior to mass spectrometric analyses.* J Am Soc Mass Spectrom, 2010. **21**(10): p. 1649-59.
99. Zhou, H., et al., *Highly specific enrichment of phosphopeptides by zirconium dioxide nanoparticles for phosphoproteome analysis.* Electrophoresis, 2007. **28**(13): p. 2201-15.
100. Sharma, K., et al., *Ultradeep human phosphoproteome reveals a distinct regulatory nature of Tyr and Ser/Thr-based signaling.* Cell Rep, 2014. **8**(5): p. 1583-94.
101. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
102. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.
103. Kanehisa, M., *The KEGG database.* Novartis Found Symp, 2002. **247**: p. 91-101; discussion 101-3, 119-28, 244-52.
104. Zeeberg, B.R., et al., *GoMiner: a resource for biological interpretation of genomic and proteomic data.* Genome Biol, 2003. **4**(4): p. R28.
105. Huang, D.W., et al., *The DAVID Gene Functional Classification Tool: a novel biological module-centric algorithm to functionally analyze large gene lists.* Genome Biol, 2007. **8**(9): p. R183.
106. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
107. Lavallee-Adam, M., et al., *PSEA-Quant: a protein set enrichment analysis on label-free and label-based protein quantification data.* J Proteome Res, 2014. **13**(12): p. 5496-509.
108. Cha, S., et al., *In situ proteomic analysis of human breast cancer epithelial cells using laser capture microdissection: annotation by protein set enrichment analysis and gene ontology.* Mol Cell Proteomics, 2010. **9**(11): p. 2529-44.
109. Obenauer, J.C., L.C. Cantley, and M.B. Yaffe, *Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs.* Nucleic Acids Res, 2003. **31**(13): p. 3635-41.
110. Goel, R., et al., *Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis.* Mol Biosyst, 2012. **8**(2): p. 453-63.

111. Lachmann, A. and A. Ma'ayan, *KEA: kinase enrichment analysis.* Bioinformatics, 2009. **25**(5): p. 684-6.
112. Casado, P., et al., *Kinase-substrate enrichment analysis provides insights into the heterogeneity of signaling pathway activation in leukemia cells.* Sci Signal, 2013. **6**(268): p. rs6.
113. Suo, S.B., et al., *PSEA: Kinase-specific prediction and analysis of human phosphorylation substrates.* Sci Rep, 2014. **4**: p. 4524.

Chapter 2

A PLC-γ1 Feedback Pathway Regulates Lck Substrate Phosphorylation at the T Cell Receptor and SLP-76 Complex

This chapter is adapted from a manuscript published in the Journal of Proteome Research (DOI: 10.1021/acs.jproteome.6b01026). The original authors are Judson Belmont, Tao Gu, Ashley Mudd, and Art Salomon.

2.1. Introduction

In response to antigenic stimulation, CD4+ T cells undergo transcriptional, biochemical, and morphological changes to enable them to carry out their immune defense functions. At the intracellular level, these functions are regulated by tightly coordinated signaling cascades. The generation of an immune response of proper strength and duration requires careful coordination of signal transduction machinery, since autoimmune or immunodeficiency disorders can occur when competing biochemical signals are improperly balanced.

Although many of the details of early T cell signaling remain to be elucidated, it is now understood that a dynamic tyrosine phosphoproteome underlies many of the significant events in early T cell activation. Initiation of T cell signaling occurs when a cognate peptide-major histocompatibility complex engages with the T cell receptor (TCR), resulting in clustering of the cytoplasmic tails of the receptor subunits as well as recruitment of the Src family kinase Lck. Once recruited there, Lck phosphorylates immunoreceptor tyrosine-based activation motifs (ITAMs) located on the ζ and CD3 subunits of the receptor.[1] The phosphorylated CD3 and ζ ITAMs provide a docking site for the SH2 domains of the Syk family kinase Zap-70, which becomes activated through both autophosphorylation and modification by Lck.[2-4] Subsequently, Zap-70 and Lck kinase activity promote the activation of downstream signaling modules through their interactions with a multiprotein signalosome complex nucleated by SH2 domain-containing leukocyte phosphoprotein of 76 kDa (SLP-76) and LAT.[5] Assembly and phosphorylation of proteins within the signalosome contributes critically to a variety of downstream effects, including mobilization of Ca^{2+}, cytoskeletal reorganization, activation of MAPK pathways, modulation of protein kinase C (PKC) signaling, and activation of transcriptional programs.[5-9]

Phospholipase C-γ1 (PLC-γ1) is a member of the Phospholipase C family of enzymes that is widely expressed throughout many tissues.[10] During T cell activation, Itk phosphorylates PLC-γ1 on Tyr^{783} to reorient the active site into an open conformation.[10-12] In this conformation,

PLC-γ1 catalyzes the cleavage of phosphatidylinositol 4,5-bisphosphate (PIP_2) to inositol 1,4,5-trisphosphate (IP_3) and diacylglycerol (DAG).[10] Both DAG and IP_3 function as essential second messengers in the transduction of T cell signaling: DAG is a cofactor for various PKC-family enzymes, while IP_3 regulates release of Ca^{2+} from the ER.[10] PLC-γ1 and its related isoform, PLC-γ2, are distinguished from the rest of the phospholipase C family by the presence of tandem SH2 domains, which are essential for PLC-γ1's functionality and activation, as well as an SH3 domain, which mediates the constitutive interaction of PLC-γ1 with the proline-rich region of SLP-76 in resting T cells.[13-15] Both PLC-γ1 and PLC-γ2 contain an N-terminal PH domain that is characteristic of most phospholipase C enzymes and allows for recruitment of the enzymes to PIP_3 in the membrane.[16] Like PLC-γ1, PLC-γ2 is also positively regulated by Itk phosphorylation of Tyr^{759} that allows for an intramolecular interaction between the phosphorylated residue and the C-terminal SH2 domain.[17, 18] Despite its structural similarities to PLC-γ1, however, PLC-γ2 is more narrowly expressed, being restricted primarily to hematopoietic lineages, and has previously been shown to incompletely compensate for loss of PLC-γ1.[10, 19]

Phospholipase C signaling has far-reaching consequences in cell spreading and metastasis as well as thymocyte development and T-cell activation. PLC-γ1 has been shown to promote chemotaxis via modulation of Rac and CDC42 activity, and elevated expression of PLC-γ1 is a commonly observed signature of metastatic tumors.[20-22] In mice, conditional deletion of PLC-γ1 in thymocytes produces a profound block in both positive and negative selection during development.[23] These mice also experience an increased prevalence of autoimmune disorders as well as a substantial reduction in peripheral T cells, reportedly due to an elevated rate of activation-induced cell death.[23] Additionally, T cells that do successfully mature in these lines exhibit defects in proliferation and cytokine production in response to antigen stimulation.[14, 23] At the intracellular level, PLC-γ1 is required in mature T cells for sustained increases in intracellular Ca^{2+} as well as activation of Erk, Jnk, and the transcription factors NFAT, AP-1, and NF-κB.[14, 23] The defect in Ca^{2+} signaling observed in PLC-γ1 cells is not absolute, as these

cells still exhibit a transient increase in Ca^{2+} in response to TCR engagement.[14] This has led to speculation that PLC-γ2 expressed in T cells may be sufficient for temporary, but not sustainable, elevation of intracellular Ca^{2+};[19] however, it is presently unknown whether the incompleteness of this compensation is due strictly to the comparatively lower levels of PLC-γ2 expressed in T cells or whether PLC-γ1 and PLC-γ2 carry out non-redundant functions in T cell signaling.[24, 25]

While the significance of PLC-γ1 in TCR signal transduction is widely recognized, its role in regulating the receptor-proximal phosphorylation events that govern the pathway is not well understood. Recent studies have revealed the existence of both positive and negative feedback pathways from components of the signalosome complex to the apical kinase in the pathway, Lck. N terminal tyrosine residues of SLP-76 regulate a negative feedback pathway early and positive feedback pathway late after receptor stimulation to the activation site on Lck, Tyr^{394}, and the substrates of active Lck, including the CD3 and ζ ITAMs and ZAP-70.[26, 27] Another study showed that Vav1 regulates a negative feedback pathway to Lck Tyr^{394} as well as the CD3 and ζ ITAMs and ZAP-70.[28] In the present study, we performed an LC-MS/MS-based quantitative phosphoproteomic analysis to identify altered phosphorylation during early T cell signaling in a PLC-γ1-deficient cell line, Jgamma1, and its PLC-γ1-reconstituted counterpart. Our data reveal a widespread dependence on PLC-γ1 for hundreds of sites across the tyrosine phosphoproteome. Notably, these sites include known regulatory residues on Lck, Zap-70, and Itk, as well as many more sites on established targets of Lck. These observations both reaffirm the centrality of PLC-γ1 to T cell functionality and suggest that events in the signaling pathway as early as the TCR stimulation-induced phosphorylation of ITAM sites are regulated by PLC-γ1. Furthermore, our data point to a previously uncharacterized mechanism through which PLC-γ1 regulates the activity and targeting of Lck.

2.2. Materials and Methods

2.2.1. Cell Culture and T Cell Stimulation

The PLC-γ1 deficient cell line Jgamma1 (ATCC#CRL-2678) and its reconstituted counterpart Jgamma1.WT (ATCC# CRL-2679)[14] were obtained from American Tissue Culture Collection (ATCC, Manassas, VA). Both cell lines were cultured in RPMI 1640 medium (Hyclone, Logan, UT) supplemented with 10% heat-inactivated FBS (Hyclone, Logan, UT), 2mM L-glutamine, 100U/ml penicillin G, and 100µg/ml streptomycin (Invitrogen, Carlsbad, CA) in a humidified incubator with 5% CO_2 at 37°C.

The TCR stimulation was performed as previously described[29] by treating with anti-CD3 and anti-CD4 (clones OKT3 and OKT4; eBioscience, San Diego, CA). For each replicate and time point, 1×10^8 cells were treated with 2.5µg/ml each of anti-CD3 and anti-CD4 antibodies for 30 seconds at 37°C. Cells were then treated with 22µg/ml of goat anti-mouse IgG (Jackson ImmunoResearch, West Grove, PA) to cross-link OKT3/4 before incubation at 37°C for 0, 1, 2, 3, 5, or 10 minutes. Stimulation was halted with the addition of lysis buffer (8M urea, 20mM HEPES, 2.5mM sodium pyrophosphate, 1mM sodium orthovanadate, 1mM β-glycerophosphate, pH=8.0) followed by a 30 minute incubation at 4°C. Lysates were sonicated at a 30watt output with 3 bursts of 30 seconds each and cleared by centrifugation at 12,000×g for 15 minutes at 4 °C.

2.2.2. Protein Reduction, Alkylation, Digestion and Peptide Immunoprecipitation

Protein reduction and alkylation were performed as previously described.[30] Trypsin digestion was also performed according to a previously detailed protocol,[30] but with TPCK-treated trypsin (Worthington, Lakewood, NJ) substituted and used in a 1:1 (w/w) trypsin:protein ratio. Tryptic peptides were acidified, cleared, and desalted using C18 Sep-Pak plus cartridges (Waters, Milford, MA) in accordance with a previously described protocol,[30] then lyophilized for 48 hours.

Prior to peptide immunoprecipitation, a 10 pmol fraction of a synthetic phosphopeptide (LIEDAEpYTAK) was spiked in to each sample to serve as a quantitation standard. Peptide immunoprecipitation was performed using a previously detailed protocol[30] employing p-Tyr-100 phosphotyrosine antibody beads (Cell Signaling Technology). After washing, peptides were eluted from the beads with 0.15% TFA and desalted via C18 Zip Tip pipette tips (Millipore Corporation Billerica, MA) as previously described.[31] Five biological replicates of each time point were prepared for the phosphoproteomic investigation of each cell line.

2.2.3. Automated nano-LC-MS

Phosphopeptides were loaded from an Agilent 1200 G1367B high performance Autosampler to a C18 precolumn in accordance with a previously detailed method,[30] with the exception that the precolumn and analytical column in the current study were packed with 2cm of 3-μm C18 particles, and an electrospray voltage of 2.0kV was applied to the flow path via platinum wire using an Upchurch MicroTee (IDEX Health & Science, Oak Harbor, WA) between the precolumn and analytical column.

MS spectra were acquired on a Q Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA) using parameters detailed previously,[32] with the exception that the ion selection abundance threshold was set at 1.0×10^3.

2.2.4. Data Analysis

Data analysis was performed using a custom phosphoproteomic software platform.[33, 34] MS/MS spectra were searched against a custom database comprised of the 87,613 forward protein sequences contained in the non-redundant human UniProt complete proteome set database (2/1/2013) and an equal number of reversed decoy entries using the Mascot algorithm (version 2.4.0, Matrix Science).[35] Peak lists were generated using Proteowizard's msconvert software (version 3.0.4888) and default parameters with the MS2Deisotope filter. The following parameters were used for the Mascot database search: trypsin enzyme cleavage specificity, 2

possible missed cleavages, 10 ppm mass tolerance for precursor ions, 50 mmu mass tolerance for fragment ions. Search parameters specified a variable modification of phosphorylation (+79.9663 Da) on serine, threonine, and tyrosine residues, a variable modification of methionine oxidation (+15.9949 Da), and a static modification of carbamidomethylation (+57.0215 Da) on cysteine. To validate sequence assignment of phosphopeptides, a logistic spectral score[33] filter was applied to attain a final estimated decoy database false discovery rate (FDR) of 1%. FDR was estimated via the decoy database approach following the collation of non-redundant data into heatmaps.[36] To ensure high confidence in the assigned positions of phosphorylation sites, the Ascore algorithm[37] was applied to each phosphopeptide spectrum match, and the top Ascore prediction was used as the reported phosphorylation site position.

2.2.5. Quantitation of Relative Phosphopeptide Abundance

Label-free quantitation of relative phosphopeptide abundance was achieved by integrating selected ion chromatogram (SIC) peak areas for all phosphopeptides. Retention time alignment of individual replicate analyses was performed as described previously, as was calculation of peak areas using a previously established in-house software package updated for compatibility with Xcalibur Development Kit 2.2 SP1 (Thermo Fisher Scientific).[38] The peak area of the spiked-in LIEDAEpYTAK phosphopeptide was used as a normalization factor for each individual SIC peak area during quantitation.

Two different heatmaps were generated for each peptide observed in the data: an abundance heatmap showing the amount of phosphopeptide detected in a cell line, and a ratio heatmap indicating the fold change in phosphopeptide abundance between the two cell lines. Abundance heatmaps were generated for the Jgamma1.WT line across the timecourse of TCR stimulation.[39] Each individual heatmap square color was generated from the average of the standard phosphopeptide-normalized SICs from five biological replicate experiments. In the abundance heatmap representation, the geometric mean of a given phosphopeptide abundance across all time points was set to the color black. A blue color represented below average

abundance, while yellow represented above average abundance for each unique phosphopeptide across the timecourse. Blank squares in the heatmap indicated that no clearly defined SIC peaks were observed for that phosphopeptide in any of the replicate analyses at that time point. *p* values were calculated from the replicate data using a two-sample t-test comparing each time point to the time point with the minimum average peak area for that phosphopeptide. To adjust for multiple hypothesis testing, *q* values were subsequently calculated for each time point using the R package QVALUE as previously described.[40, 41] A white dot on an abundance heatmap square indicated that a significant difference ($q < 0.05$) was detected for that phosphopeptide and timepoint relative to the timepoint with the minimal value.

In the ratio heatmap, the ratios of phosphopeptide abundances between the Jgamma1 and Jgamma1.WT cell lines for each timepoint within the time course of TCR stimulation were represented. For the ratio heatmap, a black color represented a ratio of 1 between the Jgamma1 and Jgamma1.WT cells at that time point. A red color represented lower abundance, while a green color represented higher abundance of the given phosphopeptide in Jgamma1 cells compared with the Jgamma1.WT cells. The magnitude of change of the heatmap color was calculated as described.[39] Two-sample t-tests were performed to identify changes in abundance between the Jgamma1 and Jgamma1.WT cells for each phosphopeptide and time point, and *q* values were subsequently calculated to control the FDR. A white dot on a heatmap square indicated that a significant change ($q < 0.05$) was observed between the replicate data from the Jgamma1 and Jgamma1.WT cells samples for that time point and phosphopeptide.

2.2.6. Hierarchical Clustering

Hierarchical clustering was performed using Cluster 3.0[42] and visualized with TreeView 3.0.[43] Input to the hierarchical clustering algorithm is a 195x6 matrix of phosphopeptide peak areas, where 195 is the number of peptides from our phosphoproteomic dataset selected for clustering, and each row consists of the log2-transformed ratios of average peptide peak areas between Jgamma1 and Jgamma1.WT at the 6 timepoints sampled. To be selected for clustering,

a peptide had to be derived from a protein containing either one or more Lck-phosphorylated sites documented in the PhosphoSitePlus database,[44] or one or more sites within an Lck kinase motif predicted at high stringency in Scansite.[45] Peptides derived from Lck are included in the analysis due to the prevalence of Lck autophosphorylation during TCR signaling. Additional peptides were included based on in-house manual curation of known Lck substrates. For hierarchical clustering, Pearson correlation coefficient was used as the distance metric, and clusters distances were assessed based on average linkage.

2.2.7. Western Blotting

Cell lysates prepared with 8 M urea were diluted 1:1 with sample loading buffer (20% v/v glycerol, 5% 2-mercaptoethanol, 4% SDS, 125 mM Tris-HCl, pH 6.8, 0.01% bromophenol blue) for each proteomic sample. Following a DC protein assay, equal amounts of protein were separated by 4-20% gradient SDS-PAGE on Precise Tris-HEPES gel (Thermo Fisher Scientific), and electroblotted onto Immobilon membranes (Millipore, Billerica, MA). Membranes were blocked in Odyssey blocking buffer for 30 minutes at RT (Li-Cor, Lincoln, NE) and then incubated with primary antibody overnight at 4°C. Membranes were washed 3 times for 5 minutes at RT in PBS/0.1% Tween-20, incubated with IgGs conjugated with an IRDye (Li-Cor) for 45 minutes, then washed again. Bands were visualized using an Odssey Imaging System (Li-Cor). Antibodies to CD3ε, PLC-γ1, Zap-70, and phospho-Zap-70 (Tyr319 and Ty493) were obtained from Cell Signaling Technologies. Antibody to TCRζ was obtained from Santa Cruz Biotechnology. Antibody to GADPH was obtained from Sigma-Aldrich. Antibody to Itk was obtained from Thermo Fisher.

2.3. Results

A quantitative, bottom-up phosphoproteomic strategy was used to compare the phosphotyrosine profiles of Jurkat PLC-γ1-deficient cells (Jgamma1) and PLC-γ1-reconstituted cells (Jgamma1.WT). After validating the PLC-γ1 deficiency and reconstitution of the Jgamma1

and Jgamma1.WT cell lines respectively (Supplemental Figure 2.1), samples were taken from each cell line at six timepoints ranging from 0-10 minutes following stimulation of the cells with OKT3 crosslink. For each time point and cell line, five biological replicates were analyzed by LC-MS/MS. The resulting phosphoproteomic data represents an unbiased, wide-scale view of the dynamic tyrosine phosphoproteome during early TCR signaling in the presence and absence of PLC-γ1.

After processing the phosphoproteomic samples and obtaining raw MS/MS spectra, high confidence sequence assignments were made as described in the Materials and Methods section. Relative abundance of each phosphopeptide at each time point was calculated by integrating SIC peak areas. A complete list of identified phosphopeptides and their associated SIC peak areas and statistics, all LC-MS-specific peptide data including MOWSE score and Ascore, and annotated MSMS spectra for all PTM-containing peptides from this study are provided as supporting information available at http://pubs.acs.org/doi/abs/10.1021/acs.jproteome.6b01026. The raw LC-MS data (Thermo Fisher raw format) and assigned MSMS spectra (pepXML) for all peptides have been submitted to the ProteomeXchange (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD005165.[46, 47]

To better visualize the phosphoproteomic data, two types of heatmaps were generated for each peptide using the in-house Peptide Depot software platform.[48] For each cell line, abundance heatmaps visualizing the average phosphopeptide abundances across the time series were created (Figure 2.1, pictured in a blue/yellow color scale). Additionally, a ratio heatmap visualizing the fold changes in phosphopeptide abundance between Jgamma1 and Jgamma1.WT at each timepoint was also created (Figure 2.1, pictured in a red/green color scale). As discussed in the Materials and Methods section, stringent statistical testing and multiple hypothesis correction were performed to determine both significant changes in phosphopeptide abundance as a function of time as well as significant changes in phosphopeptide abundance between the

two cell lines. The biological conclusions discussed herein are drawn from timepoints with $q < 0.05$, which are annotated directly on the corresponding heatmap squares with a white dot.

Overall, a total of 1557 unique tyrosine phosphorylation sites from 1012 proteins were identified at less than 1% FDR. The high quality and reproducibility of each LC-MS run were demonstrated by analyzing the pairwise correlation of SIC peak areas observed in separate replicates (average R^2 of 0.83 ± 0.05) (Supplemental Figure 2.2).

2.3.1. Analysis of KEGG-Annotated TCR Signaling Proteins

To obtain a more focused view of phosphorylation events occurring on canonical TCR signaling proteins, we examined the subset of sites in our data on proteins included in the Kyoto Encyclopedia of Genes and Genomes (KEGG) T Cell Receptor Signaling Pathway category.[91] In total, 97 unique tyrosine phosphorylation sites on 54 KEGG-annotated TCR signaling pathway proteins were observed. A selection of these sites are displayed in Figure 2.1, and in the context of the pathway in Figure 2.2. Abundance heatmaps created for these sites in the Jgamma1.WT line (Figure 2.1) indicate that induction of TCR signaling proceeds normally in PLC-γ1-reconstituted T cells and with the expected kinetics.

2.3.2. Analysis of Lck Substrates in Jgamma1

Based on previous investigations revealing the existence of multiple feedback pathways connecting canonical T cell signaling proteins back to Lck,[26-28, 92] we hypothesized that PLC-γ1 may also influence the upstream kinase's activity via an indirect mechanism. To determine if Lck activity is altered by the loss of PLC-γ1, we performed hierarchical clustering of known and predicted Lck substrates observed in our dataset (Figure 2.3). This analysis revealed divergent phosphorylation patterns at Lck substrate sites, with distinct clusters of sites exhibiting constitutively increased and decreased fold-changes in phosphorylation across the timecourse. Among the phosphopeptides derived from CD3 receptor subunits and the TCRζ chain, the majority clustered together due to their decreased abundance in Jgamma1 (Figure

2.3B, cluster v). In contrast, phosphopeptides exhibiting constitutively elevated abundance in Jgamma1 also clustered together, and notably included a number of sites located on SLP-76-associated proteins (Figure 2.3B, cluster iii). Together, these results suggest that PLC-γ1 positively regulates phosphorylation of Lck substrates at the TCR while negatively regulating phosphorylation of other Lck substrates.

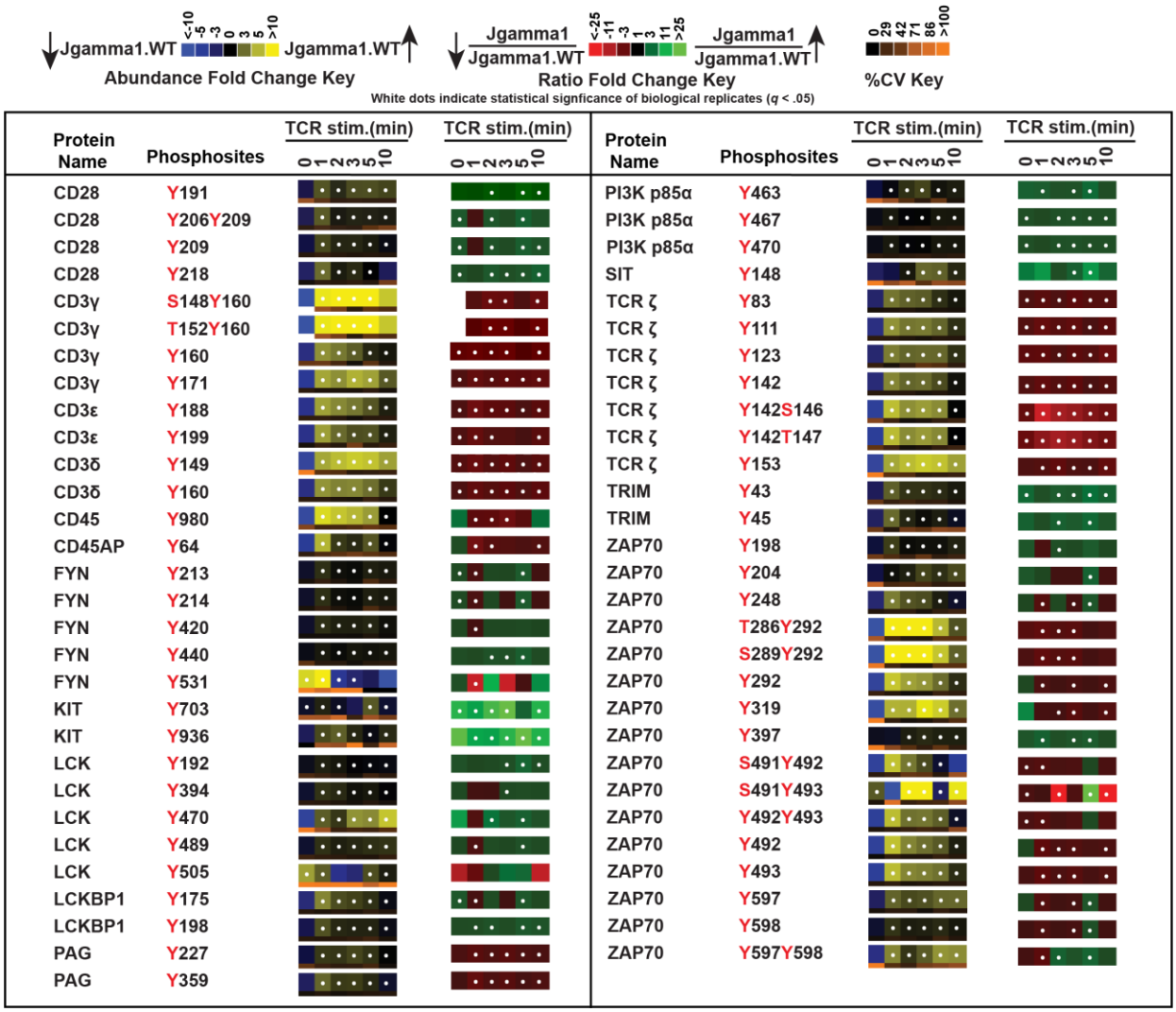


Figure 2.1: Quantitative phosphoproteomic analysis of receptor-proximal TCR signaling proteins. Heatmaps were calculated from five biological replicate experiments. Blue/yellow heatmaps represent temporal changes in phosphopeptide abundance in either the Jurkat PLC-γ1 null cell line (Jgamma1) or reconstituted cell line (Jgamma1.WT) following TCR stimulation. Black squares represent a phosphopeptide abundance equal to the geometric mean for that phosphopeptide across all time points for each cell line. Yellow and blue represent phosphopeptide abundances above and below the average, respectively. White dots in the yellow/blue heatmap squares indicate a statistically significant ($q < 0.05$) difference between the phosphopeptide abundance at the indicated timepoint and the lowest abundance of that peptide observed in the time course for that cell line. For each timepoint in which a sufficient number of replicate measurements were observed, the coefficient of variation (CV) was computed. The magnitude of the CV is indicated by the black/orange heatmap underneath each time point. Red/green heatmaps indicate the ratio between the peptide abundances measured in Jgamma1 and Jgamma1.WT cells for each phosphopeptide at each time point. Black signifies no difference in peptide abundance between the cell lines, while red and green represent reduced and

elevated phosphopeptide levels in Jgamma1 relative to Jgamma1.WT, respectively. White dots on red/green ratio heatmap squares indicate a statistically significant difference ($q < 0.05$) in the comparison between Jgamma1 cells and Jgamma1.WT cells ratios for that time point.

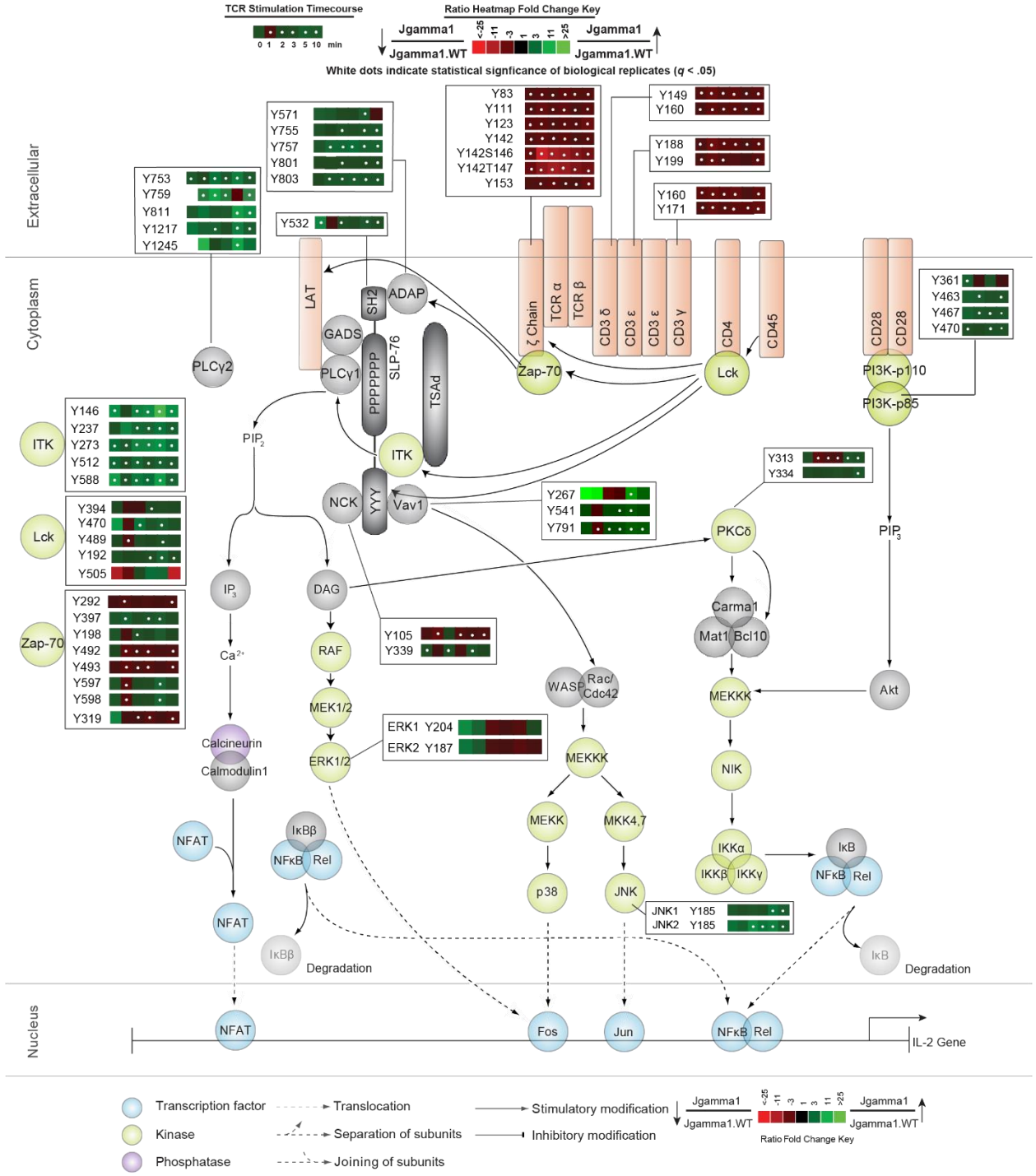


Figure 2.2: PLC-γ1 deficiency perturbs phosphorylation on canonical TCR signaling proteins. The canonical TCR signaling pathway is depicted, along with quantitative Jgamma1 to Jgamma1.WT ratio heatmaps depicting altered phosphorylation profiles of select sites. The interpretation of the heatmap color scheme as well as the white dots is the same as that described in Materials and Methods and Figure 2.1.

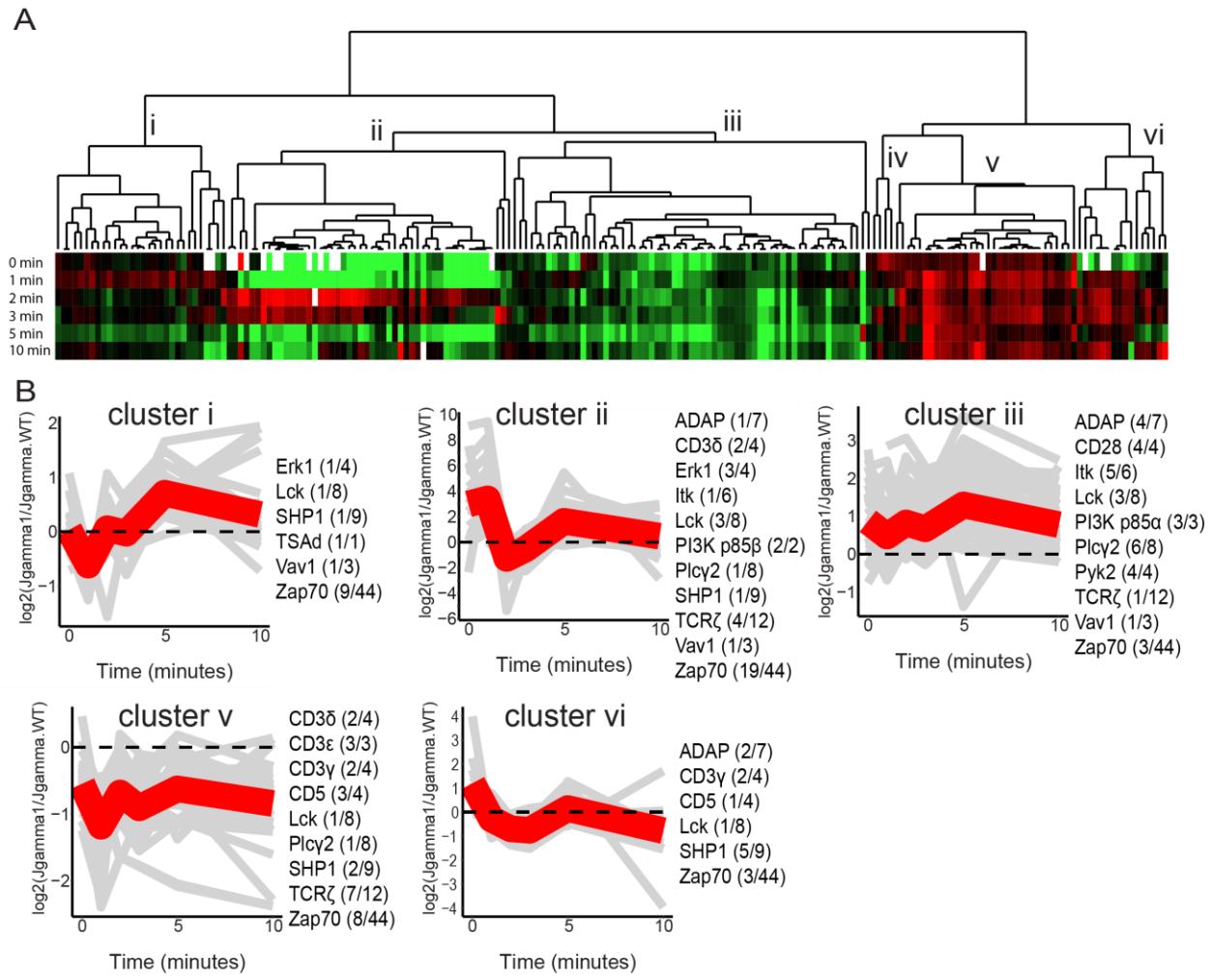


Figure 2.3: Hierarchical clustering of known and predicted Lck substrates reveals PLC-γ1-dependent programs of phosphorylation. A) Clustered heatmap of phosphopeptide abundance ratios. The clustered data consists of 195 phosphopeptides in the data derived from proteins that are known or confidently predicted Lck substrates (see Materials and Methods for details). For each peptide selected, the ratios of Jgamma1.WT to Jgamma1 abundance at each timepoint were log2 transformed. Hierarchical clustering was performed using Pearson correlation coefficient as the distance metric and average linkage to calculate cluster distances. B) Time series ratios of phosphopeptides in the five largest clusters obtained. Each plot represents the time course of ratios for each peptide in the cluster (grey series) as well as the population average for the cluster (red series). The text to the right of each plot lists the KEGG-annotated T cell receptor signaling pathway components represented in the cluster, and the proportion of total phosphopeptides of each protein observed in the cluster.

2.4. Discussion

2.4.1. TCR-CD3 ITAM and ZAP-70 Phosphorylation is Positively Regulated by PLC-γ1

Phosphorylation of CD3 and ζ subunit cytoplasmic tails by Lck is among the earliest events in the initiation of TCR signal transduction. Phosphopeptides observed in our data encompass nearly every ITAM tyrosine on the ζ chain as well as ITAM sites on CD3 δ, ε, and γ. The majority

of CD3 and ζ ITAM tyrosines exhibited constitutively and significantly decreased phosphorylation in Jgamma1 compared to Jgamma1.WT cells (Figure 2.1). Notably, phosphorylation of these residues is significantly lower in Jgamma1 at all timepoints, including samples taken prior to pathway activation at 0 minutes. Western blot confirmed that receptor subunit abundance was unchanged between Jgamma1 and Jgamma1.WT (Supplemental Figure 2.3). These data

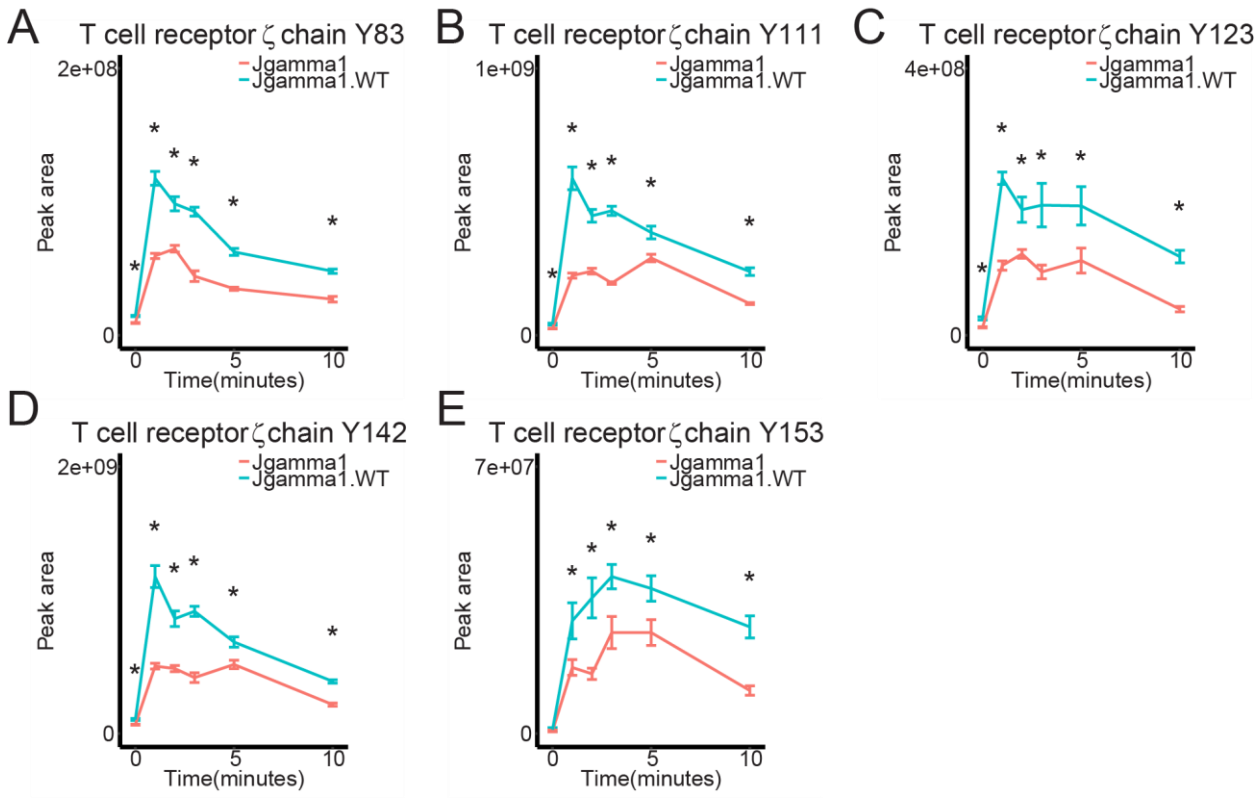


Figure 2.4: Phosphorylation kinetics of tyrosine residues located on the T cell receptor ζ chain. Phosphorylation kinetics of TCRζ tyrosines A) Tyr^{83}, B) Tyr^{111}, C) Tyr^{123}, D) Tyr^{142}, and E) Tyr^{153} are depicted during early TCR signaling. Each data point represents the mean phosphopeptide abundance measured across five biological replicate experiments, and error bars indicate the standard error of the mean. Asterisks represent a statistically significant difference ($q < 0.05$) between Jgamma1 cells and Jgamma1.WT cells at the indicated time point.

indicate that basal phosphorylation of receptor ITAM sites is disrupted in the absence of PLC-γ1, and that this disruption persists through the first 10 minutes of TCR activation. Despite the overall decrease in ITAM phosphorylation in Jgamma1, examination of ITAM phosphorylation kinetics

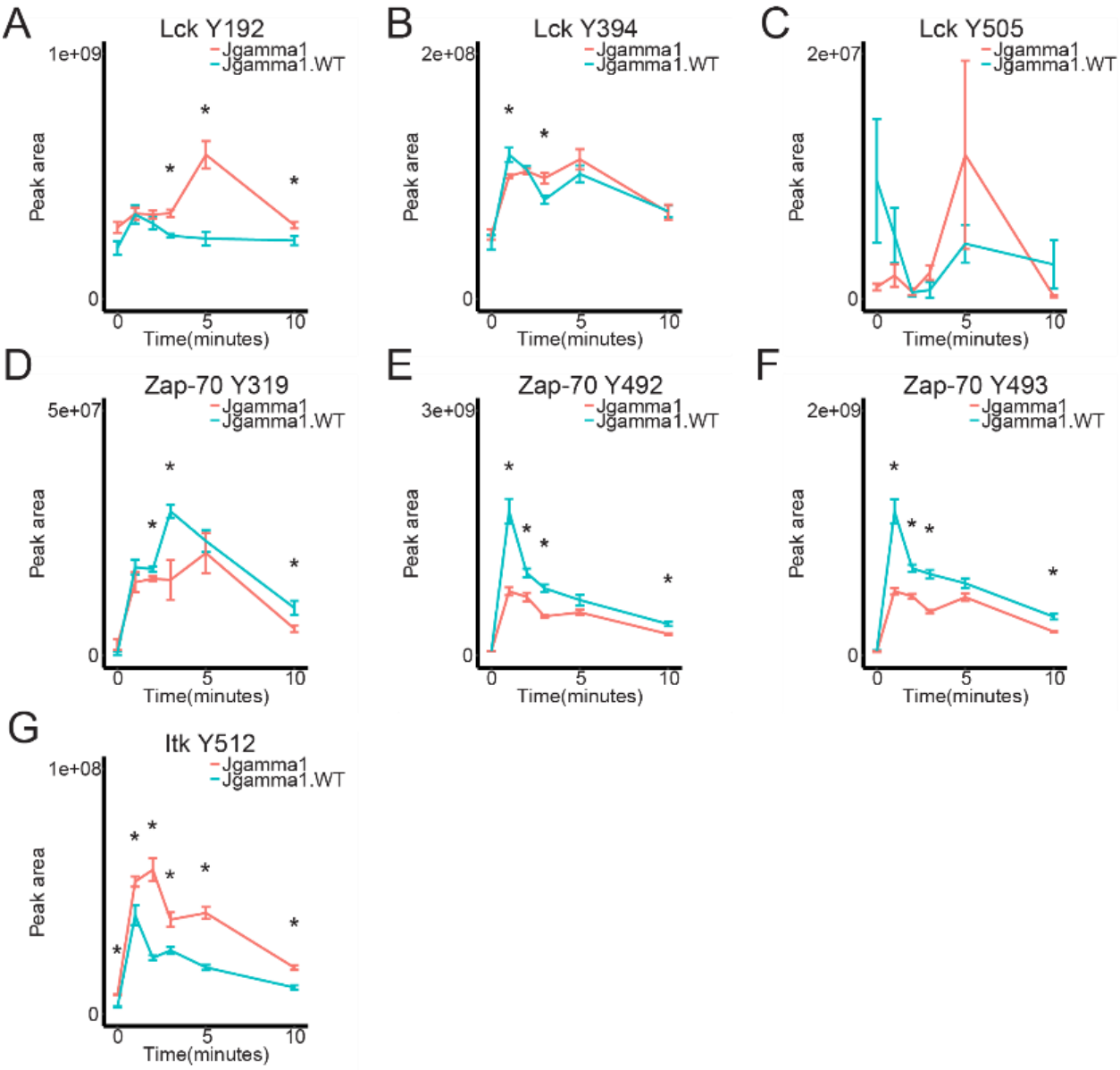


Figure 2.5: Phosphorylation kinetics of regulatory tyrosine residues on Lck, Zap-70, and Itk. Phosphorylation kinetics of Lck regulatory sites A) Tyr^{192}, B) Tyr^{394}, and C) Tyr^{505}, Zap-70 D) Tyr^{319}, E) Tyr^{492}, F) Tyr^{493}, and Itk G) Tyr^{512} are depicted during early TCR signaling. Each data point represents the mean phosphopeptide abundance measured across five biological replicate experiments, and error bars indicate the standard error of the mean. Asterisks represent a statistically significant difference ($q < 0.05$) between Jgamma1 cells and Jgamma1.WT cells at the indicated time point.

indicates that phosphorylation of these sites is still induced by OKT3 stimulation in both PLC-γ1-deficient and reconstituted cell lines (Figure 2.4).

2.4.2. TCR Complex Positive Regulation by PLC-γ1

The phosphorylated CD3 ITAMs provide a docking site for the SH2 domains of the Syk family kinase Zap-70, which becomes activated through both autophosphorylation and modification by Lck.[3, 49] Our phosphoproteomic findings revealed decreased phosphorylation of Tyr^{492} and Tyr^{493} of Zap-70, two positive regulatory residues located within the activation loop of the kinase[4, 50, 51] at multiple timepoints in Jgamma1, consistent with impaired activation of Zap-70 (Figure 2.5E-F). Additionally, we noted decreased phosphorylation of Tyr^{319} of Zap-70 in Jgamma1 (Figure 2.5D). This residue mediates Zap-70's interaction with Lck, and correlates with Zap-70 activity.[52-54] The statistically significant decreases in Tyr^{319} and Tyr^{493} phosphorylation observed in Jgamma1 were verified by Western blot (Supplemental Figure 2.4). The decreased phosphorylation of Zap-70 is consistent with the decreased phosphorylation of ITAMs in the PLC-γ1 deficient cells.

Lck phosphorylation of ITAM tyrosine residues is critical in T cell signaling.[55-57] Disruption of ITAM tyrosine phosphorylation in PLC-γ1 deficient cells could arise from decreased activation of Lck or mislocalization of the kinase away from substrate ITAMs. Tyr^{505}, a negative regulatory residue on Lck,[58, 59] revealed no statistically significant changes in phosphorylation (Figure 2.5A-C). Conversely, Tyr^{394}, which positively regulates Lck kinase activity,[58, 60] exhibited elevated phosphorylation in Jgamma1 relative to Jgammat1.WT at 3 minutes post-stimulation; however, this increase was short-lived, and was not detected 2 minutes later. Despite this evidence that cellular Lck kinase activity was slightly elevated with PLC-γ1 deficiency, the constitutive and significant decrease in phosphorylation on Lck substrate ITAM sites on ζ and

CD3δ, ε, and γ as well as the aforementioned Lck-regulated activation sites on Zap-70 indicates that phosphorylation of canonical Lck substrates is impaired in Jgamma1.

2.4.3. Regulation of Lck Specificity via Lck Tyr^{192}

Of the Lck regulatory tyrosines observed in this study, only Tyr^{192} exhibited persistent and significantly elevated phosphorylation in Jgamma1. Tyr^{192} is located in the SH2 domain of Lck, and regulates the ligand binding selectivity of the domain.[61, 62] In its Tyr^{192}-phosphorylated form, Lck demonstrates enhanced association with TSAd as well as Itk and Pyk2,[62, 63] both of which exhibited increased phosphorylation in Jgamma1 in our study (Figure 2.3, cluster iii). Phosphorylation of this site is thought to alter the substrate targeting of Lck by diverting it from its early signaling program, which includes phosphorylation of ITAMs and Zap-70, to a later one involving activation of Itk as well as effectors of negative feedback.[61, 62] Lck-deficient JCAM 1.6 cells reconstituted with mutant Lck-Y192F exhibit increased tyrosine phosphorylation on TCRζ and a greater tendency to form clusters of active Lck.[62, 64] Our observation that PLC-γ1 deficiency significantly elevated Tyr^{192} phosphorylation and decreased ITAM and Zap-70 phosphorylation supports a model wherein PLC-γ1 regulates Lck localization through Tyr^{192}.

2.4.4. SLP-76 Signalosome Component Phosphorylation is Negatively Regulated by PLC-γ1

The simultaneous observation of unimpaired Lck activity and significantly decreased phosphorylation of selected Lck substrates suggests a paradoxical outcome of PLC-γ1 deficiency on Lck activity. Yet this apparent paradox was resolved when an examination of proteins within the SLP-76 complex revealed significantly elevated phosphorylation on a multitude of sites, including multiple residues on Itk, Vav1, PLC-γ2, and ADAP (Figure 2.1, 2.3B, cluster iii). Although function and upstream kinases have not been definitively established for all the observed sites, all four proteins are known or predicted direct substrates of Lck that associate with SLP-76 and exhibit TCR stimulation-dependent phosphorylation.[17, 65-69] Among the multiple Itk tyrosine

residues exhibiting significant and constitutive elevated phosphorylation in Jgamma1, the canonical Lck substrate site, Tyr^{512}, is of particular interest (Figure 2.5G). Tyr^{512} is located within the activation loop of Itk, and phosphorylation of this residue by Lck promotes Itk's kinase activity.[66, 70] Subsequent examination confirmed that Itk abundance was unchanged between Jgamma1 and Jgamma1.WT (Supplemental Figure 2.5). In all, these data suggest that in the absence of PLC-γ1, Lck activity is shifted from the TCR to the SLP-76 complex and activation of Itk by Lck is promoted.

2.4.5. Models for PLC-γ1 Regulation of Lck through SLP-76 or TSAd

Reduced ITAM and Zap-70 phosphorylation and increased SLP-76 complex protein phosphorylation could be explained by mislocalization of Lck in PLC-γ1 deficient cells. Both PLC-γ1 and Lck have been demonstrated to bind SLP-76's proline-rich region in an SH3 domain-dependent fashion.[6, 13, 71-73] However, only PLC-γ1 has been shown to constitutively associate with SLP-76 in resting T cells.[73] In light of the phosphoproteomic data, we speculate that the absence of PLC-γ1 could allow Lck to bind the less occupied proline-rich region on SLP-76, bringing it into proximity with Itk and other SLP-76-scaffolded signaling proteins.[5] Decreased ITAM phosphorylation could thus result from depletion of Lck at the receptor. In support of this model, the Lck SH3 domain was shown previously to impact both Lck substrate selectivity and T cell development in mice using an SH3 inactivating mutant, W97A. Lck W97A mutant mice exhibit impaired early thymocyte development within the CD4+ and CD8+ lineages.[74] Lck substrate specificity was perturbed in Lck W97A thymocytes, with elevated Zap-70 phosphorylation and reduced LAT and PLC-γ1 phosphorylation in response to TCR stimulation.[74] Furthermore, Jurkat cells expressing Lck W97A showed dramatic elevation of TCRζ subunit and Zap-70 Tyr^{319} phosphorylation, opposite the patterns observed in our data (Figure 2.1).[75] These data indicate that the Lck SH3 domain decreases Lck substrate phosphorylation at the TCR, resulting in elevated SLP-76 complex phosphorylation both *in vitro* and *in vivo*.

Lck substrate specificity within the TCR and SLP-76 complex can also be regulated by the protein T cell specific adaptor (TSAd). Previous studies point to an inverse correlation between TSAd expression and TCRζ phosphorylation, with TSAd-deficient T cells exhibiting increased ζ phosphorylation[62] and TSAd-overexpressing T cells exhibiting decreased ζ phosphorylation.[76] Both Itk and Lck associate with TSAd, and this association is reinforced by Itk's phosphorylation of Tyr^{192} on Lck, which promotes the interaction of Lck's SH2 domain with TSAd.[62] Reciprocally, Lck and TSAd promote phosphorylation of Itk's Tyr^{512} residue, and TSAd is required for robust phosphorylation of Itk in Jurkat cells.[77] Our phosphoproteomic data reveal increased phosphorylation on multiple Itk residues spanning the full duration of the experiment (Figure 2.1,2.3). The elevation of phosphorylation at both Lck Tyr^{192} and Itk Tyr^{512} in PLC-γ1 deficient cells suggests that association of these proteins is increased in the absence of PLC-γ1. As a consequence of increased recruitment of Itk and Lck to TSAd, decreased phosphorylation

of CD3 and ζ ITAMs may occur as Lck is titrated away from CD3 subunits.

Models proposing sequestration of Lck on SLP-76 or TSAd are not mutually exclusive, as Lck may associate with both adaptors simultaneously via separate Src homology domains.[72, 78] Previous investigations have demonstrated that TSAd also associates directly with Nck, and in turn promotes its association with SLP-76.[79] Like Lck, Nck's associations with SLP-76 and TSAd may also be mediated by separate domains.[79-81] An intriguing possibility that is in agreement with the results of the present study as well as previous findings is that Lck could participate in a multimolecular signaling complex containing both SLP-76 and TSAd, the formation of which occurs only in late TCR signaling under normal conditions.

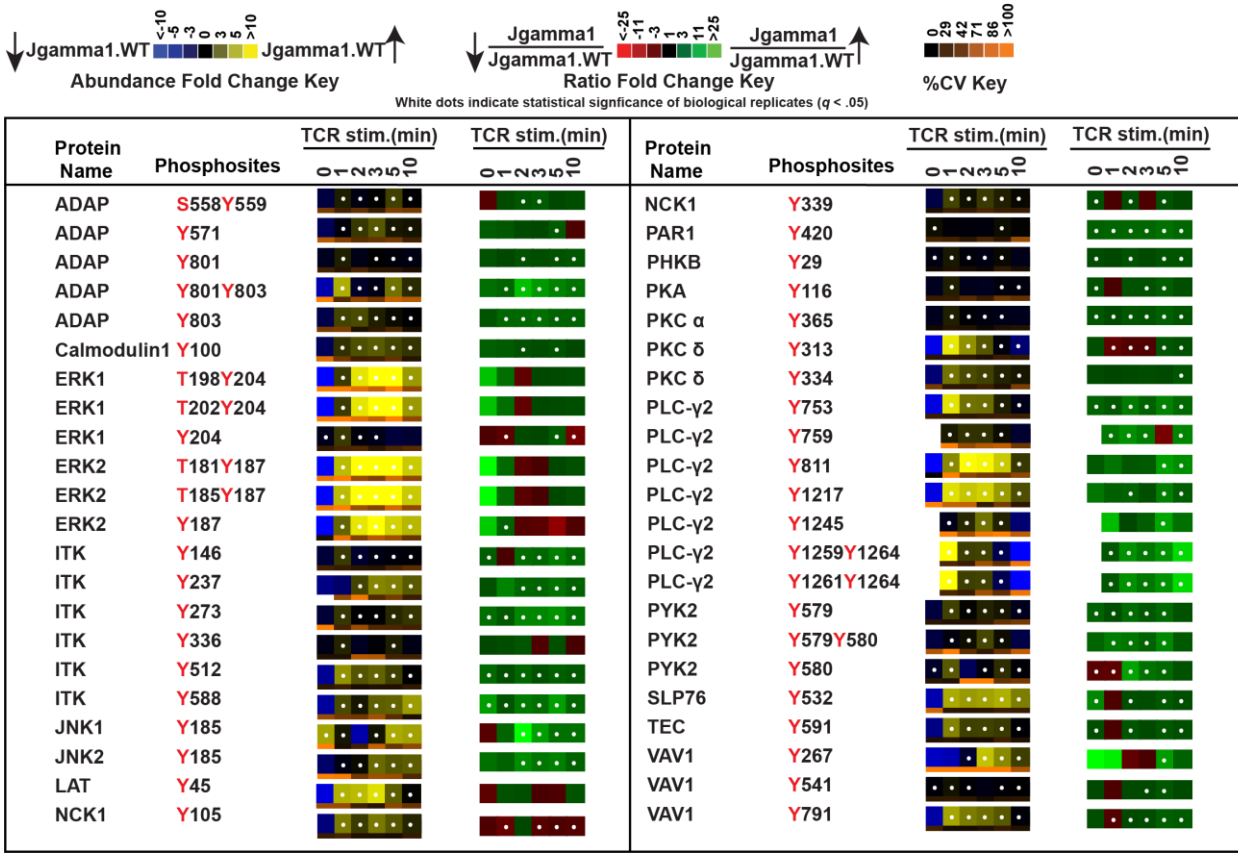


Figure 2.6: Quantitative phosphoproteomic analysis of proteins within the calcium signaling pathway and SLP-76 multimolecular complex. Heatmaps were calculated from 5 biological replicate experiments. The interpretation of the heatmap color scheme as well as the white dots is described in Materials and Methods and identical to Figure 2.1.

2.4.6. TCR- and PLC-γ1-Dependent Phosphorylation of PLC-γ2

Previous studies have suggested PLC-γ2 is capable of partially compensating for the absence of PLC-γ1.[19, 24, 25] Our phosphoproteomic data yielded peptides encompassing three functionally significant PLC-γ2 residues, Tyr^{753}, Tyr^{759}, and Tyr^{1217} (Figure 2.6). Tyr^{753} and Tyr^{759} are located between the C-terminal SH2 and the SH3 domain of PLC-γ2. Phosphorylation of these residues induces intramolecular associations of PLC-γ2's SH2 domains, forcing the active site of the enzyme into an open conformation.[12] Tyr^{1217} is located in the carboxy-terminal region of PLC-γ2, and like Tyr^{753} and Tyr^{759}, is required for maximal phospholipase activity.[82] While the role of Tec family kinases in regulating Tyr^{753} and Tyr^{759} is well established, regulation of Tyr^{1217} by the Tec family is still subject to debate.[17, 18, 82] In both the Jgamma1 and Jgamma1.WT lines, phosphorylation of these sites was induced by TCR stimulation, with Jgamma1 exhibiting significantly more phosphorylation at these residues (Figure 2.6). These results indicate both that PLC-γ2 is rapidly phosphorylated during early TCR activation and that PLC-γ1 negatively regulates the phosphorylation of PLC-γ2. This regulation may be due to excess PLC-γ1 outcompeting PLC-γ2 for access to the SLP-76 SH3 domain, which situates PLC-γ1 in proximity to Itk. In the absence of PLC-γ1, increased activation of PLC-γ2 may be sufficient to partially compensate for the loss of PLC-γ1's phospholipase activity. Consistent with a previous study of PLC-γ1 conditionally deficient mice,[23] we did not observe a significant change in the phosphorylation state of Erk's activating residues in Jgamma1. The maintenance of phosphorylation on Erk is likely due to the compensatory effects of PLC-γ2. Our findings in Jurkat cells agreed with observations of PLC-γ1 conditionally deficient mice, which also showed elevated phosphorylation of JNK (Figure 2.6).[23] Vav1 GEF/Rac recruitment may be heightened in the absence of PLC-γ1 as a result of elevated PIP_3 levels [83, 84], which could arise due to both an increase in PIP_2 as well as heightened activation of PI3k, as assessed by the increased phosphorylation of Tyr^{467} of p85α (Figure 2.1).[85-87] This mechanism has previously been established in T cells,[88] and may involve Pyk2.[89, 90] In total, the phosphoproteomic data

indicate that Lck targeting, but not its activity, is regulated by PLC-γ1 through an Erk-independent mechanism.

2.5. Conclusions

To date, numerous studies have highlighted PLC-γ1's importance in modulating MAPK pathways as well as regulating intracellular Ca^{2+}. However, the role that PLC-γ1 serves in regulating the upstream phosphorylation events that precede cleavage of PIP_2 in activated T cells is not well understood. In this study, we employed a quantitative MS/MS-based strategy to explore dynamic tyrosine phosphorylation in PLC-γ1-deficient T cells. Our results highlight a role for PLC-γ1 in regulating the activity and substrate specificity of Lck. PLC-γ1 may carry out this function by

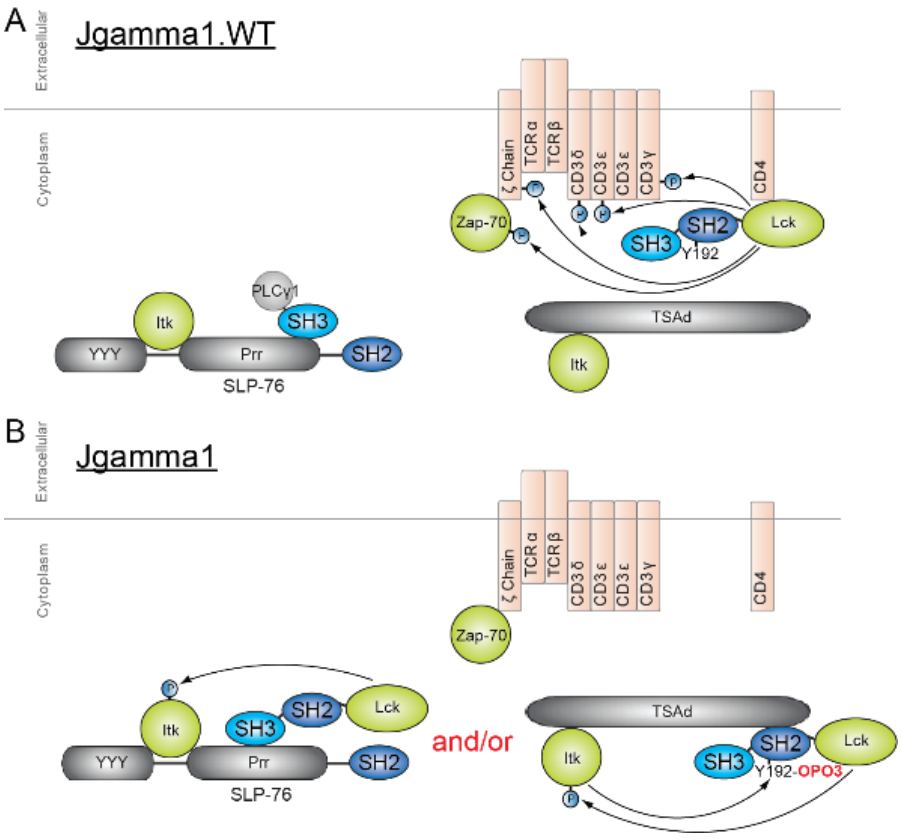


Figure 2.7: Proposed model of early Lck-mediated phosphorylation events following TCR engagement in wild-type and PLC-γ1-deficient cells. PLC-γ1 is expanded to depict a single SH3 domain only. A) In reconstituted cells, Lck associates with the cytoplasmic tail of CD4. Upon engagement of the TCR, Lck phosphorylates the cytoplasmic tails of the receptor subunits, as well as activating residues on Zap-70. B) In the absence of PLC-γ1, Lck association with the proline-rich region of SLP-76 as well as TSAd is increased, bringing it into proximity with Itk. Reciprocal phosphorylation of Itk and Lck at Tyr^{512} and Tyr^{192} respectively reinforces the association of both kinases with the adaptors.

outcompeting Lck for access to SLP-76 and Itk, by promoting the interaction of Lck Tyr^{192} and TSAd, or some combination of both mechanisms (Figure 2.7). Furthermore, our data reveal that Erk activity is not significantly diminished by the loss of PLC-γ1 during the first 10 minutes of TCR signal transduction, and that this may be due to a buffering effect of PLC-γ2 expressed in T cells. We speculate that increased activation of JNK may occur via Vav1/Rac as a consequence of elevated PIP_3 brought on by the absence of PLC-γ1. In total, our results confirm the central role PLC-γ1 plays in orchestrating the events of early TCR signal transduction, and suggest that PLC-γ1 may exert a significant regulatory influence on feedback pathways connecting the signalosome complex to Lck.

2.6. References

1. Salmond, R.J., et al., *T-cell receptor proximal signaling via the Src-family kinases, Lck and Fyn, influences T-cell activation, differentiation, and tolerance.* Immunol Rev, 2009. **228**(1): p. 9-22.
2. Chan, A.C., et al., *Activation of ZAP-70 kinase activity by phosphorylation of tyrosine 493 is required for lymphocyte antigen receptor function.* EMBO J, 1995. **14**(11): p. 2499-508.
3. Neumeister, E.N., et al., *Binding of ZAP-70 to phosphorylated T-cell receptor zeta and eta enhances its autophosphorylation and generates specific binding sites for SH2 domain-containing proteins.* Mol Cell Biol, 1995. **15**(6): p. 3171-8.
4. Kong, G., et al., *Distinct tyrosine phosphorylation sites in ZAP-70 mediate activation and negative regulation of antigen receptor function.* Mol Cell Biol, 1996. **16**(9): p. 5026-35.
5. Koretzky, G.A., F. Abtahian, and M.A. Silverman, *SLP76 and SLP65: complex regulation of signalling in lymphocytes and beyond.* Nat Rev Immunol, 2006. **6**(1): p. 67-78.
6. Yablonski, D., et al., *Uncoupling of nonreceptor tyrosine kinases from PLC-gamma1 in an SLP-76-deficient T cell.* Science, 1998. **281**(5375): p. 413-6.
7. Su, Y.W., et al., *Interaction of SLP adaptors with the SH2 domain of Tec family kinases.* Eur J Immunol, 1999. **29**(11): p. 3702-11.
8. Bubeck Wardenburg, J., et al., *Phosphorylation of SLP-76 by the ZAP-70 protein-tyrosine kinase is required for T-cell receptor function.* J Biol Chem, 1996. **271**(33): p. 19641-4.
9. Kumar, L., et al., *Differential role of SLP-76 domains in T cell development and function.* Proc Natl Acad Sci U S A, 2002. **99**(2): p. 884-9.
10. Kadamur, G. and E.M. Ross, *Mammalian phospholipase C.* Annu Rev Physiol, 2013. **75**: p. 127-54.
11. Yang, Y.R., et al., *Diverse cellular and physiological roles of phospholipase C-gamma1.* Adv Biol Regul, 2012. **52**(1): p. 138-51.
12. Gresset, A., et al., *Mechanism of phosphorylation-induced activation of phospholipase C-gamma isozymes.* J Biol Chem, 2010. **285**(46): p. 35836-47.
13. Yablonski, D., T. Kadlecek, and A. Weiss, *Identification of a phospholipase C-gamma1 (PLC-gamma1) SH3 domain-binding site in SLP-76 required for T-cell receptor-mediated activation of PLC-gamma1 and NFAT.* Mol Cell Biol, 2001. **21**(13): p. 4208-18.

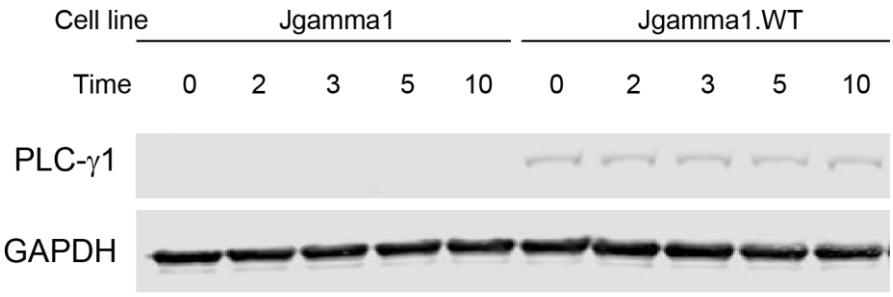
14. Irvin, B.J., et al., *Pleiotropic contributions of phospholipase C-gamma1 (PLC-gamma1) to T-cell antigen receptor-mediated signaling: reconstitution studies of a PLC-gamma1-deficient Jurkat T-cell line.* Mol Cell Biol, 2000. **20**(24): p. 9149-61.
15. Stoica, B., et al., *The amino-terminal Src homology 2 domain of phospholipase C gamma 1 is essential for TCR-induced tyrosine phosphorylation of phospholipase C gamma 1.* J Immunol, 1998. **160**(3): p. 1059-66.
16. Falasca, M., et al., *Activation of phospholipase C gamma by PI 3-kinase-induced PH domain-mediated membrane targeting.* EMBO J, 1998. **17**(2): p. 414-22.
17. Humphries, L.A., et al., *Tec kinases mediate sustained calcium influx via site-specific tyrosine phosphorylation of the phospholipase Cgamma Src homology 2-Src homology 3 linker.* J Biol Chem, 2004. **279**(36): p. 37651-61.
18. Ozdener, F., et al., *Activation of phospholipase Cgamma2 by tyrosine phosphorylation.* Mol Pharmacol, 2002. **62**(3): p. 672-9.
19. Fu, G., et al., *Phospholipase Cgamma2 plays a role in TCR signal transduction and T cell selection.* J Immunol, 2012. **189**(5): p. 2326-32.
20. Thomas, S.M., et al., *Epidermal growth factor receptor-stimulated activation of phospholipase Cgamma-1 promotes invasion of head and neck squamous cell carcinoma.* Cancer Res, 2003. **63**(17): p. 5629-35.
21. Sala, G., et al., *Phospholipase Cgamma1 is required for metastasis development and progression.* Cancer Res, 2008. **68**(24): p. 10187-96.
22. Jones, N.P. and M. Katan, *Role of phospholipase Cgamma1 in cell spreading requires association with a beta-Pix/GIT1-containing complex, leading to activation of Cdc42 and Rac1.* Mol Cell Biol, 2007. **27**(16): p. 5790-805.
23. Fu, G., et al., *Phospholipase C{gamma}1 is essential for T cell development, activation, and tolerance.* J Exp Med, 2010. **207**(2): p. 309-18.
24. Wen, R., et al., *An important role of phospholipase Cgamma1 in pre-B-cell development and allelic exclusion.* EMBO J, 2004. **23**(20): p. 4007-17.
25. Coggeshall, K.M., J.C. McHugh, and A. Altman, *Predominant expression and activation-induced tyrosine phosphorylation of phospholipase C-gamma 2 in B lymphocytes.* Proc Natl Acad Sci U S A, 1992. **89**(12): p. 5660-4.
26. Cao, L., et al., *Quantitative phosphoproteomics reveals SLP-76 dependent regulation of PAG and Src family kinases in T cells.* PLoS One, 2012. **7**(10): p. e46725.
27. Sjolin-Goodfellow, H., et al., *The catalytic activity of the kinase ZAP-70 mediates basal signaling and negative feedback of the T cell receptor pathway.* Sci Signal, 2015. **8**(377): p. ra49.
28. Helou, Y.A., A.P. Petrashen, and A.R. Salomon, *Vav1 Regulates T-Cell Activation through a Feedback Mechanism and Crosstalk between the T-Cell Receptor and CD28.* J Proteome Res, 2015.
29. Salomon, A.R., et al., *Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry.* Proc Natl Acad Sci U S A, 2003. **100**(2): p. 443-8.
30. Ji, Q., Y. Ding, and A.R. Salomon, *SRC homology 2 domain-containing leukocyte phosphoprotein of 76 kDa (SLP-76) N-terminal tyrosine residues regulate a dynamic signaling equilibrium involving feedback of proximal T-cell receptor (TCR) signaling.* Mol Cell Proteomics, 2015. **14**(1): p. 30-40.
31. Rappsilber, J., Y. Ishihama, and M. Mann, *Stop and go extraction tips for matrix-assisted laser desorption/ionization, nanoelectrospray, and LC/MS sample pretreatment in proteomics.* Analytical Chemistry, 2003. **75**(3): p. 663-670.
32. Ahsan, N. and A.R. Salomon, *Quantitative Phosphoproteomic Analysis of T-Cell Receptor Signaling.* Methods Mol Biol, 2017. **1584**: p. 369-382.

33. Yu, K., et al., *Integrated platform for manual and high-throughput statistical validation of tandem mass spectra.* Proteomics, 2009. **9**(11): p. 3115-25.
34. Yu, K. and A.R. Salomon, *HTAPP: high-throughput autonomous proteomic pipeline.* Proteomics, 2010. **10**(11): p. 2113-22.
35. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-3567.
36. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nat Methods, 2007. **4**(3): p. 207-14.
37. Beausoleil, S.A., et al., *A probability-based approach for high-throughput protein phosphorylation analysis and site localization.* Nat Biotechnol, 2006. **24**(10): p. 1285-92.
38. Demirkan, G., et al., *Phosphoproteomic profiling of in vivo signaling in liver by the mammalian target of rapamycin complex 1 (mTORC1).* PLoS One, 2011. **6**(6): p. e21729.
39. Nguyen, V., et al., *A new approach for quantitative phosphoproteomic dissection of signaling pathways applied to T cell receptor activation.* Mol Cell Proteomics, 2009. **8**(11): p. 2418-31.
40. Storey, J.D., *The positive false discovery rate: A Bayesian interpretation and the q-value.* Annals of Statistics, 2003. **31**(6): p. 2013-2035.
41. Storey, J.D. and R. Tibshirani, *Statistical significance for genomewide studies.* Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9440-5.
42. de Hoon, M.J., et al., *Open source clustering software.* Bioinformatics, 2004. **20**(9): p. 1453-4.
43. Saldanha, A.J., *Java Treeview--extensible visualization of microarray data.* Bioinformatics, 2004. **20**(17): p. 3246-8.
44. Hornbeck, P.V., et al., *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.* Nucleic Acids Res, 2015. **43**(Database issue): p. D512-20.
45. Obenauer, J.C., L.C. Cantley, and M.B. Yaffe, *Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs.* Nucleic Acids Res, 2003. **31**(13): p. 3635-41.
46. Vizcaino, J.A., et al., *ProteomeXchange provides globally coordinated proteomics data submission and dissemination.* Nat Biotechnol, 2014. **32**(3): p. 223-6.
47. Vizcaino, J.A., et al., *2016 update of the PRIDE database and its related tools.* Nucleic Acids Res, 2016. **44**(D1): p. D447-56.
48. Yu, K. and A.R. Salomon, *PeptideDepot: flexible relational database for visual analysis of quantitative proteomic data and integration of existing protein information.* Proteomics, 2009. **9**(23): p. 5350-8.
49. Wange, R.L., et al., *Activating and inhibitory mutations in adjacent tyrosines in the kinase domain of ZAP-70.* J Biol Chem, 1995. **270**(32): p. 18730-3.
50. Watts, J.D., et al., *Identification by electrospray ionization mass spectrometry of the sites of tyrosine phosphorylation induced in activated Jurkat T cells on the protein tyrosine kinase ZAP-70.* J Biol Chem, 1994. **269**(47): p. 29520-9.
51. Mege, D., et al., *Mutation of tyrosines 492/493 in the kinase domain of ZAP-70 affects multiple T-cell receptor signaling pathways.* J Biol Chem, 1996. **271**(51): p. 32644-52.
52. Di Bartolo, V., et al., *Tyrosine 319, a newly identified phosphorylation site of ZAP-70, plays a critical role in T cell antigen receptor signaling.* J Biol Chem, 1999. **274**(10): p. 6285-94.
53. Pelosi, M., et al., *Tyrosine 319 in the interdomain B of ZAP-70 is a binding site for the Src homology 2 domain of Lck.* J Biol Chem, 1999. **274**(20): p. 14229-37.
54. Williams, B.L., et al., *Phosphorylation of Tyr319 in ZAP-70 is required for T-cell antigen receptor-dependent phospholipase C-gamma1 and Ras activation.* EMBO J, 1999. **18**(7): p. 1832-44.
55. Straus, D.B. and A. Weiss, *Genetic evidence for the involvement of the lck tyrosine kinase in signal transduction through the T cell antigen receptor.* Cell, 1992. **70**(4): p. 585-93.

56. van Oers, N.S., N. Killeen, and A. Weiss, *Lck regulates the tyrosine phosphorylation of the T cell receptor subunits and ZAP-70 in murine thymocytes.* J Exp Med, 1996. **183**(3): p. 1053-62.
57. Wong, J., D. Straus, and A.C. Chan, *Genetic evidence of a role for Lck in T-cell receptor function independent or downstream of ZAP-70/Syk protein tyrosine kinases.* Mol Cell Biol, 1998. **18**(5): p. 2855-66.
58. Palacios, E.H. and A. Weiss, *Function of the Src-family kinases, Lck and Fyn, in T-cell development and activation.* Oncogene, 2004. **23**(48): p. 7990-8000.
59. Bergman, M., et al., *The human p50csk tyrosine kinase phosphorylates p56lck at Tyr-505 and down regulates its catalytic activity.* EMBO J, 1992. **11**(8): p. 2919-24.
60. Abraham, N. and A. Veillette, *Activation of p56lck through mutation of a regulatory carboxy-terminal tyrosine residue requires intact sites of autophosphorylation and myristylation.* Mol Cell Biol, 1990. **10**(10): p. 5197-206.
61. Couture, C., et al., *Regulation of the Lck SH2 domain by tyrosine phosphorylation.* J Biol Chem, 1996. **271**(40): p. 24880-4.
62. Granum, S., et al., *The kinase Itk and the adaptor TSAd change the specificity of the kinase Lck in T cells by promoting the phosphorylation of Tyr192.* Sci Signal, 2014. **7**(355): p. ra118.
63. Berg, N.N. and H.L. Ostergaard, *T cell receptor engagement induces tyrosine phosphorylation of FAK and Pyk2 and their association with Lck.* J Immunol, 1997. **159**(4): p. 1753-7.
64. Kapoor-Kaushik, N., et al., *Distinct Mechanisms Regulate Lck Spatial Organization in Activated T Cells.* Front Immunol, 2016. **7**: p. 83.
65. da Silva, A.J., et al., *Cloning of a novel T-cell protein FYB that binds FYN and SH2-domain-containing leukocyte protein 76 and modulates interleukin 2 production.* Proc Natl Acad Sci U S A, 1997. **94**(14): p. 7493-8.
66. Heyeck, S.D., et al., *Lck phosphorylates the activation loop tyrosine of the Itk kinase domain and activates Itk kinase activity.* J Biol Chem, 1997. **272**(40): p. 25401-8.
67. Han, J., et al., *Lck regulates Vav activation of members of the Rho family of GTPases.* Mol Cell Biol, 1997. **17**(3): p. 1346-53.
68. Gulbins, E., et al., *Tyrosine kinase-stimulated guanine nucleotide exchange activity of Vav in T cell activation.* Science, 1993. **260**(5109): p. 822-5.
69. Rodriguez, R., et al., *Tyrosine residues in phospholipase Cgamma 2 essential for the enzyme function in B-cell signaling.* J Biol Chem, 2001. **276**(51): p. 47982-92.
70. Berg, L.J., et al., *Tec family kinases in T lymphocyte development and function.* Annu Rev Immunol, 2005. **23**: p. 549-600.
71. Gonen, R., et al., *T cell receptor-induced activation of phospholipase C-gamma1 depends on a sequence-independent function of the P-I region of SLP-76.* J Biol Chem, 2005. **280**(9): p. 8364-70.
72. Sanzenbacher, R., D. Kabelitz, and O. Janssen, *SLP-76 binding to p56lck: a role for SLP-76 in CD4-induced desensitization of the TCR/CD3 signaling complex.* J Immunol, 1999. **163**(6): p. 3143-52.
73. Jung, S.H., et al., *Competition between SLP76 and LAT for PLCgamma1 binding in resting T cells.* Eur J Immunol, 2010. **40**(8): p. 2330-9.
74. Rudd, M.L., A. Tua-Smith, and D.B. Straus, *Lck SH3 domain function is required for T-cell receptor signals regulating thymocyte development.* Mol Cell Biol, 2006. **26**(21): p. 7892-900.
75. Li, M., et al., *The SH3 domain of Lck modulates T-cell receptor-dependent activation of extracellular signal-regulated kinase through activation of Raf-1.* Mol Cell Biol, 2008. **28**(2): p. 630-41.
76. Sundvold, V., et al., *T cell-specific adapter protein inhibits T cell activation by modulating Lck activity.* J Immunol, 2000. **165**(6): p. 2927-31.

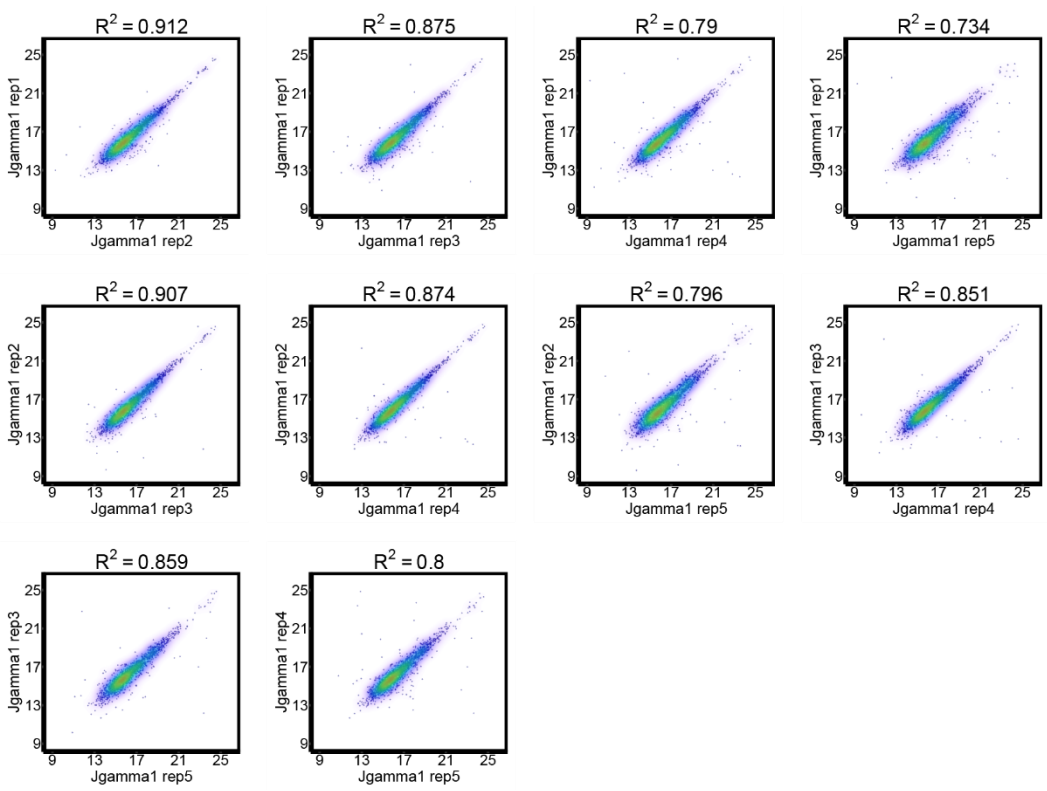
77. Berge, T., et al., *T cell specific adapter protein (TSAd) interacts with Tec kinase ITK to promote CXCL12 induced migration of human and murine T cells.* PLoS One, 2010. **5**(3): p. e9761.
78. Granum, S., et al., *Modulation of Lck function through multisite docking to T cell-specific adapter protein.* J Biol Chem, 2008. **283**(32): p. 21909-19.
79. Hem, C.D., et al., *T cell specific adaptor protein (TSAd) promotes interaction of Nck with Lck and SLP-76 in T cells.* Cell Commun Signal, 2015. **13**: p. 31.
80. Lettau, M., J. Pieper, and O. Janssen, *Nck adapter proteins: functional versatility in T cells.* Cell Commun Signal, 2009. **7**: p. 1.
81. Wunderlich, L., et al., *Association of Nck with tyrosine-phosphorylated SLP-76 in activated T lymphocytes.* Eur J Immunol, 1999. **29**(4): p. 1068-75.
82. Watanabe, D., et al., *Four tyrosine residues in phospholipase C-gamma 2, identified as Btk-dependent phosphorylation sites, are required for B cell antigen receptor-coupled calcium signaling.* J Biol Chem, 2001. **276**(42): p. 38595-601.
83. Han, J., et al., *Role of substrates and products of PI 3-kinase in regulating activation of Rac-related guanosine triphosphatases by Vav.* Science, 1998. **279**(5350): p. 558-60.
84. Das, B., et al., *Control of intramolecular interactions between the pleckstrin homology and Dbl homology domains of Vav and Sos1 regulates Rac binding.* J Biol Chem, 2000. **275**(20): p. 15074-81.
85. Lau, C., et al., *Syk associates with clathrin and mediates phosphatidylinositol 3-kinase activation during human rhinovirus internalization.* J Immunol, 2008. **180**(2): p. 870-80.
86. Warfel, N.A., M. Niederst, and A.C. Newton, *Disruption of the interface between the pleckstrin homology (PH) and kinase domains of Akt protein is sufficient for hydrophobic motif site phosphorylation in the absence of mTORC2.* J Biol Chem, 2011. **286**(45): p. 39122-9.
87. Kim, J.H., et al., *Inhibition of EGFR signaling in human prostate cancer PC-3 cells by combination treatment with beta-phenylethyl isothiocyanate and curcumin.* Carcinogenesis, 2006. **27**(3): p. 475-82.
88. Turner, M. and D.D. Billadeau, *VAV proteins as signal integrators for multi-subunit immune-recognition receptors.* Nat Rev Immunol, 2002. **2**(7): p. 476-86.
89. Collins, M., et al., *The T cell receptor-mediated phosphorylation of Pyk2 tyrosines 402 and 580 occurs via a distinct mechanism than other receptor systems.* J Leukoc Biol, 2010. **87**(4): p. 691-701.
90. Gao, C. and S.D. Blystone, *A Pyk2-Vav1 complex is recruited to beta3-adhesion sites to initiate Rho activation.* Biochem J, 2009. **420**(1): p. 49-56.
91. Kanehisa, M., *The KEGG database.* Novartis Found Symp, 2002. **247**: p. 91-101; discussion 101-3, 119-28, 244-52.
92. Helou, Y.A., et al., *ERK positive feedback regulates a widespread network of tyrosine phosphorylation sites across canonical T cell signaling and actin cytoskeletal proteins in Jurkat T cells.* PLoS One, 2013. **8**(7): p. e69641.

2.7. Supplemental Data

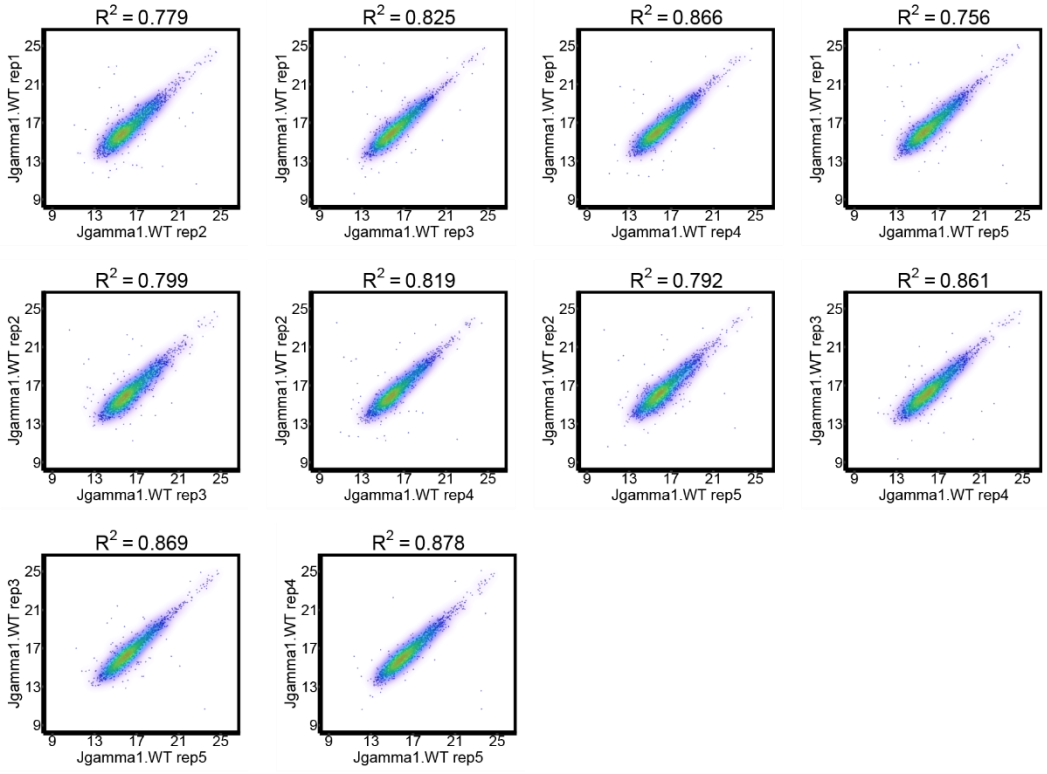


Supplemental Figure 2.1: Validation of PLC-γ1 levels in Jgamma1 and Jgamma1.WT cells.

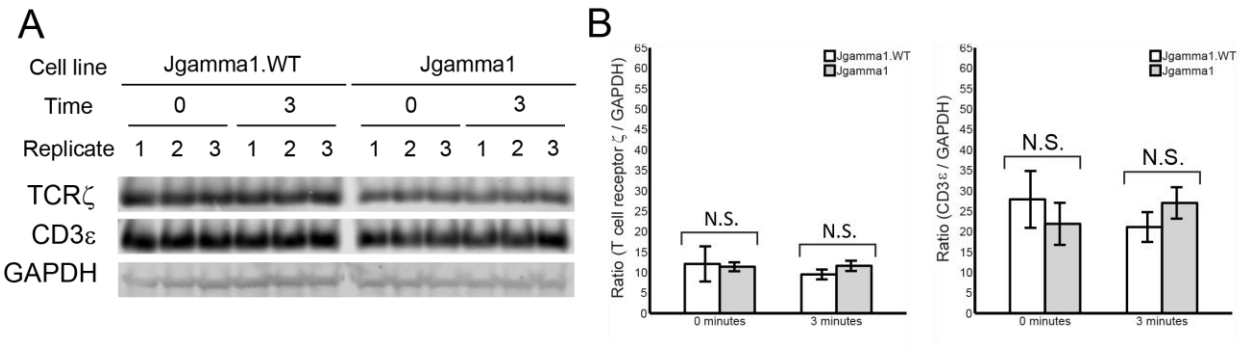
A


R2 = 0.912
R2 = 0.875
R2 = 0.79
R2 = 0.734
R2 = 0.907
R2 = 0.874
R2 = 0.796
R2 = 0.851
R2 = 0.859
R2 = 0.8
Jgamma1 rep1
Jgamma1 rep2
Jgamma1 rep3
Jgamma1 rep4
Jgamma1 rep5
9
13
17
21
25

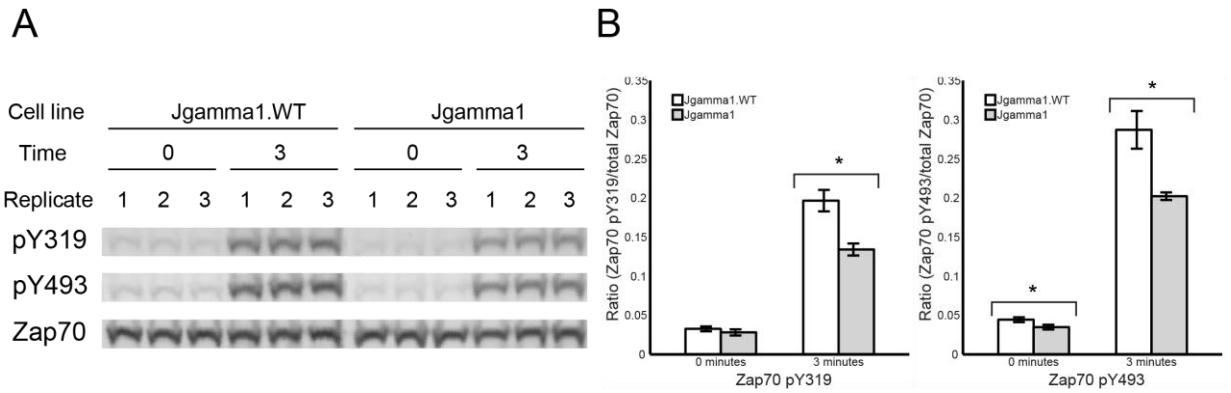
B


R2 = 0.779
R2 = 0.825
R2 = 0.866
R2 = 0.756
R2 = 0.799
R2 = 0.819
R2 = 0.792
R2 = 0.861
R2 = 0.869
R2 = 0.878
Jgamma1.WT rep1
Jgamma1.WT rep2
Jgamma1.WT rep3
Jgamma1.WT rep4
Jgamma1.WT rep5
9
13
17
21
25

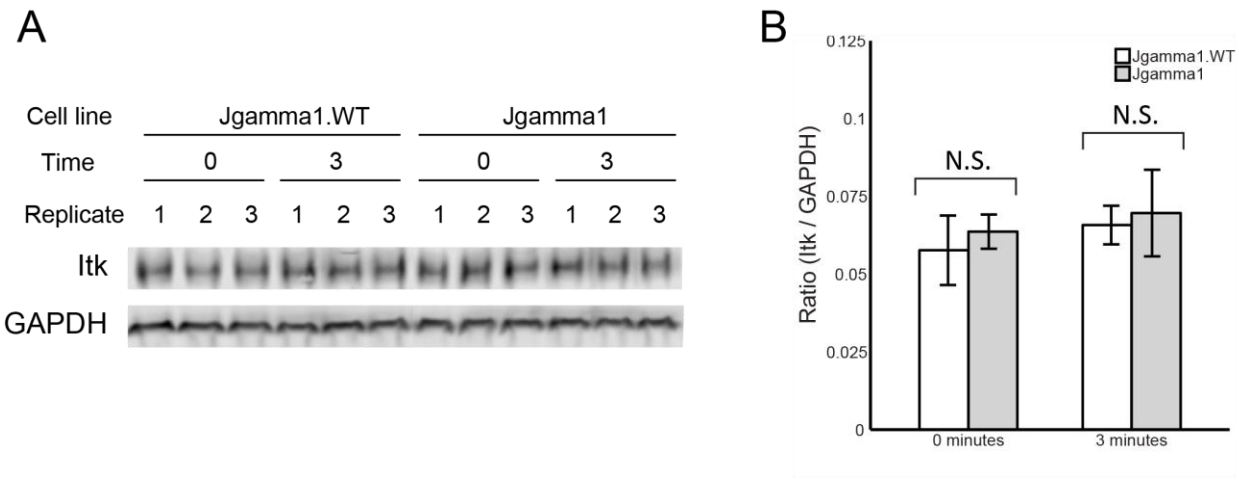
Supplemental Figure 2.2: Pairwise inter-replicate comparison of phosphopeptide peak areas observed in biological replicates. Each data point represents the peak areas observed in separate replicates of phosphopeptides with the same charge state and sequence sampled at the same timepoint. A) Jgamma1 replicate comparisons. B) Jgamma1.WT replicate comparisons.



Supplemental Figure 2.3: Validation of equivalent receptor subunit expression in Jgamma1 and Jgamma1.WT. A) For each cell line, three replicates of the 0 and 3 minute timepoints were examined. Antibodies to TCRζ and CD3ε were used, and band intensities were normalized to GAPDH. B) Quantitation of Western blot data. Error bars indicate the standard error of the mean. t test *p* values were thresholded at 0.05 when assessing significance.



Supplemental Figure 2.4: Validation of Zap-70 phosphoproteomic data. A) For each cell line, three replicates of the 0 and 3 minute timepoints were examined by Western blot. Phospho-specific antibodies to pY319 and pY493 were used, and band intensities were normalized to a non-phospho-specific Zap-70 antibody. B) Quantitation of Western blot data. Error bars indicate the standard error of the mean. * indicates $p < 0.05$.



Supplemental Figure 2.5: Validation of equivalent Itk expression in Jgamma1 and Jgamma1.WT. A) For each cell line, three replicates of the 0 and 3 minute timepoints were examined by Western blot. Antibody to Itk was used, and band intensities were normalized to GAPDH. B) Quantitation of Western blot data. Error bars indicate the standard error of the mean. t test *p* values were thresholded at 0.05 when assessing signficance.

Chapter 3

Highly Reproducible Improved Label-Free Quantitative Analysis of Cellular Phosphoproteome by Optimization of LC-MS/MS gradient and Analytical Column Construction

This chapter is adapted from a manuscript published in the Journal of Proteomics (DOI: 10.1016/j.jprot.2017.06.013). The original authors are Nagib Ahsan, Judson Belmont, Zhuo Chen, James Clifton, and Art Salomon.

3.1. Introduction

Mass spectrometry (MS)-based proteomics has played an indispensable role in wide-scale analysis of complex protein mixtures. The technology has been employed in the wide-scale characterization of post-translational modifications (PTMs) and protein interactions.[1, 2] A major challenge in proteomic identification and quantitation is the large dynamic range of protein abundance contained in cellular lysates: four orders of magnitude in yeast and at least seven orders of magnitude in human cells.[3, 4] Improvements in peptide separation technology are crucial for complex proteome characterization. To analyze the proteome with high complexity, two-dimensional separation based on gels or liquid chromatography has been used.[5-8] With these methods, LC-MS/MS is preceded by an additional mode of fractionation, such as strong cation exchange (SCX), SDS-PAGE, isoelectric focusing (IEF), and high pH RPLC. Although these methods can reduce sample complexity during LC-MS/MS and thereby increase sequencing depth, they do require more starting material (up to mg level) and each additional separation step decreases the quantitative reproducibility of replicate analyses, frustrating the extraction of biologically relevant insights.

Post translational modifications (PTMs) play a critical role in regulating many signaling pathways. Therefore, identification and quantification of post translationally modified peptides, particularly phosphorylated peptides, is highly desirable. However, identification of phosphopeptides by mass spectrometry is complicated by the high dynamic range of in vivo phosphorylation and low stoichiometry in biological systems.[9] Therefore, to date many methods have been developed to enrich phosphopeptides using a variety of approaches including ion exchange chromatography, metal oxide surfaces, immobilized metal ion affinity chromatography and enrichment by phosphospecific antibodies.[10-17] Quantitation of phosphopeptides is also very critical as variable phosphorylation alters many biological pathways and can correlate with disease states.[18] Optimization of quantitative reproducibility with each technique while

maintaining the highest possible sequencing depth will maximize the number of biologically relevant insights delivered from a proteomics experiment.

Certainly, improvements in instrument sampling rates such as with the Orbitrap mass spectrometer and improved chromatographic peptide separations have led many researchers to explore the utility of complex proteome characterization without peptide pre-fractionation.[3, 19, 20] In addition, analytical column length has been shown previously to improve sequencing depth.[3] However, compared to traditional 3 µm or 5 µm reversed-phase C18 particles, sub-2 µm C18 particles decrease the diffusion distances of solutes while minimizing the band spreading, improving column efficiency of reversed-phase separations. Another strategy is to operate a long column (50 cm or even longer), which provides more theoretical plates, with a long LC gradient (several hours or more).[21] A major technical difficulty in using sub-2 µm particles and long columns is the high column backpressure, which has been previously overcome by ultra-high pressure liquid chromatography (UHPLC) [22-24] or a column heater.[21] A true wide-scale understanding of a biological system requires statistical analysis of biological replicates to correctly understand biological variation. Mass spectrometry-based peptide quantification methods can be either label-free or incorporate a stable isotope label such as stable isotope labels with amino acids in cell culture (SILAC) [25], or isobaric tags for relative and absolute quantitation (iTRAQ).[26]

In this report, the analysis of the T cell phosphoproteome using an optimized chromatographic configuration that reduces backpressure using a fritless column packed with sub-2 µm particles is explored. Although many recent reports have impressively expanded the depth of proteomic sequencing, far fewer reports have examined the relationship between sequencing depth and reproducibility of proteomic and phosphoproteomic quantitation through replicate analyses. Pre-fractionation of proteomic samples prior to LC-MS can increase the yield of peptide identifications, however, this strategy increases the variability of the quantitative data analysis and increases the overall cost and instrument time required. For analyses without pre-

fractionation, minimization of the necessity for specialized equipment not available in many proteomic labs can provide higher accessibility of sensitive proteomic capabilities to more labs and greatly simplify the analysis.

3.2. Methods and Materials

3.2.1. T Cell Receptor Stimulation and Isolation and Purification of Phosphopeptides

The Jurkat T cell clone E6-1 was obtained from American Tissue Culture Collection (Manassas, VA) and cells were cultured and maintained according to standard protocols. For the CD3/4 stimulation, after harvest, Jurkat cells were treated with anti-CD3 and anti-CD4 antibody in PBS (clone OKT3 and OKT4; eBioscience, San Diego, CA) for 3 min at 37°C. Cell lysis, protein extraction and trypsin digestion followed by Sep-Pack C18 column purification was performed as previously described,[27] with the exception that TFA was used instead of acetic acid. Eluents containing peptides were lyophilized for 48 hours to dryness. Phosphopeptides were enriched with Titansphere Phos-TiO tips (GL Sciences, Tokyo Japan) following the manufacturer's protocol.

3.2.2. LC-MS and Peptide Assignment

Analytical columns 50 and/or 15 cm long were packed in-house with ReproSil-Pur C18-AQ 1.9 µm particles (Dr. Maisch GmbH, Ammerbuch, Germany) or 3 µm Monitor particles. Detailed protocols for column packing are provided in the supplementary methods section. Tryptic peptides were analyzed by a fully automated phosphoproteomic technology platform. [28, 29] The nanoLC-MS/MS experiments were performed with an Agilent 1200 Series Quaternary HPLC system (Agilent Technologies, Santa Clara, CA) connected to a Q Exactive mass spectrometer (Thermo Fisher Scientific, Waltham, MA) (see Supplementary Material and Methods for details). Peptide spectrum matching of MS/MS spectra from whole cell lysate Jurkat tryptic digest samples was performed against a human-specific database (UniProt; downloaded 2/1/2013) using MASCOT v. 2.4 (Matrix Science, Ltd, London W1U 7GB UK). A concatenated database containing 144,156

"target" and "decoy reversed" sequences was employed to estimate the false discovery rate (FDR).[30] Peptide assignments from the database search were filtered down to 1% false discovery rate (FDR) by a logistic spectral score, as previously described.[30, 31] To validate the position of the phosphorylation sites, the Ascore algorithm[32] was applied to all data, and the reported phosphorylation site position reflected the top Ascore prediction.

3.2.3. Quantitative and Statistical Analysis

Relative quantification of phosphopeptide abundance was performed via calculation of selected ion chromatograms (SIC) peak areas. Retention time alignment of individual replicate analyses was performed as previously described.[33] A minimum SIC peak area equivalent to the typical spectral noise level of 1000 was required of all data reported for label-free quantitation. Individual SIC peak areas were normalized to the peak area of the exogenously spiked phosphopeptide FQpSEEQQQTEDELQDK added prior to phosphopeptide enrichment and reversed-phase elution into the mass spectrometer. Quantitative analysis was applied to five biological replicate experiments from separate batches of cells. To select the phosphopeptides that show a statistically significant change in abundance between CD3/4 stimulated and unstimulated cells, q-values for multiple hypothesis tests were calculated based on p-values from two-tailed unpaired Student's t tests using the R package QVALUE as previously described.[34, 35] The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD005650.

3.3. Results and Discussion

We first compared the performance of a fritless, 50 cm analytical column with 1.9 μm particles with an integrated electrospray tip that was self-pulled or from a commercial source (New Objective, FS360-75-8-N-5-C70, Woburn, MA). Fritless columns were first employed in the capillary electrochromatography (CEC) field.[36, 37] Five different LC gradients (2 hr, 3 hr, 4.5 hr, 6 hr, 8 hr) were used to characterize equal portions of Jurkat T cell whole cell lysate. Over 23,000

unique peptides corresponding to 4,000 unique proteins were identified using either the 6 or 8 hr LC gradients at a 1% FDR estimated by decoy database approach (Supplemental Figure 3.1). The in-house pulled 50 cm-long, 1.9 µm C18 analytical column performed nearly identically to the commercially pulled column in terms of PSM yield (Supplemental Figure 3.1). Reversed-phase separations were performed with a 50 cm-long, 1.9 µm C18 analytical column using an Agilent 1200 quaternary pump operated at less than 300 bar (Supplemental Figure 3.2). To verify the stability of this fritless 50 cm, 1.9 µm C18 column, the HPLC pressure was cycled from 0 to ~260 bar a total of 100 times. The length of the packing bed and flow rate was measured after every 10 cycles (Supplemental Table 3.1). The results clearly showed that both the packing bed length and flow rate at the electrospray tip were unaltered through 100 cycles.

3.3.1. Optimized Gradient Length and Flow Rate Balance Sequencing Depth and Acquisition Time

The optimal reversed-phase gradient length of this newly constructed 50 cm-long, 1.9 µm column was examined using five different LC gradients (2 hr, 3 hr, 4.5 hr, 6 hr, 8 hr). Among the five gradients, the 6 hr gradient yielded the highest number of PSMs (30,403) which corresponded to 4679 protein groups (Supplemental Table 3.2). However, the 3 hr linear gradient also provided the best balance between sequencing depth and acquisition time yielding 88% of the maximal PSMs in half the time. Other groups have also demonstrated that over 90% of protein or peptide PSMs can be identified using a 3 hr gradient with a 50 cm column .[3] In addition, use of a 50 cm column with sub-2 µm particles produced peak capacities up to 500 within a 120 min gradient.[38] Other reports demonstrated that sequencing depth also increased with increasing peak capacity while peak capacity increased with increasing gradient length and decreasing particle diameter.[38-40] Based on these earlier studies and our preliminary data, the subsequent phosphopeptide quantitative analysis was performed using a 3 hr gradient.

3.3.2. 50 cm Column with Sub-2μm Particles Delivers Highly Reproducible Quantitation

Quantitative reproducibility is essential to maximize extraction of biologically relevant insights from proteomic experiments. To compare the quantitative reproducibility of analytical columns with either 1.9 or 3 μm particles, Jurkat T cells stimulated for 3 min with CD3/4 antibody crosslink were compared to unstimulated cells (control) after phosphopeptide enrichment by the TiO_2 method from whole cell lysates. A 50 cm column packed with 1.9 μm particles was compared to 15 and 50 cm columns constructed with 3 μm particles. The 15 cm, 3 μm column was included because this configuration represents a commonly used column employed in proteomic labs. For each specific column configuration, intensity-based label-free quantitative analysis with retention time alignment[29, 33] was applied to quantitate the five biological replicate analyses. All selected ion chromatogram (SIC) peak areas from each LCMS run were normalized to the signal intensity of an exogenously spiked phosphopeptide (FQpSEEQQQTEDELQDK) standard that was added to every sample at equal levels and co-purified with the T cell phosphopeptides. This standard peptide was quantified reproducibly in all the replicate LC-MS experiments with less than 20% CV across all replicates, highlighting the reproducibility of the quantitation in this study (Supplemental Table 3.3). A comprehensive list of all phosphorylated peptides with Mowse score >20 including reversed database hits from every replicate in the CD3/4 stimulated LC-MS analysis as well as the full detailed quantitative proteomic data is available at http://www.sciencedirect.com/science/article/pii/S1874391917302233?via%3Dihub.

Although both 50 cm column configurations were operated at a 60-100 nl/min flow rate due to the backpressure generated from the length of the column and use of a standard HPLC pump compared to 200-300nl/min for the 15 cm configuration, the FWHM medians were smaller for the 50 cm configurations (Supplemental Figure 3.4). Although the majority of peaks quantified in the study relied on greater than 10 MS scans across their selected ion chromatogram (Supplemental Figure 3.4), there was also no correlation at all between FWHM and q-value for the three column configurations tested (Supplemental Figure 3.5). In addition, the column configurations using 60-

100 nl/min also showed a high degree of reproducibility in the five biological replicates analyzed which resulted in pairwise replicate peak area R^2 correlation coefficients of ~0.9 (Supplemental Figure 3.6A-B) compared to 0.8 (Supplemental Figure 3.6C) for the 15 cm configuration. Overall, the difference in flow rate did not greatly impact the reproducibility of quantitation between 15 and 50 cm configurations. A histogram analysis of retention times and the selected ion chromatograms obtained from the 50cm 1.9 µm column showed a true distribution of peptides quantified in this experiment ranged from 50-200 min (Supplemental Figure 3.7). Furthermore, compared to the 3 µm particles, the 1.9 µm particles decreased the SIC peak spreading (Supplemental Figure 3.8). In addition, MA plot around the 0 line, especially for the 50 cm, 1.9 µm column configuration also indicates the absence of any bias in the fold changes across a range of peak areas, supporting the quality of the quantitative data (Supplemental Figure 3.9). When examining the distribution of the calculated phosphopeptide SIC peak areas, more than 30% of the identified significantly changed phosphopeptides detected in the 50 cm-long, 1.9 µm C18 column were below 1E7, while this percentage was less than 14% with the 3 µm C18 column configurations (Supplemental Table 3.4 and Supplemental Figure 3.10).

3.3.3. 50 cm Column with Sub-2 µm Particles Delivers Greater Statistical Sensitivity

Phosphoproteomic analysis revealed significant differences in phosphopeptide identification yields among the columns tested (Supplemental Table 3.5). For instance, the 50 cm, 1.9 µm particle column yielded over 23,000 unique phosphopeptide PSMs when combining both the stimulated and unstimulated samples. In contrast, a total of 15,173 and 14,494 phosphopeptide PSMs were detected using the 3 µm particle, 50 cm and 15 cm columns, respectively (Supplemental Table 3.5). At a strict cutoff of $q < 0.01$, a total of 7900, 2117 and 1391 unique phosphopeptides were significantly changed in abundance >2-fold in CD3/4 stimulated T cell compared with unstimulated cells when analyzed by 50 cm 1.9 µm, 50 cm 3 µm, and 15 cm 3 µm particle columns, respectively (Supplemental Table 3.5). These results revealed that the 1.9 µm C18 column reproducibly detected 373% more phosphopeptides with greater than 2 fold

significant change and 299% more phosphopeptides with greater than 10 fold significant change compared to the 3 µm C18 column.

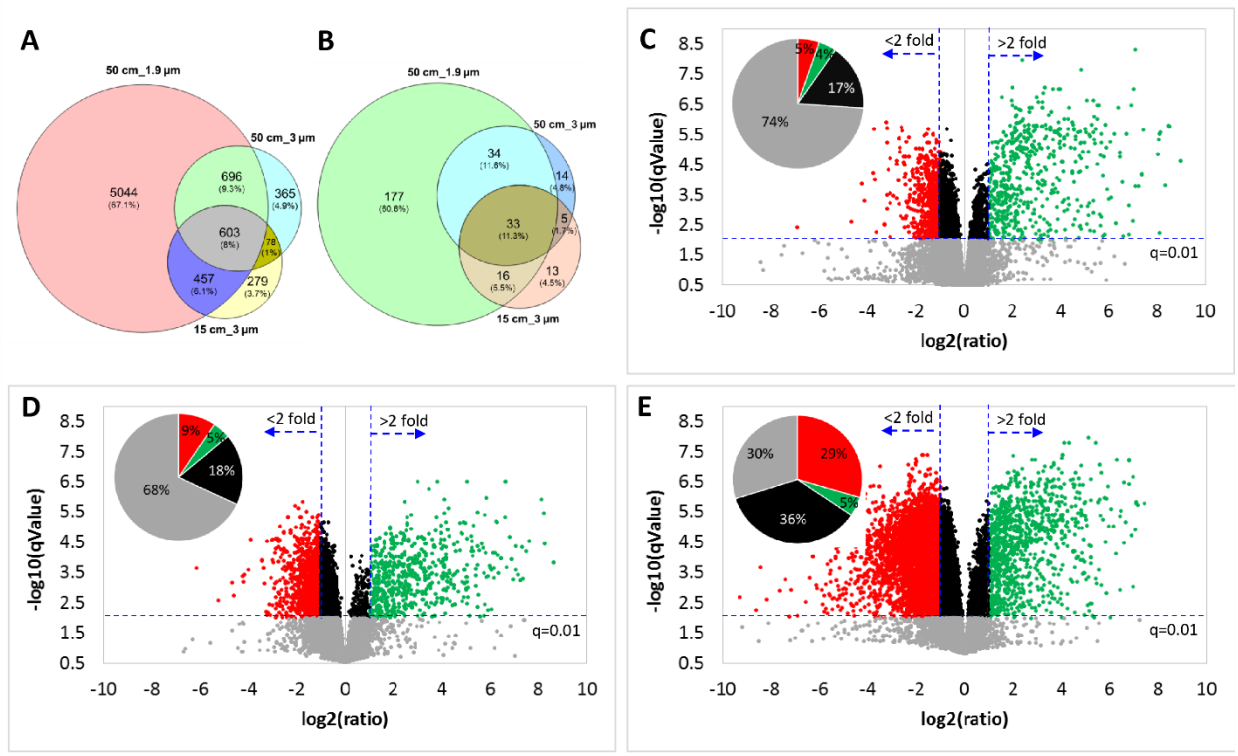


Figure 3.1: Venn diagram of the number of unique phosphorylation sites residing on significantly changed phosphopeptides and volcano plot analysis of significantly changed phosphopeptides in T cells in response to CD3/4 stimulation analyzed by the three different analytical column configurations. A and B represent the number of unique total (serine, threonine and tyrosine) and tyrosine phosphorylation sites with significantly altered phosphopeptide abundance, respectively. Volcano plot of fold change (log 2) versus q-value (-log10) of peak-area for 14,377, 15,114 and 22,945 phosphopeptides identified from 15 cm, 3.0 µm Column (C), 50 cm, 3.0 µm Column (D) and 15 cm, 1.9 µm Column (E), respectively. Red and green circles represent the significant (q > 0.01) 2-fold down and up regulated peptides, respectively compared with the untreated control cells. The black circles are significant but the fold change is lower than 2. Gray circles bellow the blue dot horizontal line (q = 0.01) are non significant. Pie chart shows percentage of the peptides significantly changed compared with the untreated cells.

The distribution of fold changes calculated between the stimulated and unstimulated cells was compared for the 1.9 µm and 3 µm column configurations (Supplemental Figure 3.11). This analysis revealed that the distribution of fold changes for the 1.9 µm configuration was broader than with the 3 µm configurations, indicating an elevated dynamic range. Overall, the 50 cm, 3 µm column missed 73% and the 15 cm, 3 µm column missed 76% of the unique phosphorylation sites residing on significantly changed phosphopeptides compared to the optimized 50 cm, 1.9 µm column (Figure 3.1A). A similar trend was observed when only unique phosphotyrosine sites

were compared (Figure 3.1B). A volcano plot analysis showed that 74%, 68% and 30% of phosphopeptides were not significantly changed in response to CD3/4 stimulation in T cells when analyzed by 3.0 µm, 15 cm, 3.0 µm, 50 cm and 1.9 µm, 50 cm column, respectively (Figure 3.1C-E). However, the percentage of elevated peptides (>2 fold, q<0.01) was very similar among the columns. Interestingly, the percentage of peptides with significantly decreased abundance (<2 fold, q<0.01) were gradually increased with increasing of column length and decreasing particle size (Figure 3.1C-E).

3.3.4. Higher Resolution Characterization of the T Cell Phosphoproteome is Enabled by the 50cm, Sub-2µm Particle Configuration

Although kinases play a prominent role in the pathway activated by crosslinking the T cell receptor, phosphatases and negative feedback pathways are also very important in these networks to regulate autoimmune disease and to control the duration of activation of T cells. Recent studies from our lab have revealed previously uncharacterized negative feedback pathways from Vav1, Zap-70, SLP-76, and PLC-γ1 to receptor proximal signaling molecules in T cells.[27, 41-45] At this point, the vast majority of the phosphorylation sites on phosphopeptides with reduced abundance are not yet biologically characterized. Future studies will determine the role, if any, of these newly discovered sites in feedback regulation of the signaling network. A comparative heatmap analysis of the KEGG T cell pathway annotated proteins observed in this study showed many novel phosphorylation sites that are significantly changed in response to T cell stimulation (Figure 3.2). The 1.9 µm, 50 cm column configuration significantly outperformed the other configurations. This configuration revealed both the expected elevation of

phoshopeptide abundance on canonical phosphorylation sites downstream of the T cell receptor as well as novel sites on previously established T cell signaling proteins.

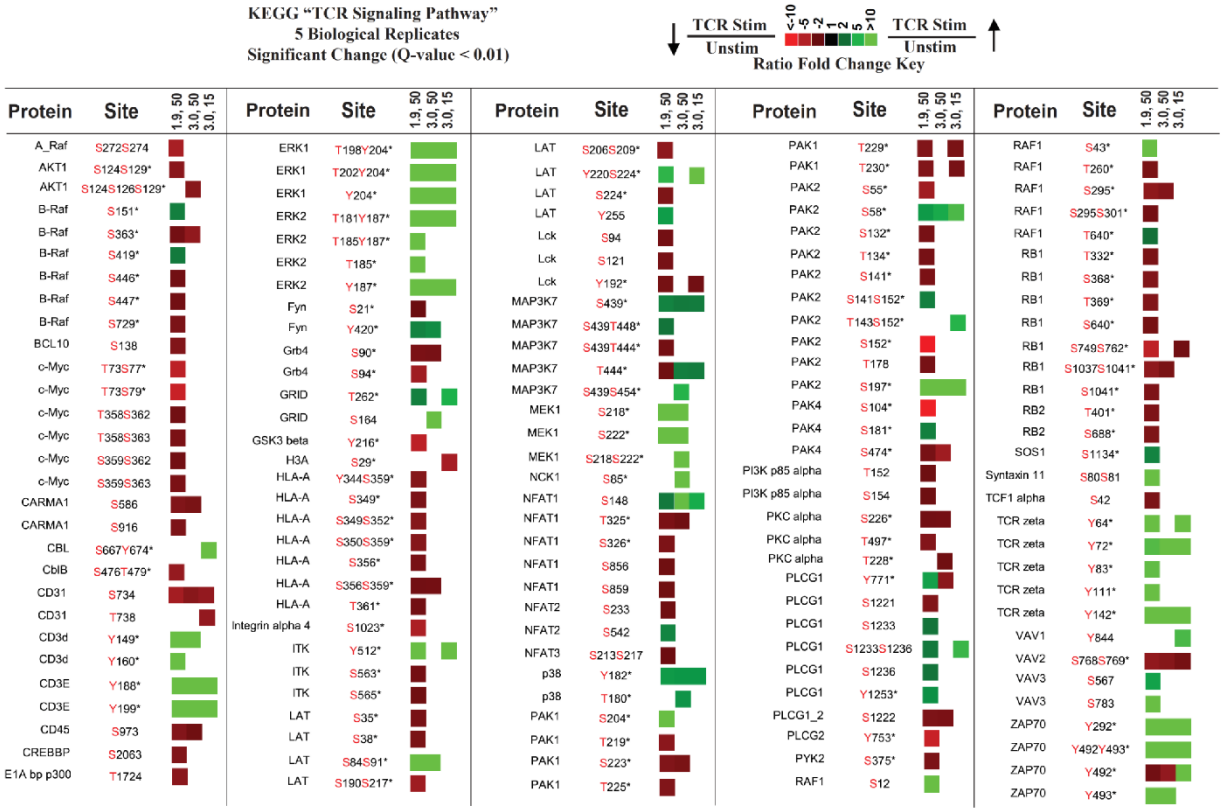


Figure 3.2: Comparative quantitative phosphoproteomic analysis of T-cells in response to CD3/4 stimulation. Quantitative changes in phosphorylation state of the proteins within the KEGG TCR signaling pathway ontology are represented as heatmaps. The results represent averages of five biological replicate experiments analyzed by three different analytical workflows. Blanks in the heatmap indicate that a clearly defined SIC peak was not observed for that phosphopeptide for that particular column. An asterisk next to the phosphorylation site signifies that this site has been categorized in the HPRD database as described previously in the literature.

Our quantitative data is consistent with previously described phosphorylation site changes on canonical T cell signaling proteins. For example, our study revealed TCR stimulation-induced elevated phosphorylation on immunoreceptor tyrosine activation motif phosphorylation sites on receptor subunits TCR ζ Y^{64}, Y^{72}, Y^{83}, Y^{111}, Y^{142}, CD3ε Y^{188}, Y^{199}, CD3δ Y^{149}, Y^{160}. Elevated ITAM phosphorylation is a hallmark of T cell activation.[46] Additionally, elevated phosphorylation was also observed on downstream canonical T cell activation sites on ITK Y^{512}, ERK1 $T^{202}Y^{204}$, Erk2 $T^{185}Y^{187}$, p38 Y^{182}, ZAP-70 Y^{292} and $Y^{492,493}$, MEK1 S^{218}, S^{222} as previously described.[27, 41-48] Although decreased phosphorylation is much less characterized within the TCR signaling pathway, this study revealed decreased phosphorylation on the Calcineurin

phosphatase substrate phosphorylation sites on NFAT1 at T^{325} and S^{326}. Calcineurin-mediated dephosphorylation of NFAT is required for its import from the cytosol into the nucleus where it regulates key transcriptional targets.[49] The canonical sites observed uniquely to be significantly elevated in abundance with the 1.9 µm, 50 cm column include CD3δ Tyr^{160}, TCR ζ $Tyr^{83,111}$, and PLC-γ1 Tyr^{1253}. The discovery power of the optimized column configuration was illustrated by a plethora of novel sites seen only with this column, significantly changed 3 min after T cell receptor stimulation on proteins previously established to play important roles in the signaling pathway including A-Raf, B-Raf, c-Myc, CARMA1, Fyn, ITK, LAT, NFAT1/2/3, PKCα, PLC-γ1/2, RAF1, and SOS1 (Figure 3.2).

3.4. Conclusions

Recently described phosphoproteomic approaches generally employ a targeted enrichment prior to LC-MS/MS analysis coupled with automated workflows. One recent study identified over 10,000 phosphopeptides and was able to decrease the variability among the samples.[10, 13, 15] Our results are in good agreement with earlier reports that demonstrated increasing column length and gradient time increased PSM yield.[3] Biological researchers are primarily concerned with detection of significant changes in the proteome or phosphoproteome following a biological stimulation or perturbation of cells through the analysis of biological replicates. Our optimized analytical column configuration significantly improved quantitative reproducibility, enabling detection of changes in the phosphoproteome through biological replicates. The improved ability of the 1.9 µm C18 configuration to detect large fold changes in phosphopeptide abundance may be attributed to this column's improved ability to reproducibly quantitate low abundance peptides.

In conclusion, the ideal column configuration reported here represents an inexpensive and highly accessible approach to analyze the phosphoproteome of cells. The use of a fritless 50 cm column packed with 1.9 um particles provides dramatic improvements in detection of significantly altered phosphopeptides without the need for specialized UHPLC or column heating apparatus.

This analysis emphasizes the importance of consideration of the quantitative reproducibility of proteomic methods in lieu of a monolithic focus on sequencing depth.

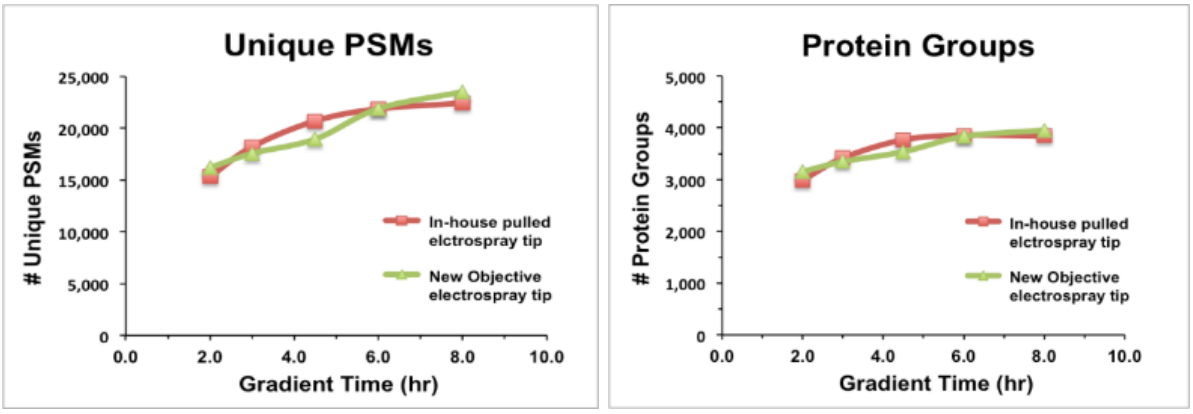
3.5. References

1. Aebersold, R. and M. Mann, *Mass spectrometry-based proteomics.* Nature, 2003. **422**(6928): p. 198-207.
2. Choudhary, C. and M. Mann, *Decoding signalling networks by mass spectrometry-based proteomics.* Nat Rev Mol Cell Biol, 2010. **11**(6): p. 427-39.
3. Pirmoradian, M., et al., *Rapid and deep human proteome analysis by single-dimension shotgun proteomics.* Mol Cell Proteomics, 2013. **12**(11): p. 3330-8.
4. Zubarev, R.A., *The challenge of the proteome dynamic range and its implications for in-depth proteomics.* Proteomics, 2013. **13**(5): p. 723-6.
5. Gautier, V., et al., *Label-free quantification and shotgun analysis of complex proteomes by one-dimensional SDS-PAGE/NanoLC-MS: evaluation for the large scale analysis of inflammatory human endothelial cells.* Mol Cell Proteomics, 2012. **11**(8): p. 527-39.
6. Gilar, M., et al., *Orthogonality of separation in two-dimensional liquid chromatography.* Anal Chem, 2005. **77**(19): p. 6426-34.
7. Gilar, M., et al., *Two-dimensional separation of peptides using RP-RP-HPLC system with different pH in first and second separation dimensions.* J Sep Sci, 2005. **28**(14): p. 1694-703.
8. Zhou, F., et al., *Online nanoflow reversed phase-strong anion exchange-reversed phase liquid chromatography-tandem mass spectrometry platform for efficient and in-depth proteome sequence analysis of complex organisms.* Anal Chem, 2011. **83**(18): p. 6996-7005.
9. Mann, M., et al., *Analysis of protein phosphorylation using mass spectrometry: deciphering the phosphoproteome.* Trends Biotechnol, 2002. **20**(6): p. 261-8.
10. Abelin, J.G., et al., *Reduced-representation Phosphosignatures Measured by Quantitative Targeted MS Capture Cellular States and Enable Large-scale Comparison of Drug-induced Phenotypes.* Mol Cell Proteomics, 2016. **15**(5): p. 1622-41.
11. Beausoleil, S.A., et al., *Large-scale characterization of HeLa cell nuclear phosphoproteins.* Proc Natl Acad Sci U S A, 2004. **101**(33): p. 12130-5.
12. Giansanti, P., et al., *Interrogating cAMP-dependent kinase signaling in Jurkat T cells via a protein kinase A targeted immune-precipitation phosphoproteomics approach.* Mol Cell Proteomics, 2013. **12**(11): p. 3350-9.
13. Humphrey, S.J., S.B. Azimifar, and M. Mann, *High-throughput phosphoproteomics reveals in vivo insulin signaling dynamics.* Nat Biotechnol, 2015. **33**(9): p. 990-5.
14. Larsen, M.R., et al., *Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns.* Mol Cell Proteomics, 2005. **4**(7): p. 873-86.
15. Post H, P.R., Fitzpatrick M. A., Garrigues L B., Wu W, MacGillavry H. D, Hoogenraad C. C, Heck A.J, Altelaar A.F, *Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons.* J Proteome Res, 2016.
16. Sugiyama, N., et al., *Phosphopeptide enrichment by aliphatic hydroxy acid-modified metal oxide chromatography for nano-LC-MS/MS in proteomics applications.* Mol Cell Proteomics, 2007. **6**(6): p. 1103-9.
17. Zhou, H., et al., *Robust phosphoproteome enrichment using monodisperse microsphere-based immobilized titanium (IV) ion affinity chromatography.* Nat Protoc, 2013. **8**(3): p. 461-80.
18. Pagel, O., et al., *Current strategies and findings in clinically relevant post-translational modification-specific proteomics.* Expert Rev Proteomics, 2015. **12**(3): p. 235-53.

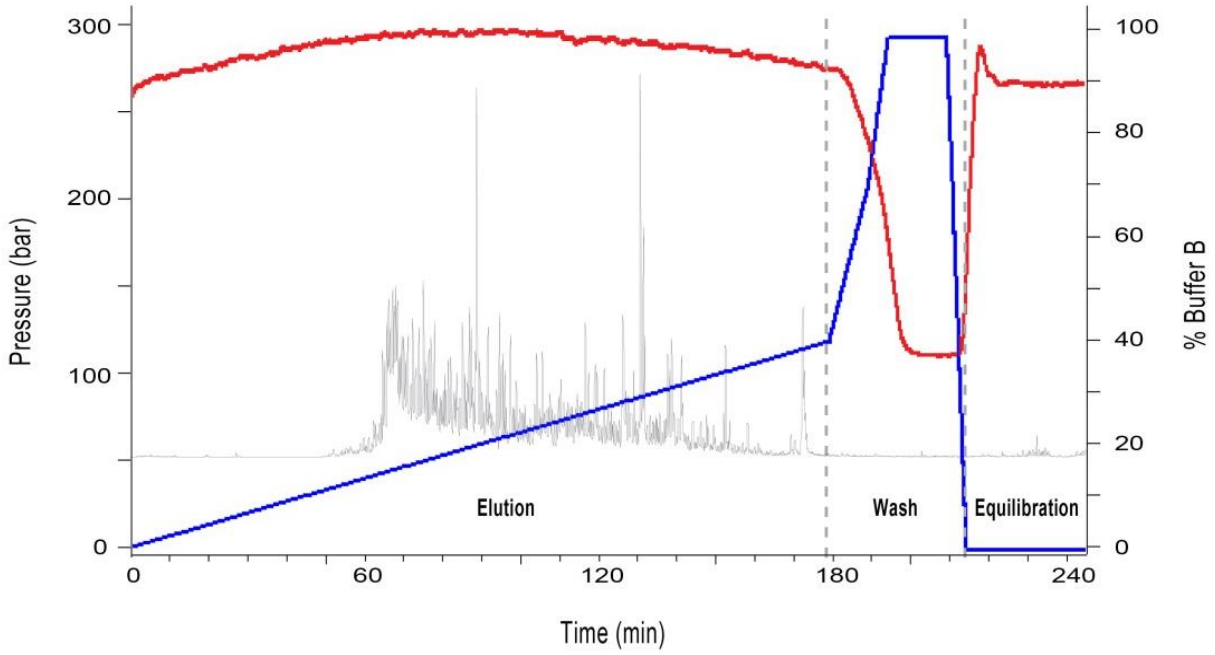
19. Cristobal, A., et al., *In-house construction of a UHPLC system enabling the identification of over 4000 protein groups in a single analysis.* Analyst, 2012. **137**(15): p. 3541-8.
20. Nagaraj, N., et al., *System-wide perturbation analysis with nearly complete coverage of the yeast proteome by single-shot ultra HPLC runs on a bench top Orbitrap.* Mol Cell Proteomics, 2012. **11**(3): p. M111 013722.
21. Thakur, S.S., et al., *Deep and highly sensitive proteome coverage by LC-MS/MS without prefractionation.* Mol Cell Proteomics, 2011. **10**(8): p. M110 003699.
22. MacNair, J.E., K.C. Lewis, and J.W. Jorgenson, *Ultrahigh-pressure reversed-phase liquid chromatography in packed capillary columns.* Anal Chem, 1997. **69**(6): p. 983-9.
23. MacNair, J.E., K.D. Patel, and J.W. Jorgenson, *Ultrahigh-pressure reversed-phase capillary liquid chromatography: isocratic and gradient elution using columns packed with 1.0-micron particles.* Anal Chem, 1999. **71**(3): p. 700-8.
24. Yin, X., et al., *Rapid and sensitive profiling and quantification of the human cell line proteome by LC-MS/MS without prefractionation.* Proteomics, 2014. **14**(17-18): p. 2008-16.
25. Ong, S.E., et al., *Stable isotope labeling by amino acids in cell culture, SILAC, as a simple and accurate approach to expression proteomics.* Mol Cell Proteomics, 2002. **1**(5): p. 376-86.
26. Ross, P.L., et al., *Multiplexed protein quantitation in Saccharomyces cerevisiae using amine-reactive isobaric tagging reagents.* Mol Cell Proteomics, 2004. **3**(12): p. 1154-69.
27. Nguyen, V., et al., *A new approach for quantitative phosphoproteomic dissection of signaling pathways applied to T cell receptor activation.* Mol Cell Proteomics, 2009. **8**(11): p. 2418-31.
28. Yu, K. and A.R. Salomon, *PeptideDepot: flexible relational database for visual analysis of quantitative proteomic data and integration of existing protein information.* Proteomics, 2009. **9**(23): p. 5350-8.
29. Yu, K. and A.R. Salomon, *HTAPP: high-throughput autonomous proteomic pipeline.* Proteomics, 2010. **10**(11): p. 2113-22.
30. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry.* Nat Methods, 2007. **4**(3): p. 207-14.
31. Yu, K., et al., *Integrated platform for manual and high-throughput statistical validation of tandem mass spectra.* Proteomics, 2009. **9**(11): p. 3115-25.
32. Beausoleil, S.A., et al., *A probability-based approach for high-throughput protein phosphorylation analysis and site localization.* Nat Biotechnol, 2006. **24**(10): p. 1285-92.
33. Demirkan, G., et al., *Phosphoproteomic profiling of in vivo signaling in liver by the mammalian target of rapamycin complex 1 (mTORC1).* PLoS One, 2011. **6**(6): p. e21729.
34. Storey, J.D., *The positive false discovery rate: A Bayesian interpretation and the q-value.* Annals of Statistics, 2003. **31**(6): p. 2013-2035.
35. Storey, J.D. and R. Tibshirani, *Statistical significance for genomewide studies.* Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9440-5.
36. Lord, G.A., et al., *Tapers and restrictors for capillary electrochromatography and capillary electrochromatography mass spectrometry.* Journal of Chromatography A, 1997. **768**(1): p. 9-16.
37. Mayer, M., et al., *Fritless capillary electrochromatography.* Electrophoresis, 1999. **20**(1): p. 43-49.
38. Fairchild, J.N., et al., *Correlation between peak capacity and protein sequence coverage in proteomics analysis by liquid chromatography-mass spectrometry/mass spectrometry.* J Chromatogr A, 2010. **1217**(29): p. 4779-83.
39. Novakova, L., et al., *High-resolution peptide separations using nano-LC at ultra-high pressure.* J Sep Sci, 2013. **36**(7): p. 1192-9.
40. Trusch, M., et al., *Improved particle-packed HPLC/MS microchips for proteomic analysis.* J Sep Sci, 2010. **33**(21): p. 3283-91.

41. Cao, L., et al., *Quantitative phosphoproteomics reveals SLP-76 dependent regulation of PAG and Src family kinases in T cells.* PLoS One, 2012. **7**(10): p. e46725.
42. Helou, Y.A., A.P. Petrashen, and A.R. Salomon, *Vav1 Regulates T-Cell Activation through a Feedback Mechanism and Crosstalk between the T-Cell Receptor and CD28.* J Proteome Res, 2015. **14**(7): p. 2963-75.
43. Ji, Q., Y. Ding, and A.R. Salomon, *SRC homology 2 domain-containing leukocyte phosphoprotein of 76 kDa (SLP-76) N-terminal tyrosine residues regulate a dynamic signaling equilibrium involving feedback of proximal T-cell receptor (TCR) signaling.* Mol Cell Proteomics, 2015. **14**(1): p. 30-40.
44. Ji, Q. and A.R. Salomon, *Wide-scale quantitative phosphoproteomic analysis reveals that cold treatment of T cells closely mimics soluble antibody stimulation.* J Proteome Res, 2015. **14**(5): p. 2082-9.
45. Sjolin-Goodfellow, H., et al., *The catalytic activity of the kinase ZAP-70 mediates basal signaling and negative feedback of the T cell receptor pathway.* Sci Signal, 2015. **8**(377): p. ra49.
46. Abraham, R.T. and A. Weiss, *Jurkat T cells and development of the T-cell receptor signalling paradigm.* Nat Rev Immunol, 2004. **4**(4): p. 301-8.
47. Brill, L.M., et al., *Robust phosphoproteomic profiling of tyrosine phosphorylation sites from human T cells using immobilized metal affinity chromatography and tandem mass spectrometry.* Anal Chem, 2004. **76**(10): p. 2763-72.
48. Salomon, A.R., et al., *Profiling of tyrosine phosphorylation pathways in human cells using mass spectrometry.* Proc Natl Acad Sci U S A, 2003. **100**(2): p. 443-8.
49. Okamura, H., et al., *Concerted dephosphorylation of the transcription factor NFAT1 induces a conformational switch that regulates transcriptional activity.* Mol Cell, 2000. **6**(3): p. 539-50.

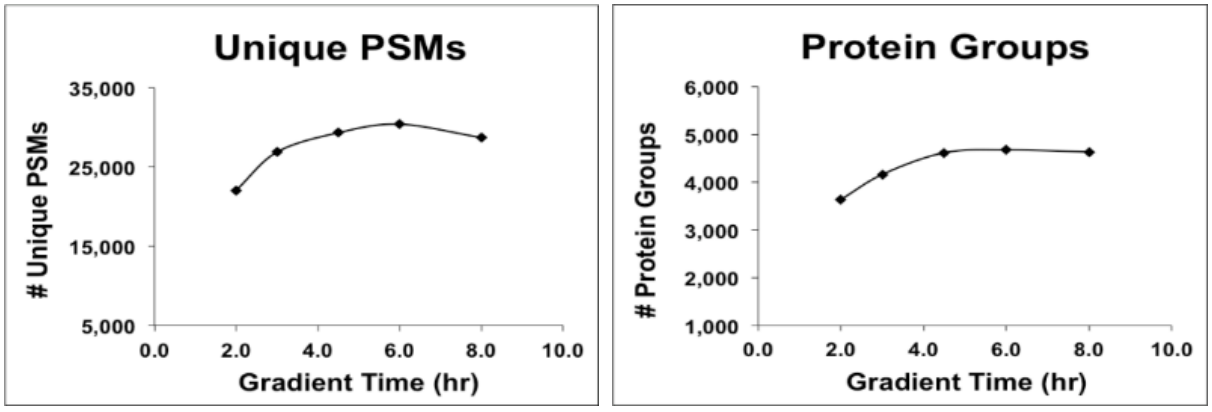
3.6. Supplemental Data



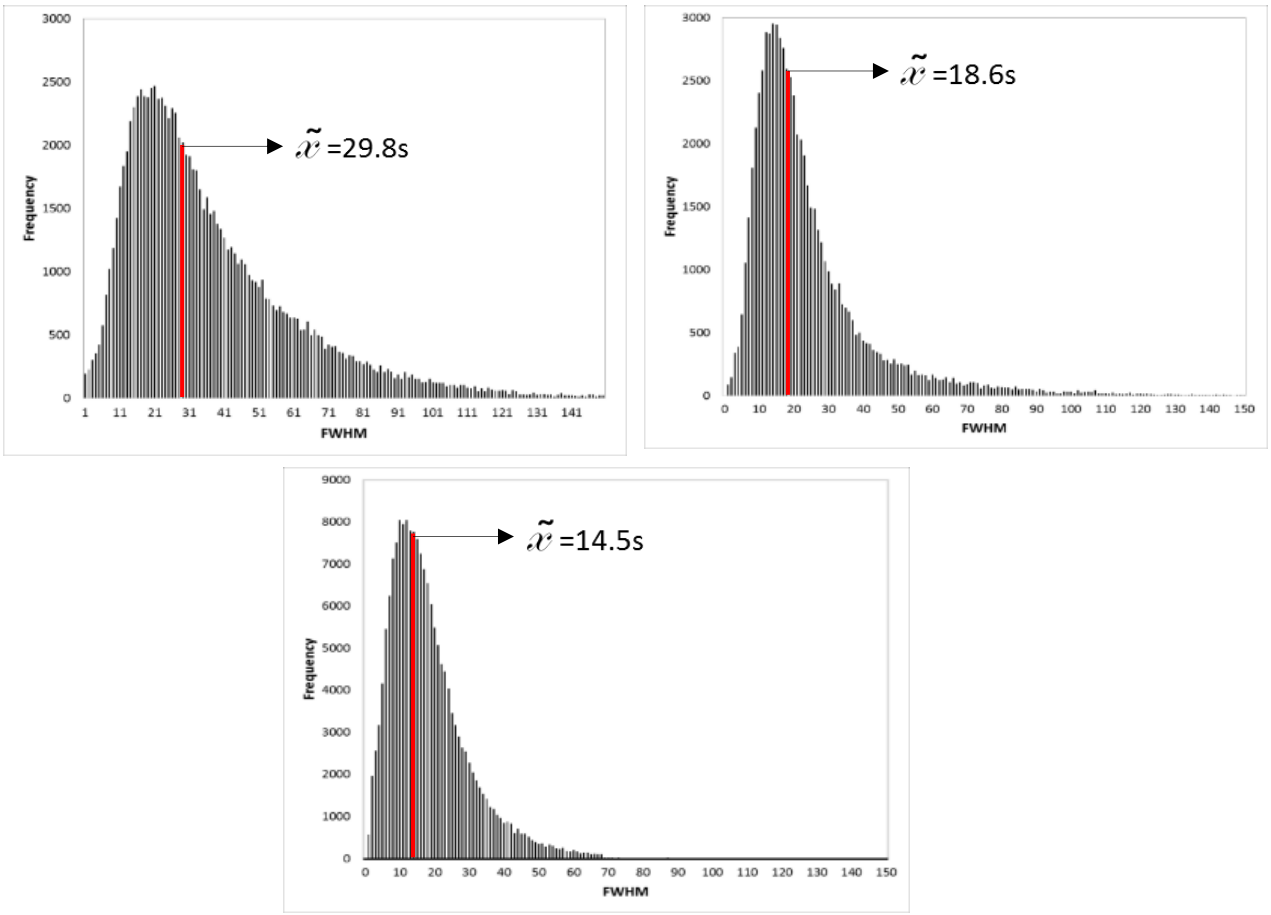
Supplemental Figure 3.1: Optimizing gradient length versus phosphoproteome coverage. Comparison of the number of identified unique peptide-spectrum matches (A) and protein groups (B) from a single-run analysis of Jurkat whole cell lysate tryptic peptides using 1.9 µm, 50 cm fritless analytical columns integrated with two different electrospray tips.



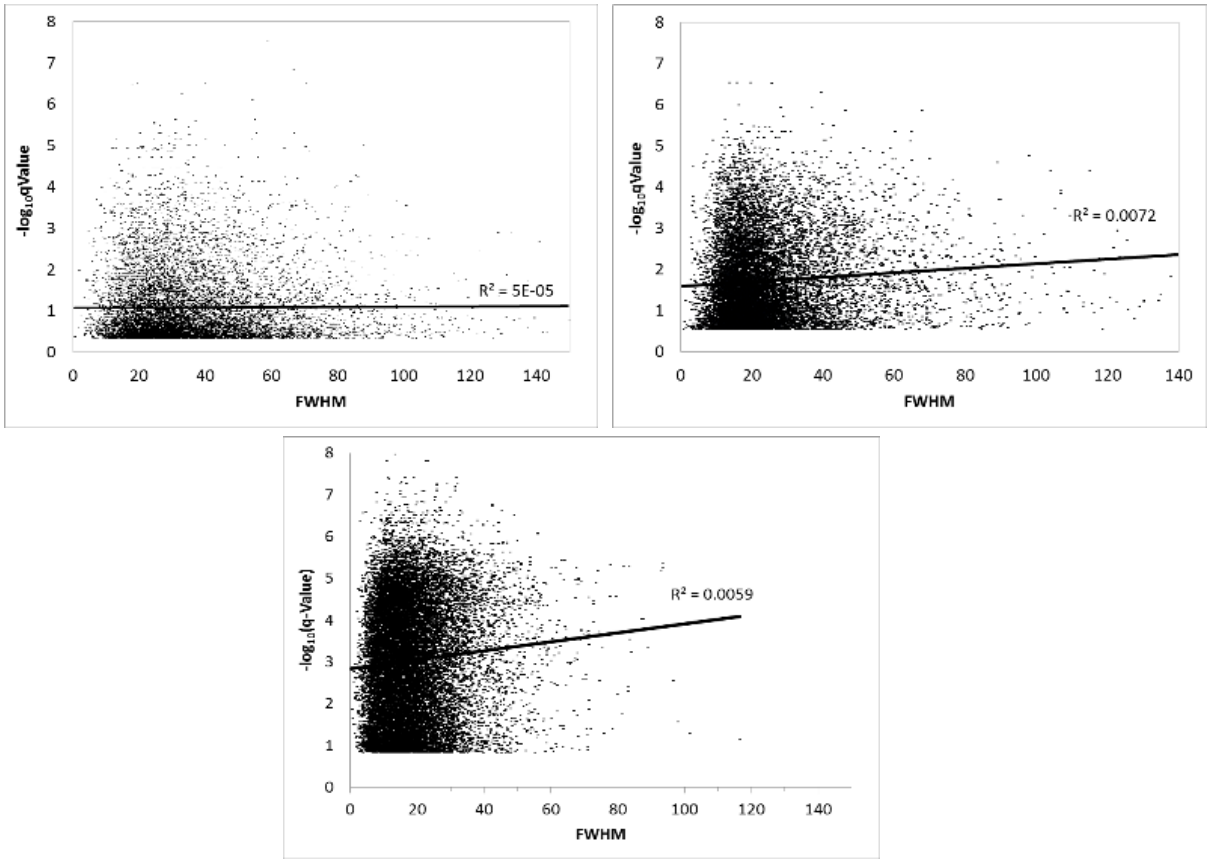
Supplemental Figure 3.2: HPLC separation profile of 50 cm-long/1.9 µm C_{18} column. The red line indicates the pressure curve during the LC-MS/MS analysis and the blue line indicates the %B at the HPLC pump. The gray line indicates the total ion chromatogram (TIC) of the LC-MS/MS analysis of Jurkat tryptic peptides.



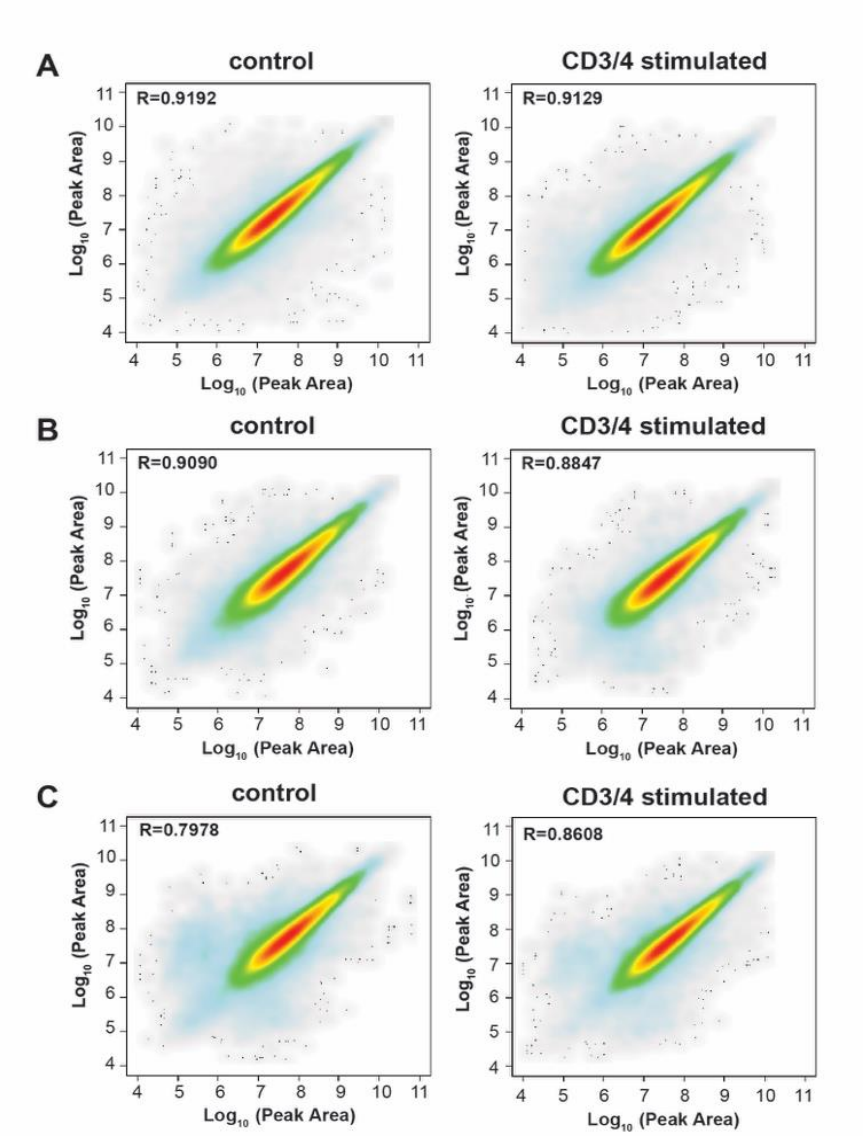
Supplemental Figure 3.3: Optimizing gradient length versus average phosphoproteome coverage in triplicate LC-MS runs. Average number of identified unique peptide-spectrum matches (A) and protein groups (B) from Jurkat tryptic peptides from a triplicate analysis as a function of gradient time. The analytical column is 50 cm-long/1.9 µm C18 column. The relative standard deviations of all the data points are below 2.5% (Supplemental Table 3.2).



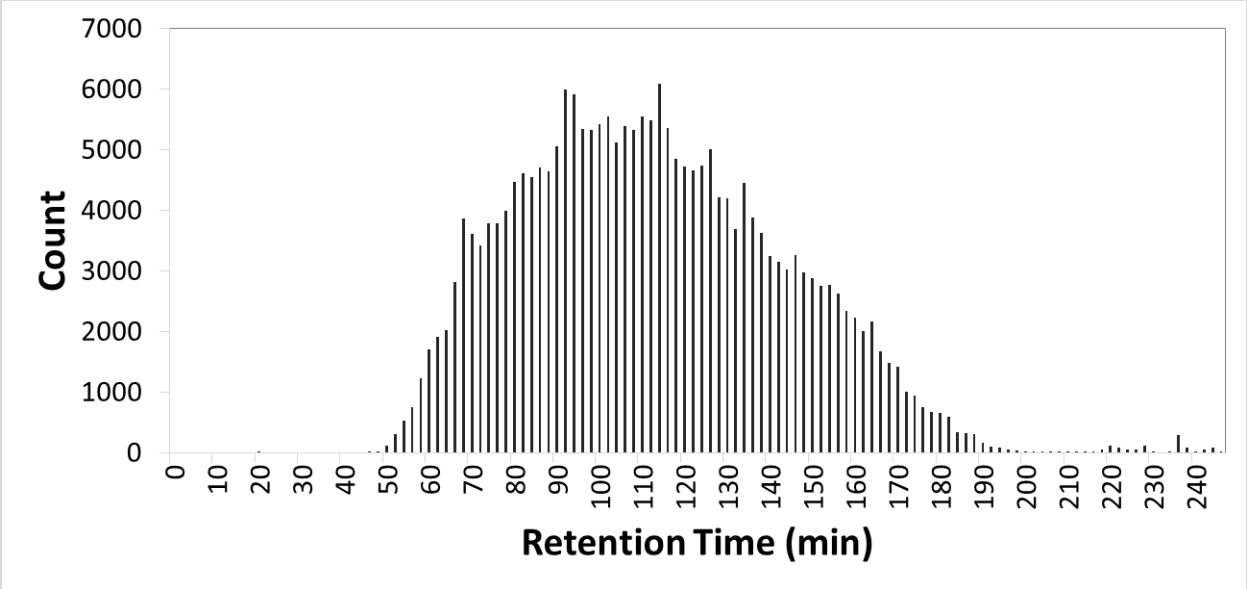
Supplemental Figure 3.4: Histogram analysis of full width at half maximum (FWHM) obtained from three different columns. A) 15 cm 3.0 µm C18 particles, B) 50 cm 3.0 µm C18 particles and C) 50 cm 1.9 µm C18 particles. The data was obtained from a 3 hr LC-MS/MS analysis of TiO2-enriched Jurkat tryptic peptides as described in the Methods section.



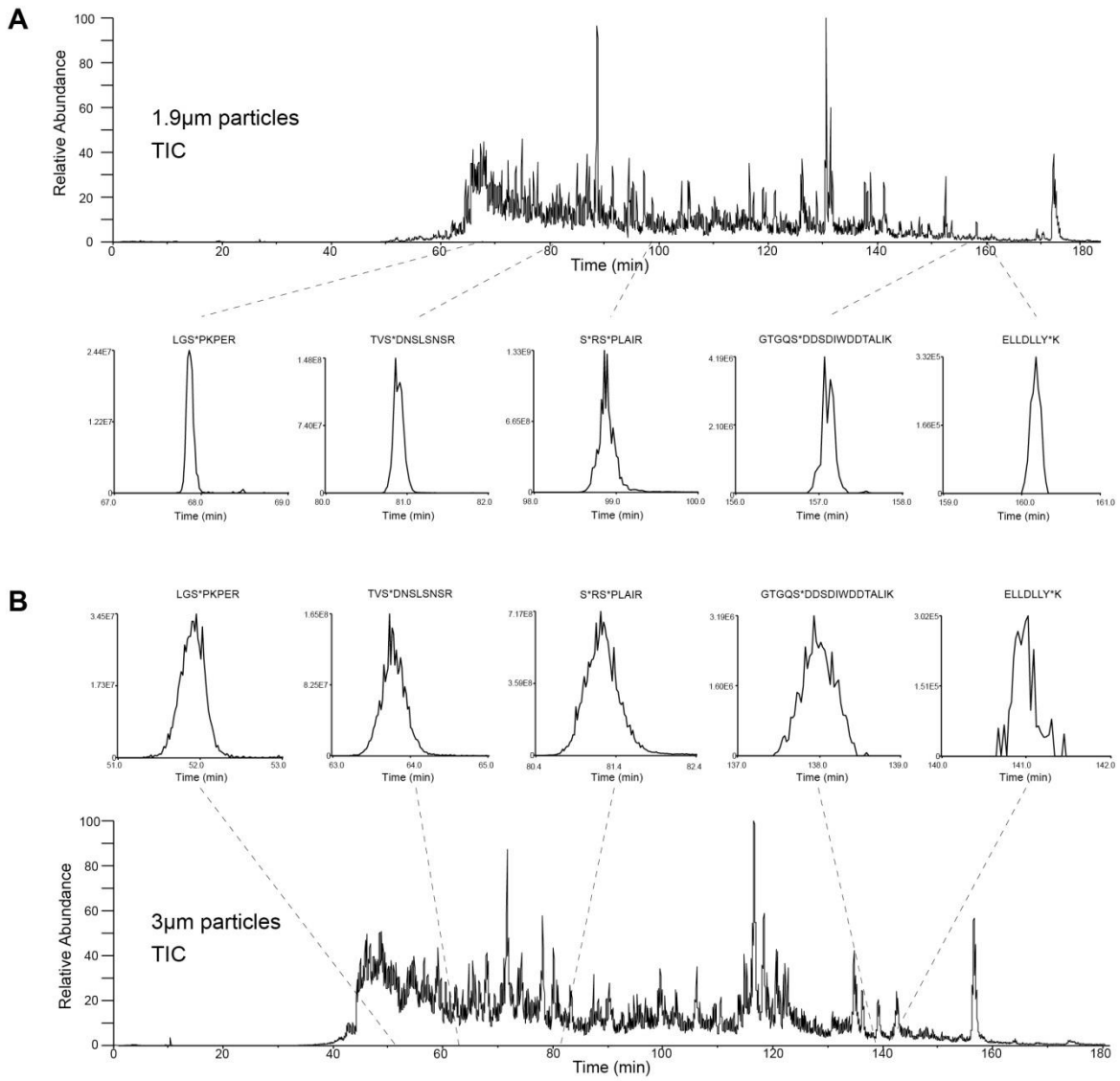
Supplemental Figure 3.5: Scatter plot analysis of q-value (-log 10) vs the width at half maximum (FWHM) obtained from three different columns. A) 15 cm 3.0 µm C18 particles, B) 50 cm 3.0 µm C18 particles and C) 50 cm 1.9 µm C18 particles . The data was obtained from a 3 hr LC-MS/MS analysis of TiO2-enriched Jurkat tryptic peptides as described in the Methods section.



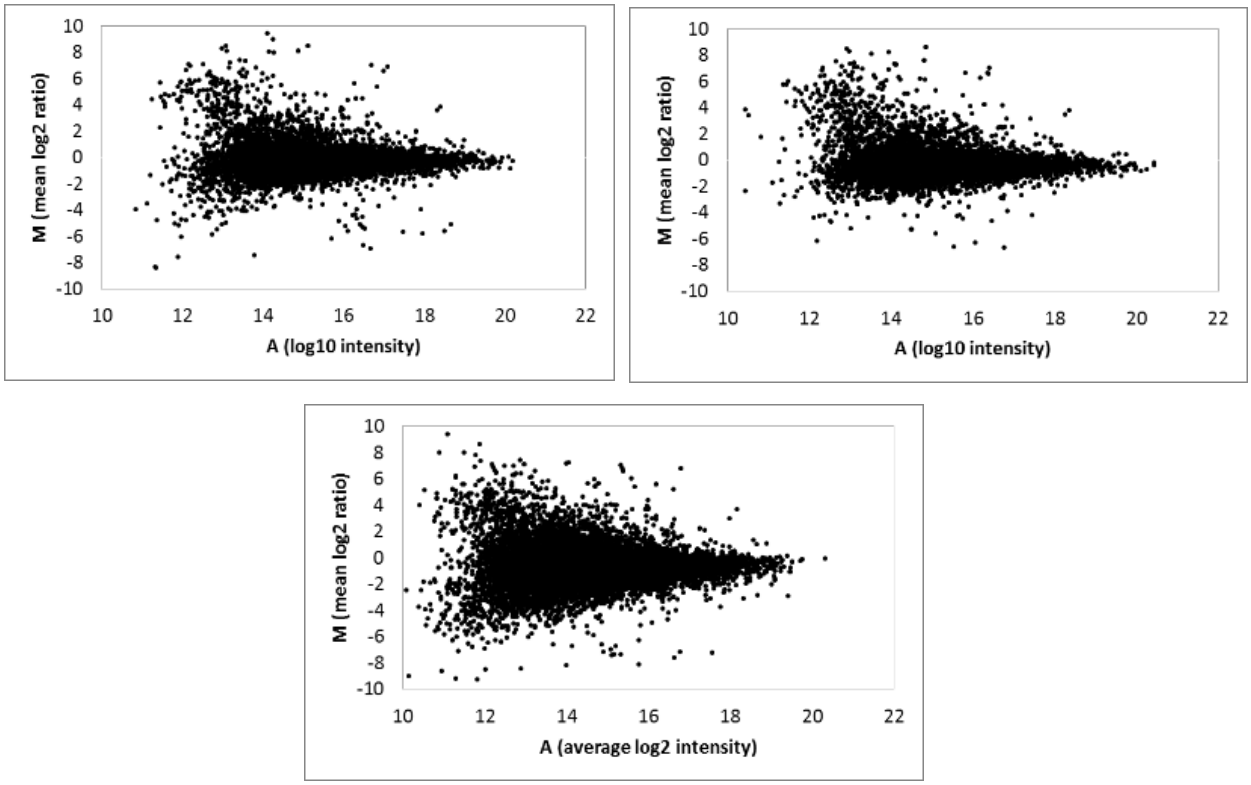
Supplemental Figure 3.6: Pairwise replicate comparison of selected ion chromatography (SIC) peak areas from different analytical columns. (A) 50 cm-long/1.9 µm C18 columns (B) 50 cm-long/3 µm C18 columns (C) 15 cm-long/3 µm C18 columns. Each dot in the scatterplot represents the log10 (normalized SIC peak area) of a single phosphopeptide in two different replicates (ten possible pairs in total: Rep1:Rep2, Rep1:Rep3, Rep1:Rep4, Rep1:Rep5, Rep2:Rep3, Rep2:Rep4, Rep2:Rep5, Rep3:Rep4, Rep3:Rep5, Rep4:Rep5.). Dot density is indicated by color (from low to high: gray, blue, green, yellow, orange and red). The sample types (CD3/4 stimulated or unstimulated (control)) and correlation coefficient are marked in the figure respectively. The calculation of correlation coefficient was based on the combination of all the 10 pairwise replicate to replicate comparisons.



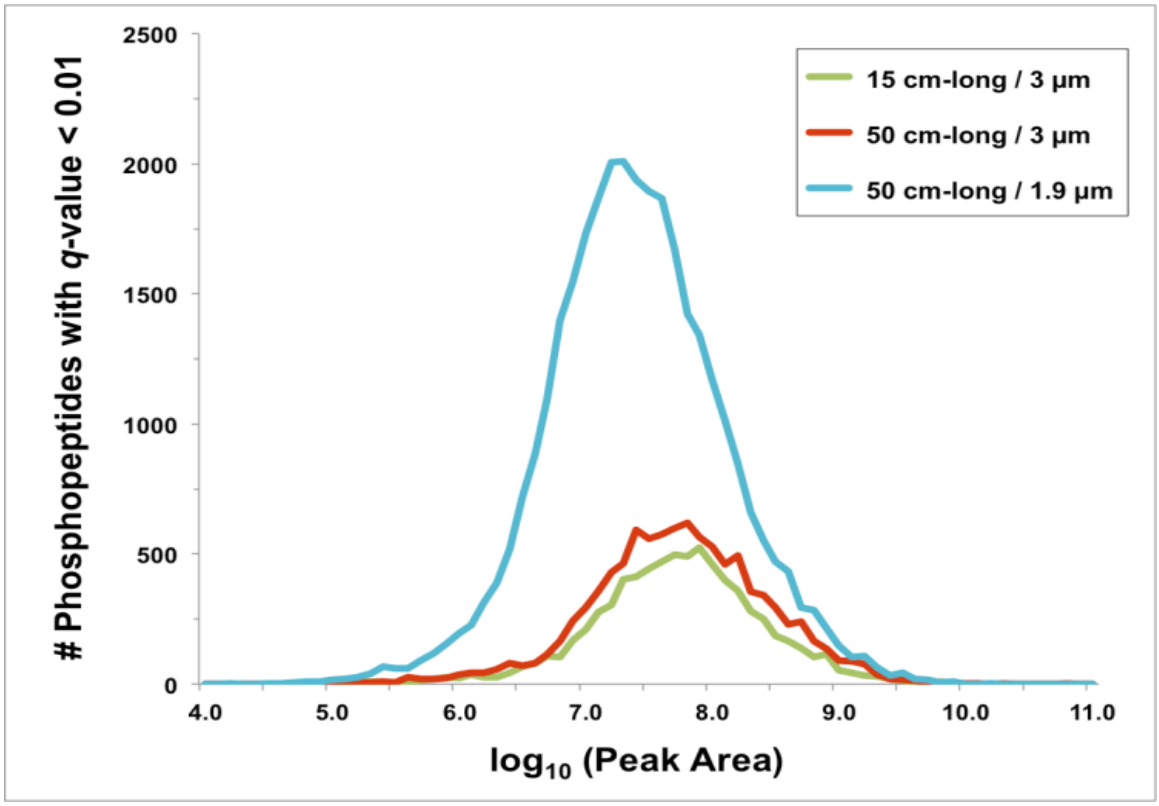
Supplemental Figure 3.7: Retention time distribution of the phosphopeptides detected using the 50 cm, 1.9 µm column configuration.



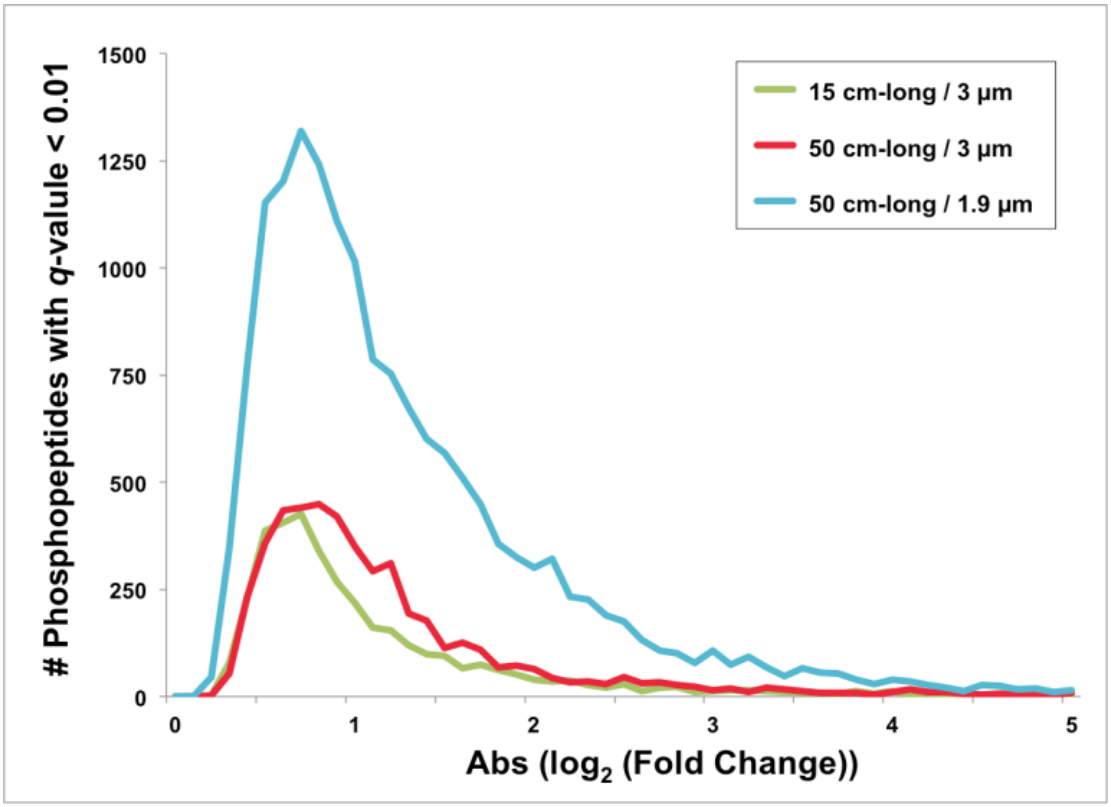
Supplemental Figure 3.8: TIC and SIC profiles obtained using 3 µm and 1.9 µm C18 particles. Comparison of total ion chromatogram (TIC) and five representative peptide selected ion chromatograms (SIC) between two 50 cm-long columns packed with (A) 1.9 µm C18 particles and (B) 3 µm C18 particles. The data was obtained from a 3 hr LC-MS/MS analysis of TiO2-enriched Jurkat tryptic peptides as described in the Methods section. SICs were randomly selected at a variety of abundance levels. Peptide sequences of five SICs are marked in the figure respectively. The retention time window (x axis range) of all the SICs are set at 2 min.



Supplemental Figure 3.9: MA plot analysis of the quantified phosphopeptides from T cells in response to CD3/4 stimulation analyzed by the three different analytical columns. A-C represents the MA-plot of 15 cm 3 µm column, 50 cm 3 µm column and 50 cm 1.9 µm column, respectively.



Supplemental Figure 3.10: Normalized peak area distribution of three different analytical columns. Peak area is the average value from the five replicate experiments. This distribution includes only the identified phosphopeptides with significant change ($q < 0.01$) when comparing the CD3/4 stimulated and unstimulated Jurkat cells.



Supplemental Figure 3.11: Fold change distribution of three different analytical columns. Fold change is defined as the ratio of normalized peptide peak areas between CD3/4 stimulated and unstimulated Jurkat cells. This distribution includes all the identified phosphopeptides with significant change (q < 0.01).

Cycle number of instantaneous pressurization and depressurization	Length of column packing bed (cm)	Flow rate at the tip of column (nl/min) [a]	LC pressure when testing flow rate (Bar)
0	50.25	55	258.62
10	50.25	60	258.89
20	50.25	60	259.93
30	50.25	60	258.31
40	50.25	55	260.14
50	50.25	60	257.64
60	50.25	60	259.62
70	50.25	55	259.27
80	50.25	60	261.37
90	50.25	60	258.26
100	50.25	55	257.74

[a] Flow rate was estimated by measuring the length of eluent (collected from column tip) in a 1~5 µl calibrated pipet during a given time window.

Supplemental Table 3.1: Stability test for in-house fabricated 50 cm-long, 1.9 µm C18 fritless column

LC gradient time (hr)	Average yield from a single run[a]		Aggregate yield from triplicate analyses	
	# Unique PSMs	# Protein groups	# Unique PSMs	# Protein groups
2	15324 ± 128	2987 ± 25	22059	3630
3	18330 ± 168	3426 ± 38	26894	4165
4.5	20706 ± 408	3767 ± 52	29327	4620
6	21876 ± 369	3856 ± 56	30403	4679
8	22427 ± 559	3843 ± 71	28716	4638

[a] The value after "±" indicates the standard deviation of the triplicate experiments.

Supplemental Table 3.2: Comparison of PSM yield of Jurkat-derived tryptic peptides using in-house fabricated 50 cm-long, 1.9 µm C18 column with different LC gradients

		Analytical Column Type		
	Replication	15 cm 3 µm	50 cm 3 µm	50 cm 1.9 µm
Unstimulated	R1	5.10E+09	3.72E+09	2.62E+09
	R2	3.82E+09	3.46E+09	2.97E+09
	R3	3.45E+09	2.90E+09	2.79E+09
	R4	3.89E+09	3.28E+09	2.95E+09
	R5	4.08E+09	3.53E+09	2.78E+09
CV (%)		15.29	9.25	5.1
CD3/4	R1	4.72E+09	3.98E+09	3.47E+09
	R2	4.35E+09	3.63E+09	4.35E+09
	R3	4.56E+09	3.74E+09	4.03E+09
	R4	4.87E+09	4.64E+09	4.37E+09
	R5	4.01E+09	5.51E+09	3.74E+09
CV (%)		7.47	18.19	9.7

Supplemental Table 3.3: Selected ion chromatogram peak areas of the exogenously spiked standard peptide used for normalization across each individual LC-MS experiment

Column type	Range of normalized peak area [a]			
	< 1E7	1E7 - 5E7	5E7 – 1E8	> 1E8
50 cm-long / 1.9 µm C_{18} column	30.2%	41.1%	12.3%	16.4%
50 cm-long / 3 µm C_{18} column	13.9%	36.8%	17.7%	31.6%
15 cm-long / 3 µm C_{18} column	13.2%	37.5%	19.7%	29.6%

[a] Peak area is the average value of five replicates. This distribution includes all the identified phosphopeptides with significant change (q < 0.01) in CD3/4 stimulated and unstimulated Jurkat cells.

Supplemental Table 3.4: Peak area distribution of different analytical columns

Column length / Particle size	Total # unique phosphopeptide-spectrum matches [a]	# Unique phosphopeptide-spectrum matches with q < 0.01 and:		
		All	> 2-fold [b]	> 10-fold [b]
50 cm / 1.9 um	23131	16106	7900	660
50 cm / 3 um	15173	4860	2117	221
15 cm / 3 um	14494	3749	1391	191

[a] Includes all identified unique phosphopeptide-spectrum matches in CD3/4 stimulated and unstimulated Jurkat from five replicates.

[b] Indicates number of significantly changed phosphopeptide-spectrum matches (q-value < 0.01) with 2- or 10-fold increase or decrease when comparing average peak area between CD3/4-simulated and -unstimulated Jurkat.

Supplemental Table 3.5: Total and unique PSMs recovered using the three analytical columns.

Chapter 4

Integrated Tools for Quantitative Post-Acquisition Analysis of Proteomic Data in a Modern LC-MS Data Management and Visualization Platform

4.1. Introduction

The complexity of modern shotgun proteomics datasets poses significant technical and analytical challenges. Development of new chromatographic techniques[1-3] as well as technical advances in mass spectrometry hardware[4, 5] have contributed to marked increases in the number of peptide-spectrum matches (PSMs) obtained in contemporary LC-MS screens. While increasingly larger datasets have spurred the innovation of new computational solutions to manage them, many proteomics workflows still suffer from a series of issues impacting the efficiency and quality of data analysis. These issues often stem from the patchwork nature of LC-MS data analysis pipelines: in order to address their scientific questions, researchers must often combine resources from many different bioinformatics packages not designed for either interoperability or ease of use for non-programmers.

Recognizing the need for all-in-one software platforms to support proteomic workflows, numerous academic and commercial entities have introduced software packages to assist with multiple stages of LC-MS data processing and analysis. PREMIER Biosoft's ProteoIQ software (http://www.premierbiosoft.com/protein_quantification_software/index.html) gives users the ability to quantitate, annotate, and compare proteomic datasets. Statistical analysis tools and ontology-based annotation are also featured in ProteoIQ. OpenMS is a versatile platform for the creation of custom LC-MS data pipelines that adapts easily to new proteomic workflows thanks to its powerful API.[6] Perseus, developed in the lab of Mattias Mann, provides a comprehensive set of biological analysis tools,[7] and is tightly integrated with the Mann lab's popular MaxQuant platform for peptide identification and quantitation.[8] The MacCoss lab's open-source tool Skyline is a flexible platform for quantifying, browsing, and comparing LC-MS datasets that benefits from a growing collection of plug-ins that expand its capabilities.[9] The success of these platforms underscores the need for flexible software solutions that support later stages of users' workflows beyond peak area quantitation. However, despite the capabilities for multiple modes of analysis offered by many of these software packages, features such as efficient

visualization of selected ion chromatogram characteristics, pathway analysis, and deep integration of protein- and amino acid site-level knowledge bases often still require either further software engineering on the part of the user or external tools.[10] Furthermore, researchers working within subspecialties of proteomics may require more application-specific software solutions in order to enjoy the same out-of-the-box functionality that existing platforms offer to other users.[11-13]

The capability of LC-MS to simultaneously quantify thousands of post-translationally modified peptides in a single run makes it a powerful method for the dissection of complex signaling pathways. In phosphoproteomic studies, the unbiased, systems-level view afforded by LC-MS can yield invaluable insight into coordinated regulation of critical kinases and phosphatases. Such studies are assisted by the wealth of sequence- and site-level information in databases such as HPRD [14], PhosphoSitePlus [15], PFAM [16], InterPro [17], and Scansite [18], which provide tremendous resources for the annotation of phosphoproteomic datasets. However, extracting maximal benefit from these resources for phosphoproteomic applications requires deeper integration with existing analytical software: while native support for GO and KEGG in bioinformatics tools has become widespread, the gene/protein-level annotations these ontologies provide are of limited use in interpreting biological function at the level of amino acid-specific signaling events, such as phosphorylation. If the quality and thoroughness of phosphoproteomic data analyses are to keep pace with technical innovations delivering increasingly greater yields of PSMs, the next generation of phosphoproteomic analysis platforms must ensure that highly granular information about sites and sequence motifs is available at users' fingertips.

Previously, we introduced the High-Throughput Autonomous Proteomic Pipeline (HTAPP), a suite of software tools designed to streamline each step of the quantitative proteomics workflow from the acquisition and preprocessing of raw spectral data to the annotation of assigned peptides and characterization of biologically interesting species.[19] Central to HTAPP is the

relational database application Peptide Depot, which handles the comparison, visualization, and biological analysis of user data.[20] In this chapter, we highlight the recent developments HTAPP has undergone to meet the evolving needs of the modern proteomics workflow, with an emphasis on the analytical modules newly available in Peptide Depot. These developments build on the previously discussed capabilities of Peptide Depot, and include the introduction of chromatogram metrics for quality control of LC-MS data, expansion of built-in statistical methods for analysis and visualization of high-dimensional data, and tools for generating and viewing tissue-specific network models from user data. The latest release of Peptide Depot also debuts extensible Set Enrichment Analysis, or xSEA, a novel enrichment method for detecting correlated phosphorylation of sites sharing either common upstream kinases or regulatory motifs. With the addition of xSEA, Peptide Depot allows users to interrogate datasets for evidence of kinase and pathway activation directly from an intuitive user interface.

4.2. Materials and Methods

4.2.1. Design of Peptide Depot

The Peptide Depot relational database component consists of a MySQL (version 5.6.17; MySQL Inc., Cupertino, CA) and FileMaker (version 15.0.3; FileMaker Inc., Santa Clara, CA) stack. The two database platforms are transparently integrated for the end user using FileMaker's external SQL sources capabilities and ODBC. The database may be accessed via a web page with any operating system or via a stand-alone FileMaker client in Windows or Mac OS X. Data are also accessible independent of the FileMaker client software via ODBC, JDBC, and the FileMaker PHP API.

Customized graphical layouts within FileMaker allows the user to explore the proteomic data and associated experimental metadata, as well as to export those data into Excel or PDF files in a format designed to meet the data reporting requirements of MIAPE-compliant peer-reviewed journals.[21, 22] Within Peptide Depot, manipulation of data, construction of

quantitative comparisons, and generation of visual displays are achieved through FileMaker scripts underlying an intuitive GUI.

4.2.2. Precursor Ion Quantitation

Quantitation routines in HTAPP are implemented in R, and make use of the Scripps Research Institute's open-source xcms package.[23] Input to the quantitation modules consists of experimental data formatted as mzXML or mzML files as well as an additional XML file describing the structure, if any, of the quantitative comparison in which the data will be included. Specifically, HTAPP's default quantitation routine draws on the xcms's *centWave* algorithm, which utilizes continuous wavelet transformation to resolve peaks. By iterating over increasingly larger wavelet scales, the software ensures an optimal yield of resolvable peaks while favoring those with a high signal:noise. To expedite users' workflows, HTAPP's quantitation modules are easily configurable by end users to run in a distributed computing environment to allow for parallelization of tasks across multiple networked multi-core CPUs.

4.2.3. Chromatogram QC Metrics

All chromatographic quality metrics are computed in R during quantitation. In total, 12 metrics are computed for each quantified SIC. For all equations below, an SIC is described by an ordered set of intensities $I_1..I_N$, where I_i is the intensity observed at index i, and an ordered set of times $T_1..T_N$, where T_i is the retention time at index i. p is the index of the maximum $I_1..I_N$, w is the width of the SIC, or T_N-T_1, and μ and σ are the mean and standard deviation of $I_1..I_N$, respectively.

(1) *Proximity to MSMS call* Difference between the center of the SIC and the retention time when the MSMS call was triggered.

(2) *Signal:noise ratio*

(3) *Jaggedness* Number of times the first derivative of the peak changes sign, with additional threshold terms to allow tolerance for minor fluctuations in peak slope. It is calculated as follows:

(a) Let s_i be the slope described by $\frac{I_{i+1}-I_i}{T_{i+1}-T_i}$. Initialize the index L to 1 so that $s_L = \frac{I_2-I_1}{T_2-T_1}$ and $I_L = I_1$. Also initialize the negative and positive thresholds r^- and r^+ respectively, where $0 < r^- < 1$ and $r^+ \geq 0$, and the Δs count to 0.

(b) For i in 2..*N-1*:

(i) Calculate the slope s_i.

(ii) If $s_i * s_L < 0$:

1. If ($s_i < 0$ AND $I_i < I_L/r^-$) OR ($s_i > 0$ AND $I_i \geq I_L * r^+$):

a. Set L to i, and increment Δs by 1.

(iii) If $s_i * s_L > 0$:

1. If ($s_i > 0$ AND $I_i > I_L$) OR ($s_i < 0$ AND $I_i < I_L$):

a. Set L to i.

(iv) If $s_i * s_L = 0$:

1. If $I_i \neq I_L$:

a. Set L to i, and increment Δs by 1.

(4) *Fluctuations from baseline* Number of times the intensity of the SIC falls to the local baseline value (zero in the case of non-noisy regions) when iterating through scans.

(5) *Maximum intensity*

(6) *Peak width* Total width of the SIC in seconds.

(7) *Peak kurtosis* Measure of the combined weight of a peak's tails in relation to the total shape of the peak. Expressed as $kurtosis = \frac{1}{N}\sum_{i-1}^{N}((\frac{I_i-\mu}{\sigma})^4) - 3$.

(8) *Peak skew* Measure of peak symmetry. Expressed as $skew = (\frac{1}{N}) \sum_{i-1}^{N} (\frac{I_i - \mu}{\sigma})^3$.

(9) *Peak significance* Ratio of the mean intensity near the center of the SIC to the mean intensity near the SIC boundaries.[24, 25]

(10) *Peak sharpness* Steepness of the peak on either side of the apex. Expressed as $sharpness = rise + fall$, where $rise = \sum_{i=2}^{p} \frac{I_i - I_{i-1}}{I_{i-1}}$ and $fall = \sum_{i=p}^{N-1} \frac{I_i - I_{i+1}}{I_{i+1}}$ [24]

(11) *Triangle peak area similarity ratio (TPASR)* Measures the similarity of the peak shape to an equilateral triangle fit over the region of the SIC. Expressed as $TPASR = \frac{|0.5 * w * I_p - \sum_{i=1}^{N} I_i|}{0.5 * w * I_p}$ [26]

(12) *Full Width at Half Maximum* Difference in retention time between the two extreme points where the SIC intensity is equal to half of its maximum value.

4.2.4. Statistical Analysis Modules

Statistical routines are performed by in-house software modules written in Java and R making use of the R package QVALUE[27] and the base R package for ANOVA and Tukey features. Input to Peptide Depot's statistical routines is provided as an XML of peptide abundances, while the output consists of a tab-delimited text file reporting both non-FDR adjusted and adjusted *p* values as well as diagnostic plots to assist the user in selecting appropriate statistical thresholds. To correct for multiple hypothesis testing, Peptide Depot allows the user to compute adjusted *p* values via the Benjamini-Hochberg or Storey method for all supported tests.[28-30] When calculating *q* values using Storey's method, a series of values between 0 and 0.95 are used by default for the tuning parameter, λ, and the smoothing method for estimating the overall proportion of truly null hypotheses (π_0) is used as previously described.[30]

4.2.5. Design of Peptide Depot

Interaction networks in Peptide Depot are modeled as undirected, unweighted graphs, $G=(V,E)$, where V={proteins in the selected quantitative comparison(s) $\cup$ user-specified anchor proteins} and E={protein-protein interactions in either HPRD or STRING between two proteins *v1* and *v2*, where $v1,v2 \subseteq V$}. Interaction networks are constructed by R scripts utilizing the igraph package.[31] Input to the R script is a text file of identifiers for all proteins which will be represented in the graph, and is collated by Peptide Depot after the user specifies the comparisons to be used. By default, protein-protein interactions that are supported only by yeast 2-hybrid data are excluded unless otherwise specified by the user. For the graphical output produced by the script, vertices are clustered by community membership as determined by random walk community detection.[32]

4.2.6. Extensible Set Enrichment Analysis (xSEA)

xSEA utilizes the enrichment algorithm developed for the Broad Institute's popular gene set enrichment analysis (GSEA) tool. GSEA was conceived as a tool for microarray data analysis to detect coordinated expression changes spread across groups of functionally related genes. Like GSEA, xSEA calculates an enrichment score measuring synchronized changes in abundance across multiple features as a function of treatment group, with a major difference being that the input of xSEA consists of experimental measurements of phosphopeptide abundance and sequence- or site-specific annotations of phosphorylated proteins. The steps of xSEA are broadly summarized as follows:

(1) *Pre-process quantitative comparison data.* Prior to running xSEA, the user selects two conditions to be compared from the xSEA tabbed menu. The quantitative input to xSEA is a tab-delimited file containing each peptide and the peak areas associated with the two conditions selected. A unique identifier is assigned to each peptide consisting of the parent protein's UniProt identifier

followed by an underscore and the single-letter amino acid codes and site numbers of any phosphorylated residues contained within the peptide.

(2) *Generate custom phosphorylation site ontologies*. As an additional input to xSEA, the user selects which annotation resource to use. Depending on the resource selected, the user may also specify thresholds on the confidence level of the annotations to be considered (e.g. z score of Scansite entries). Custom ontologies mapping phosphopeptides in the current dataset to their relevant annotations are then constructed. A peptide is mapped to a site-level annotation (an annotation concerning a single site, e.g. Scansite, PhosphoNET, PhosphoSitePlus) if there an instance of that annotation pertaining to one or more phosphorylated residues in the peptide. To facilitate site-level annotation enrichment analysis using Scansite and PhosphoNET, relational databases of site annotations built from batch analyses of the entire proteome are queried locally. A peptide is mapped to subsequence-level annotation (an annotation of features that span multiple amino acids, e.g. InterPro, PFAM) if there is an instance of that annotation pertaining to a region of the protein the peptide is derived from, and the peptide overlaps that region.

(3) *Calculate enrichment scores.* Before an enrichment score can be calculated, the phosphopeptides are first organized in an ordered list ranked by the ratio of phosphopeptide abundances measured in the two conditions of interest. For each ontology, the algorithm initializes a running-sum statistic to 0, then iterates through the ranked list sequentially. With each step, a check is performed to see if the current peptide is in the ontology under consideration. If it is, the statistic is incremented by a degree weighted by the magnitude of the ratio. If the current peptide is not in the ontology, the enrichment score is decreased. The

enrichment score is the maximum deviation from zero achieved by the running-sum statistic during its walk through the list.

(4) *Determine significance of enrichment score.* To obtain a statistical estimate of the significance of the enrichment score, xSEA uses a competitive null hypothesis approach[33] wherein a user-specified number of arbitrary ontologies are generated, with each consisting of the same number of phosphopeptides as the original ontology only chosen at random. By repeating the enrichment score calculation for each arbitrary ontology, a null distribution of enrichment scores can be built, and subsequently used to estimate the probability of the actual enrichment score calculated for the real ontology.

4.3. Results and Discussion

4.3.1. Overview of HTAPP and Peptide Depot

HTAPP is designed to assist with each step of the LC-MS experimental workflow from the time of acquisition all the way to the analysis and presentation of results. The typical lifecycle of data in HTAPP can be broadly divided into three phases (Figure 4.1). In the post-acquisition processing phase, the raw LC-MS data typically undergoes database searching, validation of MSMS spectral assignments, and quantitation. To achieve autonomous execution of each of these steps, HTAPP employs a suite of custom Java applications to automate the submission of user data to each subsequent post-acquisition process, requiring only that the user provide basic parameters for his or her analysis prior to data collection. For added efficiency, these applications are easily configurable by end users to run in a distributed computing environment, allowing for simultaneous processing of multiple samples spread over a cluster. In the second phase, LC-MS runs are archived and annotated in the relational database application Peptide Depot, which handles the comparison, visualization, and biological analysis of user data.[20] In this phase, multiple LC-MS runs can be collated into a quantitative

comparison reflecting the structure and design of the experiment. Comparisons in Peptide Depot allow users to compare and contrast multiple experimental conditions by combining quantitative data from multiple replicates, time points, and/or treatment groups into a single multidimensional data structure represented by its own graphical interface (Figure 4.2A). This customizable interface allows users to browse knowledge associated with each peptide drawn from curated external sources as well as visualize the collated quantitative data in the form of annotated heatmaps (Figure 4.2B,C).[20] Once a quantitative comparison is built, the user can access the software's integrated suite of analytical tools that assist with the third phase of the data lifecycle. Recent development on Peptide Depot has focused on expanding the breadth of analytical tools available within the platform, and particularly those designed with phosphoproteomic applications in mind.

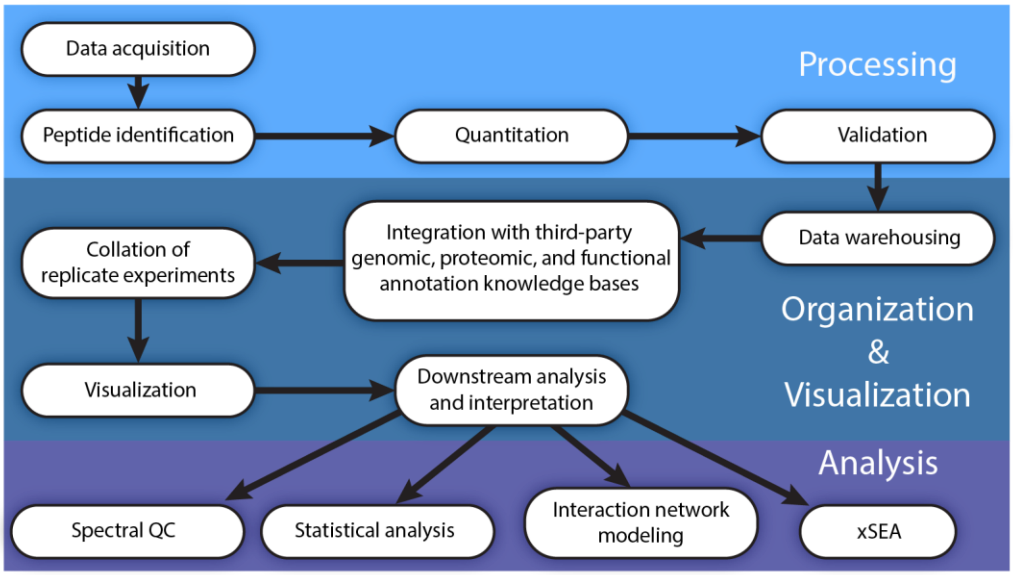


Figure 4.1: Data lifecycle in HTAPP. A typical proteomic workflow carried out in HTAPP occurs in three phases, each supported by different software modules. Post-acquisition tasks are automated to allow high-throughput and standardized processing of samples. Subsequently, data is stored in Peptide Depot for ease of browsing, annotation, visualization, and comparison of proteomic results. After collating multiple LC-MS runs into a quantitative comparison, further data analysis by a toolbox of integrated software can be performed within Peptide Depot.

4.3.2. Efficient Automatic and Manual Curation of Raw Data

The volume of data generated by contemporary LC-MS instruments in a single experiment makes thorough manual curation of spectral data logistically infeasible. As a result,

most proteomics researchers forgo manual inspection of individual SICs, or at best spot check their experiments to ensure an acceptable level of spectral quality. To empower researchers to make informed decisions about data quality, Peptide Depot now includes features to alert the user to the presence of potentially problematic spectra and facilitate the manual inspection of this data, including replicate data if present. During quantitation, HTAPP calculates 11 quality control (QC) characteristics that describe the shape, symmetry, and smoothness of selected-ion chromatogram (SIC) peaks either from label-free or stable isotope labeling experiments (see Section 4.2, Materials and Method, for details). Subsequently, all metrics are recorded in Peptide Depot, where they can be queried from a custom layout or exported in comparison reports.

Within a quantitative comparison, the user can visualize the variability of QC metrics in a heatmap form (Figure 4.2E). These heatmaps allow the user to efficiently identify outliers in the underlying data at a glance. Similar to the abundance heatmaps produced in Peptide Depot (Figure 4.2B),[34] QC metric heatmaps describe spectral variability in two dimensions, as they display information about both the average values of the SIC metrics across heatmap columns (blue-red scale heatmaps in 4.2E) as well as variance among collated replicates within a single column (black-orange scale heatmap in 4.2E). Users who are particularly concerned with variability of one or more specific QC metrics can add automated heatmap annotations to flag data points with CVs exceeding a user-specified threshold (Figure 4.2B, underneath first heatmap column). If a potentially problematic heatmap square is identified, clicking on the square launches a stand-alone chromatogram exploration tool showing the SIC of each quantified replicate (Figure 4.2D).

4.3.3. Statistical Methods and Visualization for High-Throughput Proteomic Datasets

Peptide Depot comes equipped with a toolbox of statistical methods for assessing differential abundance in quantitative comparisons. For experimental designs with a single factor of interest and two groups, users may apply a t-test and correct the resulting *p* values for

multiple hypothesis testing. Additionally, both ANOVA and Tukey's range test are available to users working with higher dimensional datasets. To provide the necessary versatility to support both labeled and label-free LC-MS data while preserving statistical power, statistical modules available in Peptide Depot can be configured for paired or unpaired data. Access to statistical

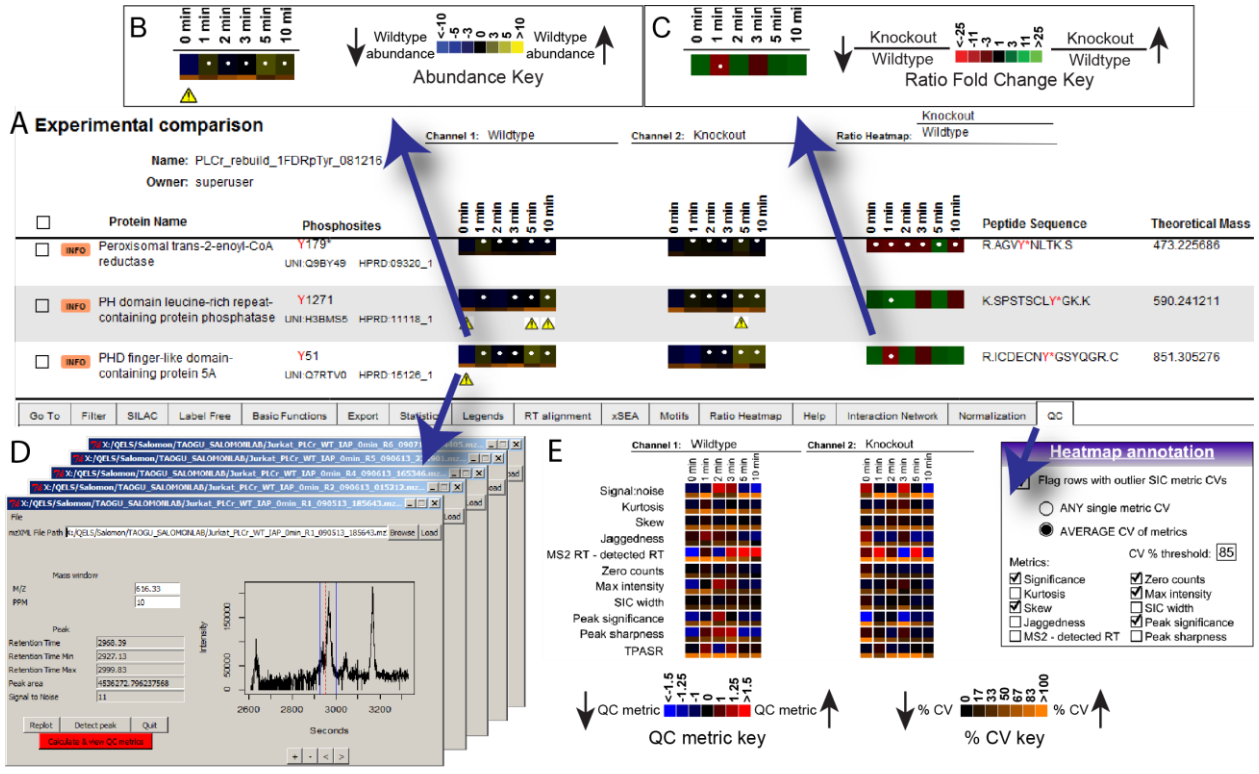


Figure 4.2: Curation and visualization of quantitative data in Peptide Depot. A) Comparison view of time series data featuring two conditions. After constructing a comparison, the quantitative data can be visualized in the layout shown using two types of heatmap displays. B) Blue-yellow abundance heatmaps indicating the peptide abundance observed at each timepoint. Black squares represent a peptide abundance equal to the geometric mean for that peptide across all time points. Yellow represents peptide abundance levels above the average, whereas blue corresponds to an abundance below the average. A separate abundance heatmap is generated for each treatment group. Black-orange heatmap bars underneath the blue-yellow heatmaps indicate the coefficient of variation (CV) of the peptide abundances of the collated replicates. White dots indicate a statistically significant difference ($q < 0.05$) between the given timepoint and minimum abundance in the row. C) Red-green ratio heatmaps indicating the ratio of peptide abundances between two user-specified conditions at each time point. Black signifies no change (i.e. ratio of 1). Red represents reduced abundance in knockout cells, while green represents elevated abundance in knockout cells. White dots on ratio heatmap squares indicate a statistically significant difference ($q < 0.05$) in the comparison between the two conditions compared for that time point. D) HTAPP's stand-alone chromatogram browser. Clicking on squares within a comparison heatmap launches the browser to reveal the quantified SICs underlying that data point. The red line indicates the position of the corresponding MSMS scan, while blue lines indicate the algorithm-detected peak boundaries. E) Chromatogram quality control menu and metric heatmaps. (Left) Blue-red heatmaps indicating the fold change of the QC metrics compared to the average value in the row. Black-orange heatmap bars underneath the blue-red heatmaps indicate the CV of the QC metrics of the collated replicates. (Right) Menu for configuring QC annotation of abundance heatmaps. The user specifies one or more QC parameters of interest and a CV threshold. Yellow caution signs in the main display (A) indicate a heatmap square with a CV higher than the user-defined threshold for the selected parameters.

analysis tools is granted via FileMaker layouts on the tabbed menu underneath the heatmap display (Figure 3A), while the statistical back-end of Peptide Depot is implemented in Java and R.

After a statistical test is performed, Peptide Depot provides the user with flexible visualization of results. To better integrate the resulting *q* values with peptide abundance data, Peptide Depot can annotate heatmap squares to indicate statistically significant results at a user-defined threshold (white dots in Figure 4.2A-C). For comparisons in which a Tukey post-hoc test was performed following ANOVA, users may instead opt to view their results in an alternate format (Figure 4.3B) that makes pairwise differences amongst multiple samples easily interpretable at a glance.

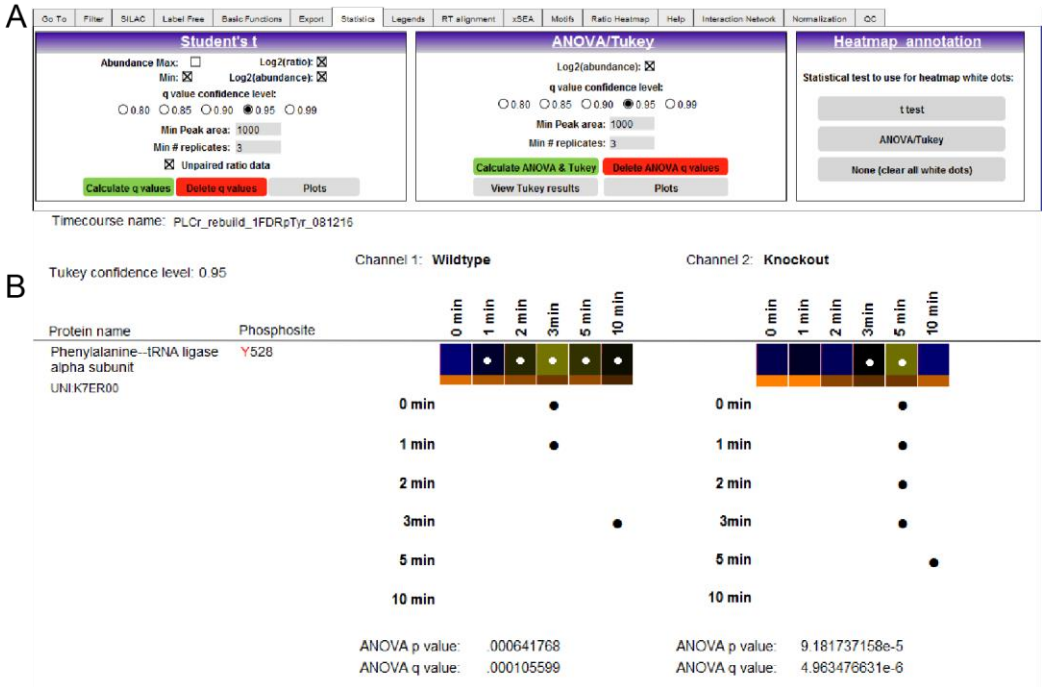


Figure 4.3: Visualization of statistical results. A) Tabbed analysis menu with statistics tab selected. After a statistical test is performed, annotation may be added to the heatmap to indicate data points where the null hypothesis was rejected at the user-specified confidence level. In (B), heatmap squares with *q* values < 0.05 are indicated with white dots. B) Alternate layout for viewing pairwise Tukey results. Black dots in the 6x6 matrix indicate significant differences in means between the intersecting timepoints.

4.3.4. Streamlined Network Analysis

The use of networks as a paradigm for modeling signaling pathways has proven useful for a variety of applications, including the identification of key signaling hubs in cancer pathways[35-38] and prediction of protein-protein interactions.[39-42] When coupled with experimental expression or abundance data, network models are useful in hypothesis generation, as they enable the researcher to annotate species in his or her dataset by their proximity to the experimentally perturbed or stimulated node in the graph, or identify changes in the network that occur adjacent to signaling hubs. In workflows employing affinity purification follow by LC-MS to identify protein complexes, partitioning peptide lists by protein interaction distance is a useful means of distinguishing known interactors from potentially novel ones. The current release of Peptide Depot enables users to generate and visualize tissue-specific proteomic network models directly from quantitative comparisons. This functionality is implemented by custom R and FileMaker scripts, and draws on Peptide Depot's tight integration with the HPRD and STRING protein-protein interaction databases. In addition to the required basic parameters for building the graph, the user may also specify a list of proteins to be used as starting points for calculating interaction distances within the network, or automatically populate this list from a set of annotations within Peptide Depot (e.g. all proteins predicted by Scansite to contain a domain of interest).

After the user specifies the required inputs, a script implemented in R generates the requested interaction network (Figure 4.4). In these networks, vertices represent proteins, while the edges signify interactions between them. As current protein-protein interaction databases are likely to contain a non-trivial number of false positives,[43] acceptable types of experimental evidence supporting each interaction can be specified by the user prior to network construction. For each protein in the dataset, the script calculates and stores the number of edges that separate it from the designated starting points as well as other properties of the network. These distance calculations are imported back into Peptide Depot, where they are archived in the

comparison. Subsequently, these distance calculations are treated as annotations in Peptide Depot, and can be used as filters to select peptides according to their parent protein's location in the network or in conjunction with other analytical tools incorporated into the software, as discussed below.

4.3.5. xSEA: Extensible Set Enrichment Analysis for Sequence-Level Features

Prediction of relative kinase activity from large phosphoproteomic datasets remains a challenging problem in the study of signal transduction. While the availability of bioinformatics databases that can assist in the identification of differential kinase activity, such as those detailing PTM and motif data, has increased,[44, 45] these resources are useful only to the extent that they can be utilized by non-bioinformatics experts to examine large datasets. To

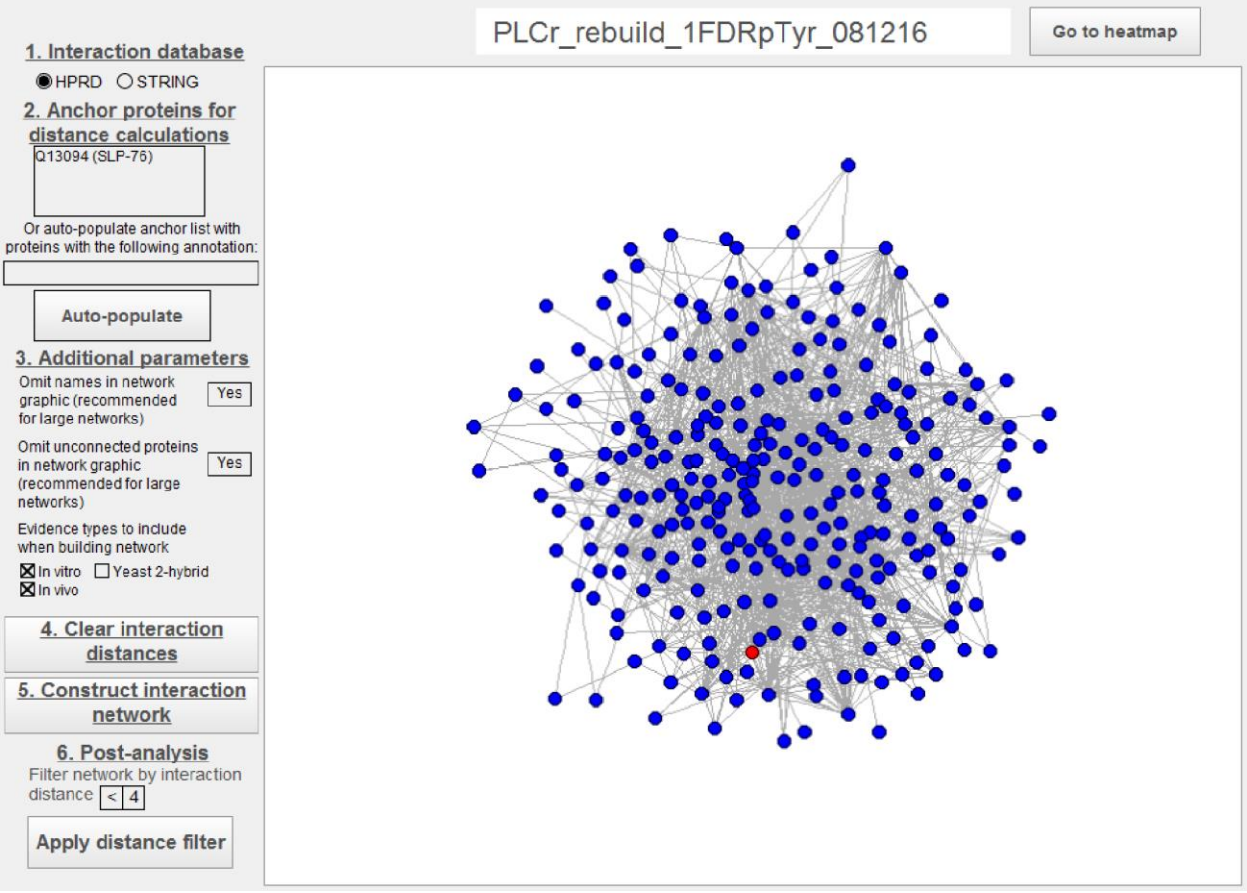


Figure 4.4: Interaction network layout and example graph. (Left) Network analysis main menu. (Right) User-generated interaction network from a timecourse dataset. The vertex in red represents SLP-76, the specified anchor protein in the network. Vertices in blue represent proteins in the input with ≤ 3 interactions separating them from SLP-76.

assist users in discovering potentially interesting patterns of differential phosphorylation, the latest release of Peptide Depot introduces a new analytical method designed to take advantage of available site- and sequence-specific annotations, called extensible set enrichment analysis (xSEA). Similar to the gene set enrichment analysis (GSEA)[46] method, xSEA identifies data points with a shared biological annotation that exhibit co-directional changes in abundance, and is particularly useful for detecting biological signals manifested by small, coordinated changes among many data points. However, xSEA differs fundamentally from gene- or protein-based enrichment analyses in the nature of the ontologies it tests for enrichment. In traditional GSEA, an ontology is a categorized list of genes or proteins that share a common feature,[46-48] as typified by the widely used GO and KEGG categories.[35, 49] In contrast, xSEA analyzes lists of phosphorylated peptides to detect trends that point to changes in kinase or phosphatase activities, and uses ontologies describing features of individual amino acid sites.

The construction of phosphosite ontologies for xSEA is straightforward in a relational database platform incorporating motif- and site-level annotations and relating them to experimentally observed peptides. In Peptide Depot, xSEA is accessible from the tabbed menu after a quantitative comparison is constructed. From there, the user can specify the parameters for the analysis, including the two conditions to be compared, the database or databases to use for constructing ontologies, thresholds on the size of the ontologies, the number of permutations to perform when estimating the null distribution of enrichment scores, and preferences regarding the presentation of the output. At present, Peptide Depot provides native support for building site-specific ontologies from Scansite, PhosphoSitePlus, PhosphoNET, PFAM, and InterPro, which are cached locally to facilitate the analysis. Custom ontology sources can also be imported as tab-delimited text files and saved for later use. Additionally, *de novo* ontologies may also be generated using other analytical tools within Peptide Depot. For example, distance calculations generated by Peptide Depot's interaction network analysis can subsequently be used in conjunction with xSEA to identify enriched phosphorylation of proteins within signaling

complexes. The xSEA pipeline utilizes custom Filemaker and Perl scripts to preprocess the data, generate all necessary input files, and launch a stand-alone client to perform the analysis (Figure 4.6).[46]

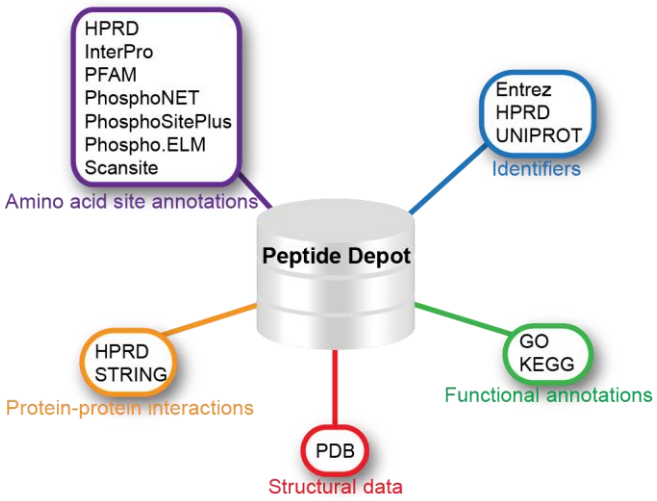


Figure 4.5: Overview of third-party knowledge bases natively integrated in Peptide Depot grouped by content.

4.3.6. Evaluation of Peptide Depot to Investigate Early Phosphoproteomic Signaling from the T Cell Receptor

To illustrate how the features of Peptide Depot come together to allow for rich analysis and interpretation of phosphoproteomics datasets, we used Peptide Depot to investigate a short time course following T cell receptor stimulation of the PLC-γ1-deficient Jgamma1 cell line and its PLC-γ1-reconstituted counterpart, Jgamma1.WT (discussed in Chapter 2). The experimental workflow is based on previously published work. Briefly, Jgamma1 and Jgamma1.WT cells were treated with OKT3/4 to induce signaling pathways downstream of the T cell receptor, and lysates were collected at six timepoints following stimulation. Prior to label-free LC-MS analysis, the proteomic samples were immunoaffinity purified to enrich for phosphotyrosine-containing peptides.

After collation of replicate experiments and assembly of the LC-MS runs into a quantitative comparison in Peptide Depot, a total of 2602 unique phosphopeptides derived from

1148 unique proteins were identified at 1% FDR. Quantitation and subsequent statistical testing of the comparison data in Peptide Depot revealed 1931 peptides exhibiting differential abundance (q<0.01) between the two cell lines at one or more timepoints in the series.

To better understand the biological consequences of PLC-γ1's absence in early T cell activation, we applied xSEA to look for functionally related and coregulated phosphotyrosine sites at each timepoint using custom ontologies constructed from InterPro annotations. In total, 16 distinct ontologies showed enrichment (q < 0.05) at one or more timepoints (Supplemental Tables 4.1 and 4.2). Intriguingly, the set of differentially enriched ontologies includes a number of ontologies representing critical kinase motifs downstream of the T cell receptor. xSEA correctly identified global decreases in phosphorylation of receptor subunits as well as protein kinase C at multiple timepoints in Jgamma1 (q < 0.01). These results are in concordance with previous studies demonstrating impaired signaling downstream of the T cell receptor in general, and downstream of Ca2+ and diacylglycerol in particular, in PLC-γ1-deficient T cells.[50] Conversely, xSEA identified increased phosphorylation of Btk/Tec-family substrates in Jgamma1 (q < 0.05). Enhanced phosphorylation of these substrates is likely mediated by Btk/Tec-family kinase Itk, which itself exhibited increased phosphorylation on activating residue Tyr^{512}[51] and may be subject to negative regulation by PLC-γ1 via clearance of the PIP_3 precursor PIP_2. Taken together, the enrichments identified by xSEA both recapitulate established T cell signaling mechanisms and provide promising avenues for future exploration.

4.4. Conclusions

Increasingly, the analysis and interpretation of large LC-MS datasets requires proficiency in a variety of computational tools and the engineering of complex pipelines to integrate them. HTAPP and Peptide Depot represent an ongoing effort to reduce the burden on researchers and expedite proteomics workflows through the automation of routine data processing tasks as well as the distillation of both popular and novel modes of analysis into a single all-in-one

software solution. To the best of our knowledge, Peptide Depot is the first free software package for LC-MS data management built around a relational database-centric architecture. This

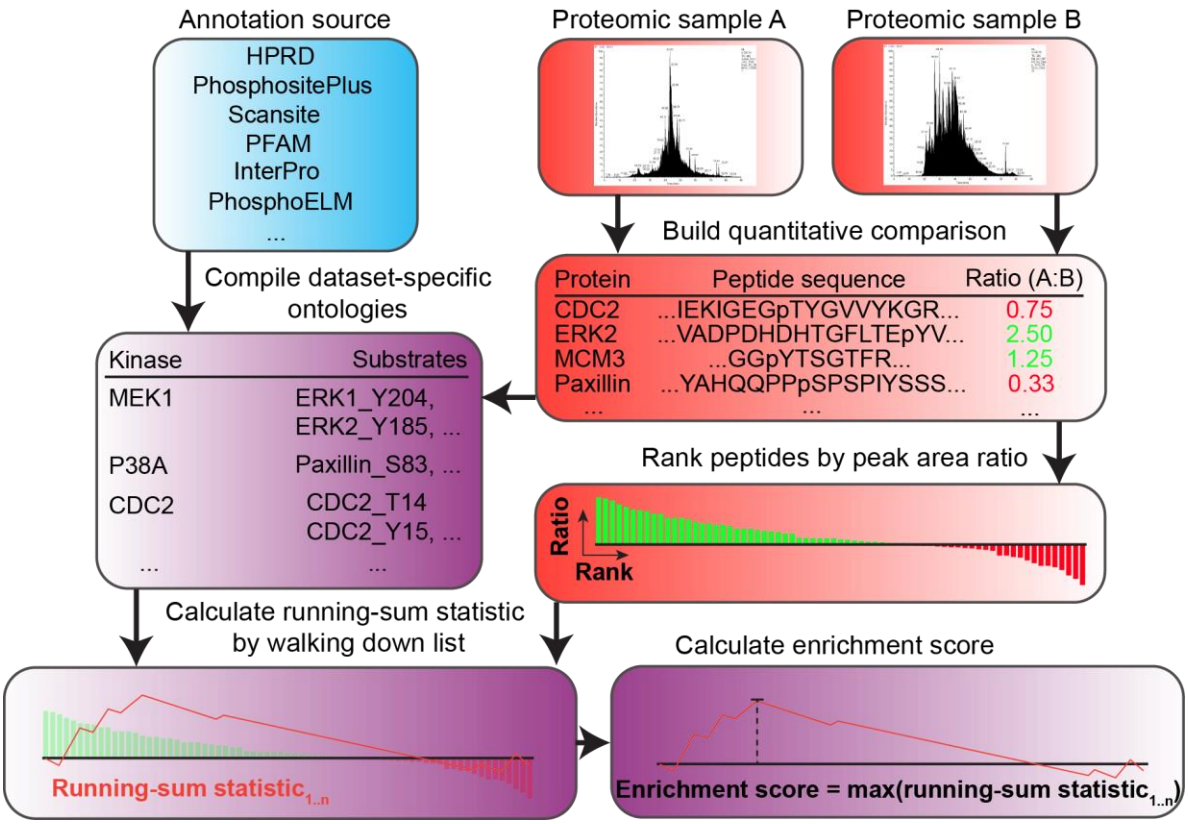


Figure 4.6: xSEA analysis workflow. The input to xSEA, a list of observed peptides ranked by peak abundance ratios in two conditions and a custom ontology list mapping observed peptides to shared annotations, is generated within Peptide Depot. For ontology, a running-sum statistic is calculated by walking down the ranked list. The maximum value of this statistic becomes the enrichment score for the annotation. This score measures the tendency of peptides in the ontology to be clustered at either end of the ratio list, indicative of coordinated differential regulation of sites represented in the set.

underlying structure affords Peptide Depot unique advantages over other proteomics software platforms, including centralized data storage, allowance and management of concurrent access, and fast, seamless integration with other third-party databases. However, for researchers already committed to alternative proteomics software platforms, the data analysis paradigms defined by our testbed Peptide Depot software can be readily incorporated within other analysis packages owing to the software's strict adherence to standard proteomic data formats and use of community-supported R packages.

Recent innovations in Peptide Depot focus on expanding the software's resources for post-acquisition analysis of data. These new resources allow users to better validate the integrity and reproducibility of data, perform and visualize quantitative comparisons, and more easily explore the biological context of their proteomic results. With the enhanced integration of third-party bioinformatics databases annotating sequence- and site-level features as well as the introduction of xSEA, the software now affords the user a greater selection of tools for the interpretation and presentation of phosphoproteomic datasets.

Currently, the usefulness of xSEA is limited only by the availability of amino acid-level annotations of protein function. Moving forward, we anticipate that the specificity and coverage of subsequence-level and phosphorylation site-specific annotations will only increase as our understanding of the phosphoproteome deepens, and that the predictive capabilities of xSEA will improve concomitantly. Furthermore, future increases in the volume and availability of such annotations will only further highlight the necessity for a data management platform such as Peptide Depot to organize these annotations. We note as well that the software infrastructure and xSEA analytical method described here provide a template that can be easily adapted to study other post-translational modifications assayable by mass spectrometry. Extending the capabilities of Peptide Depot to other post-translationally modified proteomics data requires only basic reengineering of the Peptide Depot source code as well as the availability of an annotation source relevant to the post-translational feature of interest. Thus, the recent changes introduced in Peptide Depot should ensure the continued relevance and value of the software as the state of proteomics changes in years to come.

As traditional gene and protein set enrichment methods are insufficient for deriving biological meaning from phosphoproteomic datasets, so too are protein-protein interaction networks limited in their ability to accurately describe causal relations among phosphorylated residues due to their lack of site-level resolution. The further elucidation of substrate site

repertoires for known kinases will therefore bring with it opportunities to map the connectivity of the phosphoproteome with greater resolution. Recent bioinformatics efforts to model phosphorylation-driven signaling pathways have focused increasingly on network descriptions that incorporate amino acid site-specific data.[52-54] However, we are at present unaware of any existing tools that combine phosphosite-centric network analysis and visualization with a flexible data management platform. The implementation of phosphosite-centric interaction networking tools is therefore a promising area of future development for Peptide Depot, as such a feature could build on the existing framework for network modeling and site-specific knowledge base integration.

The future evolution of Peptide Depot and HTAPP may also bring with it significant changes to the infrastructure underlying the software. The increasing demands on computational resources and infrastructural maintenance for comprehensive LC-MS data analysis have inspired a handful of recent cloud-based solutions,[55, 56] and this trend may continue as the requirements of modern data pipelines grow. The design of HTAPP and Peptide Depot already translates readily to a cloud-centric deployment, and we expect this flexibility to be crucial as the software evolves to keep pace with the challenges posed by new workflows.

4.5. References

1. Hsieh, E.J., et al., *Effects of column and gradient lengths on peak capacity and peptide identification in nanoflow LC-MS/MS of complex proteomic samples.* J Am Soc Mass Spectrom, 2013. **24**(1): p. 148-53.
2. Boersema, P.J., et al., *Evaluation and optimization of ZIC-HILIC-RP as an alternative MudPIT strategy.* J Proteome Res, 2007. **6**(3): p. 937-46.
3. Schirmer, E.C., J.R. Yates, 3rd, and L. Gerace, *MudPIT: A powerful proteomics tool for discovery.* Discov Med, 2003. **3**(18): p. 38-9.
4. Michalski, A., et al., *Mass spectrometry-based proteomics using Q Exactive, a high-performance benchtop quadrupole Orbitrap mass spectrometer.* Mol Cell Proteomics, 2011. **10**(9): p. M111 011015.
5. Andrews, G.L., et al., *Performance characteristics of a new hybrid quadrupole time-of-flight tandem mass spectrometer (TripleTOF 5600).* Anal Chem, 2011. **83**(13): p. 5442-6.
6. Rost, H.L., et al., *OpenMS: a flexible open-source software platform for mass spectrometry data analysis.* Nat Methods, 2016. **13**(9): p. 741-8.

7. Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nat Methods, 2016. **13**(9): p. 731-40.
8. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.
9. MacLean, B., et al., *Skyline: an open source document editor for creating and analyzing targeted proteomics experiments.* Bioinformatics, 2010. **26**(7): p. 966-8.
10. Gonzalez-Galarza, F.F., et al., *A critical appraisal of techniques, software packages, and standards for quantitative proteomic analysis.* OMICS, 2012. **16**(9): p. 431-42.
11. Lee, D.C., A.R. Jones, and S.J. Hubbard, *Computational phosphoproteomics: from identification to localization.* Proteomics, 2015. **15**(5-6): p. 950-63.
12. Ren, J., et al., *Computational analysis of phosphoproteomics: progresses and perspectives.* Curr Protein Pept Sci, 2011. **12**(7): p. 591-601.
13. Hu, H., K. Khatri, and J. Zaia, *Algorithms and design strategies towards automated glycoproteomics analysis.* Mass Spectrom Rev, 2016.
14. Goel, R., et al., *Human Protein Reference Database and Human Proteinpedia as resources for phosphoproteome analysis.* Mol Biosyst, 2012. **8**(2): p. 453-63.
15. Hornbeck, P.V., et al., *PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.* Nucleic Acids Res, 2012. **40**(Database issue): p. D261-70.
16. Bateman, A., et al., *The Pfam protein families database.* Nucleic Acids Res, 2004. **32**(Database issue): p. D138-41.
17. Mitchell, A., et al., *The InterPro protein families database: the classification resource after 15 years.* Nucleic Acids Res, 2015. **43**(Database issue): p. D213-21.
18. Obenauer, J.C., L.C. Cantley, and M.B. Yaffe, *Scansite 2.0: Proteome-wide prediction of cell signaling interactions using short sequence motifs.* Nucleic Acids Res, 2003. **31**(13): p. 3635-41.
19. Yu, K. and A.R. Salomon, *HTAPP: high-throughput autonomous proteomic pipeline.* Proteomics, 2010. **10**(11): p. 2113-22.
20. Yu, K. and A.R. Salomon, *PeptideDepot: flexible relational database for visual analysis of quantitative proteomic data and integration of existing protein information.* Proteomics, 2009. **9**(23): p. 5350-8.
21. Martinez-Bartolome, S., P.A. Binz, and J.P. Albar, *The Minimal Information about a Proteomics Experiment (MIAPE) from the Proteomics Standards Initiative.* Methods Mol Biol, 2014. **1072**: p. 765-80.
22. Taylor, C.F., et al., *The minimum information about a proteomics experiment (MIAPE).* Nat Biotechnol, 2007. **25**(8): p. 887-93.
23. Smith, C.A., et al., *XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification.* Anal Chem, 2006. **78**(3): p. 779-87.
24. Zhang, W. and P.X. Zhao, *Quality evaluation of extracted ion chromatograms and chromatographic peaks in liquid chromatography/mass spectrometry-based metabolomics data.* BMC Bioinformatics, 2014. **15 Suppl 11**: p. S5.
25. Lange, E., et al., *High-accuracy peak picking of proteomics data using wavelet techniques.* Pac Symp Biocomput, 2006: p. 243-54.
26. Zhang, W., et al., *MET-COFEA: a liquid chromatography/mass spectrometry data processing platform for metabolite compound feature extraction and annotation.* Anal Chem, 2014. **86**(13): p. 6245-53.
27. Dabney A, S.J., *qvalue: Q-value estimation for false discovery rate control.* 2015.

28. Benjamini, Y. and Y. Hochberg, *Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing.* Journal of the Royal Statistical Society Series B-Methodological, 1995. **57**(1): p. 289-300.
29. Storey, J.D., *The positive false discovery rate: A Bayesian interpretation and the q-value.* Annals of Statistics, 2003. **31**(6): p. 2013-2035.
30. Storey, J.D. and R. Tibshirani, *Statistical significance for genomewide studies.* Proc Natl Acad Sci U S A, 2003. **100**(16): p. 9440-5.
31. Gabor Csardi, T.N., *The igraph software package for complex network research.* InterJournal, 2006: p. 1695.
32. Pons, P. and M. Latapy, *Computing communities in large networks using random walks.* Computer and Information Sciences - Iscis 2005, Proceedings, 2005. **3733**: p. 284-293.
33. Khatri, P., M. Sirota, and A.J. Butte, *Ten years of pathway analysis: current approaches and outstanding challenges.* PLoS Comput Biol, 2012. **8**(2): p. e1002375.
34. Nguyen, V., et al., *A new approach for quantitative phosphoproteomic dissection of signaling pathways applied to T cell receptor activation.* Mol Cell Proteomics, 2009. **8**(11): p. 2418-31.
35. Ashburner, M., et al., *Gene ontology: tool for the unification of biology. The Gene Ontology Consortium.* Nat Genet, 2000. **25**(1): p. 25-9.
36. Grechkin, M., et al., *Identifying Network Perturbation in Cancer.* PLoS Comput Biol, 2016. **12**(5): p. e1004888.
37. Kreeger, P.K. and D.A. Lauffenburger, *Cancer systems biology: a network modeling perspective.* Carcinogenesis, 2010. **31**(1): p. 2-8.
38. Vandin, F., E. Upfal, and B.J. Raphael, *Algorithms for detecting significantly mutated pathways in cancer.* J Comput Biol, 2011. **18**(3): p. 507-22.
39. Altaf-Ul-Amin, M., et al., *Development and implementation of an algorithm for detection of protein complexes in large interaction networks.* BMC Bioinformatics, 2006. **7**: p. 207.
40. Bader, G.D. and C.W. Hogue, *An automated method for finding molecular complexes in large protein interaction networks.* BMC Bioinformatics, 2003. **4**: p. 2.
41. Srihari, S. and H.W. Leong, *A survey of computational methods for protein complex prediction from protein interaction networks.* J Bioinform Comput Biol, 2013. **11**(2): p. 1230002.
42. Srihari, S., et al., *Methods for protein complex prediction and their contributions towards understanding the organisation, function and dynamics of complexes.* FEBS Lett, 2015. **589**(19 Pt A): p. 2590-602.
43. Sprinzak, E., S. Sattath, and H. Margalit, *How reliable are experimental protein-protein interaction data?* J Mol Biol, 2003. **327**(5): p. 919-23.
44. Hornbeck, P.V., et al., *PhosphoSitePlus, 2014: mutations, PTMs and recalibrations.* Nucleic Acids Res, 2015. **43**(Database issue): p. D512-20.
45. *Scansite 4.0*. 2017; Available from: https://scansite4.mit.edu/4.0/#news.
46. Subramanian, A., et al., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proc Natl Acad Sci U S A, 2005. **102**(43): p. 15545-50.
47. Bard, J.B. and S.Y. Rhee, *Ontologies in biology: design, applications and future challenges.* Nat Rev Genet, 2004. **5**(3): p. 213-22.
48. Schulze-Kremer, S., *Ontologies for molecular biology and bioinformatics.* In Silico Biol, 2002. **2**(3): p. 179-93.
49. Kanehisa, M. and S. Goto, *KEGG: kyoto encyclopedia of genes and genomes.* Nucleic Acids Res, 2000. **28**(1): p. 27-30.

50. Irvin, B.J., et al., *Pleiotropic contributions of phospholipase C-gamma1 (PLC-gamma1) to T-cell antigen receptor-mediated signaling: reconstitution studies of a PLC-gamma1-deficient Jurkat T-cell line.* Mol Cell Biol, 2000. **20**(24): p. 9149-61.
51. Heyeck, S.D., et al., *Lck phosphorylates the activation loop tyrosine of the Itk kinase domain and activates Itk kinase activity.* J Biol Chem, 1997. **272**(40): p. 25401-8.
52. Ciaccio, M.F., et al., *The DIONESUS algorithm provides scalable and accurate reconstruction of dynamic phosphoproteomic networks to reveal new drug targets.* Integrative Biology, 2015. **7**(7): p. 776-791.
53. Raaijmakers, L.M., et al., *PhosphoPath: Visualization of Phosphosite-centric Dynamics in Temporal Molecular Networks.* Journal of Proteome Research, 2015. **14**(10): p. 4332-4341.
54. Narushima, Y., et al., *Quantitative phosphoproteomics-based molecular network description for high-resolution kinase-substrate interactome analysis.* Bioinformatics, 2016. **32**(14): p. 2083-2088.
55. Muth, T., et al., *ProteoCloud: a full-featured open source proteomics cloud computing pipeline.* J Proteomics, 2013. **88**: p. 104-8.
56. Slagel, J., et al., *Processing shotgun proteomics data on the Amazon cloud with the trans-proteomic pipeline.* Mol Cell Proteomics, 2015. **14**(2): p. 399-404.

4.6. Supplemental Material

Timepoint (min)	Ontology	Size	Normalized enrichment score	FDR q value
1	PROTEIN KINASE C	7	1.8830146	0.001375
1	PROTEIN KINASE C, DELTA/EPSILON/ETA/THETA TYPES	6	1.879476	6.88E-04
1	ZINC FINGER, CCCH-TYPE	10	1.7840426	0.00410031
1	SERINE/THREONINE PROTEIN KINASE	34	1.6593746	0.01522914
2	PHOSPHORYLATED IMMUNORECEPTOR SIGNALING ITAM	16	2.5047252	0
2	TYROSINE PROTEIN KINASE, SYK/ZAP-70	34	1.8178103	0.02348964
3	PHOSPHORYLATED IMMUNORECEPTOR SIGNALING ITAM	17	-2.420132	0
3	TYROSINE PROTEIN KINASE, SYK/ZAP-70	35	-1.9251502	0.00240476
5	PHOSPHORYLATED IMMUNORECEPTOR SIGNALING ITAM	17	2.7215116	0

10	PHOSPHORYLATED IMMUNORECEPTOR SIGNALING ITAM	16	2.4148872	0

Supplemental Table 4.1: Phosphopeptide ontologies derived from the InterPro sequence classification database exhibiting significant enrichment (*q* < 0.05) in PLC-γ1-reconstituted T cells using xSEA

Timepoint (min)	Ontology	Size	Normalized enrichment score	FDR q value
0	IMMUNOGLOBULIN-LIKE	9	-2.0201712	0
0	TYROSINE PROTEIN KINASE, SYK/ZAP-70	25	-1.9289101	9.36E-04
0	IMMUNOGLOBULIN I-SET	6	-1.9043673	8.70E-04
0	CD80-LIKE, IMMUNOGLOBULIN C2-SET	6	-1.8876024	0.001282574
0	IMMUNOGLOBULIN V-SET	6	-1.8775325	0.001191913
0	TYROSINE PROTEIN KINASE	41	-1.845038	0.001970793
0	IMMUNOGLOBULIN SUBTYPE 2	10	-1.7812253	0.006165983
0	IMMUNOGLOBULIN SUBTYPE	10	-1.7640222	0.006768156
0	PDZ/DHR/GLGF	11	-1.6553116	0.027119864
2	PLECKSTRIN HOMOLOGY-TYPE	7	-1.6547865	0.035562515
2	TEC/BTK	5	-1.6205082	0.026528226
3	PLECKSTRIN HOMOLOGY-TYPE	7	1.78797	0.009013287
3	TEC/BTK	5	1.7158732	0.015991662
5	TEC/BTK	5	-1.7852194	0.013491596
5	PLECKSTRIN HOMOLOGY-TYPE	7	-1.721547	0.015762588
10	PLECKSTRIN HOMOLOGY-TYPE	7	-1.5870255	0.040116817
10	PROTEIN KINASE C	7	-1.5697749	0.038455207

Supplemental Table 4.2: Phosphopeptide ontologies derived from the InterPro sequence classification database exhibiting significant enrichment (*q* < 0.05) in PLC-γ1-deficent T cells using xSEA

Chapter 5

Summary and Future Directions

5.1. Summary of This Thesis

The central challenges facing the adaptive immune system in general, and T cells in particular, are two-fold: discrimination of self- from non-self peptides, and the generation of a robust yet measured immune response. To address these challenges, T cell activation proceeds through a complex molecular dialogue in which many proteins serve dualistic roles, both transducing signals that converge on the nucleus and adjusting the cellular rheostat that generates those signals. In this system, specificity and sensitivity are emergent properties achieved through the balance between activating and inhibitory feedback signals to ensure an immune response of appropriate duration and magnitude.

Understanding the complex topology of T cell signaling networks will require a deep integration of transcriptional, translational, and post-translational knowledge. As such, the quantitative insights into proteins and their modifications provided by LC-MS have proven to be tremendous assets for deciphering the signaling pathways underlying T cell activation. In Chapter 2 of this work, we used LC-MS and bottom-up phosphoproteomics to identify a previously unknown role for PLC-γ1 in the maintenance of Lck activation. This work builds on previous data obtained by our group and others demonstrating that Lck is the convergence point of multiple modes of pathway feedback, and that proteins comprising the SLP-76 signalosome are a critical part of this feedback circuit.

While the studies presented here advance our knowledge of the molecular architecture governing T cell activation, they also highlight the importance of sequencing depth and highly reproducible quantitation for analysis of the phosphoproteome. To accompany our studies of TCR signaling, we also investigated HPLC-compatible optimizations for enhancing peak capacity and sequencing depth. Through this work, we demonstrated that method development on standard HPLC instrumentation could deliver both enhanced peak resolution and greater sensitivity to detect phosphorylation-mediated signaling events. Applying this improved workflow

to a T cell stimulation time course yielded substantial increases in the number of dynamic phosphorylation events observed in early TCR signaling. Thus, as the insights into T cell signaling presented in Chapter 2 were made possible by innovations delivering better qualitative and quantitative data, so too will new possibilities and discoveries in future phosphoproteomic studies be afforded by the methodological improvements detailed in Chapter 3.

While deeper sequencing of the proteome is an important goal, increasing the quality and quantity of large datasets without concomitantly expanding bandwidth to process and analyze such data only ensures more severe bottlenecks in analysis pipelines. The bioinformatics platform discussed in Chapter 4 catalyzes new discoveries from phosphoproteomic data in two important ways. First, deep integration of bioinformatic tools into the Peptide Depot platform expedites proteomics researchers' workflows by combining curation, annotation, quantitation, and visualization of data without the need for pipeline engineering on the part of the end user. Additionally, the pathway enrichment tool presented here, xSEA, assists users in generating new hypotheses about activation or suppression of kinase activity from their quantitative comparisons. Implementation of this new mode of analysis in Peptide Depot both ensures ease of use and facilitates construction of the required phosphosite-specific ontologies. With the software innovations discussed in this work, we aim to remove barriers that can impede efficient and insightful analysis of phosphoproteomics data.

5.2. Future Directions

5.2.1. Defining the Mechanism of PLC-γ1-Mediated Lck Regulation

In Chapter 2, we demonstrated that phosphorylation of the TCR subunits in both the basal and active states is dependent on PLC-γ1. Noting that Lck's activation did not appear to be suppressed in PLC-γ1-deficient cells, we concluded that targeting of Lck is influenced by PLC-γ1, and proposed that this change in substrate affinity is due to increased association of Lck with either SLP-76 or TSAd (Figure 2.7). Looking forward, further studies will be required to

experimentally validate these mechanisms. As the enhanced association between Lck and SLP-76 is hypothesized to involve an interaction between the proteins' SH3 and proline-rich regions respectively, targeted disruption of these domains will be informative for parsing the regulatory consequences of this interaction. The previously described SH3-inactivating mutant of Lck, Lck W97A, may play a critical role in future studies of PLC-γ1-dependent Lck localization to SLP-76, as it increases Lck's presence at the TCR and reduces phosphorylation of SLP-76-associated proteins.[1, 2] Likewise, the non-phosphorylatable Lck Y192F mutant may serve a similar role in determining if PLC-γ1 deficiency enhances the association of Lck to TSAd. Deletion of Lck and reconstitution with the aforementioned domain mutants in a *PLCG1-/-* background could provide valuable information for deciphering the precise mechanism of Lck substrate regulation by PLC-γ1.

PLC-γ1 has previously been shown to possess signaling capabilities independent of the protein's phospholipase activity.[3, 4] Intriguingly, these alternative signaling capabilities appear to be dependent on PLC-γ1's SH3 domain, which mediates the enzyme's localization to SLP-76. The requirement for phospholipase activity in regulating Lck substrate phosphorylation is not established by our study, but warrants investigation. A future TCR stimulation study comparing the Jgamma1 line to PLC-γ1(ΔSH3)- and lipase-inactive PLC-γ1-reconstituted variants would allow us to better assess PLC-γ1's role in TCR signaling and clarify important details of our existing model.

5.2.2. Expanding Sequencing Depth of the T Cell Phosphoproteome

The HPLC protocol optimization presented in Chapter 3 outlines a useful strategy for maximizing yields of assignable peaks while balancing both acquisition time and equipment costs. Based on our analysis, doubling our gradient time would only improve our yield of PSMs by roughly 5,000 (Supplemental Figure 3.1). However, further improvements in phosphoproteomic sequencing depth will be achieved as new technologies for phosphopeptide

enrichment and mass spectrometry come to maturity. Promising new methods for immunoaffinity purification such as optimized mixtures of anti-phosphotyrosine antibodies[5] and SH2-based ‘superbinder’ proteins[6] have already demonstrably enhanced sequencing depth of the tyrosine phosphoproteome, and combinations of these approaches may conceivably achieve still deeper coverage. Likewise, technical innovations improving ion delivery and instrument scan times while maintaining high resolution will enhance sensitivity and allow for the detection of lower abundance species not currently visible in LC-MS screens. Such innovations will advance phosphoproteomics towards its landmark goal of achieving a qualitative description of all phosphorylation events.

5.2.3. Enhancing Quantitative Sensitivity of Phosphoproteomic Studies

While qualitative descriptions of the phosphoproteome yield tremendous insight into potential regulatory mechanisms, our understanding of phosphorylation-driven signaling events depends critically on our ability to reproducibly quantify phosphopeptides. The study presented in Chapter 3 reveals that enhancing peak resolution improves the reproducibility of peptide quantitation. Further gains in reproducibility can be achieved through the adoption of quantitation methods such as TMT, which limits the variability associated with parallel processing of multiple samples and thus affords greater statistical sensitivity to detect differential abundance than label-free quantitation.[7] Recent phosphoproteomics work applying TMT to study long TCR stimulation time courses has already revealed differential regulation at previously uncharacterized serine and threonine residues during late TCR signaling.[8] Further application of this technology to assay phosphotyrosine peptides at earlier timepoints would likewise provide a valuable perspective on immediate TCR-proximal signaling events.

Although data collection practices are often guided by the maxim that more is better, knowing how and when to discard data may prove critical for improving the reproducibility of peptide quantitation. The QC metrics discussed in Chapter 4 provide an unbiased means to

assess features that may differentiate useful, quantifiable SICs from technical noise. Although our preliminary investigations suggest that some peaks exhibiting significant deviation from the ideal Gaussian shape are nonetheless consistently reproducible, further work is required to determine if applying thresholds on these metrics can lead to better discrimination between signal and noise, and thereby enhance statistical sensitivity.

5.2.4. Expediting Pathway Discovery from Phosphoproteomic Data

As described in Chapter 4, the goal of Peptide Depot is to allow users to interact with data more efficiently by providing an integrated suite of tools serving the needs of common phosphoproteomics workflows. To fulfill this role, it needs to constantly adapt to the new approaches and modes of analysis required by users. Future updates to Peptide Depot include support for TMT quantitation as well normalization of relative phosphopeptide abundance to estimated protein levels. As discussed above, TMT provides highly reproducible quantitation that will improve our capability to detect differential phosphorylation. Equally important, however, is our capability to accurately assess the stoichiometry of phosphorylation in studies in which data from both phosphorylation-enriched and non-enriched samples are available. This will be especially critical for TCR stimulation studies involving later time points, as changes in protein levels must be taken into consideration when examining later post-stimulation time points. The capability to analyze the normalized phosphoproteome will therefore enable us to extend our studies of TCR signaling and investigate signaling mechanisms that are only evident on longer timescales.

5.2.5. Improving Phosphosite Ontologies

In our preliminary studies using xSEA in Chapter 4, we demonstrated that the algorithm successfully detects perturbations of both known and potentially novel TCR regulatory pathways using custom ontologies constructed from PFAM and InterPro annotations. Currently, the upstream kinases for the majority of sites observed in our data are either unknown or not

systematized in a database, while our understanding of phosphatase specificity lags even further behind. The detection capabilities of xSEA are therefore limited by the current state of available annotation data, and we anticipate that this tool's usefulness will scale with the increasing coverage of functional annotation of PTM events. To this end, MS-assisted high-throughput strategies such as BioID[9] and on-bead peptide library screening[10-12] show promise for assaying *in vivo* interactions and *in vitro* substrate specificity, respectively. Increases in instrument speed and sensitivity will further extend the usefulness of these approaches, allowing researchers to more efficiently identify candidate substrates from high-throughput data that can then be further investigated by more targeted methods. Ideally, greater depth and breadth of substrate specificity data will ultimately give rise to more structured ontologies, in which effector-substrate interactions are further classified by cellular and developmental context.

5.2.6. Engineering Better Connected Platforms for Open Science

At present, the bioinformatics needs of many proteomics workflows are served by a variety of proprietary and open-source software solutions combined into in-house pipelines. As such, there is a significant advantage to tools that are both easily accessible and interface efficiently with existing resources. The current release of Peptide Depot is designed for on-premise use in a research lab's computing cluster. However, a future evolution of the software towards a cloud-based deployment would provide greater accessibility and significantly streamline the platform's deployment and maintenance. Other potential directions for our software include integration with proteomics data repositories to better assist users in utilizing and contributing to these resources. This would enable users not only to more easily satisfy publication requirements by uploading data directly to public databases, but additionally to reprocess existing public data sets using familiar tools. Taken together, these changes would

streamline the deployment and use of our platform, and enable researchers to more actively contribute to and benefit from the collective knowledge of the proteomics community.

5.3. References

1. Rudd, M.L., A. Tua-Smith, and D.B. Straus, *Lck SH3 domain function is required for T-cell receptor signals regulating thymocyte development.* Mol Cell Biol, 2006. **26**(21): p. 7892-900.
2. Li, M., et al., *The SH3 domain of Lck modulates T-cell receptor-dependent activation of extracellular signal-regulated kinase through activation of Raf-1.* Mol Cell Biol, 2008. **28**(2): p. 630-41.
3. Huang, P.S., et al., *An SH3 domain is required for the mitogenic activity of microinjected phospholipase C-gamma 1.* FEBS Lett, 1995. **358**(3): p. 287-92.
4. Xie, Z., et al., *The SH3 domain, but not the catalytic domain, is required for phospholipase C-gamma1 to mediate epidermal growth factor-induced mitogenesis.* Biochem Biophys Res Commun, 2010. **398**(4): p. 719-22.
5. van der Mijn, J.C., et al., *Evaluation of different phospho-tyrosine antibodies for label-free phosphoproteomics.* J Proteomics, 2015. **127**(Pt B): p. 259-63.
6. Bian, Y., et al., *Ultra-deep tyrosine phosphoproteomics enabled by a phosphotyrosine superbinder.* Nat Chem Biol, 2016. **12**(11): p. 959-966.
7. Ahrne, E., et al., *Evaluation and Improvement of Quantification Accuracy in Isobaric Mass Tag-Based Protein Quantification Experiments.* J Proteome Res, 2016. **15**(8): p. 2537-47.
8. Nguyen, T.D., et al., *The phosphoproteome of human Jurkat T cell clones upon costimulation with anti-CD3/anti-CD28 antibodies.* J Proteomics, 2016. **131**: p. 190-8.
9. Varnaite, R. and S.A. MacNeill, *Meet the neighbors: Mapping local protein interactomes by proximity-dependent labeling with BioID.* Proteomics, 2016. **16**(19): p. 2503-2518.
10. Garaud, M. and D. Pei, *Substrate profiling of protein tyrosine phosphatase PTP1B by screening a combinatorial peptide library.* J Am Chem Soc, 2007. **129**(17): p. 5366-7.
11. Chen, X., et al., *Determination of the substrate specificity of protein-tyrosine phosphatase TULA-2 and identification of Syk as a TULA-2 substrate.* J Biol Chem, 2010. **285**(41): p. 31268-76.
12. Xue, L. and W.A. Tao, *Current technologies to identify protein kinase substrates in high throughput.* Front Biol (Beijing), 2013. **8**(2): p. 216-227.