UNIVERSITY OF CALIFORNIA,
IRVINE

Pipeline Development for Hippo Pathway Related Protein Interactome Analysis

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biological Sciences

by

Rebecca Elizabeth Vargas

Dissertation Committee:
Associate Professor Wenqi Wang, Chair
Professor Aimee L. Edinger
Associate Professor Scott Atwood

2023

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my PI, Wenqi Wang, for his unwavering guidance, support, and mentorship throughout my journey in graduate school. Wenqi's profound knowledge, expertise, and commitment to excellence have been instrumental in shaping my academic and professional development. I am truly blessed to have had the opportunity to work under Wenqi's direction. His dedication to advancing scientific knowledge have inspired me to strive for excellence in my own work. He has consistently encouraged intellectual curiosity, critical thinking, pushing me beyond my limits and helping me discover my true potential.

I would also like to thank my current and past committee members Dr. David Fruman, Dr. Aimee Edinger, Dr. Scott Atwood, and Dr. Xiaoyu Shi. I am grateful to these scientists for their technical expertise and all our personal conversations that have guided me through my graduate journey.

I am very thankful to all the past and current members of the Wang Lab. A big thank you to Han Han and Gayoung Seo for being the best post-doctoral lab 'sibilings' I could ask for. I am grateful for all you both have taught me and the fun moments you provided in the lab. Another big thank you to Bing Yang, our lab manager whose personality always made the lab environment a much sweeter place. I also want to thank the members of the Edinger lab who became my second lab family. A big shout out to past members- Vaishali Jayashankar and Brendan Finicle for making the lab environment so special during their time here.

I would also like to thank my parents, the two people who have pushed me to move forward in my education when this type of degree was not an option for them. I will forever be grateful for the encouragement they provided over the years. I would like to thank my sister for always being curious about the "potions" I generated in lab and keeping me company during my study nights. I would like to thank my husband, Nabil, for his incredible support throughout the last few years and being so patient when I would tell him I had 5 minutes of lab work when in fact I took 5 hours. I would like to also thank my daughter Hannah for being the reason I worked so hard to finish my degree. Her smile at the end of a long day in lab made it all worthwhile. I would also like to give a big thank you to my family in-law for always supporting me. Finally, I would like to thank my friends for always being so supportive. I could not have done this without you all.

I would like to acknowledge Biorender as many figures were generated using their premium services.

Chapter Two of this dissertation is a reprint of the material as it appears in Briefings in Bioinformatics, available from https://doi.org/10.1093/bib/bbad010 used with permission from Oxford University Press. The co-authors listed on this publication are Rebecca Elizabeth Kattan, Deena Ayesh, and Wenqi Wang.

Chapter Three of this dissertation is a reprint of the material as it appears in the EMBO Journal, available from https://doi.org/10.15252/embj.2019102406, used with permission from Life Science Alliance. The co-authors listed on this publication are Rebecca E Vargas, Vy Thuy Duong, Han Han, Albert Paul Ta, Yuxuan Chen, Shiji Zhao, Bing Yang, Gayoung Seo, Kimberly Chuc, Sunwoo Oh, Amal El Ali, Olga V Razorenova, Junjie Chen, Ray Luo, Xu Li, and Wenqi Wang. Wenqi Wang the co-author listed in this publication, directed, and supervised research which forms the basis for the thesis/dissertation.

Chapter Four of this dissertation is a reprint of the material as it appears in Molecular Cell Proteomics, available from https://doi.org/10.1016/j.mcpro.2022.100195, used with permission from the American Society for Biochemistry and Molecular Biology. The co-authors listed on this publication are Rebecca Elizabeth Kattan, Han Han, Gayoung Seo, Bing Yang, Yongqi Lin Max Dotson, Stephanie Pham, Yahya Menely, Wenqi Wang Wenqi Wang the co-author listed in this publication, directed, and supervised research which forms the basis for the thesis/dissertation.

Financial support was provided by the University of California, Irvine, R01 Diversity Supplement Grant (R01GM126048), NIH Initiative for Maximizing Student Development (IMSD) Fellowship (GM055246), GAANN fellowship (P200A220015).

VITA

Rebecca Elizabeth Vargas

EDUCATION

2018 – 2023	University of California, Irvine School of Biological Sciences	PhD	Biological Sciences
2015 - 2017	California State University, Los Angeles School of Biological Sciences	Masters	Chemistry
2010 – 2015	California State University, Fullerton School of Biological Sciences	B.S.	Biochemistry

RESEARCH EXPERIENCE

2018 - 2023 **PhD student**, University of California, Irvine, Wang Lab

- Built a user-friendly model for the analysis of large proteomic datasets, leading to a first author review manuscript.
- Discovered novel binding specificity between Hippo pathway components, leading to a co-first author manuscript.
- Discovered novel interactors for the Phospholipase D superfamily and Phosphatic acid, leading to a first author manuscript.
- Collaborated with others in the Wang lab on studies related to the different regulations of the Hippo pathway, leading to multiple co-author publications.

2015 - 2017 **Graduate Researcher**, California State University, Los Angeles, Wen Lab

- Investigated the role of Anti-freeze proteins and their thermal inhibition properties

2010 - 2015 **Undergraduate Researcher**, California State University, Fullerton, Linder Lab

- Contributed to a publication of the discovery of the mechanistic details of copper uptake from mammalian cells.
- Investigated the role of an unknown copper transporters in cells.

MANUSCRIPTS UNDER PEER REVIEW, SUBMITTED, OR IN LATE-STAGE PREPARATION

1. Seo, G., Yu,C., Han, H., Xing, L., **Kattan, R.,** Kizhedathu,A., Yang, B., Buckle, A., Huang, L., and Wang, W. * (2023) The Hippo pathway noncanonically drives autophagy and cell survival in response to energy stress. ***(In revision)***
2. **Kattan, R**., Han, H., Seo, G., Ujuar, N., Mahieu, A., Ayesh,D., Yang, B.,and Wang, W (2023) Demonstrating the Huge Biological Role of the FAT family. ***(In preparation)***

PUBLICATIONS

1. **Kattan, R**., Ayesh, D., Wang,W. (2023) A user-friendly guideline for analyzing interactome-related mass spectrometry data to study protein-protein interaction networks in cells. ***Briefings in Bioinformatics.***
2. **Kattan, R**., Han, H., Seo, G., Yang, B., Wang, W. (2022) Defining the protein-protein interaction network of the human protein Phospholipase D superfamily and Phosphatidic Acid. ***Molecular and Cellular Proteomics.***
3. Han, H., Nakaoka, H., Hofman, L., Zhou, J., Yu, C., Zeng, L., Nan, J., Seo, G., **Vargas, R.,** Yang, B., Qi, R., Bardwell, L., Cho, Ken., Huang, L., Luo, R., Warrior, R., Wang, W. (2022) "The Hippo pathway kinases LATS1 and LATS2 attenuate cellular responses to heavy metals through phosphorylating MTF1". ***Nature Cell Biology.***
4. Turvey M.*, Gabriel K.*, Lee W., Lau C., Taulbee J., Kim J., Chen C., **Vargas R.,** Pham J., Majumdar S., Collins P., and Weiss G. (2022) "Single-Molecule Taq DNA polymerase Dynamics at Multiple Domains and Native Temperatures." ***Science Advances.***
5. **Vargas R**, Wang W. (2020) Significance of long non-coding RNA AGPG for the metabolism of esophageal cancer. ***Cancer Communications.***
6. Seo, G., Han, H., **Vargas, R.**,Yang B.,Li, X., & Wang, W. (2020) Proteomic analysis of the human MAP4Ks protein interaction network identifies STRIPAK complex component STRN4 as a YAP activator in the Hippo pathway. ***Cell Reports.***
7. Chen, Y., Han, H., Seo, G., **Vargas, R.**, Yang, B., Chuc, K., Zhao, H., & Wang, W. (2020) Systematic analysis of the Hippo pathway organization and oncogenic alteration in evolution. ***Scientific Reports.***
8. **Vargas, R*.**, Duong, Vy*., Han, H*., Ta, A., Chen, Y., Zhao, S., Yang, B., Seo, G., Chuc, K., Oh, S., Razorenova, O., Chen, J., Luo, R., Li, X., & Wang, W. (2020). Molecular basis of the WW domain binding specificity for the Hippo pathway. ***The EMBO Journal***. *co-first authors
9. Han, H., **Vargas, R**., Seo, G., & Wang, W. (2018). Phosphatidic acid: a lipid regulator of the Hippo pathway. ***Molecular & Cellular Oncology.***
10. Ramos, D., Mar, D., Ishida, M., **Vargas, R.**, Gaite, M., Montgomery, A., & Linder, M. C. (2016). Mechanism of Copper Uptake from Blood Plasma Ceruloplasmin by Mammalian Cells. ***PloS one***

TEACHING & PROFESSIONAL EXPERIENCE

Summer 2020/2021 **Summer Minority Science Program (MSP) Graduate Journal Club Instructor**

- In this role, I led journal club for incoming graduate students in the MSP summer program.

2020, 2021, 2023 **Teaching Assistant** to Advanced Regenerative Medicine (D133), Cell Biology (D103), and Developmental and Cell Biology Lab (D111L) courses at UCI

- In this role, I led class discussions, lab teaching, and office hours.

SELECT HONORS & AWARDS

2022 and 2023 **Brian G. Atwood '74 and Lynee H. Edminister Graduate Studies Endowment award**

- This award is the second biggest award offered to UCI PhD students for excellence in research, providing $5,000.

2021 - 2023 **ARCS Foundation Scholar**

- This award is intended to recognize and reward "academically superior doctoral students exhibiting outstanding promise as scientists, researchers, and leaders." Provided $10,000.

2019 **Joseph H. Stephens Memorial Fellowship Award**

- This award is for "displaying excellence in the field of biochemistry and molecular biology" & provided $1,000

ABSTRACT OF THE DISSERTATION

Pipeline Development for Hippo Pathway Related Protein Interactome Analysis

By

Rebecca Elizabeth Vargas

Doctor of Philosophy in Biological Sciences

University of California, Irvine, 2023

Associate Professor Wenqi Wang, Chair

This thesis presents a user-friendly pipeline for analyzing data from evolving mass spectrometry techniques, developed in response to the growing volume and complexity of data in the field of proteomics. While offering a comprehensive review of current methods for isolating and identifying interacting proteins using MS analysis, the pipeline also highlights areas for integrating other -omic resources to enhance data interpretation. This pipeline was utilized to identify novel regulators within the Hippo pathway, a central regulator of cell proliferation and organ size. A unique 9-amino acid sequence was identified, vital for interactions between WW-PxY motif containing proteins in the Hippo pathway. Furthermore, the pipeline aided in the discovery of a new YAP inhibitor, Syntaxin binding protein 4 (STXBP4), harboring this conserved sequence. The pipeline was also employed to investigate potential regulators in the PLD-PA-Hippo axis. Although no Hippo pathway regulators were found within the PLD family, the analysis unveiled significant Hippo-independent interactions. Among the interactors of six

PLD members, the E3 ubiquitin-protein ligase, PJA2, was found to bind specifically to PLD1, an oncogene. Furthermore, Phosphatidic Acid (PA) was shown to bind, co-localize, and regulate Sphingosine Kinase1 (SPHK1), linked to cell growth, proliferation, and survival. This thesis underscores the importance of continually updating bioinformatic resources to keep up with advances in data generation techniques, thus allowing the extraction of relevant biological information from intricate datasets. The studies driven by this pipeline signify just the inception, and further tools, such as additional disease -omic resources, can be integrated for comprehensive insights into many biological processes.

INTRODUCTION

Proteomics: an approach for the identification of novel biomarkers:

Proteins are essential molecules that play key roles in virtually all cellular processes, including metabolism, signaling, and gene expression. Proteomics involves the identification and quantification of proteins in biological samples, as well as the analysis of their post-translational modifications, interactions, and subcellular localization. Proteomics is a rapidly advancing field within the broader scope of molecular biology that aims to understand the structure, function, and interactions of proteins within complex biological systems. The field has become increasingly important in recent years due to the vast potential for protein-based biomarkers in disease diagnosis and drug discovery. The field of proteomics has seen significant evolution since its early days. Proteomic studies used to consist of two-dimensional gel electrophoresis, which was the primary technique for protein separation, enabling scientists to separate complex mixtures of proteins based on their isoelectric point and molecular weight. As effective as it was at the time, this method was limited by its inability to detect low-abundance, highly acidic or basic, and hydrophobic proteins.

The arrival of mass spectrometry (MS) revolutionized proteomics, providing a more precise method for protein identification, quantitation, and characterization of post-translational modifications. Advances such as tandem MS and high-resolution MS have further improved the accuracy and depth of proteomic analysis. Over the years, proteomics has grown beyond the mere cataloging of protein species within a sample to include the exploration of dynamic processes such as protein synthesis, modification, turnover, and interaction networks. Advances in mass spectrometry and bioinformatics have enabled the development of high-throughput

proteomic techniques, allowing for the analysis of thousands of proteins simultaneously (Aebersold & Mann, 2016). Biomarkers are essential for early disease detection, prognosis, and treatment, and proteomics has enabled the identification of novel biomarkers that were previously undetectable. The use of proteomics in disease research has led to significant advances in our understanding of biology and has the potential to improve cancer diagnosis and treatment.

Hand-in-hand with proteomics is bioinformatics which combines biology to analyze and interpret large-scale biological data. Bioinformatics tools have played an integral role in managing, analyzing, and interpreting the voluminous and complex data generated in proteomic studies. It has become an essential tool for proteomics research, providing powerful methods and algorithms for data processing, analysis, and interpretation **(Figure 0.1)**. Bioinformatics tools and techniques have been used for protein identification, quantification, and post-translational modification analysis in proteomics research (Calderón-González *et al*, 2016). In addition to protein analysis, bioinformatics is also essential for the analysis of protein-protein interactions, networks, and pathways. These analyses help to understand the biological functions and interactions of proteins within a biological system and can identify potential drug targets and disease biomarkers.

Furthermore, the integration of multiple -omics data, such as proteomics, genomics, and transcriptomics, is essential for the understanding of complex biological systems, and bioinformatics provides methods and tools for the integration and analysis of these datasets(Haas *et al*, 2017). Overall, bioinformatics is a rapidly growing field that has become indispensable in

proteomics research, providing powerful tools and methods for data analysis, interpretation, and integration, and driving new discoveries in biology and medicine. With current developments in single-cell proteomics and integration of multi-omics data, the field continues to evolve, promising more comprehensive understanding of biological systems and disease mechanisms. Thus far, the integration of proteomics using a user-friendly MS analysis pipeline(Kattan *et al*, 2023) has elucidated novel regulators in various biological pathways widening the scope of biological processes and the Hippo pathway(Kattan *et al*, 2022b; Seo *et al*, 2020a; Vargas *et al*, 2020a). This thesis will explore the role of proteomics in the discovery of biomarker identification and elucidate the discoveries in relevance to Hippo and non-Hippo signaling regulations.

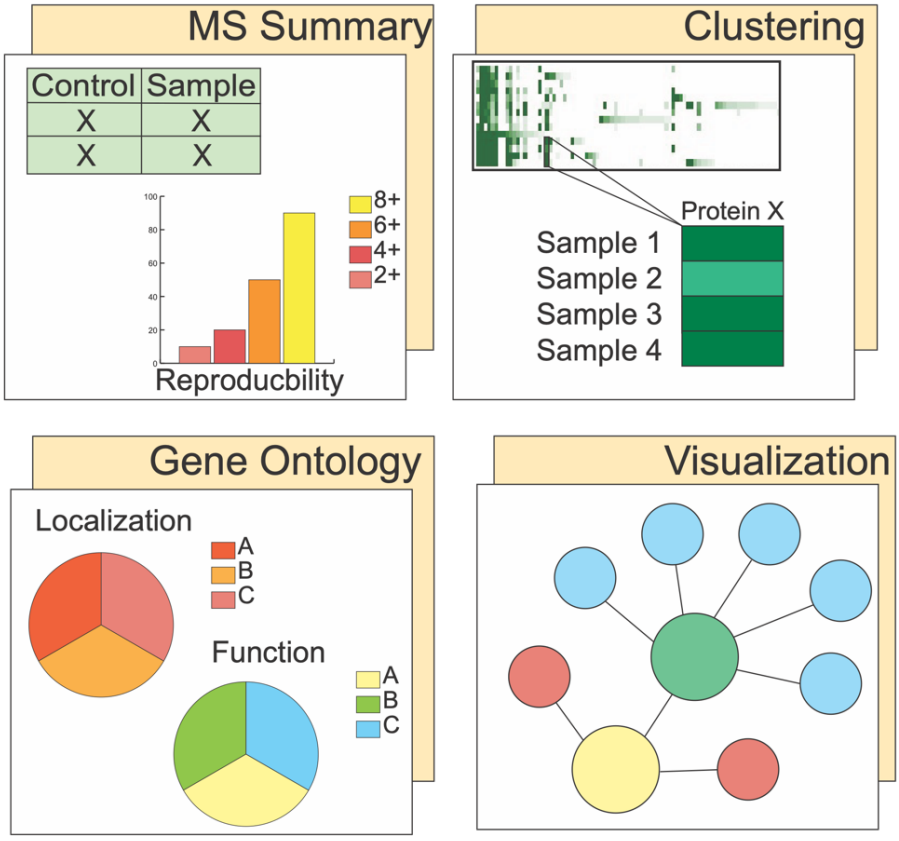


Figure 0.1: **Bioinformatics Analysis Overview.** Bioinformatics involves the analysis of mass spectrometry (MS) data to generate biologically meaningful figures to show protein overlap (clustering), biological relevance (Gene ontology) and biological network visualization.

Introduction to the mammalian Hippo pathway:

Proper organ size involves a complex array of mechanisms that must precisely regulate cell numbers during both organ development and regeneration. This tightly controlled process and its underlying mechanisms have long presented a challenge for researchers worldwide. While previous studies have proposed that organs possess information dictating their final size, the precise molecular mechanisms driving this phenomenon remain poorly understood. The discovery of the Hippo pathway highlights important mechanisms responsible for control of organ growth (Pan, 2007; Zhao et al., 2010a).The Hippo pathway is an evolutionarily conserved signaling pathway that plays a crucial role in controlling tissue growth and organ size in various organisms, including humans(Johnson & Halder, 2013; Pan, 2010b; Plouffe *et al*, 2015).

The pathway was initially identified in *Drosophila melanogaster*, where it regulates cell proliferation and apoptosis to control organ size during development(Harvey & Tapon, 2007; Xu *et al*, 1995). Overexpression of YAP, a downstream co-activator involved in the Hippo pathway, in the adult mouse liver mimics pathway inactivation in *Drosophila* and was revealed to cause a dramatic three- to four-fold increase in liver mass owing to increase of cell numbers (Camargo et al.,2007; Dong et al., 2007). Since then, studies have revealed that the Hippo pathway is also involved in other processes such as stem cell maintenance, tissue regeneration, and cancer development(Cai *et al*, 2010a; Heallen *et al*, 2011; Pobbati & Hong, 2013).

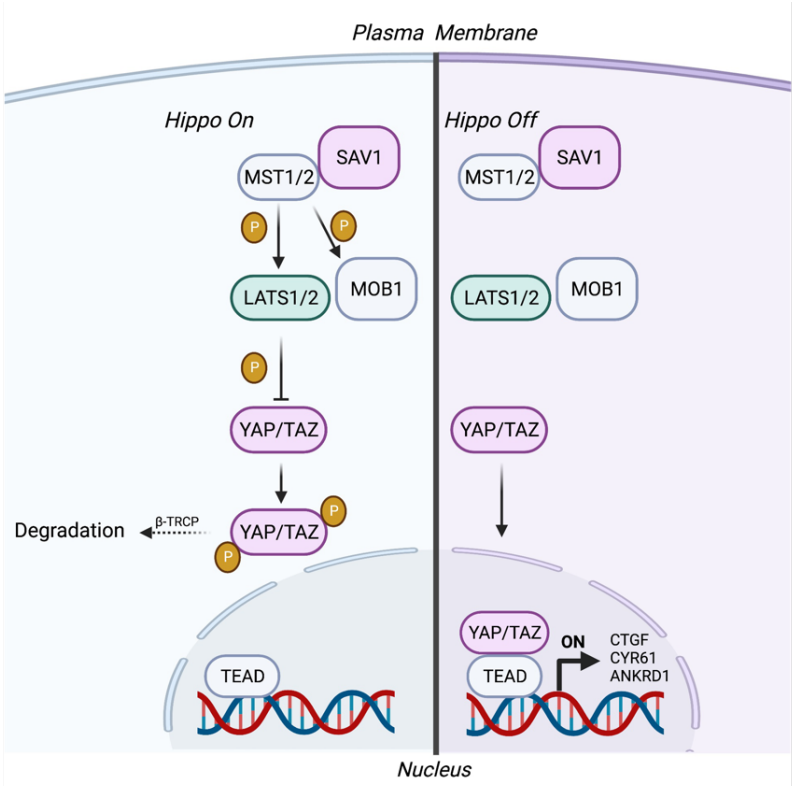


Figure 0.2. **Schematic representation of the Hippo pathway "on" and "off" states**. In the "on" state (left panel), the Hippo pathway is activated and YAP/TAZ are phosphorylated by the LATS kinase, which promotes their cytoplasmic retention and degradation. This leads to the inhibition of cell proliferation and survival, as well as the activation of apoptotic pathways. In the "off" state (right panel), the Hippo pathway is inactive and the transcriptional co-activators YAP/TAZ are translocated to the nucleus, where they interact with transcription factors to promote cell proliferation and survival.

The core components of the Hippo pathway are a series of protein kinases and transcription factors that interact with each other in a complex network of regulatory interactions **(Figure 0.2)**. In mammals, the Hippo pathway is initiated by a set of upstream signals, such as mechanical cues(Dupont *et al*, 2011), cell density ((Zhao *et al*, 2007), and growth factors(Haskins *et al*, 2014b), that activate a group of proteins called the Hippo kinases, which include MST1/2 (mammalian Ste20-like kinases 1 and 2) and LATS1/2 (large tumor suppressor kinases 1 and 2). These kinases then phosphorylate and activate a downstream transcription factor called YAP

(Yes-associated protein) and its paralog TAZ (transcriptional co-activator with nPDZ-binding motif), which are key effectors of the Hippo pathway. The phosphorylated YAP/TAZ can be recognized by 14-3-3 proteins, retained in the cytoplasm and eventually targeted by ß-TRCP E3 ligase complex for degradation. When the Hippo pathway is inactivated, unphosphorylated YAP/TAZ enter the nucleus, where they associate with TEAD ((TEA domain family member), transcriptional factors to promote the transcription of genes that are involved in proliferation and survival.

`

Physiological functions of the Hippo pathway:

From its discovery in organ size control in *Drosophila*, further physiological roles in humans including roles in development, tissue regeneration and diseases have been elucidated. It has been established that YAP/TAZ are key downstream effectors of the Hippo pathway that are important in mediating TEAD induced expression of genes involved in cell growth and proliferation(Zhao *et al*, 2008). The Hippo pathway also plays a role in development by affecting cell differentiation in tissues/organs. Upon the knockout of Mst1/2 in mice, dedifferentiation of acinar cells to ductal cells was demonstrated through a hyperactivation of YAP(Gao *et al*, 2013). Moreso, YAP has also been discovered to be required for differentiation of multiple types of tissues such as the airway of basal stem cells(Zhao *et al*, 2014), and myoblast differentiation (Nantie *et al*, 2018). The role of the Hippo pathway in tissue/organ regeneration suggests that YAP/TAZ is activated after damage various organs and tissues(Dey *et al*, 2020). An example, YAP levels in the intestinal epithelium are elevated during intestinal regeneration and upon YAP inactivation regeneration is compromised (Barry *et al*, 2013).

In cancer, aberrant activation of YAP and TAZ has been shown to promote tumor growth and metastasis by enhancing cell proliferation and survival ((Feng *et al*, 2014; Moroishi *et al*, 2015; Wang *et al*, 2010). Hippo pathway components YAP/TAZ have been identified as oncoproteins(Hall *et al*, 2010) while MST1/2 and LATS1/2 as tumor suppressors(Xia *et al*, 2002; Zhou *et al*, 2009). Various types of cancers have been related to the Hippo pathway such as uveal melanoma, mesothelioma, and ependymoma. In addition, the Hippo pathway has been shown to also influence the progression of cancer through metastasis by the control of LATS1/2(Takahashi *et al*, 2005) and YAP(Hsu *et al*, 2018). Understanding the mechanisms underlying the regulation of the Hippo pathway has become a focus of intense research, and novel therapeutic strategies targeting the pathway are being developed. The identification of key regulators and downstream targets of the pathway has led to the development of novel therapeutic strategies for the treatment of these diseases.

Protein-Protein Modulation within the Hippo pathway:

Protein-protein interactions (PPIs) are key components of the Hippo pathway, mediating the signaling cascade that ultimately leads to downstream effects. The Hippo signaling pathway is identified by the significant occurrence of WW domains (consisting of Tryptophan-Tryptophan amino acid sequences) found in the Hippo core kinase complex, the regulatory components located upstream, and the nuclear proteins situated downstream(Sudol & Harvey, 2010a). The presence of WW domains in the interactions between Hippo pathway components and other regulatory proteins offers a valuable chance to delve deeper into the molecular mechanisms involved and potentially uncover novel elements of the pathway.

WW domain is a small protein module that is characterized by the two tryptophan (W) residues separated by ~25 amino acids(Sudol *et al*, 1995c). The WW domain, which is a distinct protein module that can fold independently, has been shown to take on a structure consisting of a three-stranded β-sheet(Koepf *et al*, 1999). This domain is involved in facilitating interactions between proteins by binding to ligands that contain multiple proline residues and a minimal consensus sequence of PPXY, where X represents any amino acid and Y represents tyrosine. To date, almost 100 WW domains and about 1,900 PPXY motifs have been identified in human proteome network (Sudol & Harvey, 2010a). WW domains can be grouped into four categories based on their interactions with different peptide ligands. Group I WW domains bind to PY motifs, while Group II WW domains recognize PPLP motifs. Group III WW domains recognize PR motifs, and Group IV WW domains recognize (S/T)P motifs. Some WW domains, specifically Groups II and III, are capable of binding to multiple types of peptides, including PPLP- and PR-containing ones, as well as polyproline sequences containing glycine, methionine, or arginine. Due to their broad binding capabilities, these domains could be considered as a single group with versatile interactions (Ingham et al, 2005). Failure of their recognition is associated with human diseases such as Huntington's disease (Faber *et al*, 1998a; Passani *et al*, 2000a), Golabi-Ito-Hall Syndrome (Lubs *et al*, 2006a; Tapia *et al*, 2010a) ,muscular dystrophy (Bork & Sudol, 1994a; Rentschler *et al*, 1999b) and cancers (Chang *et al*, 2007b; Salah & Aqeilan, 2011c).

WW domain was firstly uncovered by characterizing the protein sequence of YAP, a key transcriptional co-activator downstream of the Hippo pathway(Pan, 2010b; Sudol *et al*, 1995b; Yu *et al*, 2015a). Many Hippo pathway components and regulators contain either WW domain or its proline-rich peptide ligand-"PPxY" highlighting the importance between WW domain

biochemistry and its associated cellular functions (Salah & Aqeilan, 2011b; Sudol, 2010a). YAP contains two WW domains whereas TAZ contains one WW domain used during interactions with many binding proteins that contain PPXY motifs (Liu *et al*, 2011). Within the nucleus, the WW domain of YAP/TAZ is essential for their interaction with a set of nuclear transcription factors and regulators that possess the PY motif playing a crucial role in the regulation of gene transcription(Ferrigno *et al*, 2002c; Haskins *et al*, 2014c). LATS1 kinase contains two PPXY motif (Hao *et al*, 2008a), which is important for its interaction with YAP/TAZ.

The PY motif of LATS1/2 is implicated in the phosphorylation of YAP/TAZ by LATS1/2 in the cytoplasm. This process involves the physical binding of several proteins containing the PY motif to the WW domain of YAP/TAZ, which facilitates the cytoplasmic translocation of YAP/TAZ(Sudol, 2011).Upstream Hippo pathway components, KIBRA contain two WW domains ((Kremerskothen *et al*, 2003; Tapon *et al*, 2002) and can negatively regulate YAP. (Wilson *et al*, 2014a). A number of extensive proteomic studies specifically identified a cluster of proteins containing the PY motif, such as LATS1/2, AMOTs, and PTPN14, as the interactors of the Hippo pathway's WW domain-containing components (Couzens *et al*, 2013a; Hauri *et al*, 2013a; Wang *et al*, 2014b). These findings highlight the specific binding affinity of the Hippo WW domain for these protein partners, although the precise mechanism underlying this interaction is not yet fully understood.

Regulation of the Hippo pathway in mammals:

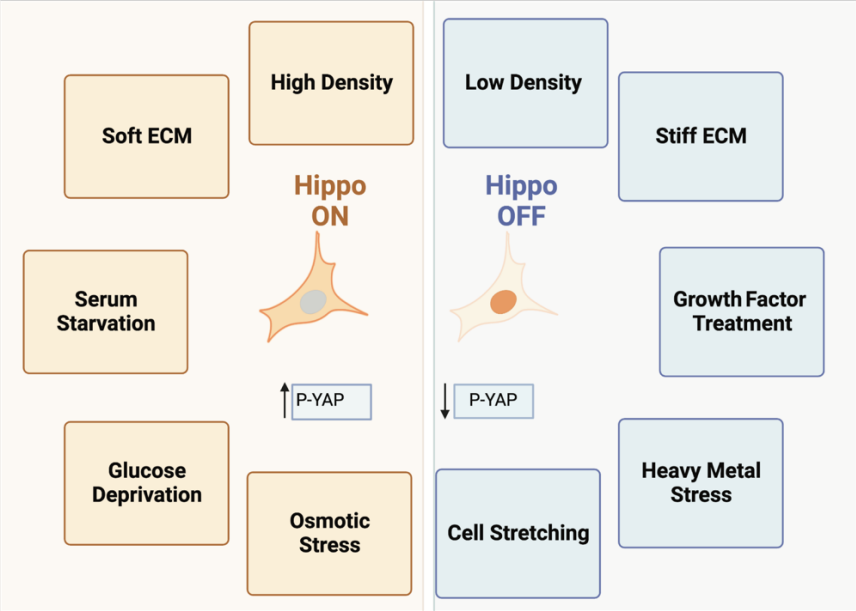


Figure 0.3: **Regulation of the Hippo pathway**. Various upstream signals turning the Hippo pathway ON (Left panel) or OFF (Right panel).

The Hippo signaling pathway is a complex system of communication with multiple parts that work together to carefully regulate the pathway. The main kinases of this pathway has been well studied in mammals. However, the specific molecular mechanisms that regulate this pathway have not been fully elucidated. Below is a summary of established Hippo pathway regulators as well as new regulators of the Hippo signaling pathway **(Figure 0.3)**.

Mechanotransduction regulation of the Hippo pathway:

The microenvironment of cells affects cell morphology and its cellular characteristics. The mechanism allowing cells to sense and engage towards mechanical cues is known as mechanotransduction which can modulate cellular characteristics such as proliferation, survival and migration (Hoffman *et al*, 2011). The way in which developing epithelial tissues interact with the extracellular matrix (ECM) is important for normal bodily functions, but it also plays a

significant role in the development of tumors and cancer metastasis. Cells are exposed to both biochemical and mechanical signals from their surroundings, including the ECM, nearby cells, and signaling molecules, which can impact their behavior and function. These signals are varied and complex, and occur *in vivo*, in living organisms. To maintain homeostasis, cells need to be capable of adjusting to mechanical forces and signals. This often leads to changes in cell behavior such as altered rates of cell division, programmed cell death, or movement (Lopez *et al*, 2008; Manninen, 2015).

Numerous mechanical factors have been recently elucidated to influence YAP/TAZ activity (Halder *et al*, 2012) such as mechanical stretching (Aragona *et al*, 2013), surface geometry (Wada *et al*, 2011),matrix rigidity(Dupont *et al.*, 2011; Thomasy *et al*, 2013), and cell density (Aragona *et al.*, 2013; Zhao *et al.*, 2007). Mechanical stretching is a phenomenon in which a force or tension is exerted on cells. This force can be transmitted through different cellular structures, such as the extracellular matrix, cell-cell contacts, and cytoskeleton, and can trigger various cellular responses. External physical stimuli such as pressure or tension can also cause mechanical stretching of cells. In certain disease conditions like cancer, abnormal accumulation of extracellular matrix components can cause tissue stiffening and change the mechanical properties leading to mechanical stretching. Manipulating the size and shape of the cell monolayer, YAP/TAZ activity is inhibited by F-actin-capping proteins during minimal mechanical forces. On the other hand, cells at the corners or edges show YAP/TAZ-dependent proliferation induced by cytoskeletal contractility(Aragona *et al.*, 2013). When cell geometry was manipulated on a single cell, the loss of cell-cell contact activated YAP/TAZ in flat spread cells however, YAP/TAZ were inactivated in round compact cells. Moreso, YAP/TAZ was

modulated upon the manipulation of matrix rigidity, where YAP/TAZ localized in the nucleus of cells seeded on stiff surfaces while YAP/TAZ was inactivated when seeded on soft surfaces (Dupont *et al.*, 2011; Wada *et al*, 2011).

Cell density influences cell-cell contact and cell morphology and have been established to play an essential role in the regulation of the Hippo signaling. At low cell densities, YAP/TAZ are localized to the nucleus driving the transcription of target genes involved in cell proliferation and survival. At high cell densities, the core kinase complex of the Hippo pathway phosphorylates YAP/TAZ leading to their cytoplasmic retention where they are later degraded (Zhao *et al.*, 2007). Additionally, cell attachment or detachment has been demonstrated to influence YAP/TAZ activity. Under normal conditions, when cells are detached from the ECM, anoikis occurs which a type of programmed cell death. However, in an oncogenic background, cells can develop a resistance to anoikis which can promote survival while circulating and migrating to other tissues leading to metastasis. During cell detachment the Hippo pathway is activated, leading to inhibition of the YAP resulting in anoikis (Zhao *et al*, 2012)

Extracellular ligands and cellular stress regulation of the Hippo pathway:

While the microenvironment affects the hippo pathway activation, growth factors have also been revealed to regulate YAP/TAZ through activation of membrane receptors and intracellular signaling pathways. G protein-coupled receptors (GPCRs) have been demonstrated to regulate the Hippo pathway by activating YAP. Serum has been demonstrated to activate YAP/TAZ in cultured cells where serum sphingosine-1-Phosphate (S1P), a ligand of GPCR, and lysophosphatidic acid (LPA) were specifically identified to inhibit LATS1/2 (Miller *et al*, 2012b;

Yu *et al*, 2012b).Specifically, S1P and LPA bound to their corresponding membrane GPCRs to activate YAP/TAZ through Rho GTPases. Further studies have revealed phosphatidic acid (PA) from phospholipase D (PLD) as a key regulator of the hippo pathway by activating YAP under various Hippo-activating conditions(Han *et al*, 2018b; Yu *et al*, 2012a). While lipid regulation of the hippo pathway has been explored, much is yet to be elucidated about the upstream regulation of lipid-Hippo axis.

Various forms of cellular stress, including hypoxia, endoplasmic reticulum (ER) stress, energy deprivation, osmotic changes, heat stress, and heavy metal stress can serve as upstream triggers for the Hippo pathway. During hypoxic conditions, YAP phosphorylation is decreased while TAZ phosphorylation is increased suggesting differential regulation of these transcriptional co-activators (Yan *et al*, 2014). Further studies have elucidated a potential role of hypoxia mediated regulation on the hippo pathway by Zyxin(Ma *et al*, 2016), SIAH2 ubiquitin E3 ligase (Ma *et al*, 2015), and hypoxia-inducible factor 1 subunit alpha (HIF-1a)(Tao *et al*, 2021). Other forms of cellular stress known to induce modulation of the Hippo pathway is ER stress by inhibiting YAP activity and enhancing apoptosis(Wu *et al*, 2015). Energy stress sensor, AMP activated protein kinase (AMPK) has been demonstrated to induce an AMPK dependent LATS activation leading to the inactivation of YAP(Wang *et al*, 2015d) via a defect of glucose metabolism. More recently, studies have interestingly identified heat stress to induce YAP activation, providing further insight in cells survival(Luo *et al*, 2020).

Osmotic stress is a condition when there is an imbalance in solute concentrations, leading to alterations in water movement across cell membranes, which can impact cellular function and

survival. Osmotic stress has been shown to regulate the Hippo pathway by modulating the phosphorylation and nuclear localization of the transcriptional co-activators YAP/TAZ. In response to osmotic stress, YAP/TAZ are phosphorylated, promoting their cytoplasmic retention and inhibiting their transcriptional activity, thereby influencing cellular responses to osmotic changes(Hong *et al*, 2020; Hong *et al*, 2017). Other studies have also identified the Hippo pathway to play a role in the regulation of heavy metal response through the modulation of metal regulatory transcription factor 1 by LATS, altering the heave metal response profile (Han *et al*, 2022a).

Regulation of the Hippo pathway by cell polarity:

Tissue integrity is contributed to the maintenance of tight junctions, which are vital cellular structures found in epithelial and endothelial tissues. Composed of transmembrane proteins, tight junctions form continuous seals between adjacent cells near their apical regions. These protein complexes create a physical barrier that prevents the cellular movement of ions, molecules, and fluids, regulating the passage of substances across biological barriers (Gonzalez-Mariscal *et al*, 2003; Tsukita *et al*, 2001). In mammals, tight junction associated scaffold protein, angiomotin (AMOT), has emerged as a critical regulator of the Hippo pathway. AMOT protein family consists of two other angiomotin-like proteins, AMOTL1 and AMOTL2, containing PDZ binding domains. The AMOT members all play an important role in cell migration and angiogenesis(Gagné *et al*, 2009). AMOT family proteins interacts directly with the WW domain of YAP/TAZ through its PPxY motif (Zhao *et al*, 2011a) and inhibits YAP/TAZ activity by promoting cytoplasmic retention. Mechanistically, AMOT binds LATS1/2 and MST1/2 to promote LATS activity which then suppresses YAP/TAZ. Additionally, α- catenin, a crucial

component of E-cadherin-catenin complex(Drees *et al*, 2005), has been shown to inhibit the nuclear localization of YAP which has been linked with the tumor suppressive effects of α-catenin (Silvis *et al*, 2011). Another adherin junction component, protein tyrosine phosphatase 14 (PTPN14), also directly interacts with YAP and sequesters it in the cytoplasm(Liu *et al*, 2013a).

Downstream regulation of the Hippo pathway:

The downstream regulation of the Hippo pathway involves the transcriptional activation or repression of specific target genes in response to cellular signals. The primary interacting partners of YAP/TAZ are TEA DNA binding domains which are transcription enhancer factors (TEFs)(Pan, 2010b). There are four homologs of TEAD that interact with YAP/TAZ through their TEAD binding domain (Vassilev *et al*, 2001) to induce proliferation. In addition to TEADs, YAP/TAZ are known to interact with other transcription factors such as runt-domain transcription factors (RUNXs), responsible for hematopoiesis and osteogenesis (Yagi *et al*, 1999) and SMADs, key players for transforming growth factor beta (TGF-β) (Ferrigno *et al*, 2002b). In association with the above transcription factors, YAP/TAZ have been reported to induce several target genes such as connective tissue growth factor (CTCF), cysteine-rich angiogenic inducer 61 (CYR61), and Ankyrin repeat domain 1 (ANKRD1) (Jiménez *et al*, 2017; Lai *et al*, 2011; Zhao *et al.*, 2008).

Introduction to the Phospholipase D family:

Phospholipase D (PLD) is a family of enzymes that hydrolyze phospholipids to produce phosphatidic acid (PA) and a corresponding alcohol. PLD isoforms are found in numerous

organisms such as viruses (Oguin *et al*, 2014b), bacteria (Pohlman *et al*, 1993b; Stuckey & Dixon, 1999), plants (Wang *et al*, 2014c) and mammals(McDermott *et al*, 2020). PLD isoforms have a variety of substrate specificities and can be regulated by various molecules. They are involved in numerous cellular processes such as receptor signaling, cytoskeletal regulation, and membrane trafficking. The PLD family is comprised of six isoforms, each with distinct structures, functions, and regulation (Frohman, 2015; Frohman *et al*, 1999) **(Figure 0.4)**.

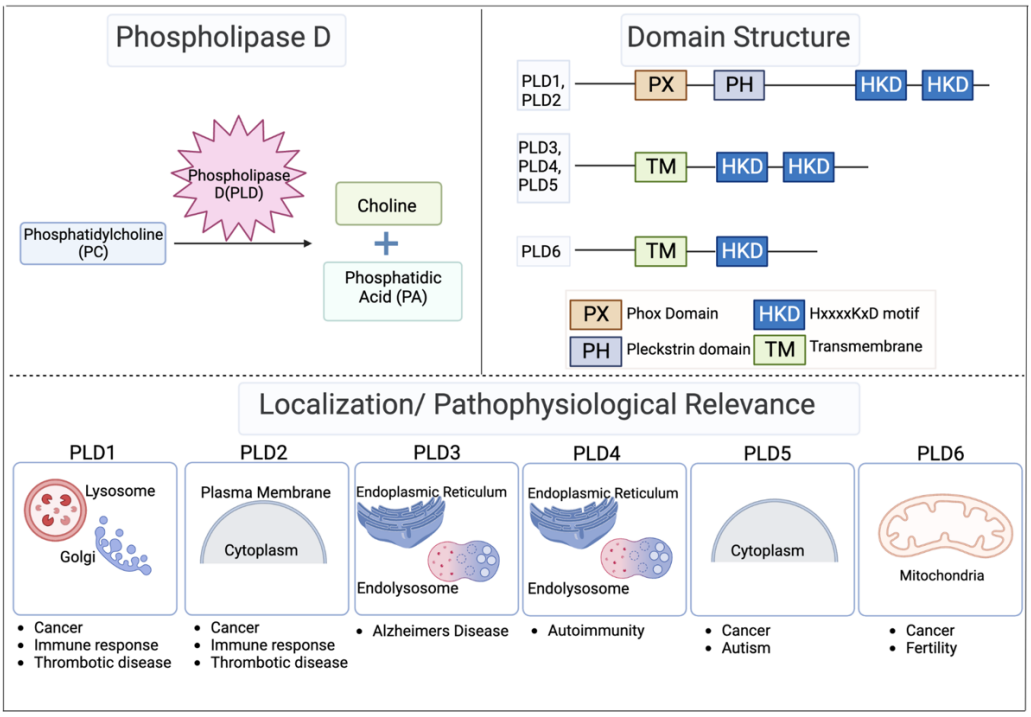


Figure 0.4: **Phospholipid D(PLD) Superfamily Overview**. PLD isoforms are composed of varying domain structures, localizations, and are relevant in various disease settings.

Phospholipase D1 and Phospholipase D2:

PLD1 and PLD2 are the two major isoforms of the PLD family with distinct roles and subcellular localizations(Jenkins & Frohman, 2005a).PLD1 is found throughout the cell, particularly in the Golgi, lysosome and perinuclear region (Freyberg *et al*, 2001a; Lucocq *et al*, 2001b).PLD2 is mostly present at the plasma membrane (Du *et al*, 2004) but has also been reported to localize in the cytosol (Colley *et al*, 1997b). However, both isoforms have been

shown to alter their localization in response to certain conditions, indicating a significant role for PA in many cellular roles(Jenkins & Frohman, 2005a; Weernink *et al*, 2007). The catalytic motif of PLD enzymes is known as HKD, which stands for HxxxxKxD, where H represents histidine, K represents lysine, D represents aspartic acid, and x can be any amino acid. Mammalian PLD1 and PLD2 both contain two HKD motifs, which are essential for their enzymatic function. This has been demonstrated by studies showing that point mutations in these motifs disrupt the activity of PLD in both *in vitro* and *in vivo* settings (Sung *et al*, 1997a). The HKD motifs are a defining feature of PLD enzymes and are critical for their catalytic activity.

PLD1 is regulated by a complex network of signaling molecules and protein-protein interactions. PLD1 can be activated by a variety of upstream signals, including G-protein coupled receptors (GPCRs), protein kinase C (PKC), small GTPases such as Rho and Arf, and tyrosine kinases such as Src and EGFR (Foster & Xu, 2003), whereas PLD2 is most likely indirectly activated by some of these noted factors. Phospholipase D1 (PLD1) has been shown to play a crucial role in the regulation of the mechanistic target of rapamycin (mTOR) pathway, which is a major regulator of cell growth and proliferation. In particular, PLD1/PLD2 has been demonstrated to activate mTOR signaling through phosphatidic acid. PA has been shown to directly bind to and activate mTOR, leading to the activation of downstream targets involved in protein synthesis and cell growth, such as p70S6 kinase (S6K) and eukaryotic initiation factor 4E-binding protein 1 (4E-BP1) (Fang *et al*, 2001b). Additionally, consistent with a competition between PA and rapamycin-FKBP12 for mTOR (Toschi *et al*, 2009), it was demonstrated that elevated levels of PLD activity conferred rapamycin resistance in human breast cancer cell lines(Chen *et al*, 2003; Yoshikawa *et al*, 2010b).

Phospholipase D3, Phospholipase D4, PLD5 and PLD6

PLD3 and PLD4 are endoplasmic reticulum (ER)–associated proteins with no classic catalytic activity(Munck *et al*, 2005b). Upon ER stress, PLD3 has been shown to increase in expression and enhance myotube formation(Osisami *et al*, 2012). Other studies have linked PLD3 and PLD4 to be transmembrane proteins localized to the endolysosomes(Fazzari *et al*, 2017; Gonzalez *et al*, 2018b).Their role in the ER is controversial with other findings highlighting PLD3 and PLD4 as 5′ exonucleases involved in inflammatory cytokine response (Gavin *et al*, 2021; Gavin *et al*, 2018b). PLD5 is another PLD family member with no catalytic activity(Dennis, 2015a) and has been reported as an oncogene involved in prostate tumorigenesis(Liu *et al*, 2021). Further studies are worth pursuing to establish the role of PLD3, PLD4, and PLD5 in their biological roles.

The final isoform, PLD6 (mitoPLD), is unique in its mitochondrial localization(Choi *et al*, 2006) and its single HKD catalytic site. PLD6 cleaves mitochondrial-specific lipid cardiolipin to generate PA and has been implicated in modulating mitochondrial morphology(Huang *et al*, 2011a). Additionally, PLD6 participates in the assembly of nuage, a distinct cytoplasmic structure called inter-mitochondrial cement or chromatoid body (Watanabe *et al*, 2011). PLD6 has also been reported to mediate MYC activity to regulate downstream co-activators YAP/TAZ in tumors(von Eyss *et al*, 2015b).

Phosphatidic acid:

As mentioned, (PLD) catalyzes the hydrolysis of phosphatidylcholine (PC) into phosphatidic acid (PtdOH) and choline. However, Phosphatidic acid (PA) is a versatile lipid that can be synthesized through pathways other than the hydrolysis of PC by PLD. In addition to the PLD-mediated PC hydrolysis, PA can be produced through alternative routes in the cell contributing to the generation of PA, which serves as an important lipid signaling molecule involved in various cellular processes. One of its primary functions is in the biosynthesis of membrane phospholipids, making it a central molecule in maintaining lipid homeostasis (Athenstaedt & Daum, 1999). Moreover, PA has gained recognition as a crucial factor in the regulation of various signaling molecules that control cell cycle progression and survival, such as mTOR, and Raf protein kinases (Kraft *et al*, 2008). As a result, the level of PA is tightly regulated to ensure proper cellular function and signaling pathways.

In conclusion, the PLD family of enzymes is comprised of several isoforms, each with distinct functions and regulation. While studies have elucidated multiple functions for each of the isoforms and PA, a complete overview of the protein-protein interactions of each isoform and PA had not been performed.

References for the Introduction

Aebersold R, Mann M (2016) Mass-spectrometric exploration of proteome structure and function. *Nature* 537: 347-355

Aragona M, Panciera T, Manfrin A, Giulitti S, Michielin F, Elvassore N, Dupont S, Piccolo S (2013) A mechanical checkpoint controls multicellular growth through YAP/TAZ regulation by actin-processing factors. *Cell* 154: 1047-1059

Athenstaedt K, Daum G (1999) Phosphatidic acid, a key intermediate in lipid metabolism. *European Journal of Biochemistry* 266: 1-16

Barry ER, Morikawa T, Butler BL, Shrestha K, De La Rosa R, Yan KS, Fuchs CS, Magness ST, Smits R, Ogino S (2013) Restriction of intestinal stem cell expansion and the regenerative response by YAP. *Nature* 493: 106-110

Bork P, Sudol M (1994) The WW domain: a signalling site in dystrophin? *Trends in biochemical sciences* 19: 531-533

Cai J, Zhang N, Zheng Y, De Wilde RF, Maitra A, Pan D (2010) The Hippo signaling pathway restricts the oncogenic potential of an intestinal regeneration program. *Genes & development* 24: 2383-2388

Calderón-González KG, Hernández-Monge J, Herrera-Aguirre ME, Luna-Arias JP (2016) Bioinformatics tools for proteomics data interpretation. *Modern Proteomics–Sample Preparation, Analysis and Practical Applications*: 281-341

Chang NS, Hsu LJ, Lin YS, Lai FJ, Sheu HM, 2007. WW domain-containing oxidoreductase: a candidate tumor suppressor, Trends in Molecular Medicine. Elsevier Current Trends, pp. 12-22.

Chen Y, Zheng Y, Foster DA (2003) Phospholipase D confers rapamycin resistance in human breast cancer cells. *Oncogene* 22: 3937-3942

Choi S-Y, Huang P, Jenkins GM, Chan DC, Schiller J, Frohman MA, 2006. A common lipid links Mfn-mediated mitochondrial fusion and SNARE-regulated exocytosis, Nature Cell Biology 2006 8:11. Nature Publishing Group, pp. 1255-1262.

Colley WC, Sung TC, Roll R, Jenco J, Hammondt SM, Altshuller Y, Bar-Sagi D, Morris AJ, Frohman MA, 1997. Phospholipase D2, a distinct phospholipase D isoform with novel regulatory properties that provokes cytoskeletal reorganization, Current Biology. Cell Press, pp. 191-201.

Couzens AL, Knight JD, Kean MJ, Teo G, Weiss A, Dunham WH, Lin Z-Y, Bagshaw RD, Sicheri F, Pawson T (2013) Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Science signaling* 6: rs15-rs15

Dennis EA (2015) Introduction to thematic review series: phospholipases: central role in lipid signaling and disease. *Journal of lipid research* 56: 1245-1247

Dey A, Varelas X, Guan KL (2020) Targeting the Hippo pathway in cancer, fibrosis, wound healing and regenerative medicine. *Nature Reviews Drug Discovery 2020 19:7* 19: 480-494

Drees F, Pokutta S, Yamada S, Nelson WJ, Weis WI (2005) α-catenin is a molecular switch that binds E-cadherin-β-catenin and regulates actin-filament assembly. *Cell* 123: 903-915

Du G, Huang P, Liang BT, Frohman MA (2004) Phospholipase D2 localizes to the plasma membrane and regulates angiotensin II receptor endocytosis. *Molecular biology of the cell* 15: 1024-1030

Dupont S, Morsut L, Aragona M, Enzo E, Giulitti S, Cordenonsi M, Zanconato F, Le Digabel J, Forcato M, Bicciato S (2011) Role of YAP/TAZ in mechanotransduction. *Nature* 474: 179-183

Faber PW, Barnes GT, Srinidhi J, Chen J, Gusella JF, MacDonald ME (1998) Huntingtin interacts with a family of WW domain proteins. *Human molecular genetics* 7: 1463-1474

Fang Y, Vilella-Bach M, Bachmann R, Flanigan A, Chen J (2001) Phosphatidic acid-mediated mitogenic activation of mTOR signaling. *Science* 294: 1942-1945

Fazzari P, Horre K, Arranz AM, Frigerio CS, Saito T, Saido TC, De Strooper B (2017) PLD3 gene and processing of APP. *Nature 2017 541:7638* 541: E1-E2

Feng X, Degese MS, Iglesias-Bartolome R, Vaque JP, Molinolo AA, Rodrigues M, Zaidi MR, Ksander BR, Merlino G, Sodhi A (2014) Hippo-independent activation of YAP by the GNAQ uveal melanoma oncogene through a trio-regulated rho GTPase signaling circuitry. *Cancer cell* 25: 831-845

Ferrigno O, Lallemand F, Verrecchia F, L'Hoste S, Camonis J, Atfi A, Mauviel A (2002a) Yes-associated protein (YAP65) interacts with Smad7 and potentiates its inhibitory activity against TGF-β/Smad signaling. *Oncogene 2002 21:32* 21: 4879-4884

Ferrigno O, Lallemand F, Verrecchia F, L'Hoste S, Camonis J, Atfi A, Mauviel A, 2002b. Yes-associated protein (YAP65) interacts with Smad7 and potentiates its inhibitory activity against TGF-β/Smad signaling, Oncogene 2002 21:32. Nature Publishing Group, pp. 4879-4884.

Foster DA, Xu L (2003) Phospholipase D in cell proliferation and cancer. *Molecular Cancer Research* 1: 789-800

Freyberg Z, Sweeney D, Siddhanta A, Bourgoin S, Frohman M, Shields D, 2001. Intracellular localization of phospholipase D1 in mammalian cells, Molecular Biology of the Cell. American Society for Cell Biology, pp. 943-955.

Frohman MA, 2015. The phospholipase D superfamily as therapeutic targets, Trends in Pharmacological Sciences. Elsevier Ltd, pp. 137-144.

Frohman MA, Sung TC, Morris AJ, 1999. Mammalian phospholipase D structure and regulation, Biochimica et Biophysica Acta - Molecular and Cell Biology of Lipids. Elsevier, pp. 175-186.

Gagné V, Moreau J, Plourde M, Lapointe M, Lord M, Gagnon E, Fernandes MJ (2009) Human angiomotin-like 1 associates with an angiomotin protein complex through its coiled-coil domain and induces the remodeling of the actin cytoskeleton. *Cell motility and the cytoskeleton* 66: 754-768

Gao T, Zhou D, Yang C, Singh T, Penzo–Méndez A, Maddipati R, Tzatsos A, Bardeesy N, Avruch J, Stanger BZ (2013) Hippo signaling regulates differentiation and maintenance in the exocrine pancreas. *Gastroenterology* 144: 1543-1553. e1541

Gavin AL, Huang D, Blane TR, Thinnes TC, Murakami Y, Fukui R, Miyake K, Nemazee D (2021) Cleavage of DNA and RNA by PLD3 and PLD4 limits autoinflammatory triggering by multiple sensors. *Nature Communications* 12: 5874

Gavin AL, Huang D, Huber C, Mårtensson A, Tardif V, Skog PD, Blane TR, Thinnes TC, Osborn K, Chong HS *et al*, 2018. PLD3 and PLD4 are single-stranded acid exonucleases that regulate endosomal nucleic-acid sensing, Nature Immunology 2018 19:9. Nature Publishing Group, pp. 942-953.

Gonzalez AC, Schweizer M, Jagdmann S, Bernreuther C, Reinheckel T, Saftig P, Damme M, 2018. Unconventional Trafficking of Mammalian Phospholipase D3 to Lysosomes, Cell Reports. Cell Press, pp. 1040-1053.

Gonzalez-Mariscal L, Betanzos A, Nava P, Jaramillo B (2003) Tight junction proteins. *Progress in biophysics and molecular biology* 81: 1-44

Haas R, Zelezniak A, Iacovacci J, Kamrad S, Townsend S, Ralser M (2017) Designing and interpreting 'multi-omic'experiments that may change our understanding of biology. *Current Opinion in Systems Biology* 6: 37-45

Halder G, Dupont S, Piccolo S (2012) Transduction of mechanical and cytoskeletal cues by YAP and TAZ. *Nature reviews Molecular cell biology* 13: 591-600

Hall CA, Wang R, Miao J, Oliva E, Shen X, Wheeler T, Hilsenbeck SG, Orsulic S, Goode S (2010) Hippo Pathway Effector Yap Is an Ovarian Cancer OncogeneYap Is an Ovarian Cancer Oncogene. *Cancer research* 70: 8517-8525

Han H, Nakaoka HJ, Hofmann L, Zhou JJ, Yu C, Zeng L, Nan J, Seo G, Vargas RE, Yang B *et al* (2022) The Hippo pathway kinases LATS1 and LATS2 attenuate cellular responses to heavy metals through phosphorylating MTF1. *Nature Cell Biology* 24: 74-87

Han H, Yang B, Nakaoka HJ, Yang J, Zhao Y, Nguyen KL, Bishara AT, Mandalia TK, Wang W (2018) Hippo signaling dysfunction induces cancer cell addiction to YAP. *Oncogene 2018 37:50* 37: 6414-6424

Hao Y, Chun A, Cheung K, Rashidi B, Yang X (2008) Tumor suppressor LATS1 is a negative regulator of oncogene YAP. *Journal of Biological Chemistry* 283: 5496-5509

Harvey K, Tapon N (2007) The Salvador–Warts–Hippo pathway—an emerging tumour-suppressor network. *Nature Reviews Cancer* 7: 182-191

Haskins JW, Nguyen DX, Stern DF (2014a) Neuregulin 1–activated ERBB4 interacts with YAP to induce Hippo pathway target genes and promote cell migration. *Science Signaling* 7: ra116-ra116

Haskins JW, Nguyen DX, Stern DF, 2014b. Neuregulin 1–activated ERBB4 interacts with YAP to induce Hippo pathway target genes and promote cell migration, Science Signaling. American Association for the Advancement of Science, pp. ra116-ra116.

Hauri S, Wepf A, van Drogen A, Varjosalo M, Tapon N, Aebersold R, Gstaiger M (2013) Interaction proteome of human H ippo signaling: modular control of the co-activator YAP 1. *Molecular systems biology* 9: 713

Heallen T, Zhang M, Wang J, Bonilla-Claudio M, Klysik E, Johnson RL, Martin JF (2011) Hippo pathway inhibits Wnt signaling to restrain cardiomyocyte proliferation and heart size. *Science* 332: 458-461

Hoffman BD, Grashoff C, Schwartz MA (2011) Dynamic molecular processes mediate cellular mechanotransduction. *Nature* 475: 316-323

Hong AW, Meng Z, Plouffe SW, Lin Z, Zhang M, Guan K-L (2020) Critical roles of phosphoinositides and NF2 in Hippo pathway regulation. *Genes & development* 34: 511-525

Hong AW, Meng Z, Yuan HX, Plouffe SW, Moon S, Kim W, Jho Eh, Guan KL (2017) Osmotic stress-induced phosphorylation by NLK at Ser128 activates YAP. *EMBO reports* 18: 72-86

Hsu PC, Miao J, Huang Z, Yang YL, Xu Z, You J, Dai Y, Yeh CC, Chan G, Liu S (2018) Inhibition of yes-associated protein suppresses brain metastasis of human lung adenocarcinoma in a murine model. *Journal of cellular and molecular medicine* 22: 3073-3085

Huang H, Gao Q, Peng X, Choi SY, Sarma K, Ren H, Morris AJ, Frohman MA (2011) piRNA-Associated Germline Nuage Formation and Spermatogenesis Require MitoPLD Profusogenic Mitochondrial-Surface Lipid Signaling. *Developmental Cell* 20: 376-387

Jenkins G, Frohman M (2005) Phospholipase D: a lipid centric review. *Cellular and Molecular Life Sciences CMLS* 62: 2305-2316

Jiménez AP, Traum A, Boettger T, Hackstein H, Richter AM, Dammann RH (2017) The tumor suppressor RASSF1A induces the YAP1 target gene ANKRD1 that is epigenetically inactivated in human cancers and inhibits tumor growth. *Oncotarget* 8: 88437

Johnson R, Halder G, 2013. The two faces of Hippo: targeting the Hippo pathway for regenerative medicine and cancer treatment, Nature Reviews Drug Discovery 2013 13:1. Nature Publishing Group, pp. 63-79.

Kattan RE, Ayesh D, Wang W (2023) Analysis of affinity purification-related proteomic data for studying protein–protein interaction networks in cells. *Briefings in Bioinformatics* 24: bbad010

Kattan RE, Han H, Seo G, Yang B, Lin Y, Dotson M, Pham S, Menely Y, Wang W (2022) Interactome Analysis of Human Phospholipase D and Phosphatidic Acid-Associated Protein Network. *Molecular & Cellular Proteomics : MCP* 21: 100195-100195

Koepf EK, Petrassi HM, Sudol M, Kelly JW (1999) WW: An isolated three-stranded antiparallel β-sheet domain that unfolds and refolds reversibly; evidence for a structured hydrophobic cluster in urea and GdnHCl and a disordered thermal unfolded state. *Protein Science* 8: 841-853

Kraft CA, Garrido JL, Fluharty E, Leiva-Vega L, Romero G (2008) Role of phosphatidic acid in the coupling of the ERK cascade. *Journal of Biological Chemistry* 283: 36636-36645

Kremerskothen J, Plaas C, Büther K, Finger I, Veltel S, Matanis T, Liedtke T, Barnekow A (2003) Characterization of KIBRA, a novel WW domain-containing protein. *Biochemical and biophysical research communications* 300: 862-867

Lai D, Ho KC, Hao Y, Yang X (2011) Taxol Resistance in Breast Cancer Cells Is Mediated by the Hippo Pathway Component TAZ and Its Downstream Transcriptional Targets Cyr61 and CTGFCyr61/CTGF Mediate TAZ-induced Taxol Resistance. *Cancer research* 71: 2728-2738

Liu C, Huang W, Lei Q (2011) Regulation and function of the TAZ transcription co-activator. *International journal of biochemistry and molecular biology* 2: 247

Liu J, Li J, Ma Y, Xu C, Wang Y, He Y (2021) MicroRNA miR-145-5p inhibits Phospholipase D 5 (PLD5) to downregulate cell proliferation and metastasis to mitigate prostate cancer. *Bioengineered* 12: 3240-3251

Liu X, Yang N, Figel SA, Wilson KE, Morrison CD, Gelman IH, Zhang J (2013) PTPN14 interacts with and negatively regulates the oncogenic function of YAP. *Oncogene* 32: 1266-1273

Lopez J, Mouw J, Weaver V (2008) Biomechanical regulation of cell orientation and fate. *Oncogene* 27: 6981-6993

Lubs H, Abidi F, Echeverri R, Holloway L, Meindl A, Stevenson R, Schwartz C (2006) Golabi-Ito-Hall syndrome results from a missense mutation in the WW domain of the PQBP1 gene. *Journal of medical genetics* 43: e30-e30

Lucocq J, Manifava M, Bi K, Roth MG, Ktistakis NT, 2001. Immunolocalisation of phospholipase D1 on tubular vesicular membranes of endocytic and secretory origin, European Journal of Cell Biology. Urban & Fischer, pp. 508-520.

Luo M, Meng Z, Moroishi T, Lin KC, Shen G, Mo F, Shao B, Wei X, Zhang P, Wei Y *et al* (2020) Heat stress activates YAP/TAZ to induce the heat shock transcriptome. *Nature Cell Biology* 22: 1447-1459

Ma B, Chen Y, Chen L, Cheng H, Mu C, Li J, Gao R, Zhou C, Cao L, Liu J (2015) Hypoxia regulates Hippo signalling through the SIAH2 ubiquitin E3 ligase. *Nature cell biology* 17: 95-103

Ma B, Cheng H, Gao R, Mu C, Chen L, Wu S, Chen Q, Zhu Y (2016) Zyxin-Siah2–Lats2 axis mediates cooperation between Hippo and TGF-β signalling pathways. *Nature communications* 7: 11123

Manninen A (2015) Epithelial polarity–generating and integrating signals from the ECM with integrins. *Experimental cell research* 334: 337-349

McDermott M, Wang Y, Wakelam M, Bankaitis V (2020) Mammalian phospholipase D: Function, and therapeutics. *Progress in lipid research* 78: 101018

Miller E, Yang J, DeRan M, Wu C, Su AI, Bonamy GM, Liu J, Peters EC, Wu X (2012) Identification of serum-derived sphingosine-1-phosphate as a small molecule regulator of YAP. *Chemistry & biology* 19: 955-962

Moroishi T, Hansen CG, Guan K-L (2015) The emerging roles of YAP and TAZ in cancer. *Nature Reviews Cancer* 15: 73-79

Munck A, Böhm C, Seibel NM, Hosseini ZH, Hampe W, 2005. Hu-K4 is a ubiquitously expressed type 2 transmembrane protein associated with the endoplasmic reticulum, The FEBS Journal. John Wiley & Sons, Ltd, pp. 1718-1726.

Nantie LB, Young RE, Paltzer WG, Zhang Y, Johnson RL, Verheyden JM, Sun X (2018) Lats1/2 inactivation reveals Hippo function in alveolar type I cell differentiation during lung transition to air breathing. *Development* 145: dev163105

Oguin TH, Sharma S, Stuart AD, Duan S, Scott SA, Jones CK, Daniels JS, Lindsley CW, Thomas PG, Brown HA (2014) Phospholipase D facilitates efficient entry of influenza virus, allowing escape from innate immune inhibition. *Journal of Biological Chemistry* 289: 25405-25417

Osisami M, Ali W, Frohman MA, 2012. A Role for Phospholipase D3 in Myotube Formation, in: Sanchis D. (Ed.) PLoS ONE. p. e33341.

Pan D (2010) The Hippo Signaling Pathway in Development and Cancer. *Developmental Cell* 19: 491-505

Passani LA, Bedford MT, Faber PW, McGinnis KM, Sharp AH, Gusella JF, Vonsattel J-P, MacDonald ME (2000) Huntingtin's WW domain partners in Huntington's disease post-mortem brain fulfill genetic criteria for direct involvement in Huntington's disease pathogenesis. *Human molecular genetics* 9: 2175-2182

Plouffe SW, Hong AW, Guan K-L (2015) Disease implications of the Hippo/YAP pathway. *Trends in molecular medicine* 21: 212-222

Pobbati AV, Hong W (2013) Emerging roles of TEAD transcription factors and its coactivators in cancers. *Cancer biology & therapy* 14: 390-398

Pohlman RF, Liu F, Wang L, Moré MI, Winans SC (1993) Genetic and biochemical analysis of an endonuclease encoded by the IncN plasmid pKM101. *Nucleic Acids Research* 21: 4867-4867

Rentschler S, Linn H, Deininger K, Bedford MT, Espanel X, Sudol M (1999) The WW domain of dystrophin requires EF-hands region to interact with β-dystroglycan.

Salah Z, Aqeilan RI (2011a) WW domain interactions regulate the Hippo tumor suppressor pathway. *Cell Death & Disease 2011 2:6* 2: e172-e172

Salah Z, Aqeilan RI, 2011b. WW domain interactions regulate the Hippo tumor suppressor pathway, Cell Death & Disease 2011 2:6. Nature Publishing Group, pp. e172-e172.

Seo G, Han H, Vargas RE, Yang B, Li X, Wang W (2020) MAP4K Interactome Reveals STRN4 as a Key STRIPAK Complex Component in Hippo Pathway Regulation. *Cell Reports* 32: 107860-107860

Silvis MR, Kreger BT, Lien W-H, Klezovitch O, Rudakova GM, Camargo FD, Lantz DM, Seykora JT, Vasioukhin V (2011) α-catenin is a tumor suppressor that controls cell accumulation by regulating the localization and activity of the transcriptional coactivator Yap1. *Science signaling* 4: ra33-ra33

Stuckey JA, Dixon JE (1999) Crystal structure of a phospholipase D family member. *Nature Structural Biology 1999 6:3* 6: 278-284

Sudol M (2010) Newcomers to the WW Domain-Mediated Network of the Hippo Tumor Suppressor Pathway. *Genes & cancer* 1: 1115-1118

Sudol M, 2011. Newcomers to the WW Domain–Mediated Network of the Hippo Tumor Suppressor Pathway:, http://dxdoiorg/101177/1947601911401911. SAGE PublicationsSage CA: Los Angeles, CA, pp. 1115-1118.

Sudol M, Bork P, Einbond A, Kastury K, Druck T, Negrini M, Huebner K, Lehman D (1995a) Characterization of the Mammalian YAP (Yes-associated Protein) Gene and Its Role in Defining a Novel Protein Module, the WW Domain ∗. *Journal of Biological Chemistry* 270: 14733-14741

Sudol M, Chen HI, Bougeret C, Einbond A (1995b) Characterization of a novel protein-binding module-the WW domain. *FEBS 15622 FEBS Letters* 369: 71-71

Sudol M, Harvey KF (2010) Modularity in the Hippo signaling pathway. *Trends in biochemical sciences* 35: 627-633

Sung T-C, Roper RL, Zhang Y, Rudge SA, Temel R, Hammond SM, Morris AJ, Moss B, Engebrecht J, Frohman MA (1997) Mutagenesis of phospholipase D defines a superfamily including a trans-Golgi viral protein required for poxvirus pathogenicity. *The EMBO journal* 16: 4519-4530

Takahashi Y, Miyoshi Y, Takahata C, Irahara N, Taguchi T, Tamaki Y, Noguchi S (2005) Down-regulation of LATS1 and LATS2 mRNA expression by promoter hypermethylation and its association with biologically aggressive phenotype in human breast cancers. *Clinical Cancer Research* 11: 1380-1385

Tao E-W, Wang H-L, Cheng WY, Liu Q-Q, Chen Y-X, Gao Q-Y (2021) A specific tRNA half, 5'tiRNA-His-GTG, responds to hypoxia via the HIF1α/ANG axis and promotes colorectal

cancer progression by regulating LATS2. *Journal of Experimental & Clinical Cancer Research* 40: 1-17

Tapia VE, Nicolaescu E, McDonald CB, Musi V, Oka T, Inayoshi Y, Satteson AC, Mazack V, Humbert J, Gaffney CJ (2010) Y65C missense mutation in the WW domain of the Golabi-Ito-Hall syndrome protein PQBP1 affects its binding activity and deregulates pre-mRNA splicing. *Journal of biological chemistry* 285: 19391-19401

Tapon N, Harvey KF, Bell DW, Wahrer DC, Schiripo TA, Haber DA, Hariharan IK (2002) salvador Promotes both cell cycle exit and apoptosis in Drosophila and is mutated in human cancer cell lines. *Cell* 110: 467-478

Thomasy SM, Morgan JT, Wood JA, Murphy CJ, Russell P (2013) Substratum stiffness and latrunculin B modulate the gene expression of the mechanotransducers YAP and TAZ in human trabecular meshwork cells. *Experimental eye research* 113: 66-73

Toschi A, Lee E, Xu L, Garcia A, Gadir N, Foster DA (2009) Regulation of mTORC1 and mTORC2 complex assembly by phosphatidic acid: competition with rapamycin. *Molecular and cellular biology* 29: 1411-1420

Tsukita S, Furuse M, Itoh M (2001) Multifunctional strands in tight junctions. *Nature reviews Molecular cell biology* 2: 285-293

Vargas RE, Duong VT, Han H, Ta AP, Chen Y, Zhao S, Yang B, Seo G, Chuc K, Oh S (2020) Elucidation of WW domain ligand binding specificities in the hippo pathway reveals STXBP 4 as YAP inhibitor. *The EMBO journal* 39: e102406

Vassilev A, Kaneko KJ, Shu H, Zhao Y, DePamphilis ML (2001) TEAD/TEF transcription factors utilize the activation domain of YAP65, a Src/Yes-associated protein localized in the cytoplasm. *Genes & development* 15: 1229-1241

von Eyss B, Jaenicke LA, Kortlever RM, Royla N, Wiese KE, Letschert S, McDuffus LA, Sauer M, Rosenwald A, Evan GI *et al*, 2015. A MYC-Driven Change in Mitochondrial Dynamics Limits YAP/TAZ Function in Mammary Epithelial Cells and Breast Cancer, Cancer Cell. Cell Press, pp. 743-757.

Wada K-I, Itoga K, Okano T, Yonemura S, Sasaki H (2011) Hippo pathway regulation by cell morphology and stress fibers. *Development* 138: 3907-3914

Wang W, Li X, Huang J, Feng L, Dolinta KG, Chen J, 2014a. Defining the Protein–Protein Interaction Network of the Human Hippo Pathway *, Molecular & Cellular Proteomics. Elsevier, pp. 119-131.

Wang W, Xiao Z-D, Li X, Aziz KE, Gan B, Johnson RL, Chen J (2015) AMPK modulates Hippo pathway activity to regulate energy homeostasis. *Nature cell biology* 17: 490-499

Wang X, Guo L, Wang G, Li M (2014b) *PLD: phospholipase Ds in plant signaling*. Springer

Wang Y, Dong Q, Zhang Q, Li Z, Wang E, Qiu X (2010) Overexpression of yes-associated protein contributes to progression and poor prognosis of non-small-cell lung cancer. *Cancer science* 101: 1279-1285

Watanabe T, Chuma S, Yamamoto Y, Kuramochi-Miyagawa S, Totoki Y, Toyoda A, Hoki Y, Fujiyama A, Shibata T, Sado T (2011) MITOPLD is a mitochondrial protein essential for nuage formation and piRNA biogenesis in the mouse germline. *Developmental cell* 20: 364-375

Weernink PAO, Han L, Jakobs KH, Schmidt M (2007) Dynamic phospholipid signaling by G protein-coupled receptors. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1768: 888-900

Wilson KE, Li Y-W, Yang N, Shen H, Orillion AR, Zhang J (2014) PTPN14 forms a complex with Kibra and LATS1 proteins and negatively regulates the YAP oncogenic function. *Journal of Biological Chemistry* 289: 23693-23700

Wu H, Wei L, Fan F, Ji S, Zhang S, Geng J, Hong L, Fan X, Chen Q, Tian J *et al*, 2015. Integration of Hippo signalling and the unfolded protein response to restrain liver overgrowth and tumorigenesis, Nature Communications. Nature Publishing Group, p. 6239.

Xia H, Qi H, Li Y, Pei J, Barton J, Blackstad M, Xu T, Tao W (2002) LATS1 tumor suppressor regulates G2/M transition and apoptosis. *Oncogene* 21: 1233-1241

Xu T, Wang W, Zhang S, Stewart RA, Yu W (1995) Identifying tumor suppressors in genetic mosaics: the Drosophila lats gene encodes a putative protein kinase. *Development* 121: 1053-1063

Yagi R, Chen L-F, Shigesada K, Murakami Y, Ito Y (1999) A WW domain-containing yes-associated protein (YAP) is a novel transcriptional co-activator. *The EMBO journal* 18: 2551-2562

Yan L, Cai Q, Xu Y (2014) Hypoxic conditions differentially regulate TAZ and YAP in cancer cells. *Archives of biochemistry and biophysics* 562: 31-36

Yoshikawa F, Banno Y, Otani Y, Yamaguchi Y, Nagakura-Takagi Y, Morita N, Sato Y, Saruta C, Nishibe H, Sadakata T *et al*, 2010. Phospholipase D Family Member 4, a Transmembrane Glycoprotein with No Phospholipase D Activity, Expression in Spleen and Early Postnatal Microglia, PLOS ONE. Public Library of Science, p. e13932.

Yu FX, Zhao B, Guan KL (2015) Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. *Cell* 163: 811-828

Yu FX, Zhao B, Panupinthu N, Jewell JL, Lian I, Wang LH, Zhao J, Yuan H, Tumaneng K, Li H *et al*, 2012a. Regulation of the Hippo-YAP Pathway by G-Protein-Coupled Receptor Signaling, Cell. Cell Press, pp. 780-791.

Yu FX, Zhao B, Panupinthu N, Jewell JL, Lian I, Wang LH, Zhao J, Yuan H, Tumaneng K, Li H *et al* (2012b) Regulation of the Hippo-YAP Pathway by G-Protein-Coupled Receptor Signaling. *Cell* 150: 780-791

Zhao B, Li L, Lu Q, Wang LH, Liu C-Y, Lei Q, Guan K-L (2011) Angiomotin is a novel Hippo pathway component that inhibits YAP oncoprotein. *Genes & development* 25: 51-63

Zhao B, Li L, Wang L, Wang C-Y, Yu J, Guan K-L (2012) Cell detachment activates the Hippo pathway via cytoskeleton reorganization to induce anoikis. *Genes & development* 26: 54-68

Zhao B, Wei X, Li W, Udan RS, Yang Q, Kim J, Xie J, Ikenoue T, Yu J, Li L (2007) Inactivation of YAP oncoprotein by the Hippo pathway is involved in cell contact inhibition and tissue growth control. *Genes & development* 21: 2747-2761

Zhao B, Ye X, Yu J, Li L, Li W, Li S, Yu J, Lin JD, Wang C-Y, Chinnaiyan AM (2008) TEAD mediates YAP-dependent gene induction and growth control. *Genes & development* 22: 1962-1971

Zhao R, Fallon TR, Saladi SV, Pardo-Saganta A, Villoria J, Mou H, Vinarsky V, Gonzalez-Celeiro M, Nunna N, Hariri LP (2014) Yap tunes airway epithelial size and architecture by regulating the identity, maintenance, and self-renewal of stem cells. *Developmental cell* 30: 151-165

Zhou D, Conrad C, Xia F, Park J-S, Payer B, Yin Y, Lauwers GY, Thasler W, Lee JT, Avruch J (2009) Mst1 and Mst2 maintain hepatocyte quiescence and suppress hepatocellular carcinoma development through inactivation of the Yap1 oncogene. *Cancer cell* 16: 425-438

CHAPTER 1

Analysis of affinity purification-related proteomic data for studying protein-protein interaction networks in cells

Rebecca Elizabeth Kattan[1], Deena Ayesh[1], and Wenqi Wang[1,*]

[1]Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA 92697, USA

*Correspondence: wenqiw6@uci.edu (W.W.)

This chapter is derived from the manuscript published in Briefings in Bioinformatics: Briefings in Bioinformatics. 2023 Mar;24(2). doi: 10.1093/bib/bbad010

1.1 Abstract

During intracellular signal transduction, protein-protein interactions (PPIs) facilitate protein complex assembly to regulate protein localization and function, which are critical for numerous cellular events. Over the years, multiple techniques have been developed to characterize PPIs to elucidate roles and regulatory mechanisms of proteins. Among them, the mass spectrometry (MS)-based interactome analysis has been increasing in popularity due to its unbiased and informative manner towards understanding PPI networks. However, with MS instrumentation advancing and yielding more data than ever, analysis of a large amount of PPI-associated proteomic data to reveal *bona fide* interacting proteins become challenging. Here, we review the methods and bioinformatic resources that are commonly used in analyzing large interactome-related proteomic data and propose a simple guideline for identifying novel interacting proteins for biological research.

1.2 Key points

- A review of the commonly used biochemical methods for identifying interacting proteins for a protein of interest
- A summary of the bioinformatic tools and resources for analyzing interactome-based mass spectrometry data
- A proposed guideline for taking proteomics as an approach to study protein-protein interactions in biological research

1.3 Keywords

Protein-protein interaction, proteomics, affinity purification, bioinformatics

1.4 Introduction

Protein-protein interactions (PPIs) mediate cellular signal transduction cascades where specific interactions between proteins define unique and robust signaling outputs for different cellular events(Pawson & Nash, 2000). Dysregulation of PPIs has been frequently associated with various human diseases(Ryan & Matthews, 2005). Therefore, better characterizing PPIs will create new opportunities for understanding normal physiology and treating human diseases.

Mass spectrometry (MS) is an instrument for analyzing small chemical and biological molecules(Aebersold & Mann, 2003), which has revolutionized scientific research in the past decades. Regarding PPI study in biological research, MS can be used in partner with multiple biochemical approaches to uncover *bona fide* interactors for proteins of interest. Commonly used methods for isolating protein complexes comprise immunoprecipitation, affinity purification(Li, 2011) and protein-proximity labeling approaches like BioID(Roux *et al*, 2012) and APEX(Lam *et al*, 2015). However, with MS technology and methods of capturing interactors continuously improving, the size and complexity of PPI-related proteomic data keep increasing. Manually converting such a large dataset to a meaningful biological list of PPIs is time and labor-consuming and often results in errors and biases.

To tackle this issue, several bioinformatic tools have been developed to facilitate the PPI-related proteomic data analysis, which include but are not limited to MS raw data filtration, gene ontology study, functional topology analysis, and PPI network visualization. These methods make processing, analyzing, and presenting large-scale MS data possible for regular biological research labs. To produce a reliable interactome dataset that can be used by other researchers, the

refined PPI network also needs to be experimentally validated and functionally characterized. However, how to integrate these tools and methods to produce meaningful and reliable interactome data for in-depth functional studies has been challenging.

Here, we review the current methods used for isolating and identifying interacting proteins by MS analysis and provide a brief summary on how to process MS raw data, interpret their biological significance, build PPI network, and characterize biological functions for newly identified interacting proteins. We hope this guideline and the related bioinformatics resources can benefit new researchers who are interested in taking MS as an approach to investigate PPIs for their biological research.

1.5 Main Review

Isolation of interacting proteins for a protein of interest

Many biochemical methods have been developed to isolate the associated protein complex for a protein of interest (i.e., bait protein)(Rao *et al*, 2014), such as immunoprecipitation (IP) through antibodies against either bait protein or an epitope tag fused with bait protein, affinity purification (AP) using single or tandem epitope tag(Li, 2011), and protein-proximity labelling approaches using the modified biotin-protein ligase tag (e.g., BioID(Roux *et al.*, 2012), APEX(Lam *et al.*, 2015)). Despite their different mechanisms, the overall goal of these purification techniques is to capture and reserve true binding proteins while minimizing the non-specific ones for bait protein.

Attempting to capture all the interacting proteins whether they are transient or stable has been a difficult task for the current purification methods. Although performing IP with a bait protein antibody (**Figure 1.1A**) can purify its associated protein complex at endogenous level, antibody availability and its IP efficiency often limit the use of this method. One way to overcome these issues is to utilize well-established antibodies for epitope tags, such as Flag, Myc, HA, GFP, and indirectly isolate the binding proteins for the bait protein fused with an epitope tag (**Figure 1.1A**). However, exogenously expressing tag-fused protein often results in non-specific and artificial binding proteins due to overexpression. In addition, both methods share common antibody leakage issue, which can affect sample preparation for MS analysis.

Affinity purification (AP) can overcome the antibody leakage issue and has been widely used for PPI studies. To achieve so, epitope tags with high binding affinity to corresponding agarose beads have been developed (**Figure 1.1B**). For example, a commonly used tag for AP is streptavidin binding protein (SBP) (**Figure 1.1B**), which has high binding affinity to streptavidin beads(Keefe *et al*, 2001). However, streptavidin beads also bind endogenous biotinylated proteins in cells, resulting in non-specific interactions(Keefe *et al.*, 2001). Maltose-binding protein (MBP)(Kellermann & Ferenci, 1982) and glutathione S-transferase (GST)(Smith & Johnson, 1988) tags are often used for protein complex purification, because they can strongly bind to the agarose beads conjugated with amylose and glutathione, respectively (**Figure 1.1B**). However, these two epitope tags are both large in protein size and can causes potential folding issues for a bait protein, thus hindering the isolation of its true binding proteins. In contrast, hexa-histidine (His) tag(Janknecht *et al*, 1991) is a small tag that binds to immobilized nickel beads (**Figure 1.1B**), which offers a better solution for AP as compared to MBP and GST tags.

However, His tag-mediated protein purification requires the step of optimizing imidazole concentration for different bait proteins, adding additional work to ensure the protein purification quality.

Recently, proximity labeling methods have been widely used for interactome study (**Figure 1.1C**), because it allows identification of transient and weak binding proteins for a bait protein. Like the previously mentioned SBP tag, this method relies on streptavidin beads to purify the newly biotinylated proteins, though the endogenous modification of biotinylation leads to the issue of isolating non-specific binding proteins.

Over time, methods for purifying protein complex associated with a protein of interest have been evolving to increase their performance and efficiency(Rigaut *et al*, 1999). Some labs incorporated different epitope tags as one for protein complex purification, allowing further elimination of non-specific binding interactions through multiple purification and washing steps(Dunham *et al*, 2012). For example, tandem affinity purification (TAP) was developed to fuse two epitope tags with one bait protein to reduce the non-specific binding proteins(Dunham *et al.*, 2012; Li, 2011). Initially, TAP was designed with two IgG-binding units of protein A of *Staphylococcus aureus* (Protein A) and the calmodulin-binding peptide (CBP)(Rigaut *et al.*, 1999) (**Figure 1.1D**), while its upgraded version includes tags like FLAG and HA in tandem(Sowa *et al*, 2009) (**Figure 1.1E**) to achieve a better purification performance.

In the past years, we have been extensively utilizing a S-Flag-SBP (SFB) triple tagged system for tandem affinity purification (**Figure 1.1F**) and revealed new regulators/effectors for multiple key

proteins involved in growth control and cancer development(Bian *et al*, 2021; Kattan *et al*, 2022a; Li *et al*, 2017; Li *et al*, 2016a; Li *et al*, 2015; Li *et al*, 2016b; Seo *et al*, 2020b; Vargas *et al*, 2020b; Wang *et al*, 2011; Wang *et al*, 2012b; Wang *et al*, 2015b; Wang *et al*, 2014a; Wang *et al*, 2015c; Wang *et al*, 2015e). As for it, SBP is a small peptide that binds effectively to streptavidin beads, which can be easily eluted with a biotin-containing solution(Keefe *et al.*, 2001); S protein is another small peptide that binds efficiently with S protein beads and can be also used for affinity purification(Kim & Raines, 1993). The Flag tag in this system is used to detect bait protein expression by Western blot and verify its localization by immunofluorescence. As a routine practice in our lab, cells stably expressing the SFB-tagged bait protein are generated by either lentiviral infection or single colony isolation. After verifying the bait protein expression and localization using anti-Flag antibody, these stable cells will be expanded to a large scale and lysed for tandem affinity purification. As shown in **Figure 1.1F**, the first step of this purification method involves the use of streptavidin beads to isolate the associated protein complex by binding to SBP tag. After biotin elution, S protein beads are used for the second around of purification. The whole process can be finished in several hours and only one buffer solution (i.e., NETN buffer) is used from beginning to end, which greatly reduce the chance of protein degradation and protein complex dissociation.

In addition to purifying the associated protein complex from cell lysates, several library screening approaches are available for identifying interacting proteins for a protein of interest, such as yeast two-hybrid (Y2H)(Young, 1998), bimolecular fluorescence complementation (BIFC)(Kerppola, 2006), which are developed based on the protein-fragment complementation strategy(Li *et al*, 2019). For Y2H screening, a transcription factor is split into the DNA-binding

domain (DBD) and activating domain (AD). If AD and DBD are brought into proximity, they can activate downstream reporter gene by binding onto its upstream activating sequence (UAS). In general, a protein of interest is fused with DBD, while its candidate interacting proteins are fused with AD and prepared as a cDNA library for screening their interactions with the bait protein. PPI indirectly connects DBD and AD to activate the transcription of reporter gene through its UAS. For BIFC, a fluorescent protein (e.g., CFP, GFP, YFP) is split into two fragments, which are respectively fused with a bait protein and its candidate interacting proteins in a format of cDNA library. The binding between bait protein and its interacting proteins will bring the two fragments of a fluorescent protein within proximity, allowing the fluorescent protein to reform and emit its fluorescent signal. Detection and quantification of such fluorescent signal can be achieved by fluorescent microscope and flow cytometry. Notably, these two assays are both performed in live cell system, making the characterization of *in vivo* PPIs possible. However, since these complementation assays require generation of two separate fusion proteins, technical issues need to be taken into consideration, which include but are not limited to the effects on protein localization/function as caused by fragment fusion, overexpression-induced artificial effects, and issues with the reforming efficiency for the split fragments from two fusion proteins. Previous studies also raise concerns regarding high false positive and false negative rates for Y2H and BIFC assays. Technically, generating the large-scale cDNA library and setting up conditions for screening is costly and labor and time-consuming, making these assays difficult to be widely used by researchers.

MS data generation, submission, processing, and quality evaluation

Upon the completion of protein complex purification, sample can be prepared in a format either on beads or in a polyacrylamide gel and processed by a MS facility. After trypsin digestion, produced peptides are eluted through high-performance liquid chromatography (HPLC), subjected to electrospray ionization, and loaded into a mass spectrometer, where peptides are detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences (i.e., protein identity) are determined by matching protein databases (e.g., UniProt) with the fragmentation pattern acquired by the software program SEQUEST. Spectral matches are filtered to contain a false discovery rate (FDR) of less than 1% at the peptide level using the target-decoy method(Elias & Gygi, 2007). The protein inference is considered followed the general rules(Nesvizhskii & Aebersold, 2005) with manual annotation based on experiences applied when necessary.

Users will then be provided with an extensive list of identified proteins from the sample, which are referred as MS raw data. Now it has become a standard practice by scientific journals that these raw data should be shared through public repositories before publication(Prince *et al*, 2004). One commonly used repository is the PRoteomics IDEntifications database (PRIDE)(Jones *et al*, 2008; Ternent *et al*, 2014), which is designed to receive raw protein and peptide files for a MS experiment. To deposit MS data to the PRIDE, users first need to gather the MS data files from the MS facility including raw, result, search, and peak files. These files are then uploaded through the ProteomeXchange (PX) submission tool(Jarnuczak & Vizcaino, 2017), where a two-step assessment process via PRIDE is provided for checking data

quality(Perez-Riverol *et al*, 2019). After submission, a PX accession number and permanent digital object identifier (DOI) will be issued for publication use.

Next, we usually use a pipeline to deconvolute the MS raw data into a short list of high confident interacting proteins (HCIPs) for a bait protein (**Figure 1.2A**). To achieve so, a web-accessible resource named the contaminant repository for affinity purification (CRAPome)(Mellacheruvu *et al*, 2013) is often used to filter out commonly identified prey proteins (i.e., non-specific binding proteins) by comparing to control experiments provided by either CRAPome or users. Control experiments are a group of unrelated MS raw datasets that are usually produced under similar experimental settings(Mellacheruvu *et al.*, 2013). Based on the quantitative comparisons of prey abundance (using spectral counts) against the prey abundances across control experiments, a significance analysis of interactome (SAINT) score(Choi *et al*, 2011) will be assigned to each prey, allowing users to generate a list of HCIPs based on a suitable cutoff value of SAINT score.

Specifically, SAINT identifies false interactions by estimating the spectral count distribution from negative controls. For experiments that are produced with multiple replicates, a probability score is assigned to estimate the false discovery rate (FDR)(Choi *et al.*, 2011), allowing users to determine the reliability of interactions. Recently, SAINT has been updated to SAINTExpress(Teo *et al*, 2014), which provides a topology-assisted probability score (TopoAvgP), incorporating the prior knowledge of the target interactome into the scoring step. In addition, SAINTExpress provides a simpler fold-change (FC) score based on the ratio of averaged normalized spectral counts between experiments and controls.

In addition to CRAPome, scoring algorithms like comparative proteomic analysis software suite (CompPASS)(Sowa *et al.*, 2009), mass spectrometry interaction statistics (MiST)(Jager *et al*, 2011) and minkowski distance-based unified scoring environment (MUSE)[18] are also available for generating HCIP list. As compared to CompPASS and MiST, SAINT can be applied to datasets of all sizes and perform filtering quantification simply using spectral counts rather than other MS parameters(Choi *et al.*, 2011). In addition, SAINT removes interactions with spectral counts less than two, making the filtering process more robust(Choi *et al.*, 2011). Different from SAINT, MiST provides a more complete dataset analysis by incorporating multiple MS parameters-based measures, such as protein abundance (i.e., peak intensities), invariability of abundance over replicated experiments (i.e., reproducibility), uniqueness of an observed interaction across all purifications (i.e., specificity)(Jager *et al.*, 2011). It is plausible to use different methods to analyze the same MS raw data, although extra bioinformatic work will be required.

For publication purpose, we usually provide a MS data filtration report as a table by detailing the numbers of total experiments, control experiments, and identified peptides and proteins, respectively (**Figure 1.2B**). In addition, the cutoff value (e.g., a SAINT score) chosen for HCIP generation and the number of total HCIPs generated in this study are often included in this report table (**Figure 1.2B**).

To evaluate the quality of the produced HCIP dataset, additional analyses are usually performed. For example, total spectral count (TSC) can be presented along with the number of HCIPs (**Figure 1.2C**), allowing readers to evaluate the HCIP rate of each sample in the dataset.

Biological replicates are required to be performed for current proteomic studies; therefore, reproducibility is another key factor to assess the variation among experiments for the bait proteins(Tabb *et al*, 2010). This can be determined by calculating the HCIP correlation *R* value under different numbered peptides in the dataset (usually from low to high), where reproducibility rate at each peptide cutoff number can be shown (**Figure 1.2D**). Another way to qualitatively evaluate the produced HCIP dataset is to compare them with some available PPI databases, such as biological general repository for interaction data sets (BioGRID)(Chatr-Aryamontri *et al*, 2017), biophysical interactions of ORFeome-based complexes (BioPlex)(Huttlin *et al*, 2015), search tool for the retrieval of interacting genes/proteins (STRING)(Szklarczyk *et al*, 2017). These databases comprise numerous reported protein interactions that have been discovered through various experimental assays including AP-MS, proximity label-MS, yeast two-hybrid, immunoprecipitation, biochemistry, immunofluorescence. Comparing HCIPs with these reported interactions not only helps to evaluate the quality of the HCIP dataset from a different perspective, but also allows to reveal new PPIs through the current interactome study for functional investigation in future.

Annotation of HCIP dataset

Once a list of HCIPs have been generated, a standard practice is to deconvolute their underlying biological connections. This step is crucial for revealing potential functional processes, signaling pathways, cellular components, and human diseases that bait proteins may be involved through their HCIPs. To achieve so, several web-based annotation tools such as the database for annotation, visualization and integrated discovery (DAVID)(Huang da *et al*, 2009), protein analysis through evolutionary relationships (PANTHER)(Mi *et al*, 2005), and Metascape(Zhou *et*

al, 2019) are available for gene ontology (GO) analysis of HCIPs. The common ontologies described in these tools include biological process, molecular function, cellular components, pathways, diseases, and tissue distributions. As for a multiple-bait interactome study, bait proteins can be either analyzed individually for its HCIPs-associated GO terms (**Figure 1.3A**) or compared globally and visualized as a heatmap (**Figure 1.3B**). The latter usually helps reveal the bait proteins who share overlapped cellular functions through their HCIPs.

To reveal the shared HCIPs across different bait proteins, clustering methods through Circos plot (**Figure 1.3C**) or heatmap (**Figure 1.3D**) are often used. On a global scale, Circos plot can provide easy visualization of relationships between different bait proteins(Krzywinski *et al*, 2009), while heatmap can show details of the overlapped HCIPs among different bait proteins to reveal potential PPI sub-networks within a multi-bait interactome study.

Notably, innovative technologies like artificial intelligence (AI) machine learning greatly facilitate the predictions of protein structures and protein-protein interactions, providing additional tools for HCIP dataset annotation. For example, AlphaFold(Bryant *et al*, 2022; Jumper *et al*, 2021) and RoseTTAFold(Baek *et al*, 2021) are both open-sourced AI tools, which can provide structural information for the identified HCIPs and predict their protein complex formations with the bait protein. In addition, TissueNet(Barshir *et al*, 2013) and integrated interactions database (IID)(Kotlyar *et al*, 2016) tools can offer tissue expression information for HCIPs, while weighted gene co-expression network analysis (WGCNA)(Zhang & Horvath, 2005) can be incorporated to annotate functional correlations between bait protein and its HCIPs.

Visualization of protein-protein interaction (PPI) network

Once HCIPs have been generated and annotated, a PPI network can be built up to provide an informatic overview of bait protein-associated interactome. Such PPI network can be generated using Cytoscape(Cline *et al*, 2007; Shannon *et al*, 2003) through various freely available plugins(Saito *et al*, 2012) or R program, a programming language for statistical computing and graphic(Jones *et al*, 2018).

Visualizing PPI network can easily present PPIs identified from different experiments and is useful when looking for unique connections and patterns among bait and prey proteins(Merico *et al*, 2009). There are different ways visualizing a PPI network along with necessary experimental and/or biological information. Specifically, PPI network can be organized in a format of prey nodes surrounding their baits, where each node simply represents prey alone (**Figure 1.4A**) or prey with additional information. Moreover, these prey nodes can be complimented with further information by adding various "visual features", such as different shapes, sizes, colors, patterns, outline thickness, to convey their experimental and/or biological details. For example, the size of the node can convey its identified TSC, where larger nodes represent preys with more TSC (**Figure 1.4B**). Nodes also can be made in different colors to represent the prey-associated GO terms, such as biological processes, localization, cellular components (**Figure 1.4C**). Another commonly presented information is to incorporate data gathered from BioGRID or STRING database to indicate the known interacting proteins within the PPI network. In addition, if reciprocal MS studies are performed using the identified HCIPs as bait proteins, the PPI network can be further enhanced through their connecting lines using either uni- or bi- directional arrows to indicate the relationship between bait proteins and its HCIPs (**Figure 1.4D**). Simultaneously

visualizing all these representations (e.g., shape, size, color, line direction) will make a PPI network summarizing all the experimental and biological information comprehensively achieved.

Designing functional assays for characterizing newly identified HCIPs

After generating a list of HCIPs and annotating their biological functions, a group of HCIPs of interest first need to be experimentally validated. Typically, immunoprecipitation and pulldown are the choice assays for assessing the interaction between bait protein and its HCIPs. As mentioned earlier, antibody-based immunoprecipitation offers a way to examine protein-protein interaction at endogenous level (**Figure 1.5A**); however, difficulties may arise in identifying a suitable antibody for use. To combat this, protein can be fused with a tag (e.g., Flag, HA, Myc, GFP, SFB), whose antibody or antibody-conjugated beads are available to help validate the protein-protein interaction at a level of overexpression (**Figure 1.5B**). In addition, immunofluorescent staining is usually used to assess the co-localization between a bait protein and its HCIPs, confirming their complex formation from a different perspective (**Figure 1.5C**).

With the interaction between bait protein and its HCIPs validated, functional significance underlying their complex formation can be further explored. For example, the binding regions between bait protein and its HCIPs can be mapped in detail through generating a series of truncation and/or deletion mutants for both bait protein and its HCIPs (**Figure 1.5D**). In addition, we usually knockout (KO) or knockdown (KD) HCIPs in functionally relevant cells to examine the potential effects on the bait protein-dependent signaling events or biological functions (**Figure 1.5E**). If confirmed, these KO or KD cells will be reconstituted with wild-type HCIP or its mutant protein that fails to bind bait protein (**Figure 1.5D**) and used to determine whether

their complex formation is required for the related cellular functions. For example, to examine the roles of one HCIP of interest in regulating the bait protein-dependent cell proliferation and migration (**Figure 1.5F**), altered proliferation and migration will be examined in the HCIP KO cells. If there is a change in cell proliferation/migration, rescue experiments will be performed by reconstituting the HCIP KO cells with wild-type HCIP and its bait protein non-binding mutants (**Figure 1.5F**). Using this strategy, we can provide both functional and mechanistic insights into the newly identified HCIPs through the interactome study.

1.6 Conclusions and perspectives

In this study, we reviewed the commonly used methods and bioinformatic resources for characterizing the interacting proteins for a protein of interest and illustrated a pipeline for analyzing the related MS data. This proposed pipeline is mainly composed of three steps. First, it assigns each identified prey with a confidence score, allowing users to generate a list of HCIPs for a bait protein. Second, it provides a series of bioinformatic resources (**Table 1.1**) for users to annotate HCIPs, build up PPI network, and visualize interactome data informatically. Third, it suggests the strategies for follow-up data validation and functional investigation for newly identified HCIPs. In the past years, we have been frequently using this pipeline to define and characterize the PPI networks for different signaling pathways and protein families(Kattan *et al.*, 2022a; Li *et al.*, 2017; Li *et al.*, 2016a; Li *et al.*, 2015; Seo *et al.*, 2020b; Vargas *et al.*, 2020b; Wang *et al.*, 2014a), fully testifying its feasibility for addressing biological questions in different fields. Here, we would like to pinpoint several key factors that may affect the outcome of the interactome analysis for researchers who may be interested in trying this method for their own studies.

First, using different cell lines may lead to the difference in the identified HCIPs due to protein abundancy variation between cells; therefore, cell line choice should be considered prior to starting. In general, cell lines should be chosen based on scientific questions and their related biological contexts, while other issues, such as cell proliferation rate, cell culture costs, the way for cell collection, are also taken into consideration. We usually use HEK293T cells for protein complex purification due to their ease of growth and collection in a large quantity, but later move to functional cell lines to study biological functions for HCIPs.

Second, technical caveats for isolating associated protein complex for a bait protein should be taken into consideration. For example, choosing appropriate tag is crucial, as it could cause protein structure change and introduce false positive/negative hits. In addition, users should be aware of the positioning of the tag (e.g., *N* terminus, *C* terminus), as it may alter bait protein cellular localization and function. Regarding this point, several methods are available to characterize PPIs without using an epitope tag. For example, thermal proximity coaggregation (TPCA) approach can be used to examine protein complex dynamics in cells(Tan *et al*, 2018). Size exclusion chromatography can separate different protein complexes based on their size(Bloustine *et al*, 2003). Combined with MS analysis, these approaches thus provide options for elucidating interacting proteins for untagged bait protein, although additional factors could be introduced to affect protein complex formation (e.g., temperature, chromatography sample preparation). These approaches can be concurrently performed when analyzing limited bait proteins, while this strategy may not be feasible for a large-scale multi-bait interactome study.

Expressing a bait protein in cells will also bring in the overexpression issue. To address it, knock-in approach can be adopted to integrate a tag into the bait protein coding region via CRISPR technique, so we can directly purify endogenous bait protein-associated protein complex. To accelerate the progress, an inducible lentiviral system can be used to establish the stable cells for a bait protein, where doxycycline concentration is optimized to make the level of exogenously expressed bait protein close to that of endogenous one(Han *et al*, 2022b; Wang *et al.*, 2012b).

In addition, the steps of cell lysing and followed protein complex purification can lead to protein degradation, loss of weak and transient interactors, and non-specific binding(Miteva *et al*, 2013). To reduce these problems, isolating associated protein complex from cells should be finished in a timely manner and avoid frequent changes of buffer systems between different steps.

Third, identifying *bona fide* interacting partners can be hindered by a vast number of contaminants (i.e., non-specific binding proteins) during purification of bait protein-associated protein complex. This issue can be solved by including a group of control experiments for MS data filtration. Before that, users are suggested to carefully examine their control experiments to ensure that they are appropriate and unrelated to their bait proteins. Another way to reduce contaminants is to apply different purification approaches (e.g., TAP and BioID) (**Figure 1**) to the same bait protein and then compare their identified HCIPs. Not only would this confirm true binding proteins for a bait protein, but also allow the user to reduce the contaminants due to technical issues, thus making the interactome analysis more robust.

Lastly, interactome data analysis has become more comprehensive with online databases and software constantly evolving, allowing the generation of more informative PPI datasets based on needs. For example, PPI dataset can be further integrated with cancer-related databases (e.g., TCGA), which can help annotate the produced HCIPs from a cancer-related perspective and provide opportunities for identifying new therapeutic strategies for cancer treatment.

Collectively, we review the commonly used methods/resources for characterizing cellular PPI networks and propose a simple pipeline for researchers to process the related large-scale MS data. As mentioned earlier, deconvoluting the MS data into a list of HCIPs and validating the HCIP dataset are just the beginning of the study. The goal of the entire work is to reveal valuable HCIPs for in-depth functional studies to advance our understanding of the mechanisms underlying the related biological questions. We hope this work would not only help alleviate the fears newcomers may face when trying to piece together bioinformatic methods/resources to analyze large-scale proteomic data, but also aid users with a user-friendly pipeline that incorporates details behind the methods/resources needed to identify *bona fide* interacting proteins for their research.

1.7Author contribution

W.W. conceived and supervised the study; R.E.K., D.A. and W.W. wrote the manuscript.

1.8 Acknowledgment

We thank all the members in the Wang lab for constructive discussion. This work was supported by NIH grants (R01GM126048, R01GM143233) and American Cancer Society Research

Scholar grants (RSG-18-009-01-CCG, TLC-21-165-01-TLC) to W.W. This work was also supported in part by NIH National Center for Research Resources and the National Center for Advancing Translational Sciences (UL1 TR001414) through the UC Irvine Institute for Clinical and Translational Science (ICTS) pilot award.

1.9 Conflict of interest

The authors declare no competing financial interests.

Table 1.1 Summary of available bioinformatic resources for analyzing interactome-related proteomics data

Name	Annotation	Website	Reference
CRAPome	Data Filtering	https://reprint-apms.org/	(Mellacheruvu *et al.*, 2013)
CompPASS	Data Filtering		(Sowa *et al.*, 2009)
MUSE	Data Filtering		(Li *et al.*, 2016a)
MiST	Data Filtering	https://modbase.compbio.ucsf.edu/mist/	(Jager *et al.*, 2011)
BioPlex	Compare Datasets	https://bioplex.hms.harvard.edu/	(Huttlin *et al.*, 2015)
BioGRID	Compare Datasets	https://thebiogrid.org/	(Chatr-Aryamontri *et al.*, 2017)
AlphaFold	Structural Analysis	https://alphafold.ebi.ac.uk/	(Bryant *et al.*, 2022; Jumper *et al.*, 2021)
RoseTTAFold	Structural Analysis	https://www.ipd.uw.edu/2021/07/rosettafold-accurate-protein-structure-prediction-accessible-	(Baek *et al.*, 2021)
TissueNet	Tissue Association for PPI	https://netbio.bgu.ac.il/tissuenet3/	(Barshir *et al.*, 2013)
IID	Condition Association for PPI	http://iid.ophid.utoronto.ca/	(Kotlyar *et al.*, 2016)
WGCNA	Explore PPI via Gene Expression Profiles		(Zhang & Horvath, 2005)
PANTHER	Gene Ontology	http://pantherdb.org/	(Mi *et al.*, 2005)
Metascape	Gene Ontology	http://metascape.org/	(Zhou *et al.*, 2019)
DAVID Bioinformatics	Gene Ontology	https://david.ncifcrf.gov/	(Huang da *et al.*, 2009)
R Studio	Visualization/Graphs Clustering/Heatmaps	https://www.rstudio.com/	Open-Source License
Cytoscape	Visualization of PPI network	https://cytoscape.org/	(Cline *et al.*, 2007; Shannon *et al.*,

Table 1.1 Summary of available bioinformatic resources for analyzing interactome-related proteomics data. The commonly used bioinformatic resources for interactome studies are listed, which include the tools for proteomics data filtration, GO analysis, PPI databases for HCIP compare, PPI network visualization.

1.11 References

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

Baek M, DiMaio F, Anishchenko I, Dauparas J, Ovchinnikov S, Lee GR, Wang J, Cong Q, Kinch LN, Schaeffer RD *et al* (2021) Accurate prediction of protein structures and interactions using a three-track neural network. *Science* 373: 871-876

Barshir R, Basha O, Eluk A, Smoly IY, Lan A, Yeger-Lotem E (2013) The TissueNet database of human tissue protein-protein interactions. *Nucleic Acids Res* 41: D841-844

Bian W, Tang M, Jiang H, Xu W, Hao W, Sui Y, Hou Y, Nie L, Zhang H, Wang C *et al* (2021) Low-density-lipoprotein-receptor-related protein 1 mediates Notch pathway activation. *Dev Cell* 56: 2902-2919 e2908

Bloustine J, Berejnov V, Fraden S (2003) Measurements of protein-protein interactions by size exclusion chromatography. *Biophys J* 85: 2619-2623

Bryant P, Pozzati G, Elofsson A (2022) Improved prediction of protein-protein interactions using AlphaFold2. *Nat Commun* 13: 1265

Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, O'Donnell L, Oster S, Theesfeld C, Sellam A *et al* (2017) The BioGRID interaction database: 2017 update. *Nucleic Acids Res* 45: D369-D379

Choi H, Larsen B, Lin ZY, Breitkreutz A, Mellacheruvu D, Fermin D, Qin ZS, Tyers M, Gingras AC, Nesvizhskii AI (2011) SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods* 8: 70-73

Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B *et al* (2007) Integration of biological networks and gene expression data using Cytoscape. *Nat Protoc* 2: 2366-2382

Dunham WH, Mullin M, Gingras AC (2012) Affinity-purification coupled to mass spectrometry: basic principles and strategies. *Proteomics* 12: 1576-1590

Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4: 207-214

Han H, Nakaoka HJ, Hofmann L, Zhou JJ, Yu C, Zeng L, Nan J, Seo G, Vargas RE, Yang B *et al* (2022) The Hippo pathway kinases LATS1 and LATS2 attenuate cellular responses to heavy metals through phosphorylating MTF1. *Nat Cell Biol* 24: 74-87

Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 4: 44-57

Huttlin EL, Ting L, Bruckner RJ, Gebreab F, Gygi MP, Szpyt J, Tam S, Zarraga G, Colby G, Baltier K *et al* (2015) The BioPlex Network: A Systematic Exploration of the Human Interactome. *Cell* 162: 425-440

Jager S, Cimermancic P, Gulbahce N, Johnson JR, McGovern KE, Clarke SC, Shales M, Mercenne G, Pache L, Li K *et al* (2011) Global landscape of HIV-human protein complexes. *Nature* 481: 365-370

Janknecht R, de Martynoff G, Lou J, Hipskind RA, Nordheim A, Stunnenberg HG (1991) Rapid and efficient purification of native histidine-tagged protein expressed by recombinant vaccinia virus. *Proc Natl Acad Sci U S A* 88: 8972-8976

Jarnuczak AF, Vizcaino JA (2017) Using the PRIDE Database and ProteomeXchange for Submitting and Accessing Public Proteomics Datasets. *Curr Protoc Bioinformatics* 59: 13 31 11-13 31 12

Jones P, Cote RG, Cho SY, Klie S, Martens L, Quinn AF, Thorneycroft D, Hermjakob H (2008) PRIDE: new developments and new datasets. *Nucleic Acids Res* 36: D878-883

Jones PJ, Mair P, McNally RJ (2018) Visualizing Psychological Networks: A Tutorial in R. *Front Psychol* 9: 1742

Jumper J, Evans R, Pritzel A, Green T, Figurnov M, Ronneberger O, Tunyasuvunakool K, Bates R, Zidek A, Potapenko A *et al* (2021) Highly accurate protein structure prediction with AlphaFold. *Nature* 596: 583-589

Kattan RE, Han H, Seo G, Yang B, Lin Y, Dotson M, Pham S, Menely Y, Wang W (2022) Interactome Analysis of Human Phospholipase D and Phosphatidic Acid-Associated Protein Network. *Mol Cell Proteomics* 21: 100195

Keefe AD, Wilson DS, Seelig B, Szostak JW (2001) One-step purification of recombinant proteins using a nanomolar-affinity streptavidin-binding peptide, the SBP-Tag. *Protein Expr Purif* 23: 440-446

Kellermann OK, Ferenci T (1982) Maltose-binding protein from Escherichia coli. *Methods Enzymol* 90 Pt E: 459-463

Kerppola TK (2006) Design and implementation of bimolecular fluorescence complementation (BiFC) assays for the visualization of protein interactions in living cells. *Nat Protoc* 1: 1278-1286

Kim JS, Raines RT (1993) Ribonuclease S-peptide as a carrier in fusion proteins. *Protein Sci* 2: 348-356

Kotlyar M, Pastrello C, Sheahan N, Jurisica I (2016) Integrated interactions database: tissue-specific view of the human and model organism interactomes. *Nucleic Acids Res* 44: D536-541

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. *Genome Res* 19: 1639-1645

Lam SS, Martell JD, Kamer KJ, Deerinck TJ, Ellisman MH, Mootha VK, Ting AY (2015) Directed evolution of APEX2 for electron microscopy and proximity labeling. *Nat Methods* 12: 51-54

Li P, Wang L, Di LJ (2019) Applications of Protein Fragment Complementation Assays for Analyzing Biomolecular Interactions and Biochemical Networks in Living Cells. *J Proteome Res* 18: 2987-2998

Li X, Han H, Zhou MT, Yang B, Ta AP, Li N, Chen J, Wang W (2017) Proteomic Analysis of the Human Tankyrase Protein Interaction Network Reveals Its Role in Pexophagy. *Cell Rep* 20: 737-749

Li X, Tran KM, Aziz KE, Sorokin AV, Chen J, Wang W (2016a) Defining the Protein-Protein Interaction Network of the Human Protein Tyrosine Phosphatase Family. *Mol Cell Proteomics* 15: 3030-3044

Li X, Wang W, Wang J, Malovannaya A, Xi Y, Li W, Guerra R, Hawke DH, Qin J, Chen J (2015) Proteomic analyses reveal distinct chromatin-associated and soluble transcription factor complexes. *Mol Syst Biol* 11: 775

Li X, Wang W, Xi Y, Gao M, Tran M, Aziz KE, Qin J, Li W, Chen J (2016b) FOXR2 Interacts with MYC to Promote Its Transcriptional Activities and Tumorigenesis. *Cell Rep* 16: 487-497

Li Y (2011) The tandem affinity purification technology: an overview. *Biotechnol Lett* 33: 1487-1499

Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, Miteva YV, Hauri S, Sardiu ME, Low TY *et al* (2013) The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* 10: 730-736

Merico D, Gfeller D, Bader GD (2009) How to visually interpret biological data using networks. *Nat Biotechnol* 27: 921-924

Mi H, Lazareva-Ulitsky B, Loo R, Kejariwal A, Vandergriff J, Rabkin S, Guo N, Muruganujan A, Doremieux O, Campbell MJ *et al* (2005) The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res* 33: D284-288

Miteva YV, Budayeva HG, Cristea IM (2013) Proteomics-based methods for discovery, quantification, and validation of protein-protein interactions. *Anal Chem* 85: 749-768

Nesvizhskii AI, Aebersold R (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* 4: 1419-1440

Pawson T, Nash P (2000) Protein-protein interactions define specificity in signal transduction. *Genes Dev* 14: 1027-1047

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M *et al* (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47: D442-D450

Prince JT, Carlson MW, Wang R, Lu P, Marcotte EM (2004) The need for a public proteomics repository. *Nat Biotechnol* 22: 471-472

Rao VS, Srinivas K, Sujini GN, Kumar GN (2014) Protein-protein interaction detection: methods and analysis. *Int J Proteomics* 2014: 147648

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17: 1030-1032

Roux KJ, Kim DI, Raida M, Burke B (2012) A promiscuous biotin ligase fusion protein identifies proximal and interacting proteins in mammalian cells. *J Cell Biol* 196: 801-810

Ryan DP, Matthews JM (2005) Protein-protein interactions in human disease. *Curr Opin Struct Biol* 15: 441-446

Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, Pico AR, Bader GD, Ideker T (2012) A travel guide to Cytoscape plugins. *Nat Methods* 9: 1069-1076

Seo G, Han H, Vargas RE, Yang B, Li X, Wang W (2020) MAP4K Interactome Reveals STRN4 as a Key STRIPAK Complex Component in Hippo Pathway Regulation. *Cell Rep* 32: 107860

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

Smith DB, Johnson KS (1988) Single-step purification of polypeptides expressed in Escherichia coli as fusions with glutathione S-transferase. *Gene* 67: 31-40

Sowa ME, Bennett EJ, Gygi SP, Harper JW (2009) Defining the human deubiquitinating enzyme interaction landscape. *Cell* 138: 389-403

Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, Santos A, Doncheva NT, Roth A, Bork P *et al* (2017) The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res* 45: D362-D368

Tabb DL, Vega-Montoto L, Rudnick PA, Variyath AM, Ham AJ, Bunk DM, Kilpatrick LE, Billheimer DD, Blackman RK, Cardasis HL *et al* (2010) Repeatability and reproducibility in proteomic identifications by liquid chromatography-tandem mass spectrometry. *J Proteome Res* 9: 761-776

Tan CSH, Go KD, Bisteau X, Dai L, Yong CH, Prabhu N, Ozturk MB, Lim YT, Sreekumar L, Lengqvist J *et al* (2018) Thermal proximity coaggregation for system-wide profiling of protein complex dynamics in cells. *Science* 359: 1170-1177

Teo G, Liu G, Zhang J, Nesvizhskii AI, Gingras AC, Choi H (2014) SAINTexpress: improvements and additional features in Significance Analysis of INTeractome software. *J Proteomics* 100: 37-43

Ternent T, Csordas A, Qi D, Gomez-Baena G, Beynon RJ, Jones AR, Hermjakob H, Vizcaino JA (2014) How to submit MS proteomics data to ProteomeXchange via the PRIDE database. *Proteomics* 14: 2233-2241

Vargas RE, Duong VT, Han H, Ta AP, Chen Y, Zhao S, Yang B, Seo G, Chuc K, Oh S *et al* (2020) Elucidation of WW domain ligand binding specificities in the Hippo pathway reveals STXBP4 as YAP inhibitor. *EMBO J* 39: e102406

Wang W, Huang J, Chen J (2011) Angiomotin-like proteins associate with and negatively regulate YAP1. *J Biol Chem* 286: 4364-4370

Wang W, Huang J, Wang X, Yuan J, Li X, Feng L, Park JI, Chen J (2012) PTPN14 is required for the density-dependent control of YAP1. *Genes Dev* 26: 1959-1971

Wang W, Li N, Li X, Tran MK, Han X, Chen J (2015a) Tankyrase Inhibitors Target YAP by Stabilizing Angiomotin Family Proteins. *Cell Rep* 13: 524-532

Wang W, Li X, Huang J, Feng L, Dolinta KG, Chen J (2014) Defining the protein-protein interaction network of the human hippo pathway. *Mol Cell Proteomics* 13: 119-131

Wang W, Li X, Lee M, Jun S, Aziz KE, Feng L, Tran MK, Li N, McCrea PD, Park JI *et al* (2015b) FOXKs promote Wnt/beta-catenin signaling by translocating DVL into the nucleus. *Dev Cell* 32: 707-718

Wang W, Xiao ZD, Li X, Aziz KE, Gan B, Johnson RL, Chen J (2015c) AMPK modulates Hippo pathway activity to regulate energy homeostasis. *Nat Cell Biol* 17: 490-499

Young KH (1998) Yeast two-hybrid: so many interactions, (in) so little time. *Biol Reprod* 58: 302-311

Zhang B, Horvath S (2005) A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* 4: Article17

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10: 1523

1.12 Figures

Figure 1.1

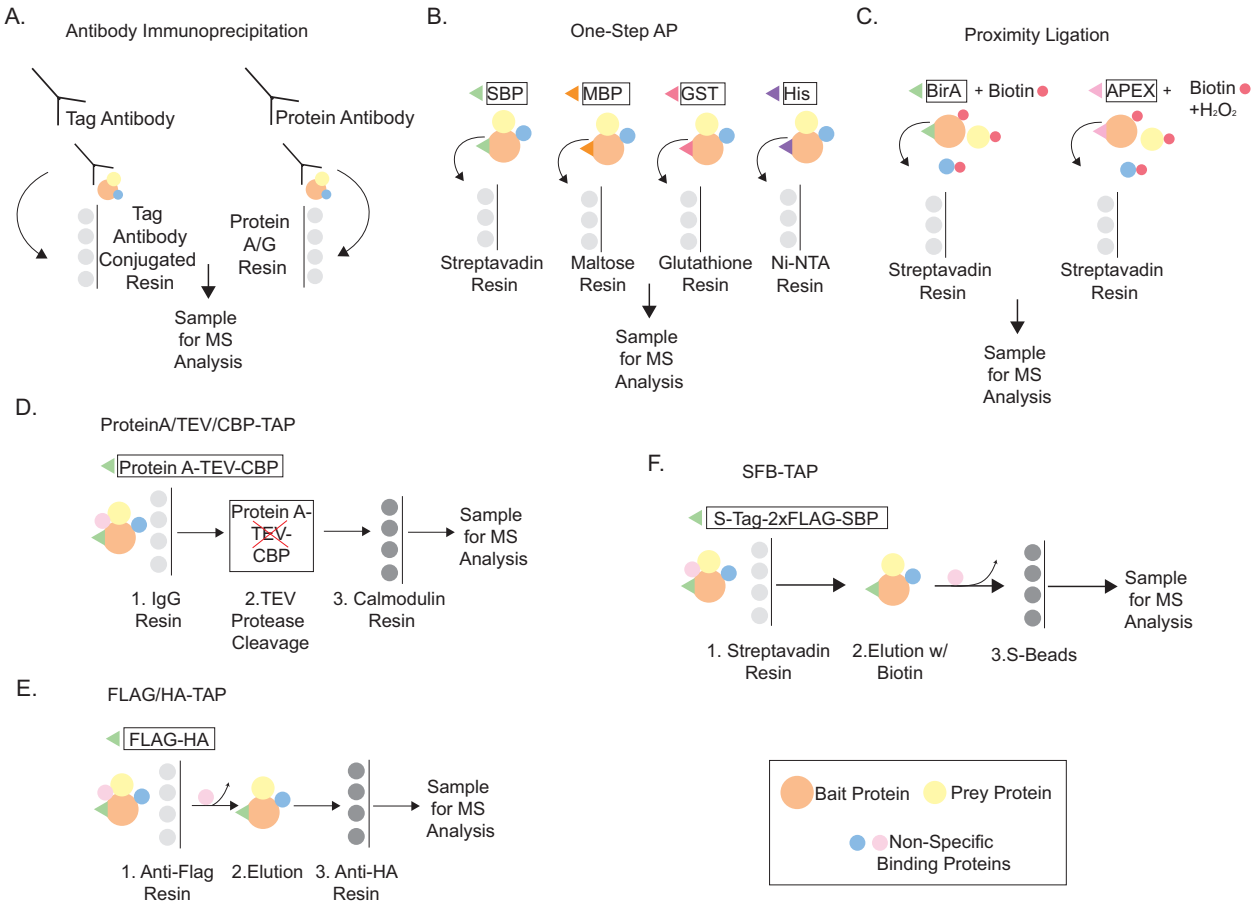


Figure 1.1. Illustration of commonly used methods for isolating associated protein complex for a protein of interest.

The commonly used methods for purifying bait protein-associated protein complex from cells for mass spectrometry (MS) analysis include antibody-based immunoprecipitation (**A**), one-step affinity purification (AP) (**B**), proximity ligation-based AP (**C**), ProteinA/TEV/CBP-based tandem affinity purification (TAP) (**D**), FLAG/HA-based TAP (**E**), and SFB-based TAP (**F**).

Figure 1.2

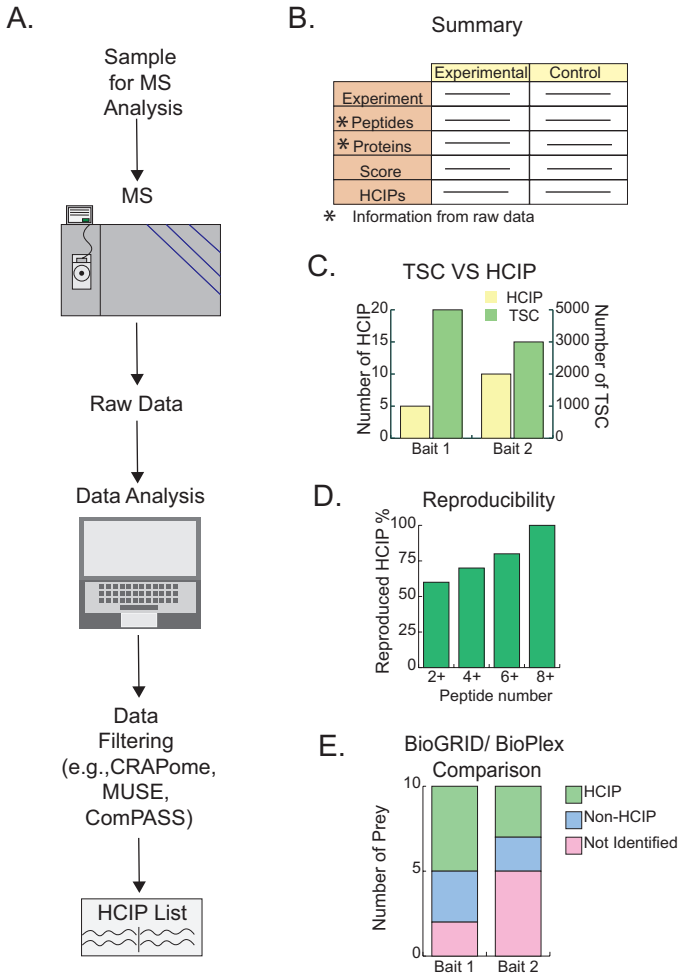


Figure 1.2. Schematic overview of the interactome-related proteomics data processing.

(**A**) Schematic workflow for the filtration of mass spectrometry (MS) data to generate a high confidence interacting proteins (HCIP) list used for further bioinformatics analysis.

(**B**) Organization of MS summary data can include number of experiments, peptide information, protein information, cutoff score for HCIP generation, and HCIP count.

(**C**) Schematic illustration of the comparison between the total spectral counts (TSCs) and HCIP number for an interactome study.

(**D**) Reproducibility rate can be illustrated based on different prey peptide cutoffs.

(**E**) HCIPs can be compared with public PPI resources like BioGRID and BioPlex to reveal known and new interactions in the produced interactome dataset.

Figure 1.3

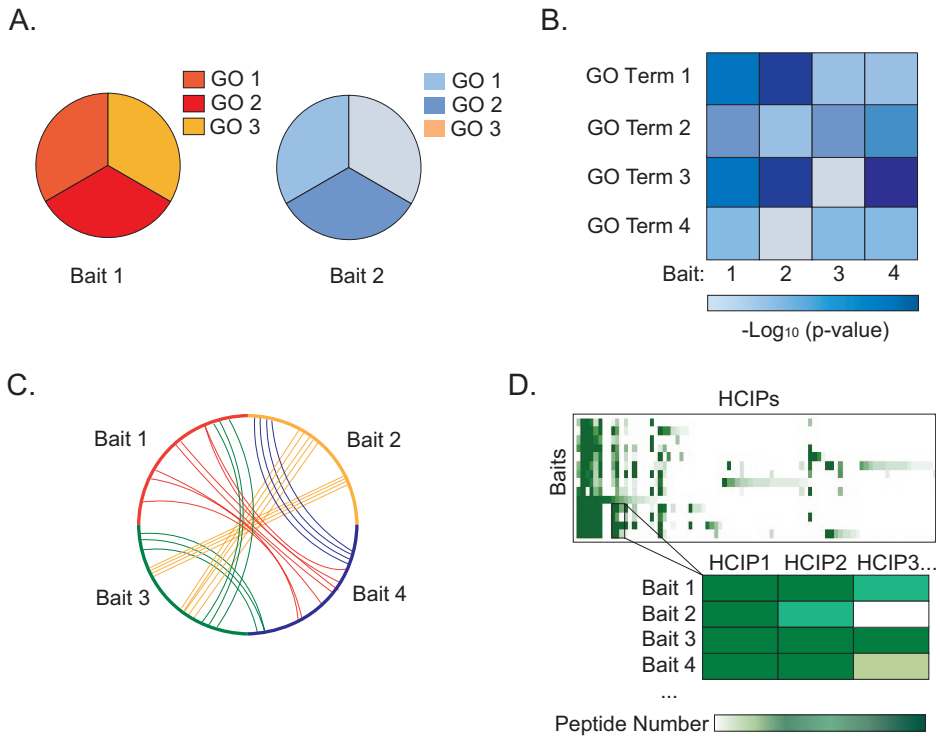


Figure 1.3. Gene ontology (GO) and topological analyses of interactome data.

(**A**) GO analysis can be applied and presented for bait proteins individually.

(**B**) GO analysis can be applied across a series of bait proteins and presented as a heatmap.

(**C**) Topological analysis of multi-bait interactome data can be presented as Circos plot to reveal the bait proteins that share overlapping HCIPs.

(**D**) Topological analysis of multi-bait interactome data can be presented as heatmap to reveal the overlapping HCIPs among bait proteins.

Figure 1.4

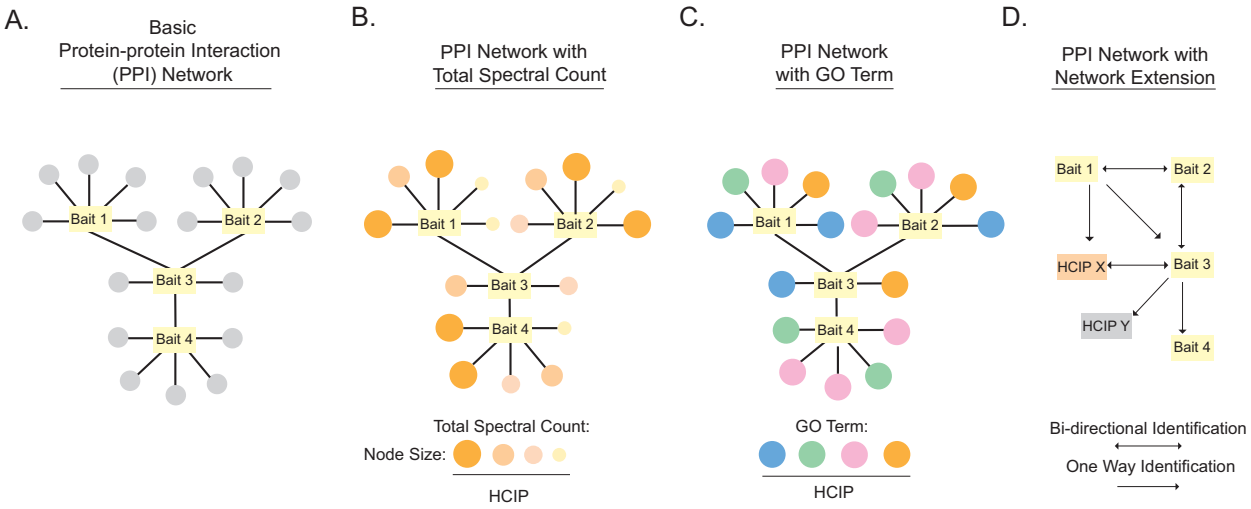


Figure 1.4. Visualization of protein-protein interaction (PPI) networks.

(**A**) A PPI network can be organized in a format of prey nodes surrounding their baits, where nodes simply represent preys alone.

(**B**) A PPI network can be organized in a format of prey nodes surrounding their baits, where nodes represent preys and their total spectral count (TSC). The size of the node conveys its identified TSC, where larger nodes represent preys with more TSC.

(**C**) A PPI network can be organized in a format of prey nodes surrounding their baits, where nodes represent preys and their GO terms that are labeled in different colors.

(**D**) The information from reciprocal MS studies using HCIPs as bait proteins can be included in the PPI network and indicated as connecting lines with either uni- or bi- directional arrows.

Figure 1.5

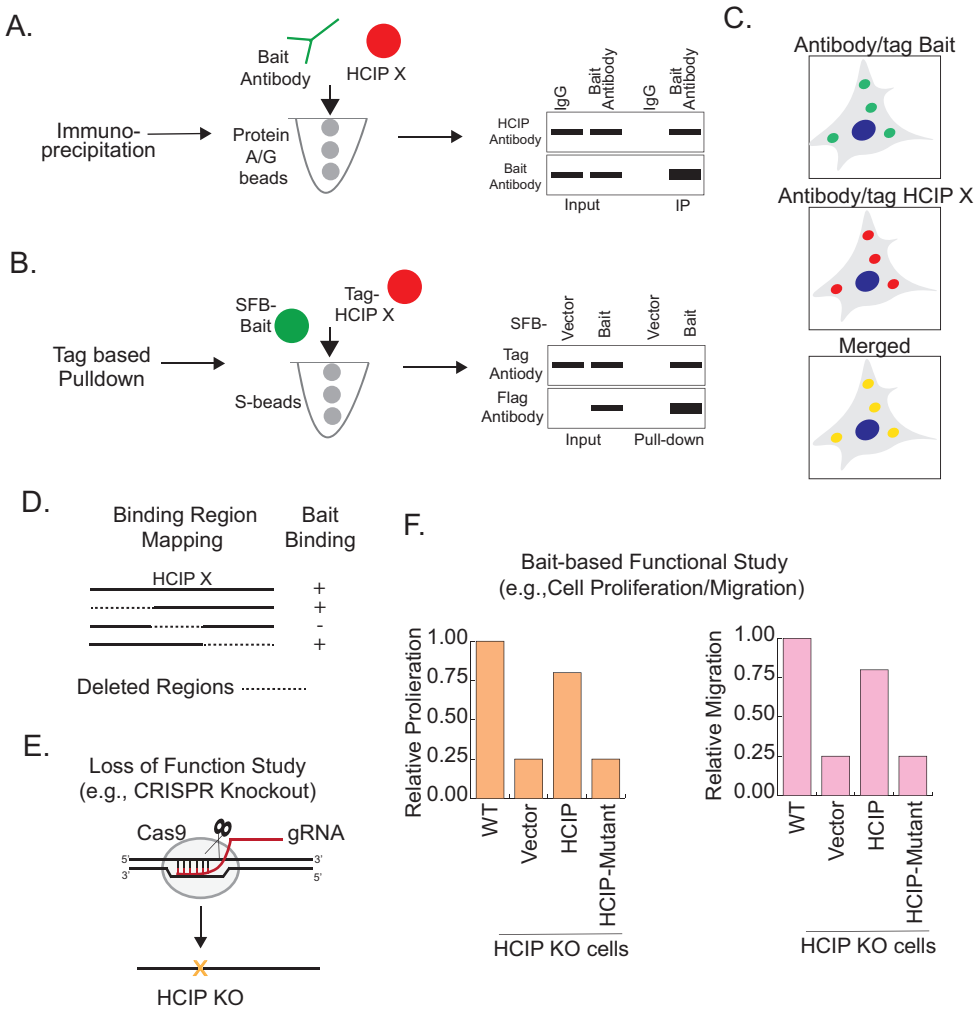


Figure 1.5. Validation and functional characterization of HCIPs produced through an interactome study.

(**A-B**) Illustration of experimental assays used for validating the complex formation between bait proteins and HCIPs. To examine the interaction between a bait protein and its HCIP, cell lysates can be subjected to immunoprecipitation (**A**) and pulldown (**B**) assays.

(**C**) Immunofluorescence assay can be used to examine the co-localization of bait proteins with their HCIPs.

(**D**) The bait protein-HCIP complex formation can be further characterized through mapping the regions required for their interaction.

(**E**) Illustration of the HCIP knockout (KO) cell generation using CRISPR/Cas9 technique.

(**F**) Rescue experiments can be performed in the HCIP KO cells to determine whether the bait protein binding is required for HCIP to regulate the related cellular functions such as cell proliferation and migration.

CHAPTER 2

Elucidation of the Hippo WW domain determinants reveals STXBP4 as a Hippo pathway regulator

Rebecca Vargas,[1,9] Vy Thuy Duong,[2,9] Han Han,[1,9] Albert Paul Ta,[1] Yuxuan Chen,[1] Shiji Zhao,[1] Bing Yang,[1] Gayoung Seo,[1] Kimberly Chuc,[1] Sunwoo Oh,[1] Amal El Ali,[3] Olga Razorenova,[3] Junjie Chen,[4,*] Ray Luo,[3,5,6,7,*] Xu Li,[8,*] and Wenqi Wang[1,10,*]

Affiliations

[1]Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA 92697, USA

[2]Department of Chemistry, University of California, Irvine, Irvine, CA 92697, USA

[3]Department of Molecular Biology and Biochemistry, University of California, Irvine, Irvine, CA 92697, USA

[4]Department of Experimental Radiation Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

[5]Department of Chemical and Biomolecular Engineering, University of California, Irvine, Irvine, CA 92697, USA

[6]Department of Materials Science and Engineering, University of California, Irvine, Irvine, CA 92697, USA

[7]Department of Biomedical Engineering, University of California, Irvine, Irvine, CA 92697, USA

[8]School of Life Sciences, Westlake University, Hangzhou, Zhejiang Province 310024, China

[9]These authors contributed equally to this work

[10]Lead Contact

[*]Correspondence: jchen8@mdandersonorg.org (J.C.), rluo@uci.edu (R.L.), lixu@westlake.edu.cn (X.L.) and wenqiw6@uci.edu (W.W.)

This chapter is derived from the manuscript accepted for publication in *The EMBO Journal:*

The EMBO Journal. 2019 Accepted for publication.

Vargas *et al*

2.1 Abstract

The Hippo pathway, which plays a critical role in organ size control and cancer, is featured with the WW domain-based protein-protein interaction. However, ~100 WW domains and 2,000 PY motif-containing peptides have been found in human proteome, raising a "WW-PY" binding specificity issue for the Hippo pathway. In this study, we demonstrated the WW domain binding specificity for the Hippo pathway components and uncovered a unique amino acid sequence required for it. By using this criterion, we identified a WW domain-containing protein, STXBP4 as a negative regulator of YAP. Mechanistically, STXBP4 assembles a protein complex comprising α-catenin and a group of Hippo PY motif-containing components/regulators to inhibit YAP, a process that is regulated by actin cytoskeleton tension. Interestingly, STXBP4 is a potential tumor suppressor for human kidney cancer, whose downregulation is correlated with YAP activation in clear cell renal cell carcinoma. Taken together, our study not only elucidates the WW domain binding specificity for the Hippo pathway, but also reveals STXBP4 as a player in actin cytoskeleton tension-mediated Hippo pathway regulation.

2.2 Key words

WW domain, PY motif, the Hippo pathway, YAP, STXBP4

2.3 Introduction

Signaling proteins often entail modular domains that facilitate protein-protein interactions to assemble functional protein complexes, control enzymatic activity and regulate protein cellular localization(Cohen *et al*, 1995; Pawson & Scott, 1997). Importantly, the recognition between domains and their peptide ligands is usually specific, thus allowing the transduction of unique information through signaling cascades(Das & Smith, 2000; Hu *et al*, 2004). The WW domain is a small protein module that is defined by the presence of two tryptophan (W) residues separated apart by ~25 amino acids(Sudol *et al*, 1995d). WW domain and its cognate proline-rich peptide motif have been identified within various protein complexes widely distributed in plasma membrane, cytoplasm and nucleus. Failure of their recognition is associated with multiple human diseases including Alzheimer's disease(Liu *et al*, 2007; Mandelkow & Mandelkow, 1998), Huntington's disease(Faber *et al*, 1998b; Passani *et al*, 2000b), Liddle Syndrome(Hansson *et al*, 1995), Golabi-Ito-Hall Syndrome(Lubs *et al*, 2006b; Tapia *et al*, 2010b), muscular dystrophy(Bork & Sudol, 1994b; Ervasti, 2007; Rentschler *et al*, 1999a) and cancers(Chang *et al*, 2007a; Salah & Aqeilan, 2011a). These facts highlight a crucial role of the WW domain-mediated protein-protein interaction in biological processes and tissue homeostasis.

WW domain was initially uncovered by characterizing the protein sequence of YAP, a key transcriptional co-activator downstream of the Hippo pathway(Jiang *et al*, 2015; Pan, 2010a; Sudol *et al*, 1995a). The Hippo pathway is a highly conserved signaling pathway involved in tissue homeostasis, organ size control and cancer development(Halder & Johnson, 2011; Jiang *et al.*, 2015; Pan, 2010a; Piccolo *et al*, 2014; Yu *et al*, 2015b). In mammals, the Hippo pathway is composed of a kinase cascade (two serine/threonine kinases, MST and LATS; and the adaptors

SAV1 for MST and MOB1 for LATS), downstream effectors (YAP and TAZ), and nuclear transcriptional factors (TEADs). MST phosphorylates and activates LATS, which in turn phosphorylates YAP and TAZ. The phosphorylated YAP/TAZ can be recognized by 14-3-3 proteins, retained in the cytoplasm and eventually targeted by □-TRCP E3 ligase complex for degradation. When the Hippo pathway is inactivated, unphosphorylated YAP/TAZ enter into the nucleus, where they associate with TEAD transcriptional factors to promote the transcription of genes that are involved in proliferation and survival.

Notably, many Hippo pathway components and regulators contain either the WW domain or its proline-rich peptide ligand, mostly "PPxY" motif (P, proline; Y, tyrosine; x, any amino acid; hereafter named as "PY" motif)(Salah & Aqeilan, 2011a; Sudol, 2010b). YAP, TAZ, SAV1 and KIBRA, an upstream component of the Hippo kinase cascade(Yu *et al*, 2010), are four known WW domain-containing components of the Hippo pathway(Salah & Aqeilan, 2011a). In the nucleus, the WW domain of YAP/TAZ is a requirement for their association with a group of nuclear transcriptional factors and regulators that contain the PY motif to regulate gene transcription (Chang *et al*, 2018; Ferrigno *et al*, 2002a; Haskins *et al*, 2014a; Liu *et al*, 2018; Qiao *et al*, 2016; Strano *et al*, 2005; Strano *et al*, 2001; Zhang *et al*, 2011). In the cytoplasm, the PY motif of LATS1/2 is involved in the LATS1/2-mediated YAP/TAZ phosphorylation(Hao *et al*, 2008b; Verma *et al*, 2018); several PY motif-containing proteins can physically bind the WW domain of YAP/TAZ and promote YAP/TAZ's cytoplasmic translocation(Chan *et al*, 2011; Espanel & Sudol, 2001; Liu *et al*, 2013b; Michaloglou *et al*, 2013; Tavana *et al*, 2016; Wang *et al.*, 2011; Wang *et al.*, 2012b; Wang *et al.*, 2014a; Zhao *et al*, 2011b). Moreover, the phosphorylated YAP/TAZ can negatively regulate Wnt pathway by forming a complex with

DVL2, which is mediated by the WW domain of YAP/TAZ and the PY motif of DVL2(Varelas *et al*, 2010). As a Hippo upstream component, KIBRA can similarly associate with several Hippo PY motif-containing proteins and negatively regulate YAP(Tavana *et al.*, 2016; Wilson *et al*, 2014b). On the other hand, several WW domain-containing proteins have been shown to modulate the Hippo pathway activity by regulating the Hippo PY motif-containing components and regulators(Abu-Odeh *et al*, 2014; Salah *et al*, 2013; Salah *et al*, 2011; Ulbricht *et al*, 2013; Wang *et al*, 2015a; Yeung *et al*, 2013). Collectively, these facts suggest that the WW domain and PY motif-mediated protein-protein interaction plays a fundamental role in building up the major framework of the Hippo pathway.

Actually, ~100 WW domains and 2,000 PY motif-containing peptides have been predicted in the human proteome(Tapia *et al.*, 2010b), raising an issue of binding specificity for the proteins containing WW domain and PY motif. Indeed, a large scale of WW domain array screen only confirmed 10% of the tested WW domain-ligand interactions(Hu *et al.*, 2004). Several large-scale proteomic studies exclusively identified a group of PY motif-containing proteins (e.g., LATS1/2, AMOTs, PTPN14) as the binding partners for the Hippo WW domain-containing components(Couzens *et al*, 2013b; Hauri *et al*, 2013b; Wang *et al.*, 2014a). These facts indicate the binding specificity for the Hippo WW domain-mediated protein-protein interaction, while the underlying mechanism is still largely unknown.

In this study, we demonstrated the WW domain binding specificity for the Hippo pathway proteins and uncovered a highly conserved amino acid sequence required for it. By using this criterion, we identified STXBP4 as a novel Hippo pathway regulator in human proteome.

Mechanistically, STXBP4 assembled a complex with α-catenin and several Hippo PY motif-containing components/regulators to negatively regulate YAP when actin cytoskeleton tension is low. Moreover, both TCGA data and tissue array studies suggested STXBP4 as a potential tumor suppressor in human kidney cancer, whose downregulation is significantly correlated with YAP activation in clear cell renal cell carcinoma. Collectively, our study not only elucidated the WW domain binding specificity for the Hippo pathway protein-protein interaction network, but also identified STXBP4 as a Hippo pathway regulator and a potential tumor suppressor in kidney cancer development.

2.4 Results

Binding specificity exists for the Hippo WW domain-containing components

We re-analyzed our previously published proteomic data(Wang *et al.*, 2014a) for four Hippo WW domain-containing components YAP, TAZ, SAV1 and KIBRA (**Figure 2.1A**), and found that most of the known Hippo PY motif-containing proteins (e.g., AMOT, AMOTL1, AMOTL2, LATS1, LATS2, PTPN14, PTPN21, WBP2) were hardly detected in the SAV1-associated protein complex (**Figure 2.1B**). Moreover, proteomic analysis of the WW domains isolated from these four Hippo components (**Figure S2.1A**) further confirmed this finding, where the WW domain of YAP, TAZ and KIBRA, but not that of SAV1, retrieved most of these known Hippo PY motif-containing proteins (**Figure 2.1B**). These data suggest that the WW domain of SAV1 is different from that of YAP, TAZ and KIBRA in associating with the known Hippo PY motif-containing proteins.

Next, we expanded our proteomic analysis for additional 22 WW domain-containing proteins (**Figure S2.1B**; **Tables S2.1-S2.3**) and examined their ability to isolate these known Hippo PY motif-containing proteins. Consistent with previous reports(Abu-Odeh *et al.*, 2014; Salah *et al.*, 2013; Ulbricht *et al.*, 2013; Wang *et al.*, 2015a; Yeung *et al.*, 2013), WWOX, BAG3 and members of the HECT family of E3 ligases NEDD4L, WWP1 and WWP2 were found to form complexes with the Hippo PY motif-containing proteins such as AMOT family proteins, CCDC85C and WBP2 (**Figure 2.1C**). However, we failed to identify these Hippo PY motif-containing proteins as the binding proteins for other tested WW domain-containing proteins (**Figure 2.1C**). Moreover, the high-confident interacting proteins (HCIPs) of the Hippo WW domain-containing components were involved in different signaling pathways from those of the control WW domain-containing proteins (**Figure 2.1D** and **Table S2.4**). We also performed proteomic analysis for the WW domains isolated from 13 randomly selected WW domain-containing proteins, and found that only 10.2% of the HCIPs were shared by the Hippo and control WW domains (**Figure S2.1C**). Taken together, these results indicate that the WW domains of the Hippo pathway components YAP, TAZ and KIBRA possess a binding specificity with the known Hippo PY motif-containing proteins.

Validation of the Hippo WW domain binding specificity

To validate our proteomic findings, we examined the interaction between a series of WW domain-containing proteins and AMOT family proteins. Unlike YAP, TAZ and KIBRA, SAV1 failed to bind AMOT and AMOTL1 (**Figure 2.1E**). Consistently, we hardly detected the association between SAV1 and LATS1 in our experimental setting (**Appendix Figure S2.7A**). Moreover, BAG3, WWOX and several members of the HECT family of E3 ligases can interact

with AMOT proteins (**Figure 2.1E**), which is consistent with our proteomic study (**Figure 2.1C**). However, other tested WW domain-containing proteins as well as their derived WW domains failed to bind AMOT family proteins (**Figures 2.1E** and 2.**1F**). These results demonstrate the WW domain binding specificity for the Hippo pathway proteins.

A highly conserved amino acid sequence is required for the Hippo WW domain binding specificity

To further explore the underlying mechanism, we analyzed the WW domain protein sequence for the Hippo pathway components as well as WWOX, BAG3 and several members of the HECT family of E3 ligases, which can bind the known Hippo PY motif-containing proteins (**Figure 2.2A**). Interestingly, in addition to the two tryptophan residues, additional 9 amino acids were found to be highly conserved among these WW domains (**Figure 2.2A**). We hypothesized that this conserved 9-amino acid sequence could be required for the specific association with the known Hippo PY motif-containing proteins.

To test this hypothesis, we examined the identified 9-amino acid sequence in the control WW domain-containing proteins that failed to bind the Hippo PY motif-containing proteins (**Figure 2.1C**) and found that their WW domains have at least one of these 9 amino acids replaced by other residues (**Figures 2.2B** and **S2.2A**). As for SAV1, the conserved glutamate residue within this 9-amino acid sequence was found changed to a serine in its WW domain (**Figure 2.2A**). Consistently, mutating either of these identified 9 amino acids to alanine dramatically disrupted the association of AMOT with TAZ (**Figure 2.2C**) or its WW domain (**Figure 2.2D**). Similar findings were also observed for both KIBRA (**Figure 2.2E**) and YAP (**Figure 2.2F**). Notably,

mutations of the G and E residues among these identified 9 amino acids are less detrimental to the Hippo WW-PY interaction as compared with other identified sites (**Figures 2.2C-2.2E**). We also tested the conservative substitution for the "E/D", "Y/F" or "F/Y" of this conserved amino acid sequence, and found that the association of AMOT with TAZ and KIBRA was not affected by these substitutions (**Appendix Figure S2.7B**). Interestingly, an interaction between SAV1 and AMOT was recovered when the unmatched serine residue was replaced by glutamate, allowing SAV1 WW domain to fit the 9-amino acid sequence criterion (**Figure 2.2G**). Taken together, these results demonstrate that the identified 9-amino acid sequence determines the WW domain binding specificity for the Hippo pathway proteins.

We also examined the Hippo WW domain-containing components in *Drosophila* and found that this 9-amino acid sequence was highly conserved in the WW domain of *Yorkie* and *Kibra*, while *Salvador* similarly contains a replacement of the conserved glutamate residue by alanine (**Appendix Figure S2.7**). By taking YAP as an example, conservation of this 9-amino acid sequence in the YAP WW domains can be even tracked to *Capsapsora owczarzaki* (**Figure S2.2B** and **Table S2.5**), an unicellular specie that is known to contain the functional Hippo pathway components(Sebe-Pedros *et al*, 2012). Interestingly, in *Capsapsora owczarzaki*, a PY motif was also identified in LATS (**Figure S2.2C**), suggesting that this conserved 9-amino acid sequence may play a crucial role for the Hippo pathway at its premetazoan origin.

Role of the 9-amino acid sequence in assembly of a specific WW-PY complex involving the Hippo pathway proteins

Next, we analyzed a NMR solution structure of the YAP-WW1 domain (the first WW domain of YAP) and SMAD7-PY motif-containing peptide complex(Aragon *et al*, 2012). Interestingly, the identified 9 amino acids form as two functional groups.

First, together with the second tryptophan (W199 of YAP-WW1), the conserved residues E178, Y188, H192 and T197 were involved in the binding interface with the SMAD7-PY motif (**Figure S2.3A**). Specifically, hydrogen bond (H-bond) formation was respectively paired between H192 (YAP-WW1 domain) and Y211 (SMAD7-PY motif), and T197 (YAP-WW1 domain) and P209 (SMAD7-PY motif) (**Figures S2.3B** and **S2.3C**). Hydrophobic contact not only existed within the intramolecular interaction between the W199 and Y188 residues of YAP1-WW domain, but also mediated their intermolecular interaction with the P208 and P209 residues within SMAD7-PY motif, respectively (**Figures S2.3B** and **S2.3C**). E178 (YAP-WW1 domain) functioned in sustaining the intermolecular contact between H192 (YAP-WW1 domain) and Y211 (SMAD7-PY motif) by forming both electrostatic and H-bonding interactions with H192 (**Figures S2.3B** and **S2.3C**).

Second, together with the first tryptophan (W177 of YAP-WW1 domain), the rest residues L173, P174, G176, F189 and P202 formed a hydrophobic cluster at the backside of the YAP-WW1/SMAD7-PY complex (**Figures S2.3A** and **S2.3C**). Although not directly interacted with SMAD7-PY motif, this hydrophobic cluster may maintain a unique YAP-WW1 domain structure to facilitate its binding with SMAD7-PY motif. Since these hydrophobic cluster residues are also frequently replaced by other amino acids in the non-Hippo WW domains (**Figures 2.2B** and **S2.2A**), we consider them as part of the determinants for the specific Hippo WW-PY recognition.

To further determine the role of this identified 9-amino acid sequence from a structure-based perspective, we mutated each of these conserved residues into alanine *in silico* and performed root-mean-square deviation (RMSD) analyses using the average unbound (apo) structure of YAP-WW1 domain as a reference. Interestingly, mutating either of the identified residues within the backside hydrophobic cluster significantly altered the YAP-WW1 protein structure as indicated by their relatively high RMSD values, while this was not the case for the residues within the binding interface with SMAD7-PY motif (**Figure S2.3D**). These results further confirm the hypothesis that the backside hydrophobic cluster may play a role in maintaining a functional YAP-WW1 structure. In addition, mutating either of the conserved residues altered the complex structure (**Figure S2.3E** and **Appendix Figure S2.8A**) and increased the average distance between YAP-WW1 domain and SMAD7-PY motif peptide (**Figure S2.3F**), indicating the intervention of their complex formation. As a control, we analyzed a NMR solution structure of the APBB3-WW domain (**Appendix Figure S2.8B**). The APBB3-WW domain failed to bind the Hippo PY motif-containing proteins (**Figure 2.1F**), since it contains two unmatched residues (as compared to the identified 9-amino acid sequence) locating in the PY motif binding interface (**Figures 2B** and **S2.2A**; **Appendix Figure S2.8B**). Consistently, the average distance between APBB3-WW domain and SMAD7-PY motif peptide is comparable to that between YAP-WW1 domain mutants and SMAD7-PY motif peptide (**Figure S2.3F**), suggesting an unstable complex formation for APBB3-WW domain and SMAD7-PY motif. Notably, the standard deviation of average distance value for both YAP-WW1 domain mutants and APBB3-WW domain complexes is relatively larger than that of the control YAP-WW domain complex (**Figure**

S2.3F), indicating a substantial movement between SMAD7-PY motif peptide and the YAP-WW1 domain mutants as well as APPB3-WW domain.

Taken together, these simulation analyses suggest that the identified 9-amino acid sequence is involved in binding PY motif and maintaining a unique WW domain structure, which both determine the Hippo WW domain binding specificity with the known Hippo PY motif-containing proteins.

Identification of STXBP4, a WW domain-containing protein, whose WW domain fits the 9-amino acid sequence criterion

Next, we searched all the WW domain-containing proteins in the human proteome and identified 12 WW domain-containing proteins whose WW domains fit such a 9-amino acid sequence (**Figure S2.4** and **Table S2.6**). Among them, role of STXBP4 in the Hippo pathway regulation has not been fully characterized (**Figure S2.4**). Although no STXBP4 ortholog is identified in *Drosophila*, this 9-amino acid sequence of the STXBP4 WW domain was largely conserved in different species (**Figure 2.3A**). Interestingly, STXBP4 can form a complex with several Hippo PY motif-containing regulators including AMOT, AMOTL2 and PTPN14 (**Figure 2.3B**). Mutating either of the conserved 9-amino acid residues diminished the interaction between STXBP4 and AMOT (**Figure 2.3C**). As expected, the association between STXBP4 and these PY motif-containing Hippo regulators are mediated by the WW domain of STXBP4 (**Appendix Figure S2.9A**) and the PY motif of these Hippo regulators (**Appendix Figure S2.9B**).

To gain a structural insight into the STXBP4 WW domain, we compared STXBP4-WW and YAP-WW1 through ensemble molecular dynamics simulations and calculating binding free energies (ΔG) using the molecular mechanics Poisson-Boltzman surface area (MM/PBSA) method. As shown in **Appendix Figure S2.9C**, the top 5 predicted clusters for the STXBP4-WW/SMAD7-PY complex is similar to those of the YAP-WW1/SMAD7-PY complex. By comparing the top one cluster for these two WW-PY complexes, we found that the identified 9-amino acid residues as well as the two tryptophan residues are similarly distributed within both the STXBP4-WW/SMAD7-PY and YAP-WW1/SMAD7-PY complexes, where they form as two groups to respectively involve in the binding with SMAD7-PY motif and assemble a supportive backside hydrophobic cluster for each WW domain (**Figure 2.3D**). The average distance between STXBP4-WW domain and SMAD7-PY motif is close to that between YAP-WW1 domain and SMAD7-PY motif with a similarly low standard deviation value (**Figure S2.3F**). Moreover, binding free energy (ΔG) from MM/PBSA calculations further indicates the similarity between YAP-WW1 and STXBP4-WW when they form as a complex with SMAD7-PY motif peptide (**Figure 2.3E**). Taken together, these data suggest that the STXBP4 WW domain possesses the Hippo WW domain binding specificity, endowing STXBP4 a potential role in the Hippo pathway.

STXBP4 is a negative regulator of YAP

To test the role of STXBP4 in regulation of the Hippo pathway, we examined YAP activation in the STXBP4 knockout (KO) cells (**Appendix Figure S2.10**). Interestingly, loss of STXBP4 significantly reduced YAP phosphorylation (**Figure 2.3F**), moved YAP into the nucleus (**Figure 3G**) and activated YAP downstream gene transcription (**Figure 2.3H**). Notably, either deleting

the WW domain or mutating the histidine residue out of the identified 9-amino acid sequence to alanine failed to rescue YAP's cytoplasmic localization (**Figure 2.3I**), suggesting that the WW domain is required for the STXBP4-mediated YAP inhibition.

The observation that STXBP4 deficiency reduced YAP phosphorylation at S127 (**Figure 2.3F**) suggests that the Hippo pathway is inhibited in the STXBP4 KO cells. Indeed, as shown in **Figure 2.3F**, loss of STXBP4 suppressed LATS phosphorylation but did not affect that of MST or its substrate MOB1. These data suggest that STXBP4 is required for LATS activation in the Hippo pathway.

Multiple upstream signaling events have been identified to regulate the Hippo pathway(Yu *et al.*, 2015b). Next, we examined the signaling context for the STXBP4-mediated Hippo pathway regulation. Interestingly, loss of STXBP4 attenuated YAP phosphorylation when actin cytoskeleton was either depolymerized or its tension was inhibited (**Figure 2.3J**); whereas, YAP was still fully phosphorylated under serum and glucose-deprived conditions (**Figure 2.3J**). These data suggest that STXBP4 is involved in the actin cytoskeleton-mediated Hippo pathway regulation.

STXBP4 involves in a protein-protein interaction network comprising multiple Hippo pathway components and regulators

To elucidate the mechanism by which STXBP4 regulates the Hippo pathway, we purified the STXBP4-associated protein complex and characterized its binding partners by mass spectrometry analysis. As shown in **Figure 2.4A**, all the AMOT family proteins were identified

to form a complex with STXBP4, which is consistent with our previous findings (**Figures 2.3B** and **Appendix Figure S2.9A**). Interestingly, we also identified α-catenin, a known Hippo upstream regulator(Feng *et al*, 2016; Rauskolb *et al*, 2014; Schlegelmilch *et al*, 2011; Vite *et al*, 2018), as a binding partner for STXBP4 (**Figure 2.4A**). STXBP4 was also reciprocally identified as a binding protein for some Hippo pathway components (e.g., LATS1, LATS2, TAZ) and regulators (e.g., AMOT, AMOTL1, AMOTL2, PTPN14)(Couzens *et al.*, 2013b; Huttlin *et al*, 2017; Wang *et al.*, 2014a) (**Figure 2.4A**). Collectively, these data suggest that STXBP4 involves in a protein-protein interaction network comprising a group of Hippo pathway components and regulators.

Notably, most of these STXBP4-associated proteins are PY-motif containing proteins (**Figure 2.4A**), suggesting that STXBP4 WW domain is required here. Since α-catenin does not contain a PY motif, we further characterized the α-catenin-binding region in STXBP4. To achieve this, a series of STXBP4 truncation and deletion mutants were generated (**Appendix Figure S2.11A**). As shown in **Appendix Figure S6B**, deletion of the 300~500 amino acid residues of STXBP4, but not its WW domain, fully abolished its association with α-catenin. Moreover, we failed to further narrow down the α-catenin binding region in STXBP4 (**Appendix Figure S2.10C**), suggesting that this identified 300~500 amino acid sequence region is required for its interaction with α-catenin.

Taken together, these data indicate that STXBP4 can form a complex with several Hippo PY motif-containing proteins and α-catenin through its WW domain and the 300~500 amino acid sequence region, respectively.\

STXBP4 functions as a scaffold protein to assemble a protein complex including α-catenin AMOT, LATS and YAP

To test this hypothesis, we performed a sequential pulldown/immunoprecipitation assay using exogenously expressed SFB-STXBP4 and Myc-α-catenin in HEK293T cells. As shown in **Figure 2.4B**, we first isolated STXBP4-associated protein complex using streptavidin beads, eluted the complex with biotin, and purified the α-catenin-associated protein complex through immunoprecipitation. This sequential purification approach can help to characterize the proteins within the STXBP4/α-catenin protein complex. Consistent with our proteomic data (**Figure 4A**), AMOT, LATS1 and YAP were all identified within the STXBP4/α-catenin protein complex (**Figure 2.4B**).

Next, we examined the role of STXBP4 in this multi-protein complex. Overexpression of STXBP4 induced the interaction of α-catenin with both AMOT and LATS1 (**Figure 2.4C**); while loss of STXBP4 largely attenuated the association of α-catenin with AMOT, LATS1 and YAP (**Figure 2.4D**). In addition, STXBP4 promoted the co-localization between AMOT and α-catenin onto cell adherens junction/membrane region, where both LATS1 and YAP were also identified (**Figure 2.4E**). These results suggest a scaffold role of STXBP4 in assembly of a protein complex containing at least α-catenin, AMOT, LATS and YAP at adherens junctions.

Both the WW domain and α-catenin association are required for the STXBP4-mediated YAP regulation

Given the potential tumor suppressive role of STXBP4 in targeting YAP, we next examined the genetic alteration of STXBP4 in the cBioportal database and found that STXBP4 alleles harbor a

series of mutations within cancer patient samples (**Appendix Figure S2.12A** and **Table S2.7**). Four missense mutations that are localized in the α-catenin-binding region (**Figure 2.4F**) disrupted the interaction between STXBP4 and α-catenin (**Figure 2.4G**). As for the STXBP4 WW domain, four out of the identified 9 amino acid residues were found mutated in oligodendrogioma (G501E), bladder urothelial carcinoma (Y513C), uterine carcinosarcoma (H517Y) and cutaneous melanoma (P527L), respectively (**Figure 2.4F**), and they all diminished the association between STXBP4 and AMOT (**Figure 2.4H**). Notably, these cancer-derived missense mutations in either α-catenin-binding region or the WW domain of STXBP4 all failed to rescue YAP's cytoplasmic localization in the STXBP4 KO cells (**Appendix Figure S2.12B**), suggesting that association with α-catenin and the Hippo PY motif-containing components/regulators is required for the STXBP4-dependent the Hippo pathway regulation.

STXBP4 functions as a potential mechano-transducer involved in actin cytoskeleton-mediated Hippo pathway regulation

Notably, α-catenin is known to play a critical role in mechanotransduction(Charras & Yap, 2018; Yonemura *et al*, 2010), and loss of STXBP4 significantly attenuated YAP phosphorylation upon disruption of actin cytoskeleton or inhibition of its tension (**Figure 2.3J**). Interestingly, depolymerization of actin cytoskeleton by latrunculin B or inhibition of its tension by blebbistatin induced the association of STXBP4 with LATS1, AMOT and α-catenin (**Figure 2.4I**). Reconstitution of STXBP4, but not its mutants with missense mutations at its α-catenin-binding region and WW domain (**Figure 2.4F**), significantly rescued YAP phosphorylation when actin cytoskeleton tension was inhibited (**Figure 2.4J**). These data indicate that the STXBP4-mediated protein complex formation with α-catenin and the Hippo PY motif-

containing proteins plays a role in actin cytoskeleton-dependent regulation of the Hippo pathway.

STXBP4 is frequently downregulated in kidney cancer correlated with YAP activation

By analyzing the cancer database, FireBrowse, a platform developed to analyze 14,729 tumor sample data generated by The Cancer Genome Atlas (TCGA), we found that the mRNA level of *STXBP4* was downregulated in all the listed kidney cancer subtypes (**Figure 2.5A**). This finding was further confirmed through a kidney tissue microarray analysis, where the expression of STXBP4 was found decreased in several types of human kidney cancer: 84.8% clear cell carcinoma, 100% papillary renal cell carcinoma, 50% chromophobe carcinoma, 66.7% carcinoma sarcomatodes and 50% high grade urothelial carcinoma of renal pelvis (**Figure 2.5B**). However, downregulation of STXBP4 was only observed in 10% normal kidney tissue (**Figure 2.5B**), suggesting an inverse correlation between STXBP4 expression and kidney cancer formation (P=2.9×10^{-20}, R=-0.41). Moreover, our TCGA data analysis indicated that low expression of STXBP4 was significantly correlated with the poor overall survival rate for the cancer patients with clear cell renal cell carcinoma (ccRCC) (**Figure 2.5C**), indicating that STXBP4 is a potential tumor suppressor in ccRCC.

YAP is highly expressed and activated in multiple major human cancer types but genetic mutation for the Hippo pathway components is hardly detected(Jiang *et al.*, 2015), suggesting that additional oncogenic alterations could lead to YAP activation for tumorigenesis. Since loss of STXBP4 activated YAP (**Figures 2.3F-2.3H**), we next examined the pathological correlation between STXBP4 and YAP using a kidney cancer tissue microarray. Consistent with previous

studies(Cao *et al*, 2014; Godlewski *et al*, 2018; Schutte *et al*, 2014), upregulation of YAP was observed in 57% (45 of 79) of ccRCC tissue samples, while only 20% (2 of 10) of normal kidney tissues showed high YAP expression (**Figures 2.5D** and **2.5E**). Moreover, an inverse correlation between STXBP4 expression and YAP nuclear enrichment was found in the tissue samples with high YAP expression (P=0.0036, R=-0.46), where 94.7% (36 of 38) of the tested tissue samples with low STXBP4 expression had high nuclear enrichment of YAP (**Figures 2.5D** and **2.5E**). However, there were still 10.6% (5 of 47) of the total tested specimens showing high STXBP4 expression but YAP nuclear enrichment (**Figure 2.5E**). These results indicate that downregulation of STXBP4 may contribute to YAP activation in a substantial fraction of ccRCC; however, YAP can still be activated in other tumors via different mechanisms.

Interestingly, although a general low expression of YAP was found in normal kidney tissues, we were still able to observe a relatively high expression of YAP in the podocytes of glomerulus region and partially in the convoluted tubule region (**Figure 2.5D**). Even though, these YAP highly expressed normal kidney regions still consistently showed a decreased STXBP4 expression level (**Figure 2.5D**), suggesting that their inverse correlation in expression could involve in normal kidney physiology.

Both the α-catenin association and functional WW domain are required for the STXBP4's tumor suppressive function in kidney cancer

To investigate the role of STXBP4 in kidney cancer, we first determined the STXBP4 expression in normal mouse kidney tissue and a group of human kidney-related cell lines. Interestingly, STXBP4 had an abundant expression in mouse kidney tissue, an embryonic kidney immortalized

cell line HEK293A and an immortalized human renal proximal tubular epithelial cell line RPTEC (**Figure 2.5F**). In contrast, STXBP4 showed moderate or low expressions in all the tested ccRCC cell lines (**Figure 2.5F**), where YAP was found majorly localized in the nucleus (**Appendix Figure S2.13A**). Overexpression of STXBP4, but not its two patient-derived missense mutants (R490C and P527L) (**Figure 2.4F**), in a ccRCC cell line 786-O (**Appendix Figure S2.13B**), significantly suppressed the xenograft tumor formation (**Figures 2.5G** and 2.**5H**). Since the R490C and P527L mutations can respectively disrupt the STXBP4's interaction with α-catenin (**Figure 2.4G**) and AMOT (**Figure 2.4H**), these results indicate that the association with α-catenin and a functional WW domain are both required for STXBP4's tumor suppressive function.

2.5 Discussion

In this study, we identified a conserved 9-amino acid sequence within the WW domain of the Hippo pathway components and regulators (**Figure 2.2**), which is required for the specific Hippo WW-PY complex formation. Notably, this identified 9-amino acid sequence has at least one residue altered in all the tested control WW domain-containing proteins (**Figures 2.2B** and **S2.2A**), which could help to explain why these control WW domain-containing proteins fail to interact with the Hippo PY motif-containing proteins (**Figures 2.1E** and **2.1F**). Since the "WW-PY" recognition is widely present in the Hippo pathway, manipulation of their recognition is likely to control the outputs of this key signaling pathway in tissue/organ growth and tumorigenesis. Thus, it would be highly exciting if this Hippo WW domain determinants could be utilized for the development of small molecules or peptides to precisely modulate YAP/TAZ activity in cancer therapy and tissue repair.

Mechanistically, the identified 9-amino acid sequence accounts for both a suitable WW domain structure and the binding interface with the PY motif peptide (**Figures S2.3A-S2.3C**), providing a structural basis for the Hippo WW domain binding specificity. Here, our study is only focused on the individual WW domain binding property. Actually, the mechanism underlying the specific "WW-PY" recognition could be more complicated given the role of WW tandem in mediating PY motif binding(Lin *et al*, 2019) and the potential homo- and hetero-dimer formations among WW domains(Sudol & Harvey, 2010b). Moreover, although our current study mostly focused on the WW domain, it is highly possible that its cognate PY motif ligand could also contribute to the specific Hippo "WW-PY" recognition. However, the PY motif is relatively short, flexible and could be easily buried into a higher level of protein structure, making it difficult to assess its role at a protein level. Thus, we did not further address this question from the PY motif-based perspective.

Among the Hippo pathway components, the SAV1 WW domain functions differently from that of YAP, TAZ and KIBRA to bind Hippo PY motif-containing proteins (**Figure 2.1**). This difference may arise from the change of one conserved glutamate residue in the identified 9-amino acid sequence for the SAV1 WW domain in both human (**Figures 2.2A** and **2.2B**) and *Drosophila* (**Appendix Figure S2.7**). Based on our E/D substitution data (**Appendix Figure S2.7B**) and the structural analysis (**Figure S2.3C**), the negative charge for this residue position could be essential. Interestingly, the substituted serine residue within the human SAV1 WW domain can be phosphorylated *in vivo* (www.phosphosite.org), suggesting that the association

between SAV1 and Hippo PY motif-containing proteins could be regulated through a yet-to-be characterized phosphorylation event.

There are only a few WW domain-containing proteins, whose WW domains fit such 9-amino acid sequence in human proteome (**Figure S2.4** and **Table S2.6**). Among them, STXBP4 was found as a negative regulator for YAP (**Figures 2.3F-2.3H**) by forming a protein complex with a series of Hippo PY motif-containing proteins and an adherens junction component, α-catenin (**Figure 2.4A**). Interestingly, STXBP4 serves as a scaffold protein in this network and transduces actin-based mechanical cues to regulate the Hippo pathway. Since α-catenin is known to play a role in both cell density and cytoskeleton tension-dependent regulation of YAP(Feng *et al.*, 2016; Rauskolb *et al.*, 2014; Schlegelmilch *et al.*, 2011; Vite *et al.*, 2018), our findings provided molecular insights into its downstream signaling events. Under the condition with low actin cytoskeleton tension, STXBP4 recruits several Hippo PY motif-containing proteins including at least AMOT, LATS to form a complex with α-catenin at adherens junction. YAP/TAZ are also within this complex based on their interaction with AMOT and LATS (**Figure S2.5**). In proximity, LATS phosphorylates and inhibits YAP. When mechanical cues increase actin cytoskeleton tension, both the adherens junction-associated α-catenin and the filament actin-bound AMOT would be affected in their conformation, resulting in the protein complex disassembly and YAP activation (**Figure S2.5**). Exactly how this α-catenin-STXBP4-Hippo PY proteins axis is coordinated with other related signaling events(Dutta *et al*, 2018; Low *et al*, 2014; Qiao *et al*, 2017) in regulating the interplay between actin cytoskeleton and the Hippo-YAP/TAZ pathway deserves further investigation.

Intriguingly, our TCGA database and tissue microarray studies suggested that STXBP4 is a potential tumor suppressor in kidney cancer (**Figures 2.5A-2.5C**) and its downregulation is significantly correlated with YAP activation in ccRCC tissues (**Figures 2.5D** and **2.5E**). YAP has been found highly expressed and activated in human kidney cancer including ccRCC(Cao *et al.*, 2014; Godlewski *et al.*, 2018; Schutte *et al.*, 2014). Here, our study identified a pathological relevance between STXBP4 and YAP, providing a potential mechanism for the YAP activation in ccRCC. Notably, a CpG island was identified in the STXBP4 promoter, suggesting that the loss of STXBP4 could occur due to its promoter methylation. In addition, STXBP4 gene alleles harbor a relative high mutation rate (13.45%) including nonsense mutation (6.92%), frameshift deletion (1.92%), in frameshift deletion (0.38%) and gene fusion (4.23%) (**Appendix Figure S2.12A**), which could also partially explain the loss of STXBP4 in cancer.

STXBP4 is originally identified as an insulin-regulated protein involved in GLUT4-mediated glucose transport in adipocyte(Min *et al*, 1999), and functions as an inhibitory protein for the SNARE complex-dependent membrane fusion(Yu *et al*, 2013). Dysregualted STXBP4 expression was associated with some SNPs in breast cancer(Caswell *et al*, 2015; Darabi *et al*, 2016; Masoodi *et al*, 2017). Recent studies also implicated the role of STXBP4 in squamous cell carcinomas, by regulating *N*-terminally truncated isoform of p63 (ΔNp63) (Otaka *et al*, 2017; Rokudai *et al*, 2018). Together with these studies, our findings in kidney cancer suggested a complex role of STXBP4 in cancer development, which could depend on tissue context.

2.6 Materials and Methods

Antibodies and chemicals

For Western blotting, anti-α-tubulin (T6199-200UL, 1:5000 dilution), anti-Flag (M2) (F3165-5MG, 1:5000 dilution), and anti-AMOTL1 (HPA001196, 1:1000 dilution) antibodies were obtained from Sigma-Aldrich. Anti-Myc (sc-40, 1:500 dilution) and anti-GFP (sc-9996, 1:1000 dilution) antibodies were purchased from Santa Cruz Biotechnology. Anti-phospho-YAP (S127) (4911S, 1:1000 dilution), anti-phospho-LATS1 (Thr1079) (8654S, 1:1000 dilution), anti-LATS1 (3477S, 1:1000 dilution), anti-phospho-MST (Thr180/Thr183) (3681S, 1:1000 dilution), anti-MST1 (3682S, 1:1000 dilution), anti-phospho-MOB1 (Thr35) (8699S, 1:1000 dilution), anti-MOB1 (3863S, 1:2000 dilution) and anti-NF2 (12896S, 1:2000 dilution) antibodies were purchased from Cell Signaling Technology. The AMOT, AMOTL2, PTPN14 and YAP polyclonal antibodies were generated as previously described(Wang *et al.*, 2011; Wang *et al.*, 2012b). The STXBP4 antiserum was raised against MBP-STXBP4 (the 251~553 amino acid residues) and polyclonal antibody was affinity-purified using an AminoLink Plus Immobilization and Purification Kit (Pierce).

For immunostaining, an anti-YAP (sc-101199, 1:200 dilution) monoclonal antibody was purchased from Santa Cruz Biotechnology. Anti-hemagglutinin (HA) polyclonal antibody (3724S, 1:3000 dilution) was obtained from Cell Signaling Technology.

For immunohistochemical staining, an anti-YAP (14074S, 1:15 dilution) monoclonal antibody was purchased from Cell Signaling Technology. The STXBP4 antiserum was raised against MBP-STXBP4 (the 1~250 amino acid residues) and polyclonal antibody (1:200 dilution) was affinity-purified using an AminoLink Plus Immobilization and Purification Kit (Pierce).Latrunculin B and blebbistatin were obtained from Sigma-Aldrich.

Constructs and viruses

Plasmids encoding the indicated genes were obtained from the Human ORFeome V5.1 library or purchased from Harvard Plasmid DNA Resource Core and Dharmacon. All constructs were generated via polymerase chain reaction (PCR) and subcloned into a pDONOR201 vector using Gateway Technology (Invitrogen) as the entry clones. For tandem affinity purification, all entry clones were subsequently recombined into a lentiviral Gateway-compatible destination vector for the expression of *C*-terminal SFB-tagged fusion proteins. Gateway-compatible destination vectors with the indicated SFB tag, HA tag, GFP tag or Myc tag were used to express various fusion proteins. PCR-mediated mutagenesis was used to generate all the indicated site mutations and internal region/domain deletion mutations.

All lentiviral supernatants were generated by transient transfection of HEK293T cells with the helper plasmids pSPAX2 and pMD2G (kindly provided by Dr. Zhou Songyang, Baylor College of Medicine) and harvested 48 hours later. Supernatants were passed through a 0.45-μm filter and used to infect cells with the addition of 8 μg/mL hexadimethrine bromide (Polybrene) (Sigma-Aldrich).

Cell culture and transfection

HEK293T, ACHN, SLR20 and UMRC6 cell lines were purchased from ATCC and kindly provided by Drs. Boyi Gan and Junjie Chen (MD Anderson Cancer Center). HEK293A cells were purchased from ThermoFisher and kindly provided by Dr. Jae-Il Park (MD Anderson Cancer Center). RPTEC, 786-O, RCC4 and UMRC2 cells were purchased from ATCC and

kindly provided by Dr. Olga Razorenova (University of California, Irvine). HEK293T, HEK293A, RCC4, UMRC2 and UMRC6 cells were maintained in Dulbecco's modified essential medium (DMEM) supplemented with 10% fetal bovine serum at 37°C in 5% CO_2 (v/v). SLR20 and 786-O cells were grown in RPMI-1640 medium supplemented with 10% fetal bovine serum at 37°C in 5% CO_2 (v/v). RPTEC cells were maintained in DMEM/F12 medium supplemented with 5 pM triiodo-L-thyronine, 10 ng/mL epidermal growth factor, 3.5 μg/mL ascorbic acid, 5 μg/mL transferrin, 5 μg/mL insulin, 25 ng/mL prostaglandin E1, 25 ng/mL hydrocortisone, 8.65 ng/mL sodium selenite and 1.2 mg/mL sodium bicarbonate at 37 °C in 5% CO_2 (v/v). All the culture media contain 1% penicillin and streptomycin. Plasmid transfection was performed using a polyethylenimine reagent.

Immunofluorescent staining

Immunofluorescent staining was performed as described previously(Wang *et al*, 2008) with minor modifications. Briefly, cells cultured on coverslips were fixed with 4% paraformaldehyde for 10 minutes at room temperature and then extracted with 0.5% Triton X-100 solution for 5 minutes. For α-catenin-related immunofluorescent staining, cells were pretreated with PBS solution containing 0.5% Trion X-100 and 1% paraformaldehyde for 4 minutes, and subjected to 4% paraformaldehyde fixation. After blocking with Tris-buffered saline with Tween 20 containing 1% bovine serum albumin, the cells were incubated with the indicated primary antibodies for 1 hour at room temperature. After that, the cells were washed and incubated with fluorescein isothiocyanate-, rhodamine- and Cy5-conjugated secondary antibodies for 1 hour. Cells were counterstained with 100 ng/mL 4′,6-diamidino-2-phenylindole (DAPI) for 2 minutes

to visualize nuclear DNA. The coverslips were mounted onto glass slides with an anti-fade solution and visualized under a Nikon Eclipse Ti spinning-disk confocal microscope.

Tandem affinity purification (TAP) of SFB-tagged protein complexes

HEK293T cells stably expressing the indicated SFB-tagged proteins were selected by culturing in medium containing 2 μg/mL puromycin and confirmed by immunostaining and Western blotting as described previously(Wang *et al.*, 2014a). For TAP, HEK293T cells were lysed in NETN buffer (with protease and phosphatase inhibitors) at 4°C for 20 minutes. The crude lysates were centrifuged at 14,000 rpm for 15 minutes at 4°C. The supernatants were incubated with streptavidin-conjugated beads (GE Healthcare) for 1 hour at 4°C. The beads were washed 3 times with NETN buffer, and bound proteins were eluted with NETN buffer containing 2 mg/mL biotin (Sigma-Aldrich) for 2 hours at 4°C. The elutes were incubated with S protein beads (Novagen) for 1 hour. The beads were washed three times with NETN buffer and subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis. Each pulldown sample was run just into the separation gel so that the whole bands could be excised as one sample and subjected to in-gel trypsin digestion and MS analysis.

Mass spectrometry (MS) analysis

The mass spectrometry was performed as described previously(Tavana *et al.*, 2016; Wang *et al.*, 2014a). Briefly, the excised gel bands described above were cut into approximately 1-mm^3 pieces. The gel pieces were then subjected to in-gel trypsin digestion(Shevchenko *et al*, 1996) and dried. Samples were reconstituted in 5 μL of high-performance liquid chromatography (HPLC) solvent A (2.5% acetonitrile, 0.1% formic acid). A nanoscale reverse-phase HPLC capillary column was created by packing 5-μm C18 spherical silica beads into a fused silica

capillary (100 μm inner diameter × ~20 cm length) with a flame-drawn tip. After the column was equilibrated, each sample was loaded onto the column via a Famos autosampler (LC Packings). A gradient was formed, and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid).

As the peptides eluted, they were subjected to electrospray ionization and then entered into an LTQ-Velos mass spectrometer (Thermo Fisher Scientific). The peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences (and hence protein identity) were determined by matching protein databases with the fragmentation pattern acquired by the software program SEQUEST (ver. 28) (Thermo Fisher Scientific). Enzyme specificity was set to partially tryptic with two missed cleavages. Modifications included carboxyamidomethyl (cysteine, fixed) and oxidation (methionine, variable). Mass tolerance was set to 0.5 Da for precursor ions and fragment ions. The database searched was UniProt. Spectral matches were filtered to contain a false discovery rate of less than 1% at the peptide level using the target-decoy method(Elias & Gygi, 2007), and the protein inference was considered followed the general rules(Nesvizhskii & Aebersold, 2005), with manual annotation based on experiences applied when necessary. This same principle was used for isoforms when they were present in the database. The longest isoform was reported as the match.

Bioinformatic analysis

The full-length YAP, TAZ, SAV1 and KIBRA dataset was retrieved from a previous study(Wang *et al.*, 2014a). The TAP-MS dataset for a group of full-length WW domain-

containing proteins randomly selected from human proteome and the WW domains isolated from these proteins as well as the four Hippo pathway WW components (YAP, TAZ, SAV1 and KIBRA) were newly generated in this study. We combined these two datasets and assigned quality scores to the identified protein-protein interactions using MUSE algorithm as previously described(Li *et al.*, 2016a), where a group of unrelated TAP-MS experiments (1,806 experiments using stably expressed TAP-tagged protein baits and 20 experiments using empty vector baits) were included as a control group. Through it, we considered any interaction with a MUSE score of at least 0.9 and raw spectra count greater than 1 to be a high-confident interacting protein (HCIP). The overall HCIP reproducibility rate was close to 85%, which increased when the cutoff peptide number increased. The full-length WW domain-containing proteins and their corresponded WW domains shared 47.5% HCIPs and only 10.2% overlapped HCIPs were identified for the WW domains isolated from the Hippo WW domain-containing components and the control ones (**Figure S2.1C**).

The WW domain-containing proteins' interactomes were enriched in signaling pathways, biological processes and diseases using the HCIPs identified in our studies. The *P* values were estimated using the Knowledge Base provided by Ingenuity Pathway software (Ingenuity Systems, www.ingenuity.com), which contains findings and annotations from multiple sources including the Gene Ontology database, KEGG pathway database, and Panther pathway database. Only statistically significant correlations ($P < 0.05$) are shown. The –log (*P* value) for each function and related HCIPs is listed.

Screen of human WW domain-containing proteins using the identified Hippo WW domain binding criterion

All the WW domain-containing proteins were retrieved from human proteome using a Simple Modular Architecture Research Tool (SMART) (http://smart.embl-heidelberg.de) and the WW domain-containing protein list was further refined in Uniprot (https://www.uniprot.org). Based on the definition, the WW domain-containing proteins are defaulted with two tryptophan (W) residues as separated by 20-22 amino acids within the sequence. All the WW domain sequences were downloaded from Uniprot and subjected to scan with the identified 9-amino acid sequence manually. The list of all the human WW domain-containing proteins and the searching result are listed in **Table S2.6**.

Gene inactivation by CRISPR/Cas9 system

To generate the STXBP4 knockout cells, five distinct single-guide RNAs (sgRNA) were designed by CHOPCHOP website (https://chopchop.rc.fas.harvard.edu), cloned into lentiGuide-Puro vector (Addgene plasmid # 52963), and transfected into HEK293A cells with lentiCas9-Blast construct (Addgene plasmid # 52962). The next day, cells were selected with puromycin (2 μg/ml) for two days and subcloned to form single colonies. Knockout cell clones were screened by Western blotting to verify the loss of STXBP4 expression and their genomic editing was further confirmed by sequencing (**Appendix Figure S2.10**).

The sequence information for sgRNAs used for STXBP4 knockout cell generation is as follows:

STXBP4_sgRNA1: AGACTTAATGTTGAGGCTTG;

STXBP4_sgRNA2: GGCTTGGTGTTGTTCCTTTG;

STXBP4_sgRNA3: TGCTTTCACCAAAGTAGCCT;

STXBP4_sgRNA4: GGAAACAGGCCTTGGCCTGA;

STXBP4_sgRNA5: AGGTACTAGGAGGAATTAAC.

RNA extraction, reverse transcription and real-time PCR

RNA samples were extracted with TRIzol reagent (Invitrogen). Reverse transcription assay was performed using the Script Reverse Transcription Supermix Kit (Bio-Rad) according to the manufacturer's instructions. Real-time PCR was performed using Power SYBR Green PCR master mix (Applied Biosystems). For quantification of gene expression, the $2^{-\Delta\Delta Ct}$ method was used. *GAPDH* expression was used for normalization.

The sequence information for each primer used for gene expression analysis is as follows:

CTGF-Forward: 5′-CCAATGACAACGCCTCCTG-3′;

CTGF-Reverse: 5′-GAGCTTTCTGGCTGCACCA-3′;

CYR61-Forward: 5′-AGCCTCGCATCCTATACAACC-3′;

CYR61-Reverse: 5′-GAGTGCCGCCTTGTGAAAGAA-3′;

ANKRD1-Forward: 5′-CACTTCTAGCCCACCCTGTGA-3′;

ANKRD1-Reverse: 5′-CCACAGGTTCCGTAATGATTT-3′.

Molecular dynamics simulations

All simulations were conducted using the AMBER18 molecular dynamics suite(Case *et al*, 2005; Case, 2017; Gotz *et al*, 2012; Salomon-Ferrer *et al*, 2013). Initial parameterization of complexes

and apo conformations was conducted with the LeAP module in AMBER18, using the protein force field *ff*14SB(Maier *et al*, 2015). YAP-WW1 domain bound to SMAD7-PY motif-containing peptide was initially parameterized using the PDB structure, 2LTW. The SMAD7-PY motif-containing peptide structure was removed from 2LTW and docked to the STXBP4-WW domain (PDB: 2YSG) to form a complex (STXBP4-WW/SMAD7-PY). In the *N*-terminal sequence of STXBP4, four non-native residues (GSSG) were removed prior to docking and formation of the complex STXBP4-WW/SMAD7-PY to maintain consistent residue number with the YAP-WW1 domain. To generate the mutant complexes, all the conserved residues from 2LTW were mutated into alanine using Modeller v9.21(Fiser *et al*, 2000; Marti-Renom *et al*, 2000; Sali & Blundell, 1993; Webb & Sali, 2016), and initial docked poses between mutated YAP-WW1 domains and SMAD7 were generated using the HADDOCK docking program(de Vries *et al*, 2007; Dominguez *et al*, 2003) prior to simulation (**Table S2.1**). This docking procedure was also repeated for the APBB3-WW/SMAD7-PY simulations. An apo form of SMAD7-PY and YAP-WW1 (wild-type domain mutants: L173A/P174A, G176A, W177A, E178A, Y188A, F189A, H192A, T197A, W199A, P202A) were also derived from PDB structure 2LTW, for simulations (**Table S2.1**). Neutralized with either Na^+ or Cl^- counter ions, systems were solvated using a 10 Å buffer of TIP3P waters in a truncated octahedron box. All complexes and apo forms were minimized in a two-step process using the PMEMD program to remove any steric clashes and overlaps. Complexes were heated to 300K for 100 ps in the canonical (NVT) ensemble and equilibrated for 10 ns at 300K in the isothermal-isobaric (NPT) ensemble. Production runs were generated using the accelerated CUDA version of PMEMD(Gotz *et al.*, 2012; Salomon-Ferrer *et al.*, 2013) in the NVT ensemble with 2-fs time

steps at 300K, until MM/PBSA calculations converged. **Table S2.1** outlines the complete simulation conditions for each complex and apo structure.

The MM/PBSA module in AMBER18(Botello-Smith & Luo, 2015; Cai *et al*, 2010b; Miller *et al*, 2012a; Wang *et al*, 2017a; Wang *et al*, 2016a; Wang *et al*, 2012a; Wang & Luo, 2010; Wang *et al*, 2006) was employed to calculate the binding free energies (ΔG) of wild-type and mutant complexes. Calculations do not take into consideration entropy; however, all complexes retain SMAD7-PY as a common binder meeting the necessary requirements for MM/PBSA calculation and comparison. Convergence of both YAP-WW1 and STXBP4-WW complex simulations was determined via cumulative average calculations of ΔG values and timeframes for all subsequent analyses (e.g. clustering, averaging, RMSD, etc.) of each complex were determined based on this metric.

Utilizing the AMBER post-processing program (CPPTRAJ)(Roe & Cheatham, 2013) module in the AMBER18 package, clustering was performed for each complex using only the Cα atoms in SMAD7-PY motif-containing peptide. We chose to cluster using SMAD7-PY motif-containing peptide that coordinates upon observation of the relative stability of both wild-type YAP-WW1 and STXBP4-WW domains. For wild-type complexes (YAP-WW1 or STXBP4-WW bound to SMAD7-PY), all frames were incorporated to generate representative clusters, and only the top 5 clusters are displayed (**Appendix Figure S2.9C**). Conformations were clustered using the hierarchical agglomerative clustering algorithm (average-linkage), with 2.33 Å criteria set as the minimum distance between clusters. Average structures were calculated from only converged timeframes indicated in **Table S2.1**. Using only Cα atoms, the conformation with the smallest

RMSD to the average structure was used to represent the average conformation (**Figures S2.3A, S2.3E** and **3D**). Hydrogen bonds were quantified using the Baker-Hubbard(Baker & Hubbard, 1984) criteria and the MDTraj(McGibbon *et al*, 2015) python module. Ionic salt bridge interactions were determined with a distance criterion(Barlow & Thornton, 1983) (6 Å) between centers of charged groups (positively charged atoms from basic residues Arg, Lys, His: NH*, NZ*, NE2; regions of partial positive charge from His: NE2, HE*, CE1, HD2; negatively charged atoms from acidic residues Glu and Asp: OE*, OD*). Hydrophobic interactions were also measured via a distance criterion of 3.9 Å between carbon atoms. Initially identified in WT YAP-WW1/SMAD7-PY simulations, four intermolecular residue pairs (P208-W199, P209-T197, Y211-H192, P209-Y188) and their Cα atoms were used to calculate the average distance (AD) values in frames outlined in the simulation conditions table (**Figure S2.3F** and **Table S2.1 S1**). This AD calculation procedure was repeated for all complex simulations (SMAD7-PY bound to YAP-WW1 mutants, STXBP4-WW, and APBB3-WW), with Cα atoms of residues in equivalent positions of YAP-WW1 residues.

Xenograft Assays

Athymic nude (nu/nu) mouse strain was used for the xenograft tumor assay in this study. Four-week-old female nude mice were purchased from Jackson Laboratory (002019) and kept in a pathogen-free environment. The xenograft tumor experiments were followed institutional guidelines, approved by the Institutional Animal Care and Use Committee of the University of California, Irvine, and performed under veterinary supervision. The indicated 786-O cells (2×10^6) were subcutaneously injected into the nude mice. After 60 days' adaptation, mice were euthanized, and tumor weights were analyzed.

Immunohistochemical analysis

The kidney tissue array (BC07115a) was purchased from US Biomax, Inc. According to the Declaration of Specimen Collection provided by US Biomax, each specimen collected from any clinic was consented by both hospital and individual.

The kidney tissue array was deparaffinized and rehydrated. The antigens were retrieved by applying Unmask Solution (Vector Laboratories) in a steamer for 40 min. To block endogenous peroxidase activity, the sections were treated with 3% hydrogen peroxide for 30 min. After 1 hour of pre-incubation in 10% goat serum to prevent non-specific staining, the samples were incubated with an antibody at 4 ^{0}C overnight. The sections were incubated with SignalStain Boost detection reagent at room temperature for 30 min. Color was developed with SignalStain 3,3'-diaminobenzidine chromogen-diluted solution (all reagents were obtained from Cell Signaling Technology). Sections were counterstained with Mayer hematoxylin. To quantify the results, a total score of protein expression was calculated from both the percentage of immunopositive cells and immunostaining intensity. High and low protein expressions were defined using the mean score of all samples as a cutoff point. Pearson chi-square analysis test was used for statistical analysis of the correlation of STXBP4 with tissue type (normal versus cancer) and the correlation between STXBP4 and YAP.

TCGA database analysis

Dataset for STXBP4 was downloaded from the Cancer Genome Atlas (TCGA) data portal (https://portal.gdc.cancer.gov/). The mRNA expression and clinical data of *STXBP4* were

analyzed in TCGA-KIRC project. The mRNA levels of *STXBP4* was categorized into high and low expression groups based on the median value. The correlation between *STXBP4* expression and patient survival rate was analyzed. Total 611 patient samples were analyzed.

Quantification and statistical analysis

Each experiment was repeated twice or more, unless otherwise noted. There were no samples or animals excluded for the analyses in this study. As for the mouse experiments, there was no statistical method used to predetermine sample size. We assigned the animals randomly to different groups. A laboratory technician was blinded to the group allocation and tumor collections during the animal experiments as well as the data analyses. The Student's *t*-test was used to analyze the differences between groups. Data were analyzed by Student's *t*-test or Pearson chi-square analysis. SD was used for error estimation. A P value < 0.05 was considered statistically significant.

2.7 Data availability

The MS proteomic data have been deposited in the ProteomeXchange Consortium database (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository(Vizcaino *et al*, 2013) with the dataset identifier PXD004649. The detailed project information is as follows:

Project Name: Human WW-domain containing proteins TAP-LC-MSMS

Project accession: PXD004649

Project DOI: 10.6019/PXD004649

Reviewer account username: reviewer38029@ebi.ac.uk

Password: eavjPdCz

2.8 Author contributions

W.W. conceived and supervised the study. R. L. designed and supervised the simulation analyses. R.V., H.H., A.P.T., S.Z., B.Y., G.S., S.O., K.C. and W.W. performed the experiments. V.T.D., A.E.A and R.L. performed the simulation analyses. Y.C. performed TCGA data and evolutionary analyses. J.C., X.L. and W.W. performed the proteomic and bioinformatic analyses. O.R. provided key reagents and revised the manuscript. V.T.D., J.C., X.L., R.L. and W.W. wrote the manuscript.

2.9Acknowledgments

We thank Drs. Steven Gygi and Ross Tomaino (Taplin Mass Spectrometry Facility, Harvard Medical School) for help with the mass spectrometry analysis and Dr. Chao Wang (MD Anderson Cancer Center) for the insightful discussion. This work was supported in part by a NIH grant (GM126048), an American Cancer Society Research Scholar grant (RSG-18-009-01-CCG), and an Anti-Cancer Challenge pilot project from the Chao Family Comprehensive Cancer Center (P30 CA062203) to W.W.; a NIH grant to R. L. (GM130367); and a Department of Defense Era of Hope Research Scholar Award to J.C. (W81XWH-09-1-0409). R.V. is supported by a NIH Initiative for Maximizing Student Development (IMSD) Fellowship (GM055246). V.T.D is supported by a Mathematical, Computational and Systems Biology Predoctoral NIH Training Grant (T32 EB009418-08).

2.10 Conflict of interest

The authors declare no competing financial interests.

2.11 References

Abu-Odeh M, Bar-Mag T, Huang H, Kim T, Salah Z, Abdeen SK, Sudol M, Reichmann D, Sidhu S, Kim PM *et al* (2014) Characterizing WW domain interactions of tumor suppressor WWOX reveals its association with multiprotein networks. *The Journal of biological chemistry* 289: 8865-8880

Aragon E, Goerner N, Xi Q, Gomes T, Gao S, Massague J, Macias MJ (2012) Structural basis for the versatile interactions of Smad7 with regulator WW domains in TGF-beta Pathways. *Structure* 20: 1726-1736

Baker EN, Hubbard RE (1984) Hydrogen bonding in globular proteins. *Prog Biophys Mol Biol* 44: 97-179

Barlow DJ, Thornton JM (1983) Ion-pairs in proteins. *J Mol Biol* 168: 867-885

Bork P, Sudol M (1994) The WW domain: a signalling site in dystrophin? *Trends in biochemical sciences* 19: 531-533

Botello-Smith WM, Luo R (2015) Applications of MMPBSA to Membrane Proteins I: Efficient Numerical Solutions of Periodic Poisson-Boltzmann Equation. *J Chem Inf Model* 55: 2187-2199

Cai Q, Hsieh MJ, Wang J, Luo R (2010) Performance of Nonlinear Finite-Difference Poisson-Boltzmann Solvers. *J Chem Theory Comput* 6: 203-211

Cao JJ, Zhao XM, Wang DL, Chen KH, Sheng X, Li WB, Li MC, Liu WJ, He J (2014) YAP is overexpressed in clear cell renal cell carcinoma and its knockdown reduces cell proliferation and induces cell cycle arrest and apoptosis. *Oncol Rep* 32: 1594-1600

Case DA, Cheatham TE, 3rd, Darden T, Gohlke H, Luo R, Merz KM, Jr., Onufriev A, Simmerling C, Wang B, Woods RJ (2005) The Amber biomolecular simulation programs. *J Comput Chem* 26: 1668-1688

Case DAD, T.A.; Cheatham III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R.C.; Zhang, W.;Merz, K.M., et al. (2017) *AMBER 2017 Reference Manual*. University of California, San Francisco

Caswell JL, Camarda R, Zhou AY, Huntsman S, Hu D, Brenner SE, Zaitlen N, Goga A, Ziv E (2015) Multiple breast cancer risk variants are associated with differential transcript isoform expression in tumors. *Hum Mol Genet* 24: 7421-7431

Chan SW, Lim CJ, Chong YF, Pobbati AV, Huang C, Hong W (2011) Hippo pathway-independent restriction of TAZ and YAP by angiomotin. *The Journal of biological chemistry* 286: 7018-7026

Chang L, Azzolin L, Di Biagio D, Zanconato F, Battilana G, Lucon Xiccato R, Aragona M, Giulitti S, Panciera T, Gandin A *et al* (2018) The SWI/SNF complex is a mechanoregulated inhibitor of YAP and TAZ. *Nature* 563: 265-269

Chang NS, Hsu LJ, Lin YS, Lai FJ, Sheu HM (2007) WW domain-containing oxidoreductase: a candidate tumor suppressor. *Trends Mol Med* 13: 12-22

Charras G, Yap AS (2018) Tensile Forces and Mechanotransduction at Cell-Cell Junctions. *Curr Biol* 28: R445-R457

Cohen GB, Ren R, Baltimore D (1995) Modular binding domains in signal transduction proteins. *Cell* 80: 237-248

Couzens AL, Knight JD, Kean MJ, Teo G, Weiss A, Dunham WH, Lin ZY, Bagshaw RD, Sicheri F, Pawson T *et al* (2013) Protein interaction network of the mammalian Hippo pathway reveals mechanisms of kinase-phosphatase interactions. *Sci Signal* 6: rs15

Darabi H, Beesley J, Droit A, Kar S, Nord S, Moradi Marjaneh M, Soucy P, Michailidou K, Ghoussaini M, Fues Wahl H *et al* (2016) Fine scale mapping of the 17q22 breast cancer locus using dense SNPs, genotyped within the Collaborative Oncological Gene-Environment Study (COGs). *Sci Rep* 6: 32512

Das S, Smith TF (2000) Identifying nature's protein Lego set. *Advances in protein chemistry* 54: 159-183

de Vries SJ, van Dijk AD, Krzeminski M, van Dijk M, Thureau A, Hsu V, Wassenaar T, Bonvin AM (2007) HADDOCK versus HADDOCK: new features and performance of HADDOCK2.0 on the CAPRI targets. *Proteins* 69: 726-733

Dominguez C, Boelens R, Bonvin AM (2003) HADDOCK: a protein-protein docking approach based on biochemical or biophysical information. *J Am Chem Soc* 125: 1731-1737

Dutta S, Mana-Capelli S, Paramasivam M, Dasgupta I, Cirka H, Billiar K, McCollum D (2018) TRIP6 inhibits Hippo signaling in response to tension at adherens junctions. *EMBO Rep* 19: 337-350

Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4: 207-214

Ervasti JM (2007) Dystrophin, its interactions with other proteins, and implications for muscular dystrophy. *Biochim Biophys Acta* 1772: 108-117

Espanel X, Sudol M (2001) Yes-associated protein and p53-binding protein-2 interact through their WW and SH3 domains. *The Journal of biological chemistry* 276: 14514-14523

Faber PW, Barnes GT, Srinidhi J, Chen J, Gusella JF, MacDonald ME (1998) Huntingtin interacts with a family of WW domain proteins. *Hum Mol Genet* 7: 1463-1474

Feng X, Liu P, Zhou X, Li M-T, Li F-L, Wang Z, Meng Z, Sun Y-P, Yu Y, Xiong Y *et al* (2016) Thromboxane A2 Activates YAP/TAZ Protein to Induce Vascular Smooth Muscle Cell Proliferation and Migration. *Journal of Biological Chemistry* 291: 18947-18958

Ferrigno O, Lallemand F, Verrecchia F, L'Hoste S, Camonis J, Atfi A, Mauviel A (2002) Yes-associated protein (YAP65) interacts with Smad7 and potentiates its inhibitory activity against TGF-beta/Smad signaling. *Oncogene* 21: 4879-4884

Fiser A, Do RK, Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9: 1753-1773

Godlewski J, Kiezun J, Krazinski BE, Kozielec Z, Wierzbicki PM, Kmiec Z (2018) The Immunoexpression of YAP1 and LATS1 Proteins in Clear Cell Renal Cell Carcinoma: Impact on Patients' Survival. *Biomed Res Int* 2018: 2653623

Gotz AW, Williamson MJ, Xu D, Poole D, Le Grand S, Walker RC (2012) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J Chem Theory Comput* 8: 1542-1555

Halder G, Johnson RL (2011) Hippo signaling: growth control and beyond. *Development* 138: 9-22

Hansson JH, Schild L, Lu Y, Wilson TA, Gautschi I, Shimkets R, Nelson-Williams C, Rossier BC, Lifton RP (1995) A de novo missense mutation of the beta subunit of the epithelial sodium channel causes hypertension and Liddle syndrome, identifying a proline-rich segment critical for regulation of channel activity. *Proc Natl Acad Sci U S A* 92: 11495-11499

Hao Y, Chun A, Cheung K, Rashidi B, Yang X (2008) Tumor suppressor LATS1 is a negative regulator of oncogene YAP. *The Journal of biological chemistry* 283: 5496-5509

Haskins JW, Nguyen DX, Stern DF (2014) Neuregulin 1-activated ERBB4 interacts with YAP to induce Hippo pathway target genes and promote cell migration. *Sci Signal* 7: ra116

Hauri S, Wepf A, van Drogen A, Varjosalo M, Tapon N, Aebersold R, Gstaiger M (2013) Interaction proteome of human Hippo signaling: modular control of the co-activator YAP1. *Mol Syst Biol* 9: 713

Hu H, Columbus J, Zhang Y, Wu D, Lian L, Yang S, Goodwin J, Luczak C, Carter M, Chen L *et al* (2004) A map of WW domain family interactions. *Proteomics* 4: 643-655

Huttlin EL, Bruckner RJ, Paulo JA, Cannon JR, Ting L, Baltier K, Colby G, Gebreab F, Gygi MP, Parzen H *et al* (2017) Architecture of the human interactome defines protein communities and disease networks. *Nature* 545: 505-509

Jiang L, Kon N, Li T, Wang S-J, Su T, Hibshoosh H, Baer R, Gu W (2015) Ferroptosis as a p53-mediated activity during tumour suppression. *Nature* 520: 57-62

Li X, Tran KM, Aziz KE, Sorokin AV, Chen J, Wang W (2016) Defining the Protein-Protein Interaction Network of the Human Protein Tyrosine Phosphatase Family. *Mol Cell Proteomics* 15: 3030-3044

Lin Z, Yang Z, Xie R, Ji Z, Guan K, Zhang M (2019) Decoding WW domain tandem-mediated target recognitions in tissue growth and cell polarity. *Elife* 8

Liu F, Li B, Tung EJ, Grundke-Iqbal I, Iqbal K, Gong CX (2007) Site-specific effects of tau phosphorylation on its microtubule assembly activity and self-aggregation. *The European journal of neuroscience* 26: 3429-3436

Liu H, Dai X, Cao X, Yan H, Ji X, Zhang H, Shen S, Si Y, Zhang H, Chen J *et al* (2018) PRDM4 mediates YAP-induced cell invasion by activating leukocyte-specific integrin beta2 expression. *EMBO Rep* 19

Liu X, Yang N, Figel SA, Wilson KE, Morrison CD, Gelman IH, Zhang J (2013) PTPN14 interacts with and negatively regulates the oncogenic function of YAP. *Oncogene* 32: 1266-1273

Low BC, Pan CQ, Shivashankar GV, Bershadsky A, Sudol M, Sheetz M (2014) YAP/TAZ as mechanosensors and mechanotransducers in regulating organ size and tumor growth. *FEBS Lett* 588: 2663-2670

Lubs H, Abidi FE, Echeverri R, Holloway L, Meindl A, Stevenson RE, Schwartz CE (2006) Golabi-Ito-Hall syndrome results from a missense mutation in the WW domain of the PQBP1 gene. *J Med Genet* 43: e30

Maier JA, Martinez C, Kasavajhala K, Wickstrom L, Hauser KE, Simmerling C (2015) ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J Chem Theory Comput* 11: 3696-3713

Mandelkow EM, Mandelkow E (1998) Tau in Alzheimer's disease. *Trends Cell Biol* 8: 425-427

Marti-Renom MA, Stuart AC, Fiser A, Sanchez R, Melo F, Sali A (2000) Comparative protein structure modeling of genes and genomes. *Annu Rev Biophys Biomol Struct* 29: 291-325

Masoodi TA, Banaganapalli B, Vaidyanathan V, Talluri VR, Shaik NA (2017) Computational Analysis of Breast Cancer GWAS Loci Identifies the Putative Deleterious Effect of STXBP4 and ZNF404 Gene Variants. *J Cell Biochem* 118: 4296-4307

McGibbon RT, Beauchamp KA, Harrigan MP, Klein C, Swails JM, Hernandez CX, Schwantes CR, Wang LP, Lane TJ, Pande VS (2015) MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophys J* 109: 1528-1532

Michaloglou C, Lehmann W, Martin T, Delaunay C, Hueber A, Barys L, Niu H, Billy E, Wartmann M, Ito M *et al* (2013) The tyrosine phosphatase PTPN14 is a negative regulator of YAP activity. *PLoS One* 8: e61916

Miller BR, 3rd, McGee TD, Jr., Swails JM, Homeyer N, Gohlke H, Roitberg AE (2012) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput* 8: 3314-3321

Min J, Okada S, Kanzaki M, Elmendorf JS, Coker KJ, Ceresa BP, Syu LJ, Noda Y, Saltiel AR, Pessin JE (1999) Synip: a novel insulin-regulated syntaxin 4-binding protein mediating GLUT4 translocation in adipocytes. *Mol Cell* 3: 751-760

Nesvizhskii AI, Aebersold R (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* 4: 1419-1440

Otaka Y, Rokudai S, Kaira K, Fujieda M, Horikoshi I, Iwakawa-Kawabata R, Yoshiyama S, Yokobori T, Ohtaki Y, Shimizu K *et al* (2017) STXBP4 Drives Tumor Growth and Is Associated with Poor Prognosis through PDGF Receptor Signaling in Lung Squamous Cell Carcinoma. *Clin Cancer Res* 23: 3442-3452

Pan D (2010) The hippo signaling pathway in development and cancer. *Dev Cell* 19: 491-505

Passani LA, Bedford MT, Faber PW, McGinnis KM, Sharp AH, Gusella JF, Vonsattel JP, MacDonald ME (2000) Huntingtin's WW domain partners in Huntington's disease post-mortem brain fulfill genetic criteria for direct involvement in Huntington's disease pathogenesis. *Hum Mol Genet* 9: 2175-2182

Pawson T, Scott JD (1997) Signaling through scaffold, anchoring, and adaptor proteins. *Science* 278: 2075-2080

Piccolo S, Dupont S, Cordenonsi M (2014) The biology of YAP/TAZ: hippo signaling and beyond. *Physiol Rev* 94: 1287-1312

Qiao Y, Chen J, Lim YB, Finch-Edmondson ML, Seshachalam VP, Qin L, Jiang T, Low BC, Singh H, Lim CT *et al* (2017) YAP Regulates Actin Dynamics through ARHGAP29 and Promotes Metastasis. *Cell Rep* 19: 1495-1502

Qiao Y, Lin SJ, Chen Y, Voon DC, Zhu F, Chuang LS, Wang T, Tan P, Lee SC, Yeoh KG *et al* (2016) RUNX3 is a novel negative regulator of oncogenic TEAD-YAP complex in gastric cancer. *Oncogene* 35: 2664-2674

Rauskolb C, Sun S, Sun G, Pan Y, Irvine KD (2014) Cytoskeletal tension inhibits Hippo signaling through an Ajuba-Warts complex. *Cell* 158: 143-156

Rentschler S, Linn H, Deininger K, Bedford MT, Espanel X, Sudol M (1999) The WW domain of dystrophin requires EF-hands region to interact with beta-dystroglycan. *Biol Chem* 380: 431-442

Roe DR, Cheatham TE, 3rd (2013) PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J Chem Theory Comput* 9: 3084-3095

Rokudai S, Li Y, Otaka Y, Fujieda M, Owens DM, Christiano AM, Nishiyama M, Prives C (2018) STXBP4 regulates APC/C-mediated p63 turnover and drives squamous cell carcinogenesis. *Proc Natl Acad Sci U S A* 115: E4806-E4814

Salah Z, Aqeilan RI (2011) WW domain interactions regulate the Hippo tumor suppressor pathway. *Cell death & disease* 2: e172

Salah Z, Cohen S, Itzhaki E, Aqeilan RI (2013) NEDD4 E3 ligase inhibits the activity of the Hippo pathway by targeting LATS1 for degradation. *Cell Cycle* 12: 3817-3823

Salah Z, Melino G, Aqeilan RI (2011) Negative regulation of the Hippo pathway by E3 ubiquitin ligase ITCH is sufficient to promote tumorigenicity. *Cancer Res* 71: 2010-2020

Sali A, Blundell TL (1993) Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234: 779-815

Salomon-Ferrer R, Gotz AW, Poole D, Le Grand S, Walker RC (2013) Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J Chem Theory Comput* 9: 3878-3888

Schlegelmilch K, Mohseni M, Kirak O, Pruszak J, Rodriguez JR, Zhou D, Kreger BT, Vasioukhin V, Avruch J, Brummelkamp TR *et al* (2011) Yap1 acts downstream of alpha-catenin to control epidermal proliferation. *Cell* 144: 782-795

Schutte U, Bisht S, Heukamp LC, Kebschull M, Florin A, Haarmann J, Hoffmann P, Bendas G, Buettner R, Brossart P *et al* (2014) Hippo signaling mediates proliferation, invasiveness, and metastatic potential of clear cell renal cell carcinoma. *Transl Oncol* 7: 309-321

Sebe-Pedros A, Zheng Y, Ruiz-Trillo I, Pan D (2012) Premetazoan origin of the hippo signaling pathway. *Cell reports* 1: 13-20

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68: 850-858

Strano S, Monti O, Pediconi N, Baccarini A, Fontemaggi G, Lapi E, Mantovani F, Damalas A, Citro G, Sacchi A *et al* (2005) The transcriptional coactivator Yes-associated protein drives p73 gene-target specificity in response to DNA Damage. *Mol Cell* 18: 447-459

Strano S, Munarriz E, Rossi M, Castagnoli L, Shaul Y, Sacchi A, Oren M, Sudol M, Cesareni G, Blandino G (2001) Physical interaction with Yes-associated protein enhances p73 transcriptional activity. *The Journal of biological chemistry* 276: 15164-15173

Sudol M (2010) Newcomers to the WW Domain-Mediated Network of the Hippo Tumor Suppressor Pathway. *Genes Cancer* 1: 1115-1118

Sudol M, Bork P, Einbond A, Kastury K, Druck T, Negrini M, Huebner K, Lehman D (1995a) Characterization of the mammalian YAP (Yes-associated protein) gene and its role in defining a novel protein module, the WW domain. *The Journal of biological chemistry* 270: 14733-14741

Sudol M, Chen HI, Bougeret C, Einbond A, Bork P (1995b) Characterization of a novel protein-binding module--the WW domain. *FEBS letters* 369: 67-71

Sudol M, Harvey KF (2010) Modularity in the Hippo signaling pathway. *Trends Biochem Sci* 35: 627-633

Tapia VE, Nicolaescu E, McDonald CB, Musi V, Oka T, Inayoshi Y, Satteson AC, Mazack V, Humbert J, Gaffney CJ *et al* (2010) Y65C missense mutation in the WW domain of the Golabi-

Ito-Hall syndrome protein PQBP1 affects its binding activity and deregulates pre-mRNA splicing. *The Journal of biological chemistry* 285: 19391-19401

Tavana O, Li D, Dai C, Lopez G, Banerjee D, Kon N, Chen C, Califano A, Yamashiro DJ, Sun H *et al* (2016) HAUSP deubiquitinates and stabilizes N-Myc in neuroblastoma. *Nature Medicine* 22: 1180-1186

Ulbricht A, Eppler FJ, Tapia VE, van der Ven PF, Hampe N, Hersch N, Vakeel P, Stadel D, Haas A, Saftig P *et al* (2013) Cellular mechanotransduction relies on tension-induced and chaperone-assisted autophagy. *Curr Biol* 23: 430-435

Varelas X, Miller BW, Sopko R, Song S, Gregorieff A, Fellouse FA, Sakuma R, Pawson T, Hunziker W, McNeill H *et al* (2010) The Hippo pathway regulates Wnt/beta-catenin signaling. *Dev Cell* 18: 579-591

Verma A, Jing-Song F, Finch-Edmondson ML, Velazquez-Campoy A, Balasegaran S, Sudol M, Sivaraman J (2018) Biophysical studies and NMR structure of YAP2 WW domain - LATS1 PPxY motif complexes reveal the basis of their interaction. *Oncotarget* 9: 8068-8080

Vite A, Zhang C, Yi R, Emms S, Radice GL (2018) alpha-Catenin-dependent cytoskeletal tension controls Yap activity in the heart. *Development* 145

Vizcaino JA, Cote RG, Csordas A, Dianes JA, Fabregat A, Foster JM, Griss J, Alpi E, Birim M, Contell J *et al* (2013) The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res* 41: D1063-1069

Wang C, Greene D, Xiao L, Qi R, Luo R (2017) Recent Developments and Applications of the MMPBSA Method. *Front Mol Biosci* 4: 87

Wang C, Nguyen PH, Pham K, Huynh D, Le TB, Wang H, Ren P, Luo R (2016) Calculating protein-ligand binding affinities with MMPBSA: Method and error analysis. *J Comput Chem* 37: 2436-2446

Wang C, Zhang W, Yin MX, Hu L, Li P, Xu J, Huang H, Wang S, Lu Y, Wu W *et al* (2015) Suppressor of Deltex mediates Pez degradation and modulates Drosophila midgut homeostasis. *Nat Commun* 6: 6607

Wang J, Cai Q, Xiang Y, Luo R (2012a) Reducing grid-dependence in finite-difference Poisson-Boltzmann calculations. *J Chem Theory Comput* 8: 2741-2751

Wang J, Luo R (2010) Assessment of linear finite-difference Poisson-Boltzmann solvers. *J Comput Chem* 31: 1689-1698

Wang J, Wang W, Kollman PA, Case DA (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model* 25: 247-260

Wang W, Chen L, Ding Y, Jin J, Liao K (2008) Centrosome separation driven by actin-microfilaments during mitosis is mediated by centrosome-associated tyrosine-phosphorylated cortactin. *J Cell Sci* 121: 1334-1343

Wang W, Huang J, Chen J (2011) Angiomotin-like proteins associate with and negatively regulate YAP1. *The Journal of biological chemistry* 286: 4364-4370

Wang W, Huang J, Wang X, Yuan J, Li X, Feng L, Park JI, Chen J (2012b) PTPN14 is required for the density-dependent control of YAP1. *Genes Dev* 26: 1959-1971

Wang W, Li X, Huang J, Feng L, Dolinta KG, Chen J (2014) Defining the protein-protein interaction network of the human hippo pathway. *Mol Cell Proteomics* 13: 119-131

Webb B, Sali A (2016) Comparative Protein Structure Modeling Using MODELLER. *Curr Protoc Protein Sci* 86: 2 9 1-2 9 37

Wilson KE, Li YW, Yang N, Shen H, Orillion AR, Zhang J (2014) PTPN14 forms a complex with Kibra and LATS1 proteins and negatively regulates the YAP oncogenic function. *The Journal of biological chemistry* 289: 23693-23700

Yeung B, Ho KC, Yang X (2013) WWP1 E3 ligase targets LATS1 for ubiquitin-mediated degradation in breast cancer cells. *PLoS One* 8: e61027

Yonemura S, Wada Y, Watanabe T, Nagafuchi A, Shibata M (2010) alpha-Catenin as a tension transducer that induces adherens junction development. *Nat Cell Biol* 12: 533-542

Yu FX, Zhao B, Guan KL (2015) Hippo Pathway in Organ Size Control, Tissue Homeostasis, and Cancer. *Cell* 163: 811-828

Yu H, Rathore SS, Shen J (2013) Synip arrests soluble N-ethylmaleimide-sensitive factor attachment protein receptor (SNARE)-dependent membrane fusion as a selective target membrane SNARE-binding inhibitor. *The Journal of biological chemistry* 288: 18885-18893

Yu J, Zheng Y, Dong J, Klusza S, Deng WM, Pan D (2010) Kibra functions as a tumor suppressor protein that regulates Hippo signaling in conjunction with Merlin and Expanded. *Dev Cell* 18: 288-299

Zhang X, Milton CC, Poon CL, Hong W, Harvey KF (2011) Wbp2 cooperates with Yorkie to drive tissue growth downstream of the Salvador-Warts-Hippo pathway. *Cell Death Differ* 18: 1346-1355

Zhao B, Li L, Lu Q, Wang LH, Liu CY, Lei Q, Guan KL (2011) Angiomotin is a novel Hippo pathway component that inhibits YAP oncoprotein. *Genes Dev* 25: 51-63

2.12 Figures

Figure 2.1

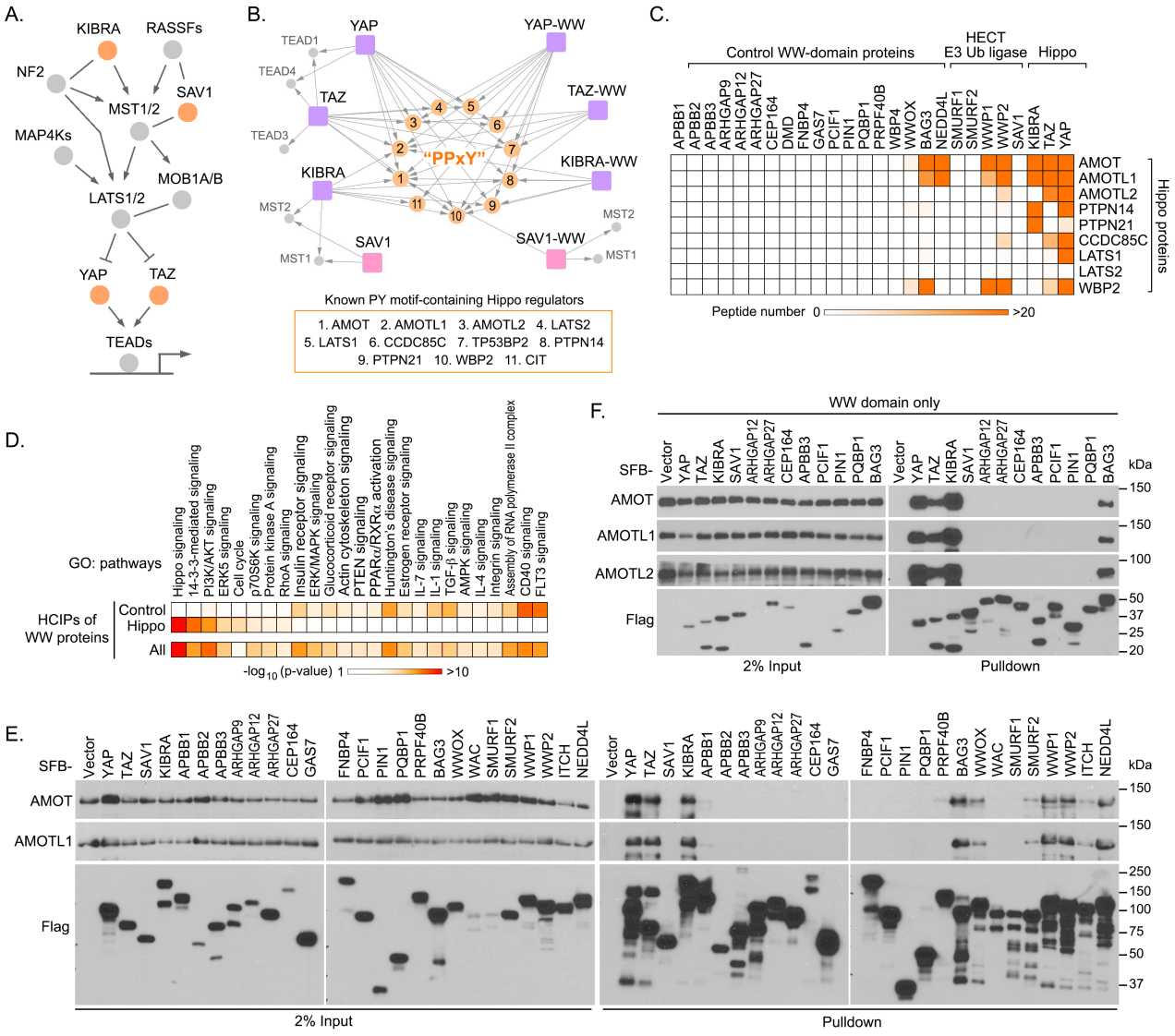


Figure 2.1. The Hippo WW domain shows binding specificity with the known Hippo PY motif-containing proteins.

(**A**) Schematic illustration of the human Hippo pathway, where the Hippo WW domain-containing components are highlighted.

(**B**) A summary map of cytoscape-generated merged interaction network for the Hippo WW domain-containing components and their WW domains.

(**C**) The Hippo WW domain-containing proteins show binding specificity to the known Hippo PY motif-containing proteins. TAP-MS analysis of a series of WW domain-containing proteins were performed and their binding with the indicated Hippo PY motif-containing proteins was summarized in a heatmap.

(**D**) The HCIPs for the Hippo WW domain-containing proteins were involved in different signaling pathways compared to those retrieved from the control WW domain-containing proteins. Gene Ontology analysis was performed.

(**E**) Validation of the binding specificity for the Hippo WW domain-containing proteins. HEK293T cells were transfected with the indicated SFB-tagged constructs and subjected to the pulldown assay.

(**F**) Validation of the binding specificity for the derived WW domains from the Hippo WW domain-containing proteins. HEK293T cells were transfected with the indicated SFB-tagged constructs and subjected to the pulldown assay.

Figure 2.2

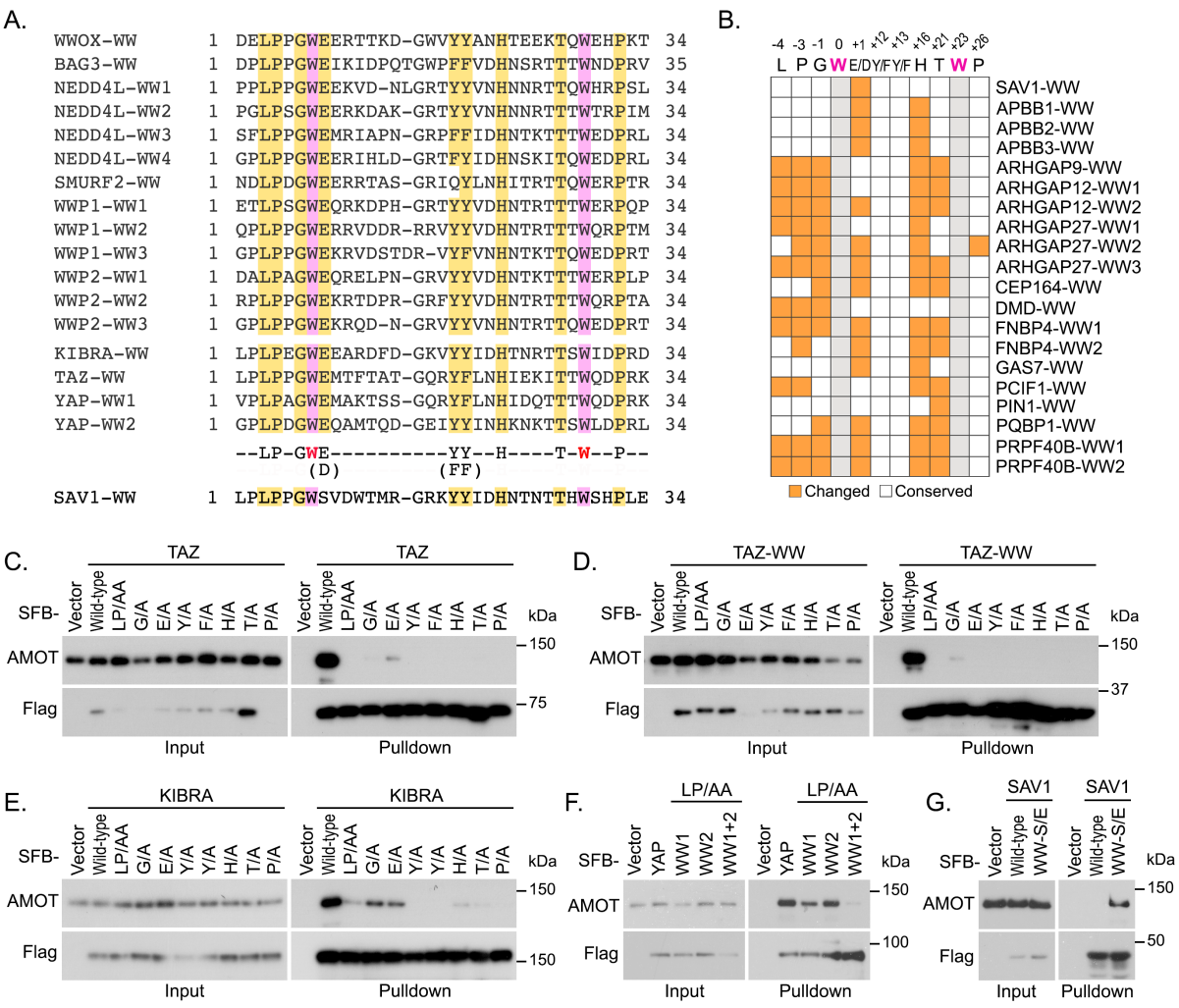


Figure 2.2. Identification of a conserved 9-amino acid sequence that determines the Hippo WW domain binding specificity.

(**A**) Sequence alignment of the WW domains derived from the WW domain-containing proteins that are known to bind the Hippo PY motif-containing proteins. The two conserved tryptophan restudies were highlighted in purple. Additional conserved amino acid residues were highlighted in yellow.

(**B**) Summary of the residue difference in the identified 9-amino acid sequence for the control WW domains. The conserved two tryptophan residues are labelled in grey; the changed residues are labelled in orange; and the unchanged residues are labelled in white.

(**C-G**) Validation of the identified 9-amino acid sequence in determining the Hippo WW domain binding specificity. The requirement of the identified 9-amino acid sequence for AMOT association was respectively examined for TAZ (**C**), TAZ-WW domain (**D**), KIBRA (**E**), YAP (**F**) and SAV1 (**G**). HEK293T cells were transfected with the indicated SFB-tagged constructs and subjected to the pulldown assay.

Figure 2.3

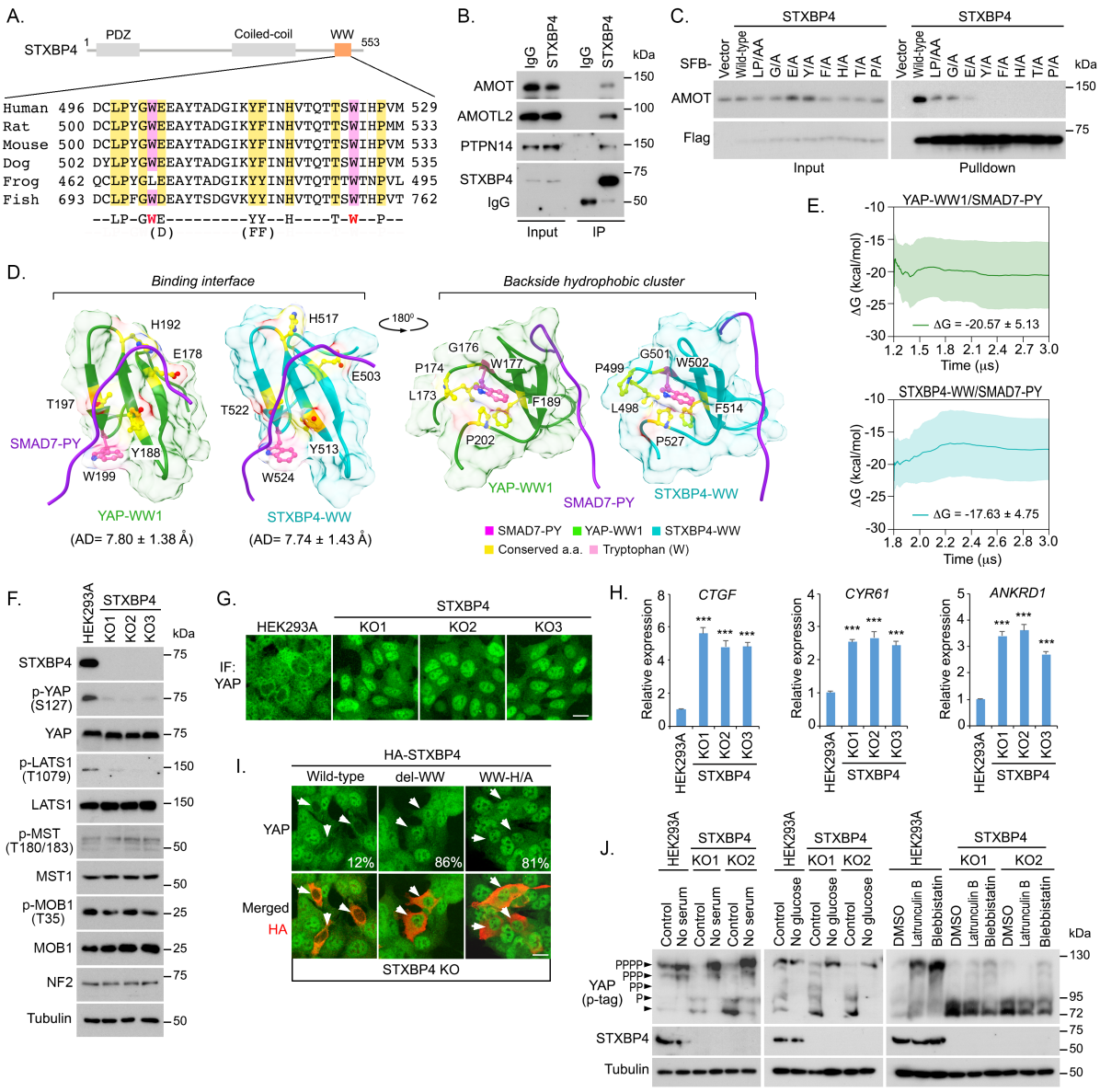


Figure 2.3. STXBP4 is a Hippo pathway regulator, which contains a WW domain that fits the criterion of the Hippo WW domain binding specificity.

(**A**) Schematic illustration of STXBP4 protein, where the identified 9-amino acid sequence of STXBP4-WW domain was aligned across the indicated species.

(**B**) STXBP4 forms a complex with several Hippo PY motif-containing proteins. Immunoprecipitation was performed with STXBP4 antibody.

(**C**) The identified 9-amino acid sequence is required for the association between STXBP4 and AMOT. HEK293T cells were transfected with the indicated STXBP4 mutants and subjected to the pulldown assay.

(**D**) Structural comparison between the YAP-WW1/SMAD7-PY and STXBP4-WW/SMAD7-PY complexes. The identified 9 amino acid residues were indicated for both complexes.

(**E**) The YAP-WW1/SMAD7-PY and STXBP4-WW/SMAD7-PY complexes show similar cumulative average trend and average binding free energy *(ΔG*) within standard deviation (the shaded region) of one another.

(**F**) Loss of STXBP4 inhibits YAP phosphorylation and LATS activation. Western blotting was performed with the indicated antibodies.

(**G** and **H**) Loss of STXBP4 activates YAP. STXBP4 deficiency promotes YAP nuclear translocation (**G**) and YAP downstream gene transcription (mean ± s.d., n=3 biological replicates) (**H**). Scale bar, 20 μm.*** $p < 0.001$ (Student's *t*-test).

(**I**) WW domain is required for the STXBP4-mediated YAP cytoplasmic translocation. STXBP4 KO cells were transfected with the indicated STXBP4 constructs and immunofluorescent staining was performed. HA-positive cells (arrows) from ~30 different views (~200 cells in total) were randomly selected and quantified for YAP localization. Percentage of HA-positive cells with nuclear YAP enrichment is shown. Scale bar, 20 μm.

(**J**) Loss of STXBP4 attenuates YAP phosphorylation as induced by actin cytoskeleton inhibition. The indicated cells were subjected to serum starvation (treatment with no-serum medium for 12 hours), glucose starvation (treatment with no-glucose medium for 6 hours) or

actin inhibition (treatment with 0.5 μg/mL latrunculin B or 5 μM blebbistatin for 30 min). YAP phosphorylation was detected using phospho-tag gel, where the YAP phosphorylation level was indicated.

Figure 2.4

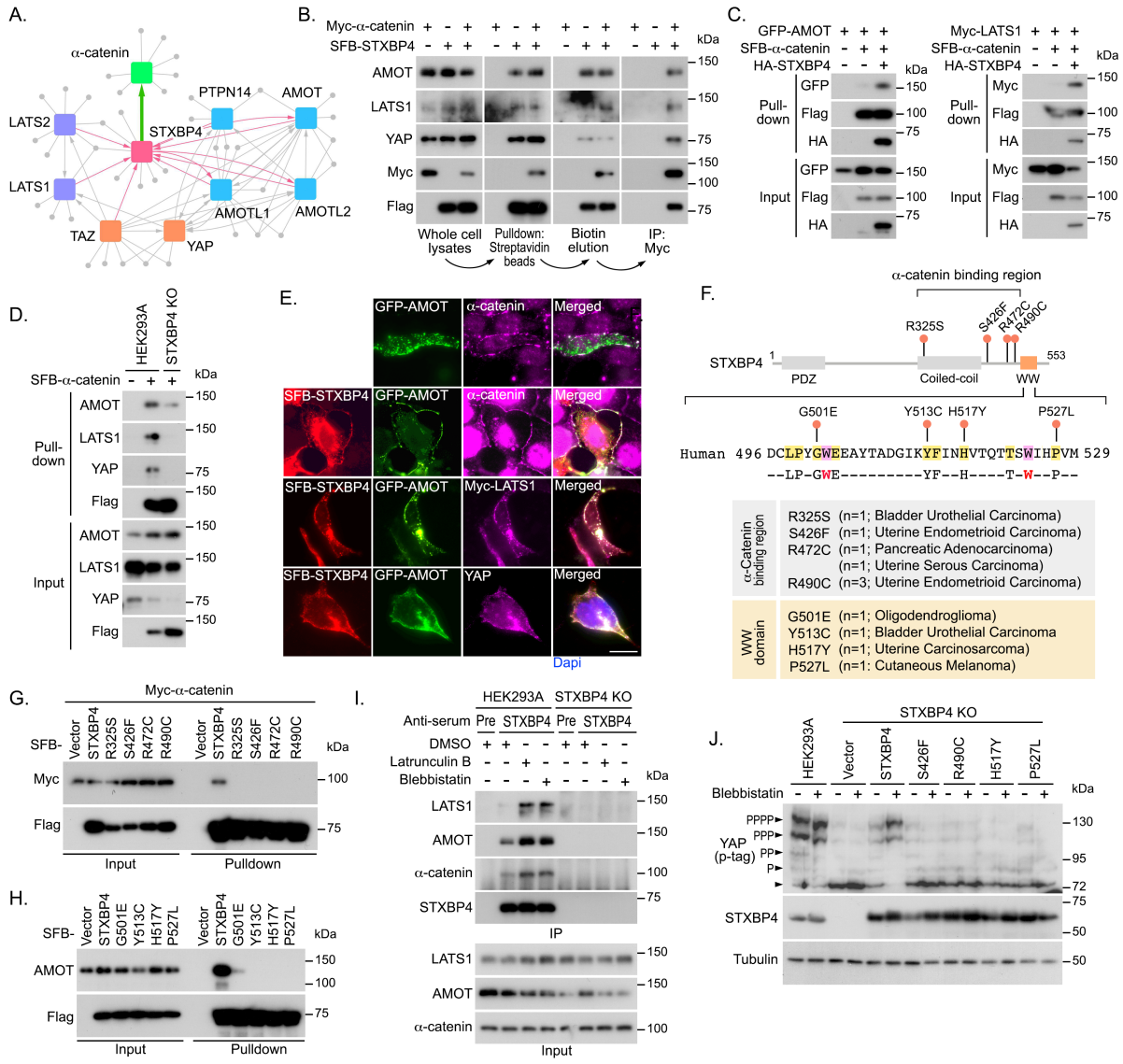


Figure 24. STXBP4 functions in the actin cytoskeleton tension-mediated Hippo pathway regulation by forming a complex with α-catenin and a group of Hippo PY motif-containing proteins.

(**A**) A summary map of cytoscape-generated protein-protein interaction network for STXBP4, α-catenin and a group of Hippo pathway proteins.

(**B**) STXBP4 forms a protein complex with α-catenin and a group of Hippo pathway proteins.

(**C**) STXBP4 promotes the association of α-catenin with AMOT and LATS1. HEK293T cells were transfected with the indicated SFB-tagged constructs and subjected to the pulldown assay.

(**D**) Loss of STXBP4 diminishes the association of α-catenin with AMOT, LATS1 and YAP. HEK293A and STXBP4 KO cells were transfected with the SFB-tagged α-catenin construct and subjected to the pulldown assay.

(**E**) STXBP4 induces the co-localization between α-catenin and AMOT as well as LATS1 and YAP. HEK293A cells were transfected with the indicated constructs and immunofluorescence was performed. Scale bar, 20 μm.

(**F-H)** Identification of several STXBP4 missense mutations that disrupt its interaction with α-catenin and AMOT. The missense mutations within the STXBP4 α-catenin-binding region and the 9-amino acid sequence of the STXBP4 WW domain were indicated and annotated (**F**). The identified missense mutations respectively disrupted the STXBP4- α-catenin (**G**) and STXBP4-AMOT (**H**) complex formation.

(**I**) Inhibition of actin cytoskeleton promotes the STXBP4-associated protein complex formation. HEK293A and the STXBP4 KO cells were subjected to immunoprecipitation using pre-immune serum and anti-STXBP4 serum under the indicated treatments.

(**J**) The missense mutations of STXBP4 (**F**) diminished the ability of STXBP4 to rescue YAP phosphorylation in the STXBP4 KO cells with low actin cytoskeleton tension. YAP phosphorylation

Figure 2.5

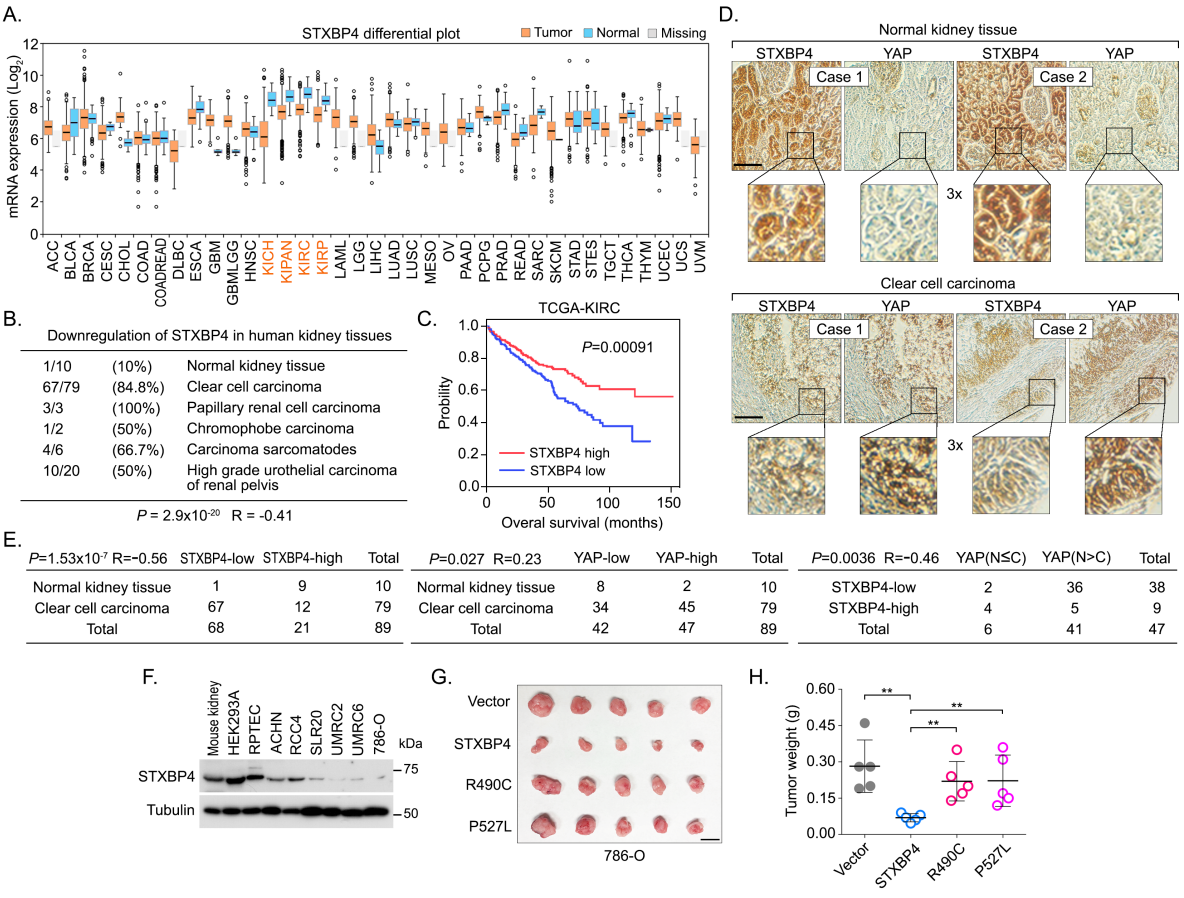


Figure 2.5. STXBP4 is a tumor suppressor in human kidney cancer.

(**A** and **B**) STXBP4 is downregulated in human kidney cancer. The mRNA level of *STXBP4* is analyzed in the Firebrowse web database (http://firebrowse.org) (**A**), where 14,729 tumor sample data generated by TCGA were included. The first quartile, median and third quartile values were indicated as the boxplots. Outliers were plotted as individual points. Error bars indicated the standard deviation above and below the mean of the data. The expression of STXBP4 was also examined using kidney tissue microarray, where percentage of the indicated tissue samples with downregulated STXBP4 was shown (**B**). The *p* value was calculated by using the paired Student's *t*-test.

(**C**) Kaplan-Meier curves of overall survival of patients with ccRCC is stratified by *STXBP4* expression level. Clinical data of *STXBP4* were analyzed in TCGA-KIRC project containing total 611 patient samples. The *p* value was calculated by using the Log-rank (Mantel-Cox) test.

(**D**) Immunohistochemical staining of STXBP4 and YAP were performed in a kidney cancer tissue microarray, where the indicated regions in the box were shown three times enlarged. Brown staining indicates positive immunoreactivity. Scale bar, 100 μm.

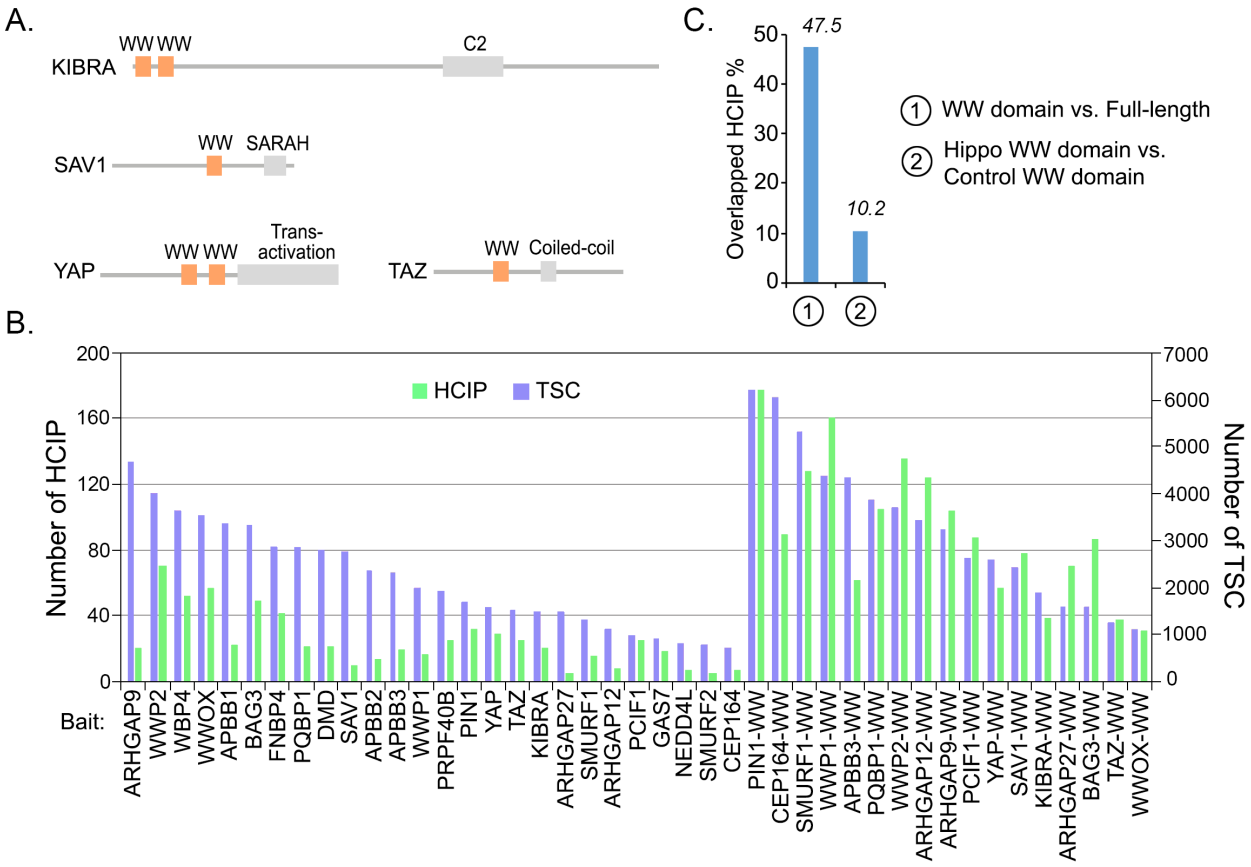
(**E**) Correlation analyses between STXBP4 and YAP in human normal kidney and clear cell carcinoma samples are shown as tables. Statistical significance was determined by chi-square test. R, correlation coefficient. N, nuclear localization. C, cytoplasmic localization.

(**F**) STXBP4 expression is examined in a panel of ccRCC cell lines by Western Blotting.

(**G** and **H**) Both the association with α-catenin and the functional WW domain are required for the STXBP4's tumor suppressive function in 786-O cells. Overexpression of STXBP4, but not the indicated STXBP4 missense mutants, significantly suppressed the 786-O cell xenograft tumor formation. Xenograft tumors are shown in (**G**), and the tumor weight is quantified in (**H**) (n = 5 mice, mean ± s.d.). ** $p < 0.01$ (Student's *t*-test). Scale bar, 1 cm.

2.13 Supplementary Figures

Supplementary Fig. S2.1



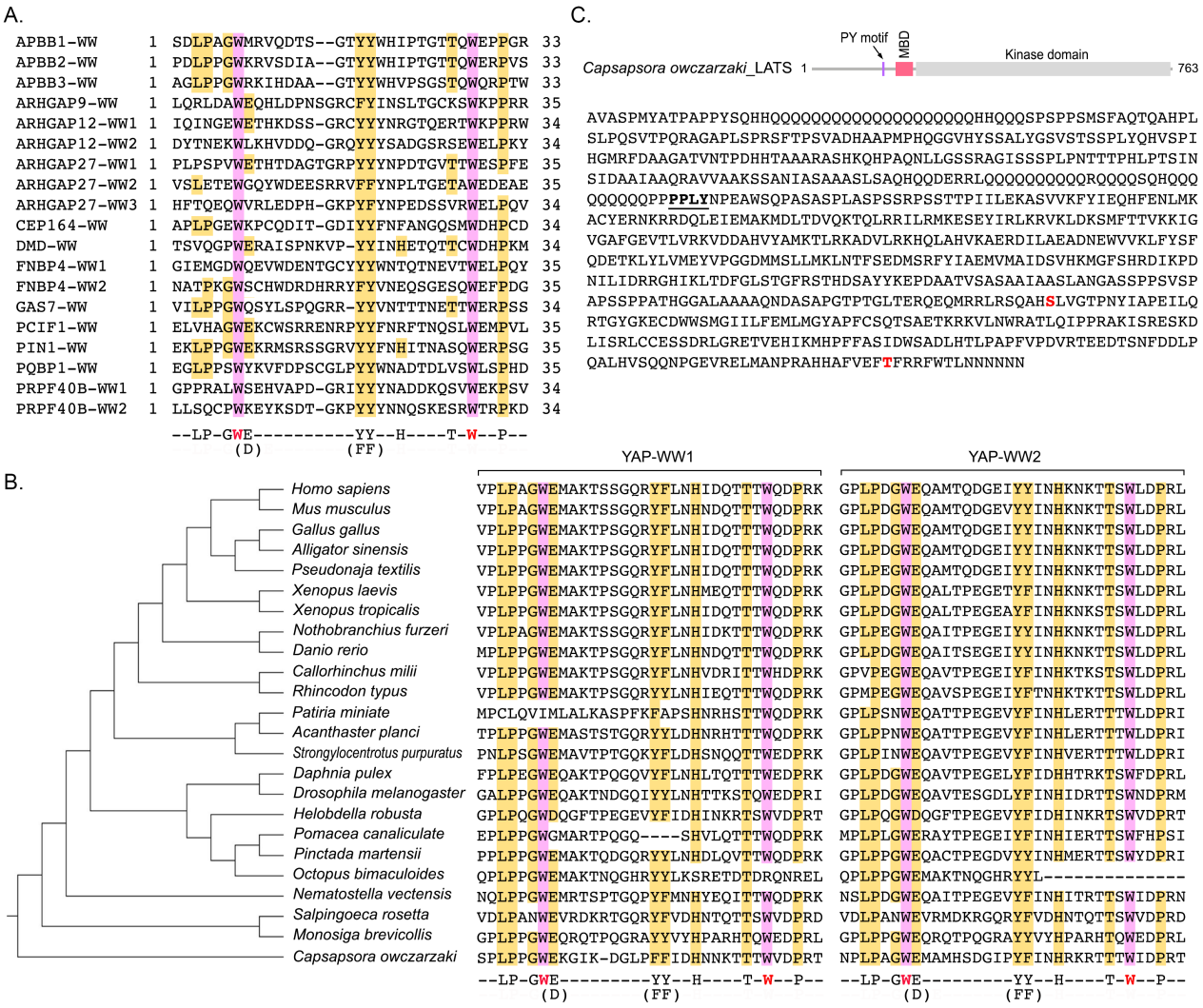
Supplementary Figure S2.1: Proteomic analysis of the WW-containing proteins.

(**A**) Schematic illustration of the Hippo WW domain-containing components.

(**B**) The total spectral counts (TSCs) and corresponding numbers of HCIPs for the indicated proteomic experiments were listed.

(**C**) The overlapped HCIP rate was compared for the full-length protein vs. its WW domain, and Hippo WW domains vs. control WW domains, respectively.

Supplementary Fig. S2.2



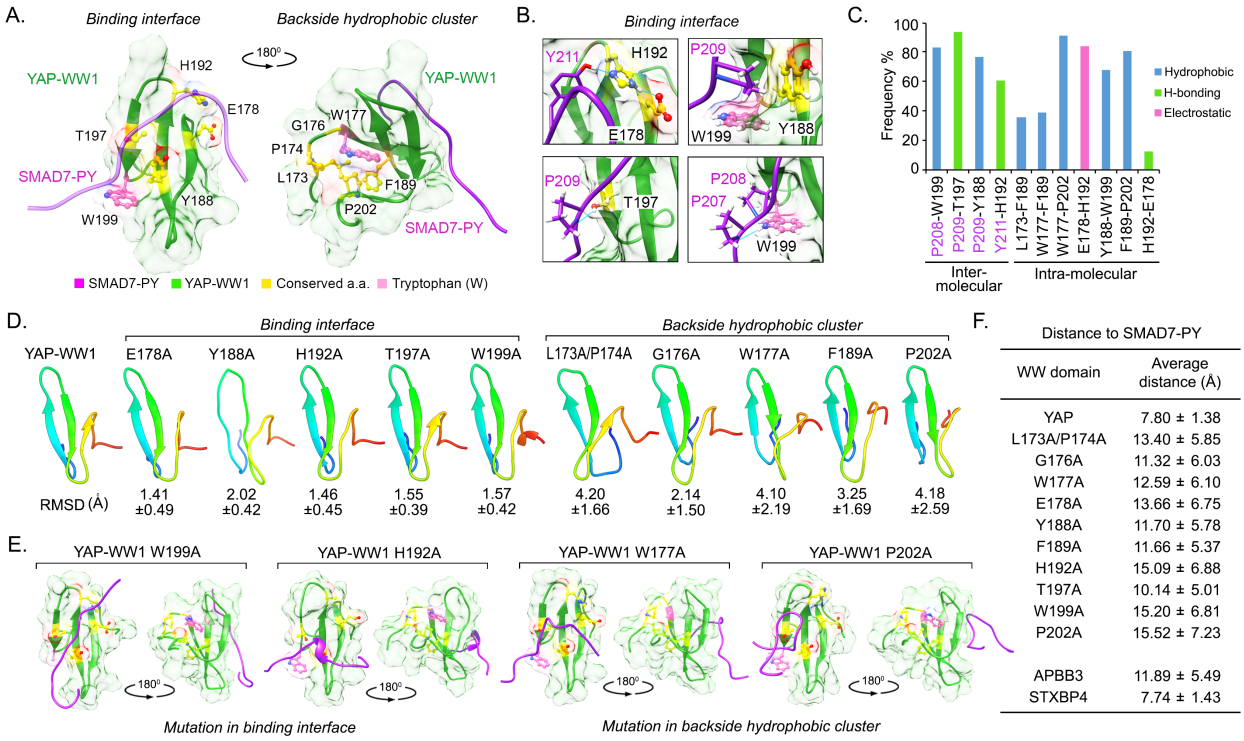
Supplementary Figure S2.2: Analyses of the identified 9-amino acid sequence in both control WW domains and evolution.

(**A**) Sequence alignment of the WW domains derived from the control WW domain-containing proteins that cannot bind the Hippo PY motif-containing proteins. The two conserved tryptophan restudies were highlighted in purple, and the identified 9 amino acid residues were highlighted in yellow.

(**B**) Evolutionary analysis of the YAP WW domains. The identified 9-amino acid sequence is highlighted in the two YAP WW domains derived from the indicated species.

(**C**) A PY motif is identified in *Capsapsora owczarzaki* LATS. Schematic illustration of the *Capsapsora owczarzaki* LATS protein, where the PY motif is indicated. MBD, MOB1-binding domain. The PY motif is underlined in the *Capsapsora owczarzaki* LATS protein sequence, where the auto-phosphorylation site (S586) and the phosphorylation site (T750) in the hydrophobic motif were shown in red.

Supplementary Fig. S2.3



F.

Distance to SMAD7-PY	
WW domain	Average distance (Å)
YAP	7.80 ± 1.38
L173A/P174A	13.40 ± 5.85
G176A	11.32 ± 6.03
W177A	12.59 ± 6.10
E178A	13.66 ± 6.75
Y188A	11.70 ± 5.78
F189A	11.66 ± 5.37
H192A	15.09 ± 6.88
T197A	10.14 ± 5.01
W199A	15.20 ± 6.81
P202A	15.52 ± 7.23
APBB3	11.89 ± 5.49
STXBP4	7.74 ± 1.43

Supplementary Figure S2.3: Structural analysis of the identified 9-amino acid sequence.

(**A**) Illustration of the identified 9 amino acid residues in the average YAP-WW1/SMAD7-PY structure. The initial structure was derived from NMR solution structure (2LTW). SMAD7-PY peptide was adjusted to 50% transparence to show the residue details on the binding interface.

(**B**) Four contact regions within the YAP-WW1/SMAD7-PY complex were shown in details from the representative top cluster structures. Residues from SMAD7-PY motif peptide were labeled in purple. Hydrogen bond is indicated in blue line.

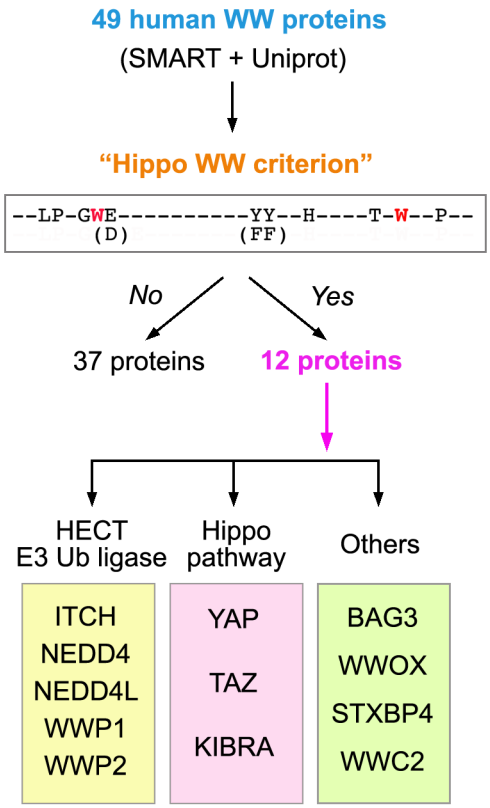
(**C**) The binding type and the corresponding frequency rate were shown for the indicated inter- and intra-molecular residue pairs.

(**D**) Simulation analysis of apo YAP-WW1 domain and its indicated mutants. RMSD value for each mutant simulation (referenced against the average apo YAP-WW1 domain) was shown.

(**E**) Average structures of the indicated YAP-WW1 mutant/SMAD7-PY complexes.

(**F**) The average distance between SMAD7-PY motif peptide and the indicated WW domains was summarized in a table.

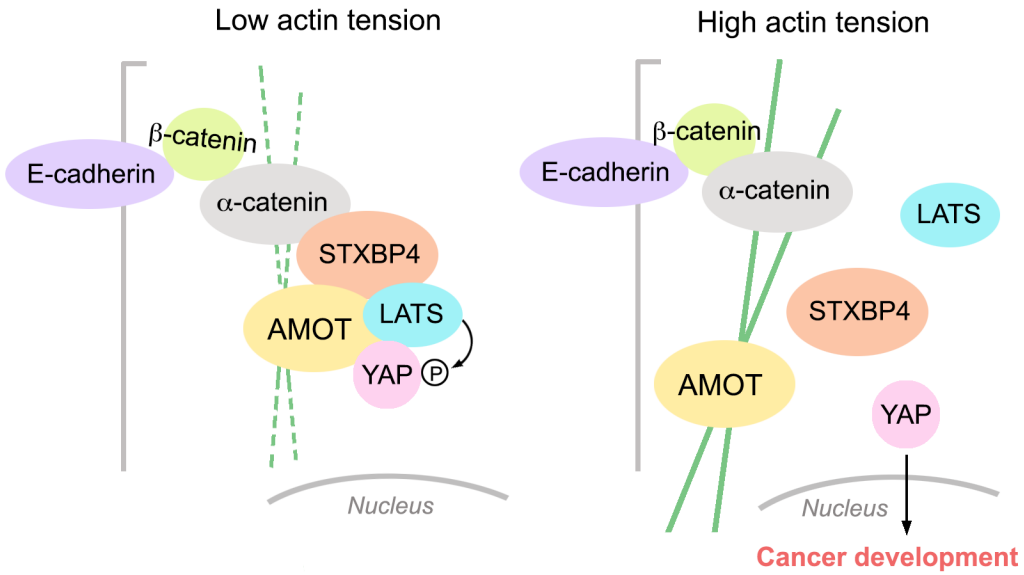
Supplementary Fig. S2.4



Supplementary Fig. S2.4: Schematic illustration of the human proteome search for the WW domain-containing proteins that fit the Hippo WW domain 9-amino acid sequence criterion.

The identified 9-amino acid sequence was subjected to the 49 WW domain-containing proteins in human proteome. Total 12 WW domain-containing proteins were found to contain the WW domains fitting the Hippo WW domain criterion.

Supplementary Fig. S2.5

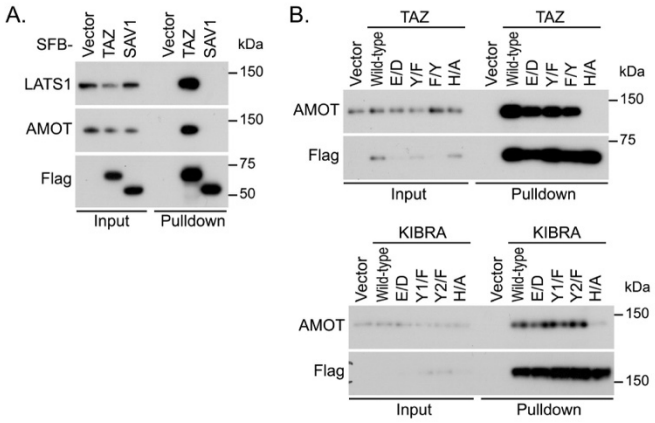


Supplementary Fig. S2.5: A proposed model for the STXBP4-mediated Hippo pathway regulation in response to actin cytoskeleton tension change.

Under low actin tension, STXBP4 assembles a protein complex comprising α-catenin, AMOT, LATS and YAP to promote YAP phosphorylation and cytoplasmic retention. When actin cytoskeleton tension increases, the STXBP4-centered protein complex is dissembled, resulting in YAP's dephosphorylation and nuclear translocation as well as the subsequent cancer development.

2.1Appendix Figure

Appendix Figure S2.6

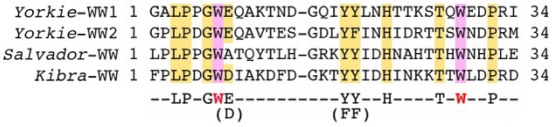


Appendix Figure S2.6 Characterization of the Hippo WW domain binding specificity.

(**A**) Hippo pathway components TAZ but not SAV1 interacts with AMOT and LATS1. HEK293T cells were transfected with the indicated SFB-tagged constructs and subjected to the pulldown assay.

(**B**) Examination of the conservative substitution mutations for the identified 9-amino acid sequence. HEK293T cells were transfected with the indicated SFB-tagged constructs and subjected to the pulldown assay. The tandem tyrosine residues within the 9-amino acid sequence of KIBRA were indicated as Y1 and Y2, respectively.

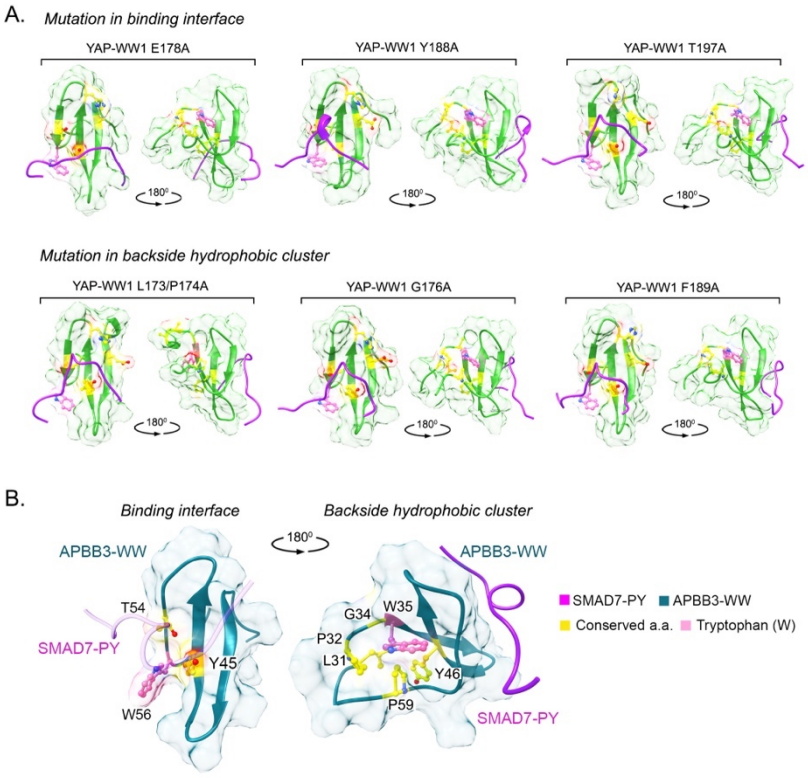
Appendix Figure. S2.7



Appendix Figure. S2.7. Examination of the identified 9-amino acid sequence for the *Drosophila* Hippo pathway components.

Sequence alignment of the WW domains derived from the *Drosophila* Hippo WW domain-containing components. The two conserved tryptophan restudies were highlighted in purple. As compared with the 9-amino acid sequence, the conserved amino acid residues were highlighted in yellow.

Appendix Figure. S2.8

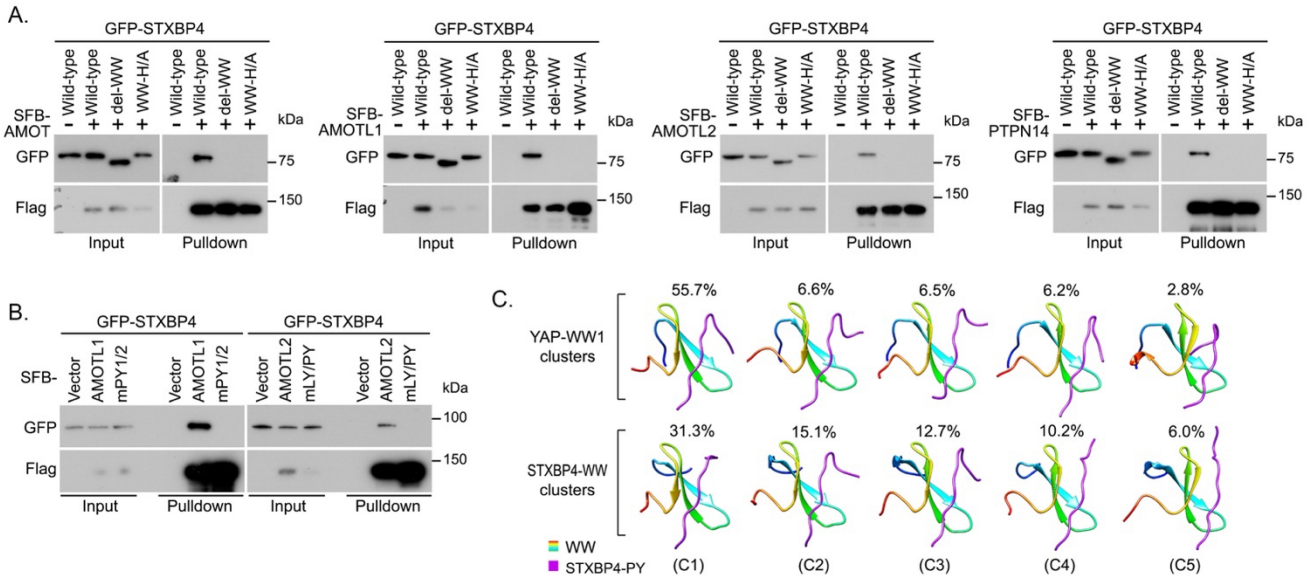


Appendix Figure. S2.8. Characterization of the identified 9-amino acid sequence through simulation analyses.

(**A**) Simulation analysis of the indicated YAP-WW1 mutant/SMAD7-PY complexes.

(**B**) Illustration of the identified 9-amino acid sequence in the APBB3-WW/SMAD7-PY complex. The NMR solution structure of the APBB3-WW domain (2YSC) was used for simulation. SMAD7-PY peptide was adjusted to 50% transparence to show the residue details on the binding interface.

Appendix Figure. S2.9

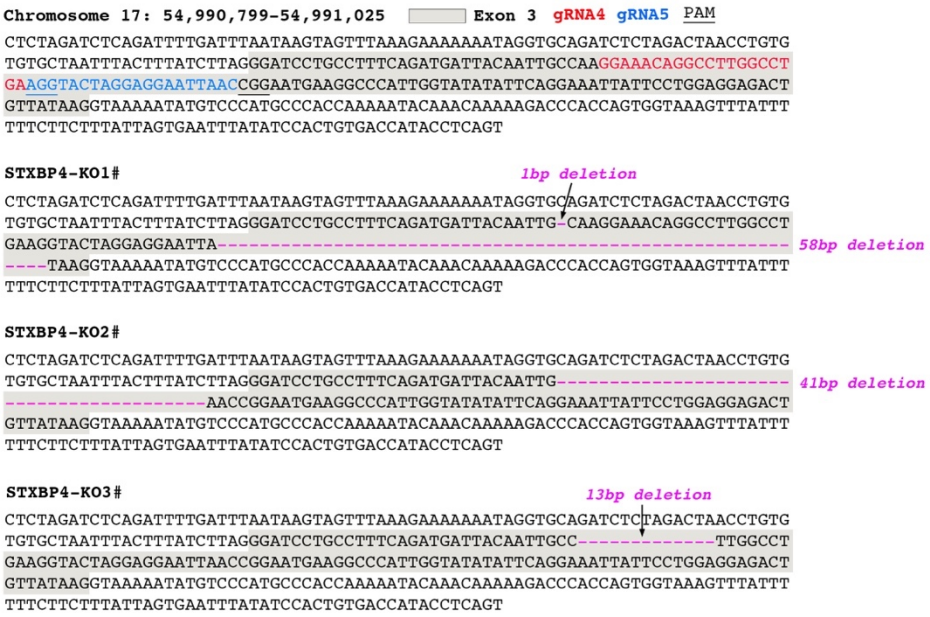


Appendix Figure. S2.9. STXBP4 associates with the Hippo PY motif-containing proteins.

(**A** and **B**) The association between STXBP4 and the indicated Hippo PY motif-containing proteins is mediated by the WW domain (**A**) and PY motif (**B**). HEK293T cells were transfected with the indicated constructs and subjected to the pulldown assay.

(**C**) Simulation of the STXBP4-WW and SMAD7-PY complex structure. The top five WW-PY structure clusters were shown for both YAP-WW1/SMAD7-PY and STXBP4-WW/SMAD7-PY complexes. The frequency rate was shown for each cluster. C, cluster.

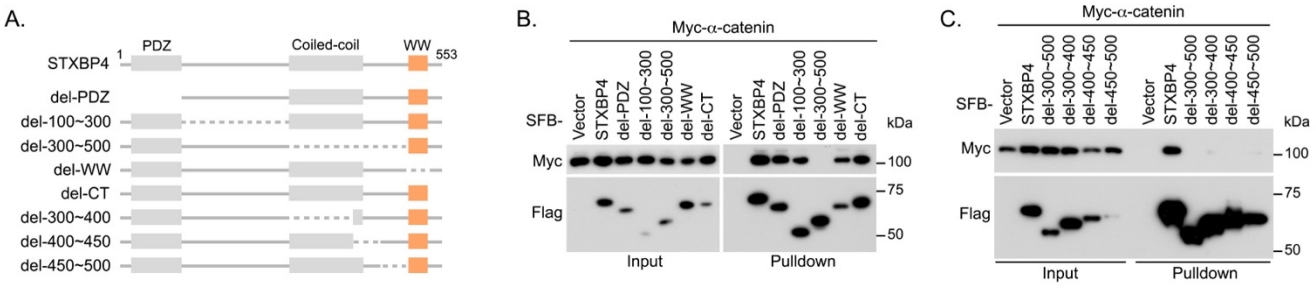
Appendix Figure. S2.10



Appendix Figure. S2.10 Genomic DNA sequencing results for the STXBP4 knockout (KO) cell lines as generated via CRISPR/Cas9.

Among the five designed guide RNAs (gRNAs), only the gRNA4 and gRNA5-targeted regions showed genomic editing for all the three STXBP4 KO cell lines. The genomic editing details for each STXBP4 KO cell line were indicated.

Appendix Figure. S2.11

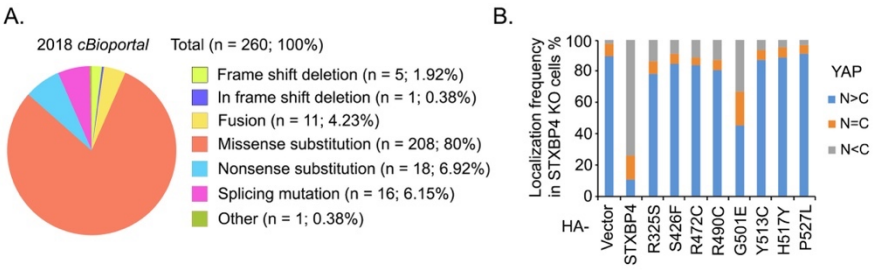


Appendix Figure. S2.11. STXBP4 interacts with α-catenin.

(**A**) Schematic illustration of a series of STXBP4 protein truncation and deletion mutants used in this study.

(**B** and **C**) Mapping the α-catenin binding region in STXBP4. An internal region (300~500 residues) of STXBP4 is required to associate with α-catenin (**B**), and we failed to further narrow down the binding region within the 300~500 residues (**C**).

Appendix Figure. S2.12

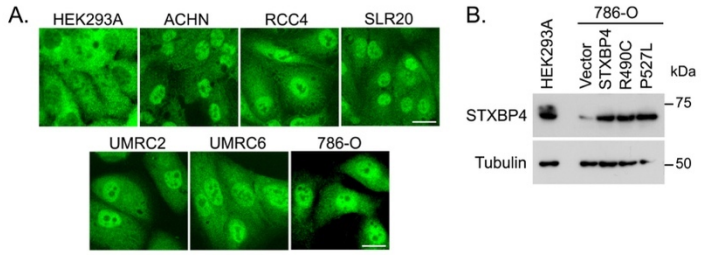


Appendix Figure. S2.12 STXBP4 binds α-catenin and AMOT to regulate YAP.

(**A**) Summary of STXBP4 mutations in cBioportal web database (http://www.cbioportal.org).

(**B**) Interactions with α-catenin and the Hippo PY motif-containing proteins are both required for the STXBP4-mediated YAP suppression. The indicated STXBP4 mutants were expressed in the STXBP4 KO cells and immunofluorescent staining was performed. HA-positive cells from ~30 different views (~200 cells in total) were randomly selected and quantified for YAP localization. N, nuclear localization. C, cytoplasmic localization.

Appendix Figure. S2.13



Appendix Figure. S2.13. STXBP4 is a potential tumor suppressor in kidney cancer.

(**A**) YAP is highly enriched in the tested ccRCC cancer cell lines. YAP cellular localization is detected by immunofluorescent staining. Scale bar, 20 μm.

(**B**) STXBP4 protein expression is examined in the 786-O cells that were transduced with STXBP4 and its cancer-derived mutants.

Appendix Table S2.1: Simulation Conditions

Structure	# of Simulations	PDB ID	Temperature (K)	Start – End Time per sim. (μs)	Ions & Waters
(**WT**) YAP- WW1 & SMAD7	3	2LTW	300	0.4–1	1 Na+ 3147waters
(**WT**) STXBP4-WW & SMAD7	3	2YSG, 2LTW	300	0.6–1	3 Na+ 4222waters
(**WT**) APBB3-WW & SMAD7	3	2YSC, 2LTW	300	0-1	1 Cl- 3897-5433waters
(**Mutant**) YAP-WW1 L173A/P174A & SMAD7	3	2LTW	300	0-1	1 Na+ 3048-4496waters
(**Mutant**) YAP-WW1 G176A & SMAD7	3	2LTW	300	0-1	1 Na+ 2880-3528waters
(**Mutant**) YAP-WW1 W177A & SMAD7	3	2LTW	300	0-1	1 Na+ 3235-4118waters
(**Mutant**) YAP-WW1 E178A & SMAD7	3	2LTW	300	0-1	3365-3546waters
(**Mutant**) YAP-WW1 Y188A & SMAD7	3	2LTW	300	0-1	1 Na+ 3292-4430waters
(**Mutant**) YAP-WW1 F189A & SMAD7	3	2LTW	300	0-1	1 Na+ 3260-3831waters
(**Mutant**) YAP-WW1 H192A & SMAD7	3	2LTW	300	0-1	1 Na+ 2916-4162waters
(**Mutant**) YAP-WW1 T197A & SMAD7	3	2LTW	300	0-1	1 Na+ 2872-3139waters
(**Mutant**) YAP-WW1 W199A & SMAD7	3	2LTW	300	0-1	1 Na+ 3608-3974waters

(**Mutant**) YAP-WW1 P202A & SMAD7	3	2LTW	300	0-1	1 Na+ 2789-3887waters
(**WT**) apo YAP-WW1	3	2LTW	300	0-1	2870waters
(**Mutant**) apo YAP-WW1 L173A/P174A	3	2LTW	300	0-1	2669-2739waters
(**Mutant**) apo YAP-WW1 G176A	3	2LTW	300	0-1	2761-2998 waters
(**Mutant**) apo YAP-WW1 W177A	3	2LTW	300	0-1	2839-2988 waters
(**Mutant**) apo YAP-WW1 E178A	3	2LTW	300	0-1	1Cl-2813-3088 waters
(**Mutant**) apo YAP-WW1 Y188A	3	2LTW	300	0-1	2661-3097waters
(**Mutant**) apo YAP-WW1 F189A	3	2LTW	300	0-1	2882-2996 waters
(**Mutant**) apo YAP-WW1 H192A	3	2LTW	300	0-1	2834-2974 waters
(**Mutant**) apo YAP-WW1 T197A	3	2LTW	300	0-1	2822-3022 waters
(**Mutant**) apo YAP-WW1 W199A	3	2LTW	300	0-1	2788-2934 waters
(**Mutant**) apo YAP-WW1 P202A	3	2LTW	300	0-1	2773-3008 waters

Appendix Table S2.1. List of simulation conditions.

Summary of all the structural simulation conditions and time/frames used for all the indicated analyses. The start and end times for the YAP-WW1 and STXBP4-WW complexes were determined via convergence analyses of MM/PBSA calculations.

2.15 Supplementary Electronic Tables

Table S2.1 Protein identification list for the WW domain-containing protein proteomic study

Table S2.2 Peptide identification list for the WW domain-containing protein proteomic study

Table S2.3 HCIP list for the WW domain-containing protein proteomic study

Table S2.4 Gene Ontology analysis for the HCIPs

Table S2.5 List of YAP protein sequences used for the evolutionary analysis

Table S2.6 List of the human WW domain-containing proteins

Table S2.7 List of the STXBP4-related cancer mutations based on cBioportal

CHAPTER 3

Interactome analysis of human phospholipase D and phosphatidic acid-associated protein network

Rebecca Elizabeth Kattan[1], Han Han[1], Gayoung Seo[1], Bing Yang[1], Yongqi Lin[1], Max Dotson[1], Stephanie Pham[1], Yahya Menely[1], and Wenqi Wang[1,*]

[1]Department of Developmental and Cell Biology, University of California, Irvine, Irvine, CA 92697, USA

[*]Correspondence: wenqiw6@uci.edu (W.W.)

This chapter is derived from the manuscript published in Molecular & Cellular Proteomics: Mol Cell Proteomics. 2022 February 2022 (21)2, doi.org/10.1016/j.mcpro.2022.100195

3.1 Abstract

Mammalian phospholipase D (PLD) enzyme family consists of six members. Among them, PLD1/2/6 catalyze phosphatidic acid (PA) production, while PLD3/4/5 have no catalytic activities. Deregulation of the PLD-PA lipid signaling has been associated with various human diseases including cancers. However, a comprehensive analysis of the regulators and effectors for this crucial lipid metabolic pathway has not been fully achieved. Using a proteomic approach, we defined the protein interaction network for the human PLD family of enzymes and PA, and revealed diverse cellular signaling events involving them. Through it, we identified PJA2 as a novel E3 ubiquitin ligase for PLD1 involved in control of the PLD1-mediated mTOR signaling. Additionally, we showed that PA interacted with and positively regulated sphingosine kinase 1 (SPHK1). Taken together, our study not only generates a rich interactome resource for further characterizing the human PLD-PA lipid signaling, but also connects this important metabolic pathway with numerous biological processes.

3.2 Key words

PLD, PA, PJA2, SPHK1, S1P

3.3 Abbreviations

PLD (phospholipase D), PA (phosphatidic acid), CRAPome (Contaminant Repository for Affinity Purification), SAINT (Significance Analysis of INTeractome), HCIP (high confidence interacting protein), PPI (protein-protein interaction), SFB (S tag-Flag tag-SBP tag), TAP-MS (tandem affinity purification coupled with mass spectrometry), mTOR (mechanistic target of rapamycin), S6K (S6 kinase), SPHK1 (sphingosine kinase 1), S1P (sphingosine-1-phosphatate).

3.4 Introduction

The PLD enzyme family is known for its catalytic action to produce phosphatidic acid (PA) (Billah *et al*, 1989). Originally identified in *salmonella typhimurium* (Pohlman *et al*, 1993a), PLDs display conserved functions in numerous organisms (Jenkins & Frohman, 2005b). The mammalian PLD enzymes comprise six members (i.e. PLD1-6), sharing the conserved HKD motif. The HKD motif contains the sequence HxKxxxxD (H: Histidine; K: lysine; D: Aspartic acid; x: any amino acid) and is essential for the catalytic activity of PLDs (Sung *et al*, 1997b). Additional domains found in some PLD family members like PH and PX domains are responsible for their membrane localization (Stahelin *et al*, 2004).

Among the human PLD enzymes, PLD1 and PLD2 are ubiquitously expressed in multiple tissues and play vital roles in various physiological processes, including receptor-mediated endocytosis, cell migration, cytoskeletal organization (Colley *et al*, 1997a; Kang *et al*, 2014). Deregulation of these two PLD members have been associated with different human diseases (Min *et al*, 2001; Saito *et al*, 2007). PLD1 and PLD2 are 50% identical and share similar regulations and functions. For example, they both can be activated by ARF family GTPases (Cockcroft *et al*, 1994; Hammond *et al*, 1995) and PI(4,5)P2 (Hammond *et al*, 1997). PLD1 is localized in perinuclear structures such as Golgi, early endosomes, late endosomes, and lysosomes (Colley *et al.*, 1997a; Dall'Armi *et al*, 2010; Freyberg *et al*, 2001b; Lucocq *et al*, 2001a); while PLD2 is found in the plasma membrane, exosomes, Golgi, and cytoplasm (Freyberg *et al*, 2002; Laulagnier *et al*, 2004). These findings suggest their roles in organelles-based cellular events. Additionally, PLD1/2 have been reported to regulate several growth-related signaling pathways through its lipid product PA such as the mammalian target of rapamycin (mTOR) (Fang *et al*,

2003) and the Hippo pathway (Han *et al*, 2018a), highlighting their crucial roles in growth control and cancer development. As for other PLD family members, PLD3 and PLD4 are endoplasmic reticulum (ER)-associated proteins with no classic catalytic activity (Munck *et al*, 2005a; Yoshikawa *et al*, 2010a). Instead, PLD3 and PLD4 show 5' exonuclease activity involved in inflammatory cytokine response (Gavin *et al*, 2018a). PLD5 is another PLD family member with no catalytic activity (Dennis, 2015b), and has been reported as an oncogene involved in prostate tumorigenesis (Liu *et al.*, 2021). PLD6 is localized on mitochondria and regulates mitochondrial fusion (Zhang *et al*, 2016). PA produced by PLD6 on the mitochondrial surface is converted to diacylglycerol to affect mitochondrial dynamics (Huang *et al*, 2011b). PLD6 also plays a role in the Myc-mediated AMPK activation, which in turn inhibits YAP/TAZ in the Hippo pathway (von Eyss *et al*, 2015a).

As the major lipid product generated by PLD family enzymes, PA is found in cell plasma membrane and vesicles, and has been extensively studied for its roles in vesicular trafficking, cytoskeletal dynamics, and endocytosis (Bader & Vitale, 2009; Bowling *et al*, 2021). Interestingly, studies also revealed PA as a key signaling molecule involved in many signaling events via forming a complex with proteins. For example, PA directly binds and activates mTOR (Fang *et al*, 2001a; Veverka *et al*, 2008). PA inhibits the Hippo pathway by interacting with its two components LATS and NF2 (Han *et al.*, 2018a). MAPK cascade is regulated by PA through its physical interaction with Raf-1 (Rizzo *et al*, 2000). PA binds KIF5B to regulate MT1-MMP in cancer metastasis (Wang *et al*, 2017b).

Pathologically, the PLD enzymes have been explored for their critical roles in numerous human diseases, including influenza (Oguin *et al*, 2014a), neurodegenerative disorders (Cai *et al*, 2006), autism (Anney *et al*, 2010), and fertility issues (Huang *et al.*, 2011b). In addition, enhanced PLD expression and activity are associated with many types of human cancer (Bruntz *et al*, 2014; Gozgit *et al*, 2007; Henkels *et al*, 2013b; Liu *et al*, 2015), implicating their potential roles as biomarkers and targets for cancer treatment. Therefore, further characterizing the regulators and effectors for this key lipid enzyme family and its lipid product PA will not only provide novel insights into the PLD-PA lipid pathway in normal physiology, but also reveal potential therapeutic targets for treating human diseases like cancer.

In this study, we conducted a large-scale proteomic study for the human PLD family enzymes and PA, and defined the protein interaction network for this important lipid metabolic pathway. Through it, we connected the PLD enzymes and PA with various biological processes and discovered interacting proteins for PLDs and PA that may exert their cellular functions. Our functional studies further identified PJA2 as a novel E3 ubiquitin ligase for PLD1 and a negative regulator of mTOR signaling. Moreover, by characterizing the PA-sphingosine kinase 1 (SPHK1) lipid-protein complex, we established PA as a lipid regulator of SPHK1 to positively drive the SPHK1-dependent sphingosine-1-phosphate (S1P) production. Taken together, our interactome analysis of the PLDs and PA-associated protein interaction network provides a rich resource for further exploration of this key lipid metabolic pathway in various signaling events and biological processes.

3.5 Results

Mapping the PLDs and PA-associated protein network

To gain an overview of the human PLD family enzymes, we first examined their subcellular localization under normal culture condition. As shown in **Figure 3.1A**, a significant portion of PLD1 was localized on lysosomes, which is consistent with previous studies (Dall'Armi *et al.*, 2010; Freyberg *et al.*, 2001b; Lucocq *et al.*, 2001a). PLD2 was majorly localized in the cytoplasm with partial membrane association (**Figures 1A** and **S1A**). Similar to previous reports (Munck *et al.*, 2005a; Yoshikawa *et al.*, 2010a), PLD3 and PLD4 were found as ER-localized proteins (**Figure 3.1A**). In addition, previously uncharacterized PLD5 was found as a cytoplasmic protein, while PLD6 was mostly localized on mitochondria (**Figure 3.1A**). These findings show distinct subcellular localizations for the human PLD enzymes.

To elucidate the intricate protein-protein interaction network for the PLD enzymes, we generated HEK293A cells stably expressing each of the six PLD members fused with a *C*-terminal SFB triple tags (i.e. S tag-Flag tag-SBP tag) through lentiviral infection and puromycin selection (**Figure 3.1B**). These PLD stable cells then underwent two independent tandem affinity purification (TAP) experiments after verification of bait protein expression and cellular localization. The associated proteins with each PLD enzyme were further examined using mass spectrometry (MS). In addition, the PA-bound proteins were isolated from HEK293A cell lysates using PA-conjugated agarose beads and analyzed by MS (**Figure 3.1B)**. A complete list of the peptides and proteins identified in this study can be found in **Tables S1** and **S2**, respectively.

To refine the identified PLDs-interacting proteins, a group of 72 unrelated TAP-MS experiments performed under similar purification conditions were included as controls. Moreover, two independent pulldown experiments using agarose beads only were performed as controls for the identified PA-interacting proteins. We assigned a SAINT score to each identified hit and considered any interaction with a SAINT score at least 0.80 and raw spectra counts at least 2 as a high confidence interacting protein (HCIP) (**Figure 3.1B**). Through this filtration, a total of 303 HCIPs among the 6586 preys were identified for the PLD-PA lipid pathway (**Figure 3.1C** and **Table S3**). The HCIP number and total spectral count (TSC) number for each PLD member and PA were summarized in **Figure 3.1D**. Data reproducibility by comparing the two biological repeats for the PLDs and PA proteomic studies suggested the reliability of the dataset (**Figure 3.1E**). Gene Ontology (GO) analysis indicated that the identified HCIPs of PLDs and PA are widely distributed with different subcellular localizations and are involved in numerous cellular functions (**Figures 1F** and **1G**; **Table S3**). Since the cells used for our proteomic study were grown under normal culture condition, the identified HCIPs should be considered as the basal-state interactome for the human PLD-PA lipid pathway. In addition, our proteomic study was performed in HEK293A cells, thus it may not recapitulate specific interacting proteins and unique cellular functions for some PLD proteins in specialized cells.

We next compared our PLDs HCIP list with the data reported in the BioPlex Interactome database (https://bioplex.hms.harvard.edu). Of 45 protein interactions identified for PLD1 in BioPlex, only 1 were identified as HCIP in our PLD1 interactome study (**Figure S3.1B** and **Table S3.4**). Of 69 interactions identified for PLD2 in BioPlex, 10 were identified as HCIPs in our PLD2 interactome study (**Figure S3.1B** and **Table S3.4**). Of 34 interactions identified for

PLD6 in BioPlex, 2 were identified as HCIPs in our PLD6 interactome study (**Figure S3.1B** and **Table S3.4**). Actually, several of the interactions reported in BioPlex were also identified in our data set but did not pass the stringent filtering criteria used in our study (8 interactions for PLD1, 16 interaction for PLD2, and 4 interactions for PLD6). In addition, the PLD3 interactome study in BioPlex only reported one interaction that was not identified in our data set (**Figure S3.1B** and **Table S3.4**). The interactome analyses of PLD4 and PLD5 were not included in BioPlex, although PLD5 was revealed as a prey protein in some interactome studies there (**Figure S3.1B** and **Table S3.4**). Moreover, we identified additional high-confidence interactions with PLD proteins that were not reported in BioPlex, further extending our knowledge of the PLDs-associated protein network. Taken together, these results suggest that our PLDs proteomic data set not only reproduced previous findings but also provided an additional collection of candidate HCIPs for further validation.

Overview of the protein interaction landscape for the PLD-PA lipid pathway

Interestingly, metascape analysis revealed that many HCIPs were shared among PLD family enzymes like PLD3, PLD4 and PLD6; while this was not the case for PLD1 or PA (**Figure 3.2A**). We also connected the HCIPs of PLDs and PA with various biological processes via GO analysis. As shown in **Figure 3.2B** and **Table S3.5**, although the HCIPs of PLDs and PA were functionally categorized into multiple biological processes, they were highly enriched in protein modification, metabolism, and organelle-based cellular functions. As for each PLD member, the HCIPs of PLD3 and PLD4 were majorly involved in protein folding and ER-associated functions (**Figure 3.2B**), which was consistent with the ER localization of these two PLD members (**Figure 3.1A**). PLD3 also played roles in lipid biosynthesis and nuclear envelop organization,

while PLD4 specifically regulated exocytosis (**Figure 3.2B**), suggesting their unique functions through their binding proteins. Although localized on mitochondria (**Figure 3.1A**), PLD6 could play a role in ER-related cellular events, because its HCIPs shared some biological processes with those of PLD3/4 (**Figure 3.2B**). In addition, this GO analysis suggested roles of PLD1 and PA in synaptic vesicle transport and vesicle protein recruitment, respectively (**Figure 3.2B**).

Based on these findings, we further examined the PLDs and PA-associated protein interaction network through a prey-oriented hierarchical clustering analysis. As shown in **Figure 3.2C**, some common subsets were revealed between different PLD family members. For example, PLD3 and PLD4 interacted with a wide range of ER proteins such protein disulfide-isomerase family members PDIA3/4/6 (cluster 1) (**Figure 3.2D**). Additionally, PLD3, PLD4 and PLD6 shared HCIPs involved in ER protein folding, including OS9, P4HB, CHPF2, HYOU1, UGGT1, GANAB (cluster 2 and cluster 3) (**Figure 3.2D**). These findings suggested that PLD3, PLD4 and PLD6 may play roles in ER-associated protein quality control and stress response. Interestingly, both PLD3 and PLD4 interacted with a group of lysosome-associated transmembrane proteins TM9SF1-3 (cluster 2) (**Figure 3.2D**), suggesting potential roles of PLD3 and PLD4 in mediating the crosstalk between ER and lysosomes. Moreover, PLD2 and PLD6 were strongly associated with tubulin subunits (i.e. TUBA1C, TUBB2B and TUBB4B) (cluster 3) (**Figure 3.2D**), indicating that their functions may involve the microtubule-based events. Again, we hardly identified significant HCIPs clusters of PLD1 or PA with other PLD members (**Figure 3.2C**), consistently showing relatively unique binding proteins for PLD1 and PA.

Next, we organized the PLDs and PA-associated protein network individually and functionally by clustering the HCIPs based on their cellular localization **(Figure 3.3**). This comparative interaction network organization highlighted the different groups of HCIPs for each PLD member and PA and implicated their roles in the related biological processes. We also compared the HCIPs for individual PLD enzyme and PA, and revealed a group of HCIPs as shared binding proteins among PLDs and PA (**Figure S3.2**), which is consistent with the prey-oriented hierarchical clustering analysis (**Figure 3.2C**). Taken together, these results show that PLD enzymes and PA share common and unique binding proteins, implicating their both redundant and specific functions within various cellular events.

Validation of the PLD3/4/6-associated protein interaction network

By examining the PLD3 and PLD4-associated proteins (**Figure 3.3** and **Table S3.3**), we found that they reciprocally identified each other (**Figure 3.4A**). Indeed, pulldown experiments showed that PLD3 and PLD4 not only strongly interacted with each other, but also bound their own, among the six PLD members (**Figure 3.4B**). In addition to some shared HCIPs, PLD3 and PLD4 also specifically interacted with some ER proteins (**Figure 3.4A**). Among them, the PLD3-PNPLA6 (**Figure 3.4C**) and PLD4-RPN2/TMX3 (**Figure 3.4D**) complexes were validated through pulldown assays.

Notably, several lysosome-associated TM9SF family proteins (i.e. TM9SF1-4) were identified as shared binding partners for PLD3 and PLD4 (**Figure 3.4A**). Among them, the associations of TM9SF1 and TM9SF3 with PLD3 (**Figures 4E** and **4F**) and PLD4 (**Figures 4G** and **4H**) were experimentally confirmed. Since TM9SFs and PLD3/4 are transmembrane proteins localized on

lysosomes and ER, respectively, their interaction could play a role in ER-lysosomes contact formation. Interestingly, overexpression of PLD3 but not vector control significantly promoted the co-localization between ER and lysosomes as visualized by mCherry-ER-3 and LAMP1, respectively (**Figures 4I** and **4J**). These data implicate a potential role of the PLD3/4-TM9SFs complex in mediating the contact formation between ER and lysosomes.

Unexpectedly, although PLD6 is majorly localized on mitochondria (**Figure 3.1A**), our proteomic study revealed multiple ER proteins as its binding proteins (**Figure 3.3**). Indeed, among the tested HCIPs (**Figure 3.4K**), the association of PLD6 with ER proteins ERLEC1, HSPA5 and P4HB were confirmed through pulldown assays (**Figure 3.4L**), which was comparable to that between PLD6 and two mitochondrial proteins ACOT9 and CPT1A (**Figure 3.4L**). Thus, these data indicate a potential role of PLD6 in ER-related functions, which deserves further investigation.

Characterization of the PLD1/2/5-based protein interaction network uncovers PJA2 as an E3 ubiquitin ligase for PLD1

Upon analyzing the HCIPs of PLD5, we identified a Hedgehog pathway regulator ANKMY2 (Somatilaka *et al*, 2020) as a shared interacting protein with PLD1 and PLD2 (**Figure 3.5A**). Although pulldown experiments confirmed the interaction between PLD1/2/5 and ANKMY2 (**Figure 5B**), PLD2 strongly interacted with ANKMY2 as compared to PLD1 and PLD5 (**Figure 5B**). This result was consistent with the observation that PLD2 was the only PLD member identified in the ANKMY2 reverse TAP-MS study (**Figure 5A**). In addition to ANKMY2, several PLD5 HCIPs, including PCCB (**Figure 5C**), GNAI3 (**Figure 5D**), TTC1 (**Figure 5E**)

and UNC45A (**Figure 5F**), were further validated for their association with PLD5 through pulldown assays. These results suggest a variety of cellular events that may involve PLD5, providing functional insights into this previously uncharacterized PLD member.

PLD1 plays a key role in PA production and dysregulated PLD1 expression has been associated with many oncogenic signaling events and observed in various human cancers (Bruntz *et al.*, 2014; Henkels *et al.*, 2013b; Min *et al.*, 2001). Interestingly, an E3 ubiquitin ligase PJA2 was identified as a binding partner for PLD1 (**Figures 3.3** and **5A**). Therefore, we further examined whether PJA2 could target PLD1 protein stability and modulate its-associated signaling events. Indeed, PJA2 specifically interacted with PLD1 among the six PLD members (**Figure 5G**). Overexpression of PJA2 reduced PLD1 protein expression (**Figure 5H**) but did not affect that of PLD2 (**Figure 5I**). Moreover, expression of PJA2, but not its ring domain mutant $PJA2^{RM}$, dramatically induced PLD1 ubiquitination (**Figure 5J**). On the other hand, depletion of PJA2 stabilized PLD1 and another known PJA2 substrate MOB1 (Lignitto *et al*, 2013) (**Figure 5K**), and also largely reduced the ubiquitination of PLD1 (**Figure 5L**). Given the critical role of PLD1 in activating mTOR signaling (Fang *et al.*, 2003), we examined whether PJA2 could modulate mTOR activity. Indeed, as shown in **Figure 5K**, loss of PJA2 promoted the mTOR-dependent S6K phosphorylation. Taken together, these data demonstrate that PJA2 serves as an E3 ubiquitin ligase for PLD1 and controls PLD1-downstream signaling events.

Analysis of the PA interactome reveals PA as a positive regulator of SPHK1

Interestingly, our proteomic study showed that the PA HCIPs have diverse cellular localizations, including cytoplasm, nucleus, cytoskeleton, plasma membrane, ER, peroxisome, mitochondria,

endosome and lysosome (**Figure 3**). Based on this finding, we performed the PA-beads pulldown assay and confirmed the association between PA and its HCIPs such as HSDL2 (peroxisome) (**Figure 3.6A**), HSD17B4 (peroxisome) (**Figure 3.6B**), VIM (cytoskeleton) (**Figure 3.6C**), GDAP2 (lysosome) (**Figure 3.6D**), and ETV6 (nucleus) (**Figure 3.6E**). These results suggest potential roles of PA in the cellular events involving these interacting proteins.

Among the PA HCIPs (**Figure 3.3** and **Table S3.3**), SPHK1 is a sphingosine kinase, whose product spingosine-1-phosphate (S1P) has been discovered as a key signaling molecule involved in various growth-related cellular events such as cell proliferation, survival, migration, transformation (Olivera & Spiegel, 1993). Although the physical interaction between PA and SPHK1 has been previously reported (Delon *et al*, 2004), the functional significance underlying this lipid-protein complex formation has not been fully elucidated.

To address it, we further characterized the interaction between PA and SPHK1. As shown in **Figure 3.6F**, PA specifically interacted with SPHK1 but not SPHK2. Moreover, SPHK1 kinase dead mutant (Chan *et al*, 2012; Kim & Sieburth, 2018; Pitson *et al*, 2002) failed to bind PA (**Figure 3.6G**), while its two hydrophobic patch mutants (i.e. L280Q and F283A/L284Q), which were previously reported to disrupt the association between SPHK1 and liposome (Shen *et al*, 2014), can still bind PA (**Figure 3.6G**). We also examined the cellular localization of SPHK1 and PA. As shown in **Figure 3.6H**, SPHK1 as well as its two hydrophobic patch mutants were co-localized with PA on the punctate structures in the cytoplasm, which are known as Rab5-positive early endosomes and endocytic intermediates (Shen *et al.*, 2014). In contrast, SPHK2 and SPHK1 kinase dead mutant (KD-5A), which cannot bind PA (**Figures 6F** and **6G**), failed to

localize on the PA-positive punctate structures (**Figure 3.6H**), suggesting that PA binding is required in this process.

Since SPHK1 kinase dead mutant (KD-5A) failed to bind PA (**Figure 3.6G**), it raised the possibility that PA could regulate SPHK1 kinase activity. To test it, we isolated SPHK1 from the serum starved HEK293A cells, incubated it with PA, and subjected it to *in vitro* kinase assay using sphingosine as a substrate. As shown in **Figure 3.6I**, PA incubation largely enhanced the ability of SPHK1 to phosphorylate sphingosine to produce S1P. Moreover, treatment with PLD1/2 inhibitors FIPI and CAY10594 significantly reduced the SPHK1-induced S1P production (**Figure 3.6J**), suggesting that targeting PA production inhibited SPHK1 kinase activity *in vivo*. Taken together, these data demonstrate PA as a positive regulator of SPHK1, where PA binds SPHK1 and regulates its cellular localization and kinase activity.

3.6 Discussion

In this study, we defined the protein interaction landscape for the human PLD family enzymes and their lipid product PA, and identified over 300 HCIPs for them, which greatly expanded our knowledge of this lipid metabolic pathway in diverse signaling events and cellular functions.

Since some PLD members comprised similar protein domains (**Figure 3.1B**) and showed similar subcellular localizations (**Figure 3.1A**), PLD enzymes were found to share some binding partners despite having the unique interacting proteins of their own (**Figures 3.2, 3.3** and **S2**), providing us an opportunity to further explore both generic and specific functions/regulations for the PLD enzymes. Indeed, our validation studies not only revealed several specific protein

complexes for PLD members such as PLD3-PLD4 (**Figure 3.4B**), PLD3-PNPLA6 (**Figure 3.4C**), PLD4-RPN2 (**Figure 3.4D**), PLD4-TMX3 (**Figure 3.4D**), PLD1-PJA2 (**Figure 5G**); but also discovered TM9SFs and ANKMY2 as shared binding partners for PLD3/PLD4 (**Figures 4A** and **4E-4H**) and PLD1/PLD2/PLD5 (**Figures 5A** and **5B**), respectively.

Interestingly, our functional studies also connected the PLD family members to several unexpected cellular events. For example, although PLD3 and PLD4 are ER-bound proteins (**Figure 3.1A**), their ER-related functions remain largely unknown. Notably, our study uncovered many key players in ER protein quality control as binding partners for PLD3 and PLD4 (**Figures 3.2D** and **3.3**), suggesting potential roles of PLD3 and PLD4 in regulating ER stress response. In addition, PLD3 and PLD4 formed a complex with lysosome proteins TM9SFs (**Figures 4A** and **4E-4H**), and overexpression of PLD3 induced the contact formation between ER and lysosomes (**Figures 4I** and **4J**). Since PLD3/4 and TM9SFs are transmembrane proteins on ER and lysosomes, respectively, following work will be focused on determining whether their interaction is directly mediated by their cytosolic regions or through additional "ligand factors". In addition, the signaling contexts as well as the functional significance underlying the PLD3/4-TM9SFs complex formation deserve further investigation. Strikingly, our data also revealed the hetero- and homo- dimerization occurring between PLD3 and PLD4 (**Figure 3.4B**). Therefore, it will be highly interesting to elucidate the role of such dimer formation in regulating the ER-related functions of PLD3/4 as mentioned above.

Our proteomic analysis uncovered PJA2 as an E3 ubiquitin ligase for PLD1 **(Figures 5G-5L**), whose loss increased mTOR signaling (**Figure 5K**). Although PLD1 also functions as a negative

regulator of the Hippo pathway (Han *et al.*, 2018a), we did not observe a dramatic change of YAP phosphorylation or subcellular localization in the PJA2 KO HEK293A cells. This could be explained by the fact that PJA2 deficiency also stabilized MOB1, a key Hippo pathway adaptor, to activate Hippo signaling (Lignitto *et al.*, 2013) and compromise the effect as caused by the increased expression of PLD1. Actually, in addition to PLD1 and MOB1, PJA2 has been identified as an E3 ubiquitin ligase for other key signaling proteins, including PKA (Lignitto *et al*, 2011), CDK5R1 (Sakamaki *et al*, 2014), KSR1 (Rinaldi *et al*, 2016; Zhao *et al*, 2021), HIV-1 protein Tat (Faust *et al*, 2017), MFHAS1 (Zhong *et al*, 2017), and TCF/LEF1 (Song *et al*, 2018), highlighting a complex downstream signaling network as controlled by PJA2.

Many studies including ours have revealed PA as a key signaling molecule through its physical interaction with proteins, which prompted us to use a proteomic approach to investigate the PA-associated proteins. Through it, we identified a group of proteins as PA binding proteins (**Figures 3.3** and **3.6A-3.6F**), linking the PLD-PA lipid pathway to additional cellular events and biological processes. Among them, we experimentally confirmed the lipid-protein interaction between PA and SPHK1, which has been previously discovered using different approaches (Delon *et al.*, 2004; Shen *et al.*, 2014). Notably, our functional study further demonstrated PA as a key regulator of SPHK1 to modulate its cellular localization and kinase activity (**Figures 3.6G-3.6J**), providing novel mechanistic insights into the regulation of this key lipid kinase. Given the crucial roles of SPHK1 in cell proliferation, survival, migration, differentiation, and cancer development (Heffernan-Stroud & Obeid, 2013; Shida *et al*, 2008), it will be highly important to determine whether these SPHK1-related cellular events also involve the PLD-PA lipid pathway.

Notably, although our proteomic study successfully identified a group of PA-binding proteins, we failed to discover some previously reported ones such as PI4P5K, Raf-1, SHP-1, mTOR (Stace & Ktistakis, 2006). This could be explained by the limitation of the PA-conjugated agarose beads. Based on the information provided by the manufacturer, the PA beads used in our study contain the PA with one C-16 carbon chain and one C-6 carbon chain, where the C-6 chain is modified for agarose bead conjugation. Therefore, such conjugation with agarose beads may affect the binding of PA with some proteins and affect the PA membrane accessibility. In addition, our stringent washing process could also interfere with the capture of some transient and/or weak interacting proteins for PA.

In summary, our proteomic study of the PLDs and PA-centered protein interaction network not only reveals a number of cellular functions/regulations for the PLDs/PA-related signaling events, but also generates a comprehensive interactome resource for further characterization of this important lipid metabolic pathway in various biological processes.

3.7 Experimental Procedures

Cell lines

HEK293T cells (a female cell line, ATCC: CRL-3216) were purchased from ATCC and kindly provided by Dr. Junjie Chen (MD Anderson Cancer Center). HEK293A cells (a female cell line, Thermo Fisher Scientific: R70507) cells were purchased from Thermo Fisher Scientific and kindly provided by Dr. Jae-Il Park (MD Anderson Cancer Center). HEK293A and HEK293T cells were maintained in Dulbecco's modified essential medium (DMEM) supplemented with

10% bovine growth serum and 1% penicillin and streptomycin at 37°C in 5% CO_2 (v/v). Plasmid transfection was performed using a polyethylenimine reagent.

Antibodies and chemicals

For Western blotting, anti-Flag (M2) (F3165-5MG, 1:5000 dilution) and anti-α-tubulin (T6199-200UL, 1:5000 dilution) monoclonal antibodies were obtained from Sigma-Aldrich. An anti-Myc (sc-40, 1:500 dilution) monoclonal antibody was purchased from Santa Cruz Biotechnology. An anti-hemagglutinin (HA) monoclonal antibody (MMS-101P, 1:3000 dilution) was obtained from BioLegend. Anti-PJA2 (#40180, 1:1000 dilution), anti-MOB1 (# 3863S, 1:2000 dilution), and anti-phospho-p70 S6 Kinase (Thr389) (#9234S,1:1000 dilution) polyclonal antibodies were obtained from Cell Signaling Technology. An anti-sphingosine-1-phosphate (S1P) (Z-P300; 1:500 dilution) monoclonal antibody was obtained from Echelon Biosciences. For immunostaining, anti-HA (3724S, 1:3000 dilution) and anti-LAMP1 (9091S, 1:400 dilution) polyclonal antibodies were obtained from Cell Signaling Technology. An anti-Flag (F7425-.2MG, 1:5000 dilution) polyclonal antibody was purchased from Sigma-Aldrich. An anti-TOM20 (612278, 1:200 dilution) monoclonal antibody was purchased from BD Biosciences. For chemicals, FIPI hydrocholoride hydrate (#F5807) was obtained from Sigma-Aldrich. CAY10594 (#13207) was obtained from Cayman chemical.

Constructs and viruses

Plasmids encoding the indicated genes were obtained from the Human ORFeome V5.1 library or purchased from DNASU Plasmid Repository. All constructs were generated via polymerase

chain reaction (PCR) and sub-cloned into a pDONOR201 vector using Gateway Technology (Thermo Fisher Scientific) as entry clones. For tandem affinity purification, all entry clones were subsequently recombined into a lentiviral gateway-compatible destination vector for the expression of *C*-terminal SFB-tagged fusion proteins. Gateway-compatible destination vectors with the indicated SFB tag, HA tag, Myc tag, mCherry tag were used to express various fusion proteins.mCherry-Lysosomes-20 (plasmid #55073) and mCherry-ER-3 (plasmid #55041) were obtained from Addgene. PJA2 plasmid was kindly provided by Antonio Feliciello (University Federico II, Italy). RFP-PASS construct was kindly provided by Dr. Guangwei Du (University of Texas Health Science Center at Houston).PCR-mediated site-directed mutagenesis was used to generate amino acid mutations for PJA2 and SPHK1. PJA2 ring domain mutation (RM) was generated by mutating sites Cys634 and Cys671 to Ala (Lignitto *et al.*, 2013). SPHK1 kinase dead (KD-5A) mutant was generated by mutating sites Ser165, Gly166, Asp167, Gly168, and Leu169 to Ala (Chan *et al.*, 2012; Kim & Sieburth, 2018; Pitson *et al.*, 2002). The two SPHK1 hydrophobic patch mutants were generated by respectively mutating the site Leu280 to Ala, and Phe283 to Ala/Leu284 to Gln (Shen *et al.*, 2014). All lentiviral supernatants were generated by transient transfection of HEK293T cells with the helper plasmids pSPAX2 and pMD2G (kindly provided by Dr. Zhou Songyang, Baylor College of Medicine) and harvested 48 hours later. Supernatants were passed through a 0.45-μm filter and used to infect cells with the addition of 8 μg/mL hexadimethrine bromide (Polybrene) (Sigma-Aldrich).

Purification of PLDs and PA-associated protein complexes

HEK293A cells stably expressing SFB-tagged PLD proteins were isolated by culturing in medium containing 2 μg/mL puromycin and validated by immunostaining and Western blotting

as described previously (Wang *et al.*, 2014a). For tandem affinity purification, the HEK293A stable cells were lysed in 10 mL NETN buffer (100 mM NaCl; 20 mM Tris-HCl, pH 8.0; 0.5 mM EDTA; 0.5 % Nonidet P-40) with protease and phosphatase inhibitors at 4 ^{0}C for 20 minutes. The crude lysates were centrifuged at 14,000 rpm at 4 ^{0}C for 15 minutes. The supernatants were incubated with 100 µL streptavidin-conjugated beads (GE Healthcare) at 4 ^{0}C for 6 hours. The beads were then washed 3 times with 10 mL NETN buffer, and bound proteins were eluted with 1.5 mL NETN buffer containing 2 mg/mL biotin (Sigma-Aldrich) at 4 ^{0}C for 12 hours. The elutes were incubated with 25 µL S protein beads (Novagen) at 4 ^{0}C for 4 hours. The beads were then washed 3 times with 1mL NETN buffer and subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis. Each sample was run into the separation gel for a short distance, so that the whole bands could be excised as one sample for in-gel trypsin digestion and mass spectrometry analysis. To isolate PA-associated protein complex, HEK293A cells were lysed in 10 mL NETN buffer with protease and phosphatase inhibitors at 4 ^{0}C for 20 minutes. The crude lysates were centrifuged at 14,000 rpm at 4 ^{0}C for 15 minutes. The supernatants were incubated with 30 µL control beads (#P-B000, Echelon Biosciences) or PA beads (#P-B0PA, Echelon Biosciences) at 4 ^{0}C for 12 hours. The beads were then washed 5 times with 1mL NETN buffer and subjected to sodium dodecyl sulfate polyacrylamide gel electrophoresis. Each sample was run into the separation gel for a short distance, so that the whole bands could be excised as one sample for in-gel trypsin digestion and mass spectrometry analysis.

Mass spectrometry analysis

The mass spectrometry was performed by the Taplin Mass Spectrometry Facility (Harvard Medical School) as described previously (Li *et al.*, 2016a; Wang *et al.*, 2014a). Briefly, the

excised gel bands described above were cut into approximately 1-mm^3 pieces. The gel pieces were then subjected to in-gel trypsin digestion and dried (Shevchenko *et al.*, 1996). Samples were reconstituted in 5 µL of high-performance liquid chromatography (HPLC) solvent A (2.5% acetonitrile, 0.1% formic acid). The facility packed a nanoscale reverse-phase HPLC capillary column by packing 5-µm C18 spherical silica beads (Thermo Fisher Scientific) into a fused silica capillary (100 µm inner diameter × ~20 cm length) with a flame-drawn tip. After the column was equilibrated, each sample was loaded onto the column via a Famos autosampler (LC Packings). A gradient was formed, and peptides were eluted with increasing concentrations of solvent B (97.5% acetonitrile, 0.1% formic acid). As the peptides eluted, they were subjected to electrospray ionization and then entered into an LTQ Orbitrap Elite mass spectrometer (Thermo Fisher Scientific). The peptides were detected, isolated, and fragmented to produce a tandem mass spectrum of specific fragment ions for each peptide. Peptide sequences (and hence protein identity) were determined by matching protein databases with the fragmentation pattern acquired by the software program SEQUEST (ver. 28) (Thermo Fisher Scientific). Enzyme specificity was set to partially tryptic with 2 missed cleavages. Modifications included carboxyamidomethyl (cysteines, fixed) and oxidation (methionine, variable). Mass tolerance was set to 5 ppm for precursor ions and 0.5 Da for fragment ions. The mass spectrometry peak list was generated by ReADW.exe (version 4.3.1) using human as species. Peptide sequences were searched using Uniprot database (https://www.uniprot.org/proteomes/?query=taxonomy:9606) downloaded on June 20, 2017. The number of entries within this database searched are 160, 020 (half forward and half reverse and 168 contaminants added). Spectral matches were filtered to contain a false discovery rate of less than 1% at the peptide level using the target-decoy method (Elias & Gygi, 2007), and the protein inference was considered followed the general rules (Nesvizhskii &

Aebersold, 2005), with manual annotation based on experiences applied when necessary. This same principle was used for isoforms when they were present in the database. The longest isoform was reported as the match.

Bioinformatic analysis

As for the TAP-MS data analysis, a group of unrelated TAP-MS experiments (i.e. 72 experiments using stably expressed SFB-tagged proteins as baits) were included as a control group. As for the PA beads-pulldown MS data, two control beads-pulldown MS data were included as a control group. Using the "sensitivity$^{1/2}$-(1-specificity)$^{1/2}$" measurement as determined by the CRAPome databases (Mellacheruvu *et al.*, 2013), we considered any interaction with a SAINT score of at least 0.8 and raw spectra count of at least 2 to be a high-confident interacting protein (HCIP) (Wang *et al.*, 2014a). The Gene Ontology (GO) *p* values were estimated using metascape (www.metascape.org) (Zhou *et al.*, 2019), which contains findings and annotations from multiple sources including the Gene Ontology database, KEGG pathway database, and Panther pathway database. Only statistically significant correlations ($p < 0.05$) are shown. The $-\log_{10}$ (*p* value) for each biological process with related HCIPs is listed.

Experimental Design and Statistical Rationale

To apply SAINT analysis through CRAPome (Contaminant Repository for Affinity Purification, https://reprint-apms.org/), we first gathered information about baits and preys including the spectra counts and assignment of control baits. We reorganized the data to the format compatible to the SAINT program and used two-pool analysis, which recognizes the control group as a separate pool. We did not remove outlier data points. The statistics used to assess accuracy and

significance of measurements was referred to the SAINT algorithms, where SAINT score ≥ 0.80 was taken as the threshold required for the data quantification, as indicated by the SAINT method (Choi *et al.*, 2011). In total, 6586 interactions were identified in 14 experiments total which includes biological replicates for each PLD enzyme and PA. There was a total of 74 unrelated negative control experiments that were used to carry out filtration via CRAPome. A total of 403 interactions passed a probability score of ≥ 0.80. To evaluate the specificity of our results, we overlapped our SAINT interactors amongst the six different isoforms/PA and considered the overlapped ones to be "false positives" therefore designating 303 interactions as HCIPs. The PLDs and PA interactomes were generated using information from Ingenuity pathway software (Ingenuity Systems, www.ingenuity.com) for function/localization and Metascape (www.metascape.org) for gene ontology (GO). We used -log(*p* value) of individual functions to generate heatmap where a *p* value <0.05 was considered statistically significant. The heatmap for the GO: biological processes and hierarchical clustering was generated by Rstudio.

Immunofluorescent staining

Immunofluorescent staining was performed as described previously (Wang *et al.*, 2008). Briefly, cells cultured on coverslips were fixed with 4% paraformaldehyde for 10 minutes at room temperature and then extracted with 0.5% Triton X-100 solution for 5 minutes. After blocking with Tris-buffered saline with Tween 20 containing 1% bovine serum albumin, the cells were incubated with the indicated primary antibodies for 1 hour at room temperature. After that, the cells were washed and incubated with fluorescein isothiocyanate, rhodamine, or Cy5-conjugated secondary antibodies for 1 hour. Cells were counterstained with 100 ng/mL 4′,6-diamidino-2-phenylindole (DAPI) for 2 minutes to visualize nuclear DNA. The cover slips were mounted

onto glass slides with an anti-fade solution and visualized under a Nikon Ti2-E inverted microscope.

Gene inactivation by CRISPR/Cas9 system

To generate the PJA2 knockout cells, five distinct single-guide RNAs (sgRNA) were designed by CHOPCHOP website (https://chopchop.rc.fas.harvard.edu), cloned into lentiGuide-Puro vector (Addgene plasmid # 52963), and transfected into HEK293A cells with lentiCas9-Blast construct (Addgene plasmid # 52962). The next day, cells were selected with puromycin (2 μg/mL) for two days and sub-cloned to form single colonies. Knockout cell clones were screened by Western blotting to verify the loss of PJA2 expression. The oligo sequence information of sgRNAs used for knockout cell generation is listed in the **Table S3.6**.

SPHK1 kinase assay

SFB-tagged SPHK1 was purified from serum-starved HEK293A cells using S protein beads; washed with NETN buffer, washing buffer (40 mM HEPES, 250 mM NaCl) and kinase assay buffer (30 mM HEPES, 50 mM potassium acetate, 5 mM MgCl2); and subjected to *in vitro* kinase assay in the presence of cold ATP (500 μM), sphingosine (1 mM), and PA (300 μM). The reaction mixture was incubated at 37 ^{0}C for 30 min and terminated with 1 M HCl. S1P was extracted with chloroform and dotted on nitrocellulose membrane. The S1P-dotted nitrocellulose membrane was blocked in 3% BSA at 4 ^{0}C overnight and followed by incubation with anti-S1P monoclonal antibody at 4 ^{0}C overnight. To examine the roles of PLD inhibitors in regulating SPHK1 activity, SFB-SPHK1 was purified from HEK293A cells treated with DMSO, 30 μM FIPI or 20 μM CAY10594 overnight using S protein beads, similarly washed as described above,

and subjected to the kinase assay in the presence of cold ATP (500 μM) and sphingosine (1 mM). S1P dot-blot assay was performed.

3.8 Author contributions

W.W. conceptualization, supervision, methodology, writing-original draft preparation, writing-reviewing and editing, funding acquisition; R.E.K. investigation, data curation, validation, formal analysis, writing-original draft preparation, writing-reviewing and editing; H.H., G.S., B.Y., Y.L., M.D., S.P. and Y.M. validation, data curation.

3.9 Acknowledgments

We thank Dr. Ross Tomaino (Taplin Mass Spectrometry Facility, Harvard Medical School) for the mass spectrometry analysis. We thank Drs. Lan Huang (University of California, Irvine) and Xu Li (Westlake University) for the technical help. This work was supported by an NIH grant (R01 GM126048) and an American Cancer Society Research Scholar grant (RSG-18-009-01-CCG) to WW. WW is a member of the Chao Family Comprehensive Cancer Center (P30CA062203) at University of California, Irvine.

3.10 Declaration of interests

The authors declare no competing financial interests.

3.11 Data and Code Availability

The proteomic data have been deposited in the ProteomeXchange Consortium database (http://proteomecentral.proteomexchange.org/) via the PRIDE partner repository (Perez-Riverol

et al., 2019) with the project identifier PXD029183. The detailed project information is as follows:

Project Name: Interactome analysis of human phospholipase D and phosphatidic acid-associated protein network

Project accession: PXD029183

Project DOI: 10.6019/PXD029183

Reviewer account details:

Username: reviewer_pxd029183@ebi.ac.uk

Password: WVaFkBux

3.12 References:

Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, Sykes N, Pagnamenta AT *et al* (2010) A genome-wide scan for common alleles affecting risk for autism. *Hum Mol Genet* 19: 4072-4082

Bader MF, Vitale N (2009) Phospholipase D in calcium-regulated exocytosis: lessons from chromaffin cells. *Biochim Biophys Acta* 1791: 936-941

Billah MM, Eckel S, Mullmann TJ, Egan RW, Siegel MI (1989) Phosphatidylcholine hydrolysis by phospholipase D determines phosphatidate and diglyceride levels in chemotactic peptide-stimulated human neutrophils. Involvement of phosphatidate phosphohydrolase in signal transduction. *J Biol Chem* 264: 17069-17077

Bowling FZ, Frohman MA, Airola MV (2021) Structure and regulation of human phospholipase D. *Adv Biol Regul* 79: 100783

Bruntz RC, Lindsley CW, Brown HA (2014) Phospholipase D signaling pathways and phosphatidic acid as therapeutic targets in cancer. *Pharmacol Rev* 66: 1033-1079

Cai D, Netzer WJ, Zhong M, Lin Y, Du G, Frohman M, Foster DA, Sisodia SS, Xu H, Gorelick FS *et al* (2006) Presenilin-1 uses phospholipase D1 as a negative regulator of beta-amyloid formation. *Proc Natl Acad Sci U S A* 103: 1941-1946

Chan JP, Hu Z, Sieburth D (2012) Recruitment of sphingosine kinase to presynaptic terminals by a conserved muscarinic signaling pathway promotes neurotransmitter release. *Genes Dev* 26: 1070-1085

Choi H, Larsen B, Lin ZY, Breitkreutz A, Mellacheruvu D, Fermin D, Qin ZS, Tyers M, Gingras AC, Nesvizhskii AI (2011) SAINT: probabilistic scoring of affinity purification-mass spectrometry data. *Nat Methods* 8: 70-73

Cockcroft S, Thomas GM, Fensome A, Geny B, Cunningham E, Gout I, Hiles I, Totty NF, Truong O, Hsuan JJ (1994) Phospholipase D: a downstream effector of ARF in granulocytes. *Science* 263: 523-526

Colley WC, Sung TC, Roll R, Jenco J, Hammond SM, Altshuller Y, Bar-Sagi D, Morris AJ, Frohman MA (1997) Phospholipase D2, a distinct phospholipase D isoform with novel regulatory properties that provokes cytoskeletal reorganization. *Curr Biol* 7: 191-201

Dall'Armi C, Hurtado-Lorenzo A, Tian H, Morel E, Nezu A, Chan RB, Yu WH, Robinson KS, Yeku O, Small SA *et al* (2010) The phospholipase D1 pathway modulates macroautophagy. *Nat Commun* 1: 142

Delon C, Manifava M, Wood E, Thompson D, Krugmann S, Pyne S, Ktistakis NT (2004) Sphingosine kinase 1 is an intracellular effector of phosphatidic acid. *J Biol Chem* 279: 44763-44774

Dennis EA (2015) Introduction to Thematic Review Series: Phospholipases: Central Role in Lipid Signaling and Disease. *J Lipid Res* 56: 1245-1247

Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* 4: 207-214

Fang Y, Park IH, Wu AL, Du G, Huang P, Frohman MA, Walker SJ, Brown HA, Chen J (2003) PLD1 regulates mTOR signaling and mediates Cdc42 activation of S6K1. *Curr Biol* 13: 2037-2044

Fang Y, Vilella-Bach M, Bachmann R, Flanigan A, Chen J (2001) Phosphatidic acid-mediated mitogenic activation of mTOR signaling. *Science* 294: 1942-1945

Faust TB, Li Y, Jang GM, Johnson JR, Yang S, Weiss A, Krogan NJ, Frankel AD (2017) PJA2 ubiquitinates the HIV-1 Tat protein with atypical chain linkages to activate viral transcription. *Sci Rep* 7: 45394

Freyberg Z, Bourgoin S, Shields D (2002) Phospholipase D2 is localized to the rims of the Golgi apparatus in mammalian cells. *Mol Biol Cell* 13: 3930-3942

Freyberg Z, Sweeney D, Siddhanta A, Bourgoin S, Frohman M, Shields D (2001) Intracellular localization of phospholipase D1 in mammalian cells. *Mol Biol Cell* 12: 943-955

Gavin AL, Huang D, Huber C, Martensson A, Tardif V, Skog PD, Blane TR, Thinnes TC, Osborn K, Chong HS *et al* (2018) PLD3 and PLD4 are single-stranded acid exonucleases that regulate endosomal nucleic-acid sensing. *Nat Immunol* 19: 942-953

Gozgit JM, Pentecost BT, Marconi SA, Ricketts-Loriaux RS, Otis CN, Arcaro KF (2007) PLD1 is overexpressed in an ER-negative MCF-7 cell line variant and a subset of phospho-Akt-negative breast carcinomas. *Br J Cancer* 97: 809-817

Hammond SM, Altshuller YM, Sung TC, Rudge SA, Rose K, Engebrecht J, Morris AJ, Frohman MA (1995) Human ADP-ribosylation factor-activated phosphatidylcholine-specific

phospholipase D defines a new and highly conserved gene family. *J Biol Chem* 270: 29640-29643

Hammond SM, Jenco JM, Nakashima S, Cadwallader K, Gu Q, Cook S, Nozawa Y, Prestwich GD, Frohman MA, Morris AJ (1997) Characterization of two alternately spliced forms of phospholipase D1. Activation of the purified enzymes by phosphatidylinositol 4,5-bisphosphate, ADP-ribosylation factor, and Rho family monomeric GTP-binding proteins and protein kinase C-alpha. *J Biol Chem* 272: 3860-3868

Han H, Qi R, Zhou JJ, Ta AP, Yang B, Nakaoka HJ, Seo G, Guan KL, Luo R, Wang W (2018) Regulation of the Hippo Pathway by Phosphatidic Acid-Mediated Lipid-Protein Interaction. *Mol Cell* 72: 328-340 e328

Heffernan-Stroud LA, Obeid LM (2013) Sphingosine kinase 1 in cancer. *Adv Cancer Res* 117: 201-235

Henkels KM, Boivin GP, Dudley ES, Berberich SJ, Gomez-Cambronero J (2013) Phospholipase D (PLD) drives cell invasion, tumor growth and metastasis in a human breast cancer xenograph model. *Oncogene* 32: 5551-5562

Huang H, Gao Q, Peng X, Choi SY, Sarma K, Ren H, Morris AJ, Frohman MA (2011) piRNA-associated germline nuage formation and spermatogenesis require MitoPLD profusogenic mitochondrial-surface lipid signaling. *Dev Cell* 20: 376-387

Jenkins GM, Frohman MA (2005) Phospholipase D: a lipid centric review. *Cell Mol Life Sci* 62: 2305-2316

Kang DW, Choi KY, Min DS (2014) Functional regulation of phospholipase D expression in cancer and inflammation. *J Biol Chem* 289: 22575-22582

Kim S, Sieburth D (2018) Sphingosine Kinase Activates the Mitochondrial Unfolded Protein Response and Is Targeted to Mitochondria by Stress. *Cell Rep* 24: 2932-2945 e2934

Laulagnier K, Grand D, Dujardin A, Hamdi S, Vincent-Schneider H, Lankar D, Salles JP, Bonnerot C, Perret B, Record M (2004) PLD2 is enriched on exosomes and its activity is correlated to the release of exosomes. *FEBS Lett* 572: 11-14

Li X, Tran KM, Aziz KE, Sorokin AV, Chen J, Wang W (2016) Defining the Protein-Protein Interaction Network of the Human Protein Tyrosine Phosphatase Family. *Mol Cell Proteomics* 15: 3030-3044

Lignitto L, Arcella A, Sepe M, Rinaldi L, Delle Donne R, Gallo A, Stefan E, Bachmann VA, Oliva MA, Tiziana Storlazzi C *et al* (2013) Proteolysis of MOB1 by the ubiquitin ligase praja2 attenuates Hippo signalling and supports glioblastoma growth. *Nat Commun* 4: 1822

Lignitto L, Carlucci A, Sepe M, Stefan E, Cuomo O, Nistico R, Scorziello A, Savoia C, Garbi C, Annunziato L *et al* (2011) Control of PKA stability and signalling by the RING ligase praja2. *Nat Cell Biol* 13: 412-422

Liu J, Li J, Ma Y, Xu C, Wang Y, He Y (2021) MicroRNA miR-145-5p inhibits Phospholipase D 5 (PLD5) to downregulate cell proliferation and metastasis to mitigate prostate cancer. *Bioengineered* 12: 3240-3251

Liu M, Du K, Fu Z, Zhang S, Wu X (2015) Hypoxia-inducible factor 1-alpha up-regulates the expression of phospholipase D2 in colon cancer cells under hypoxic conditions. *Med Oncol* 32: 394

Lucocq J, Manifava M, Bi K, Roth MG, Ktistakis NT (2001) Immunolocalisation of phospholipase D1 on tubular vesicular membranes of endocytic and secretory origin. *Eur J Cell Biol* 80: 508-520

Mellacheruvu D, Wright Z, Couzens AL, Lambert JP, St-Denis NA, Li T, Miteva YV, Hauri S, Sardiu ME, Low TY *et al* (2013) The CRAPome: a contaminant repository for affinity purification-mass spectrometry data. *Nat Methods* 10: 730-736

Min DS, Kwon TK, Park WS, Chang JS, Park SK, Ahn BH, Ryoo ZY, Lee YH, Lee YS, Rhie DJ *et al* (2001) Neoplastic transformation and tumorigenesis associated with overexpression of phospholipase D isozymes in cultured murine fibroblasts. *Carcinogenesis* 22: 1641-1647

Munck A, Bohm C, Seibel NM, Hashemol Hosseini Z, Hampe W (2005) Hu-K4 is a ubiquitously expressed type 2 transmembrane protein associated with the endoplasmic reticulum. *FEBS J* 272: 1718-1726

Nesvizhskii AI, Aebersold R (2005) Interpretation of shotgun proteomic data: the protein inference problem. *Mol Cell Proteomics* 4: 1419-1440

Oguin TH, 3rd, Sharma S, Stuart AD, Duan S, Scott SA, Jones CK, Daniels JS, Lindsley CW, Thomas PG, Brown HA (2014) Phospholipase D facilitates efficient entry of influenza virus, allowing escape from innate immune inhibition. *J Biol Chem* 289: 25405-25417

Olivera A, Spiegel S (1993) Sphingosine-1-phosphate as second messenger in cell proliferation induced by PDGF and FCS mitogens. *Nature* 365: 557-560

Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M *et al* (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47: D442-D450

Pitson SM, Moretti PA, Zebol JR, Zareie R, Derian CK, Darrow AL, Qi J, D'Andrea RJ, Bagley CJ, Vadas MA *et al* (2002) The nucleotide-binding site of human sphingosine kinase 1. *J Biol Chem* 277: 49545-49553

Pohlman RF, Liu F, Wang L, More MI, Winans SC (1993) Genetic and biochemical analysis of an endonuclease encoded by the IncN plasmid pKM101. *Nucleic Acids Res* 21: 4867-4872

Rinaldi L, Delle Donne R, Sepe M, Porpora M, Garbi C, Chiuso F, Gallo A, Parisi S, Russo L, Bachmann V *et al* (2016) praja2 regulates KSR1 stability and mitogenic signaling. *Cell Death Dis* 7: e2230

Rizzo MA, Shome K, Watkins SC, Romero G (2000) The recruitment of Raf-1 to membranes is mediated by direct interaction with phosphatidic acid and is independent of association with Ras. *J Biol Chem* 275: 23911-23918

Saito M, Iwadate M, Higashimoto M, Ono K, Takebayashi Y, Takenoshita S (2007) Expression of phospholipase D2 in human colorectal carcinoma. *Oncol Rep* 18: 1329-1334

Sakamaki J, Fu A, Reeks C, Baird S, Depatie C, Al Azzabi M, Bardeesy N, Gingras AC, Yee SP, Screaton RA (2014) Role of the SIK2-p35-PJA2 complex in pancreatic beta-cell functional compensation. *Nat Cell Biol* 16: 234-244

Shen H, Giordano F, Wu Y, Chan J, Zhu C, Milosevic I, Wu X, Yao K, Chen B, Baumgart T *et al* (2014) Coupling between endocytosis and sphingosine kinase 1 recruitment. *Nat Cell Biol* 16: 652-662

Shevchenko A, Wilm M, Vorm O, Mann M (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal Chem* 68: 850-858

Shida D, Takabe K, Kapitonov D, Milstien S, Spiegel S (2008) Targeting SphK1 as a new strategy against cancer. *Curr Drug Targets* 9: 662-673

Somatilaka BN, Hwang SH, Palicharla VR, White KA, Badgandi H, Shelton JM, Mukhopadhyay S (2020) Ankmy2 Prevents Smoothened-Independent Hyperactivation of the Hedgehog Pathway via Cilia-Regulated Adenylyl Cyclase Signaling. *Dev Cell* 54: 710-726 e718

Song Y, Lee S, Kim JR, Jho EH (2018) Pja2 Inhibits Wnt/beta-catenin Signaling by Reducing the Level of TCF/LEF1. *Int J Stem Cells* 11: 242-247

Stace CL, Ktistakis NT (2006) Phosphatidic acid- and phosphatidylserine-binding proteins. *Biochim Biophys Acta* 1761: 913-926

Stahelin RV, Ananthanarayanan B, Blatner NR, Singh S, Bruzik KS, Murray D, Cho W (2004) Mechanism of membrane binding of the phospholipase D1 PX domain. *J Biol Chem* 279: 54918-54926

Sung TC, Roper RL, Zhang Y, Rudge SA, Temel R, Hammond SM, Morris AJ, Moss B, Engebrecht J, Frohman MA (1997) Mutagenesis of phospholipase D defines a superfamily including a trans-Golgi viral protein required for poxvirus pathogenicity. *EMBO J* 16: 4519-4530

Veverka V, Crabbe T, Bird I, Lennie G, Muskett FW, Taylor RJ, Carr MD (2008) Structural characterization of the interaction of mTOR with phosphatidic acid and a novel class of inhibitor: compelling evidence for a central role of the FRB domain in small molecule-mediated regulation of mTOR. *Oncogene* 27: 585-595

von Eyss B, Jaenicke LA, Kortlever RM, Royla N, Wiese KE, Letschert S, McDuffus LA, Sauer M, Rosenwald A, Evan GI *et al* (2015) A MYC-Driven Change in Mitochondrial Dynamics Limits YAP/TAZ Function in Mammary Epithelial Cells and Breast Cancer. *Cancer Cell* 28: 743-757

Wang W, Chen L, Ding Y, Jin J, Liao K (2008) Centrosome separation driven by actin-microfilaments during mitosis is mediated by centrosome-associated tyrosine-phosphorylated cortactin. *J Cell Sci* 121: 1334-1343

Wang W, Li X, Huang J, Feng L, Dolinta KG, Chen J (2014) Defining the protein-protein interaction network of the human hippo pathway. *Mol Cell Proteomics* 13: 119-131

Wang Z, Zhang F, He J, Wu P, Tay LWR, Cai M, Nian W, Weng Y, Qin L, Chang JT *et al* (2017) Binding of PLD2-Generated Phosphatidic Acid to KIF5B Promotes MT1-MMP Surface Trafficking and Lung Metastasis of Mouse Breast Cancer Cells. *Dev Cell* 43: 186-197 e187

Yoshikawa F, Banno Y, Otani Y, Yamaguchi Y, Nagakura-Takagi Y, Morita N, Sato Y, Saruta C, Nishibe H, Sadakata T *et al* (2010) Phospholipase D family member 4, a transmembrane

glycoprotein with no phospholipase D activity, expression in spleen and early postnatal microglia. *PLoS One* 5: e13932

Zhang Y, Liu X, Bai J, Tian X, Zhao X, Liu W, Duan X, Shang W, Fan HY, Tong C (2016) Mitoguardin Regulates Mitochondrial Fusion through MitoPLD and Is Required for Neuronal Homeostasis. *Mol Cell* 61: 111-124

Zhao Z, Zhu L, Xing Y, Zhang Z (2021) Praja2 suppresses the growth of gastric cancer by ubiquitylation of KSR1 and inhibiting MEK-ERK signal pathways. *Aging (Albany NY)* 13: 3886-3897

Zhong J, Wang H, Chen W, Sun Z, Chen J, Xu Y, Weng M, Shi Q, Ma D, Miao C (2017) Ubiquitylation of MFHAS1 by the ubiquitin ligase praja2 promotes M1 macrophage polarization by activating JNK and p38 pathways. *Cell Death Dis* 8: e2763

Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK (2019) Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. *Nat Commun* 10: 1523

3.13 Figures

3.1

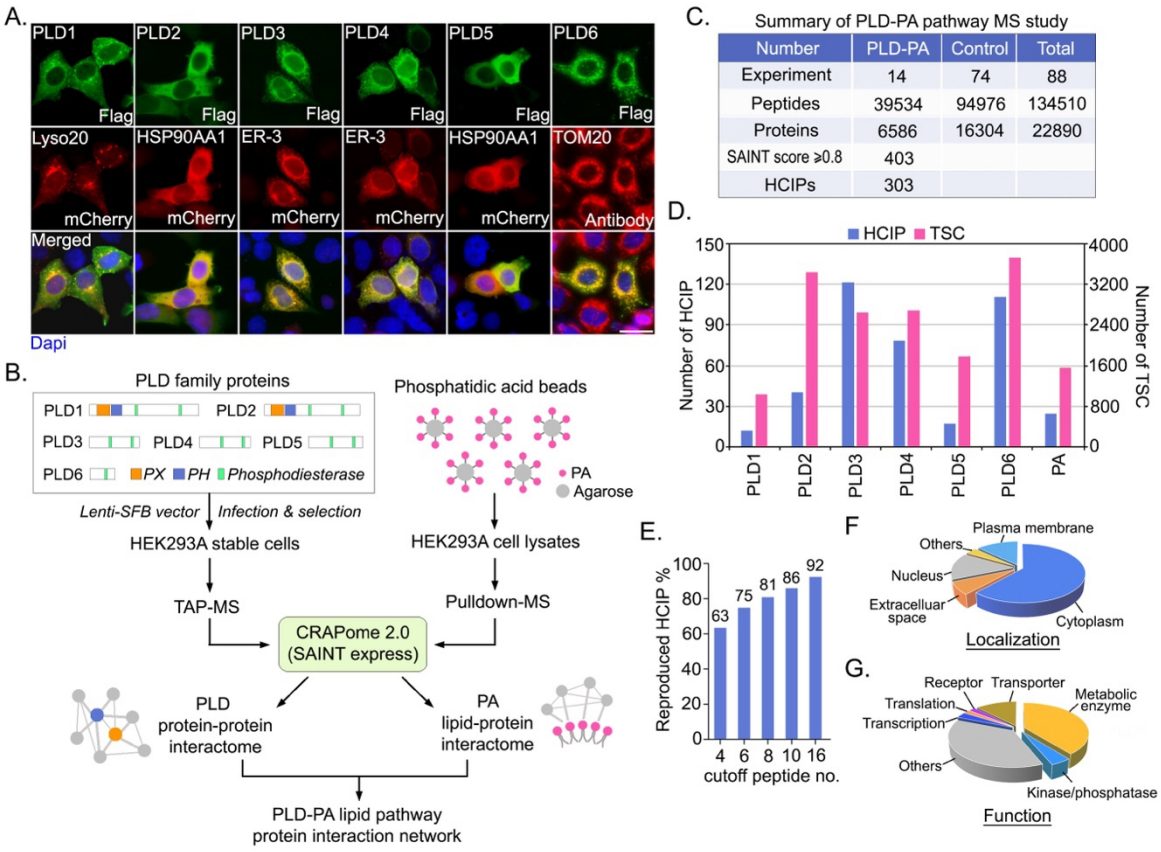


Figure 3.1. Proteomic analysis of the phospholipase D (PLD) and phosphatidic acid (PA)-associated protein network.

(**A**) PLD family enzymes show different subcellular localizations. Immunofluorescent staining was performed using the indicated antibodies. Scale bar, 30 μm.

(**B**) Schematic illustration of the major steps used in the proteomic study of human PLDs and PA-associated protein complexes. PLDs were constructed into a *C*-terminal SFB–tag fused lentiviral vector. HEK293A cells stably expressing each PLD bait protein were generated via lentiviral infection and puromycin selection, and subjected to the TAP-MS analysis. For isolation of the PA-associated protein complex, HEK293A cells were subjected to pulldown assays using

the PA-conjugated agarose beads. All the MS results were processed via the CRAPome 2.0 statistical model (i.e. SAINT express).

(**C**) Summary of the PLD-PA lipid pathway MS study. Experimental information and total numbers of peptides and proteins identified in the MS analysis were shown. A SAINT score ≥ 0.8 was used as the cutoff to identify HCIPs.

(**D**) The total spectral counts (TSCs) and corresponding numbers of HCIPs for PLDs and PA. Data reproducibility between two biological MS experiments for PLD members and PA are evaluated using the number of peptide spectrum matches (PSMs).

(F-G) Gene Ontology (GO) annotations of the identified HCIPs of PLDs and PA. Cellular localization (**F**) and cellular functions (**G**) for the HCIPs of PLDs and PA were shown.

Figure 3.2

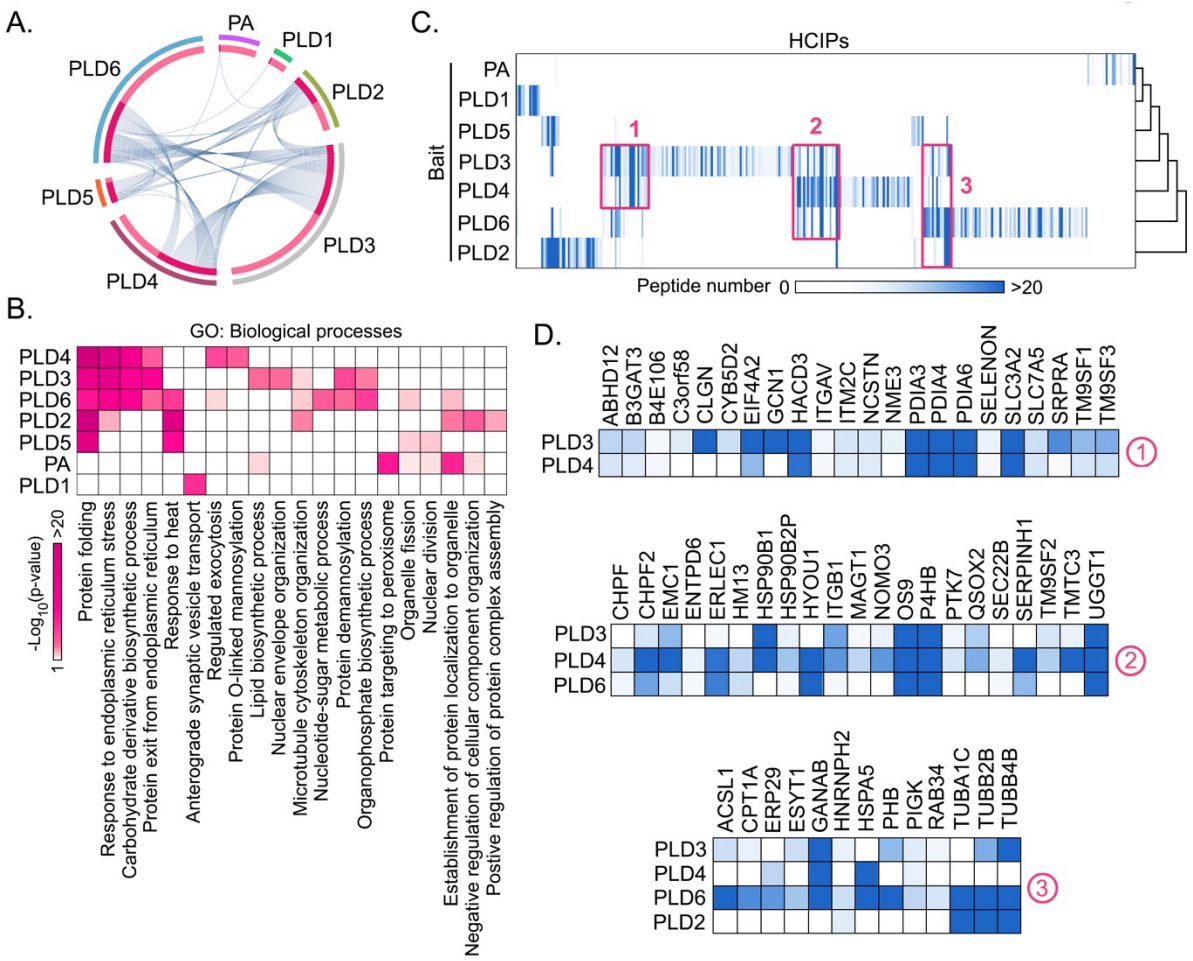


Figure 3.2. Hierarchical clustering analysis of the HCIPs generated for the PLD-PA lipid pathway.

(**A**) The HCIPs shared between PLDs and PA were analyzed by www.metascape.org.

(**B**) The HCIPs of PLDs and PA were subjected to GO analysis. Through it, the biological processes of the PLDs/PA-HCIPs were shown as a heatmap.

(**C-D**) A heatmap was generated from hierarchical clustering of 303 HCIPs for PLDs and PA. Three prominent HCIP clusters were manually selected (**C**) and enlarged below (**D**). The color of squares in the heatmap indicates the identified HCIP peptide number for each bait.

Figure 3.3

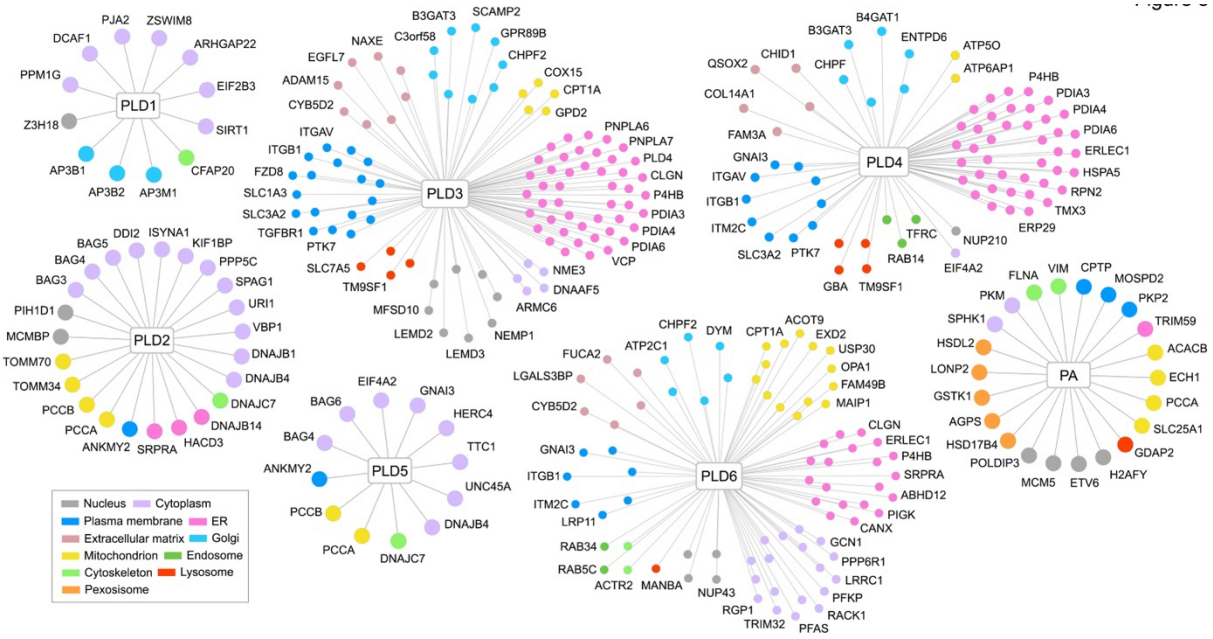


Figure 3.3. Interaction maps of the human PLDs and PA-centered protein interaction network.

As for each PLD family member and PA, HCIPs are grouped based on their cellular localization according to the GO analysis and literature search, and visualized in different colors.

Figure 3.4

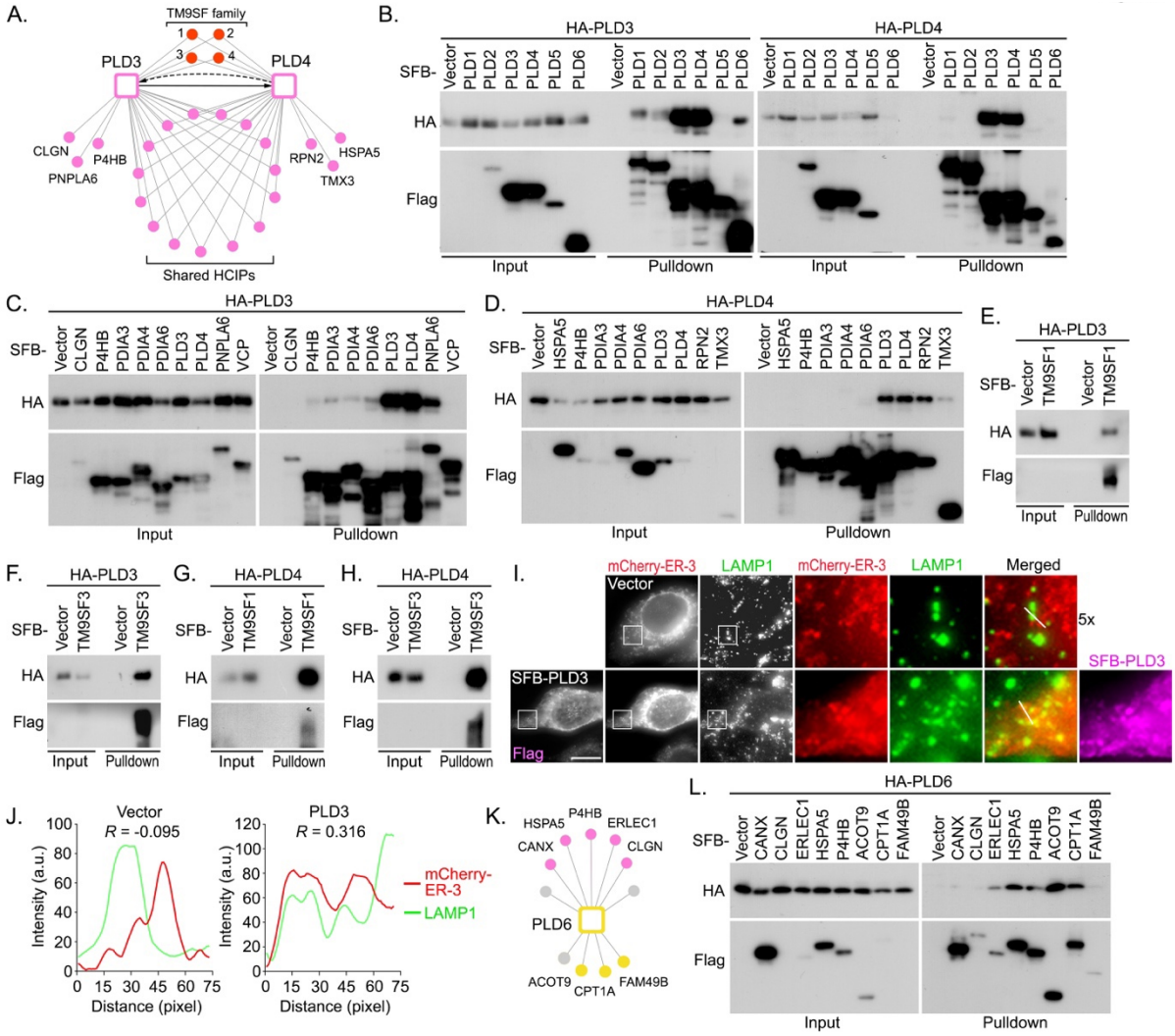


Figure 3.4. Validation of the PLD3/PLD4/PLD6-associated protein network.

(**A**) A summary of the protein-protein interaction network with the selected HCIPs of PLD3 and PLD4.

(**B**) Validation of the interaction between PLD3 and PLD4 in the PLD family. HEK293T cells were transfected with the constructs encoding the indicated proteins and subjected to pulldown assay using S protein beads.

(**C**) Validation of the interaction between PLD3 and its selected HCIPs. HEK293T cells were transfected with the constructs encoding the indicated proteins and subjected to pulldown assay using S protein beads.

(**D**) Validation of the interaction between PLD4 and its selected HCIPs. HEK293T cells were transfected with the constructs encoding the indicated proteins and subjected to pulldown assay using S protein beads.

(**E-H**) PLD3/4 interact with TM9SF1/3. HEK293T cells were transfected with the constructs encoding SFB-tagged TM9SF1 or TM9SF3 and HA-tagged PLD3 (**E, F**) or PLD4 (**G, H**), and subjected to pulldown assay using S protein beads.

(**I-J**) PLD3 drives ER-lysosome contact formation. Immunofluorescent staining was performed using the indicated antibodies (**I**). The indicated area were 5 times enlarged (5x). Scale bar, 15 μm. The fluorescent intensity correlation (R) between ER-3 (red) and LAMP1 (green) were analyzed (**J**) along the lines as indicated in (**I**).

(**K**) A summary of the protein-protein interaction network for the selected HCIPs of PLD6.

(**L**) Validation of the interaction between PLD6 and its selected HCIPs. HEK293T cells were transfected with the constructs encoding the indicated proteins and subjected to pulldown assay using S protein beads.

Figure 3.5

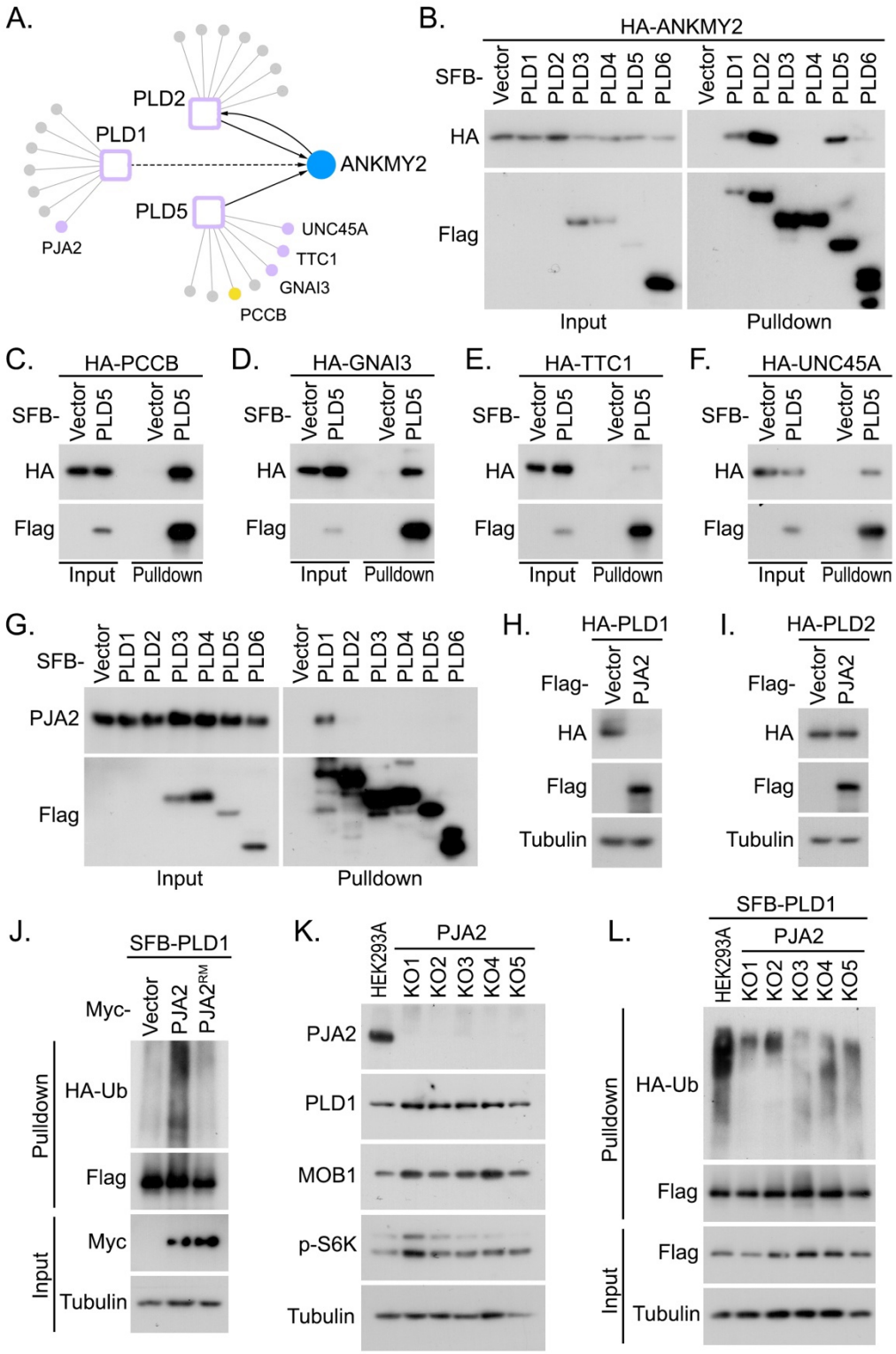

A.
PLD2
PLD1
ANKMY2
PLD5
PJA2
UNC45A
TTC1
GNAI3
PCCB
B.
HA-ANKMY2
SFB-
Vector
PLD1
PLD2
PLD3
PLD4
PLD5
PLD6
HA
Flag
Input
Pulldown
C.
HA-PCCB
D.
HA-GNAI3
E.
HA-TTC1
F.
HA-UNC45A
G.
PJA2
H.
HA-PLD1
Flag-
Tubulin
I.
HA-PLD2
J.
SFB-PLD1
Myc-
PJA2RM
Pulldown
HA-Ub
Input
Myc
K.
HEK293A
PJA2
KO1
KO2
KO3
KO4
KO5
PLD1
MOB1
p-S6K
L.

Figure 3.5. Characterization of PLD1/PLD2/PLD5-associated protein network reveals PJA2 as an E3 ubiquitin ligase for PLD1.

(**A**) A summary of the protein-protein interaction network with the selected HCIPs of PLD1/2/5. ANKMY2 was indicated as the shared HCIP for these three PLD proteins.

(**B**) ANKMY2 interacts with PLD1/2/5. HEK293T cells were transfected with the constructs encoding the indicated proteins and subjected to pulldown assay using S protein beads.

(**C-F**) Validation of the interaction between PLD5 and its selected HCIPs. Constructs encoding SFB-tagged PLD5 and HA-tagged PCCB (**C**), GNAI3 (**D**), TTC1 (**E**) and UNC45A (**F**) were transfected into HEK293T cells. Pulldown experiments were performed using S protein beads.

(**G**) PJA2 specifically interacts with PLD1 among the PLD members. HEK293T cells were transfected with the constructs encoding the indicated proteins and subjected to pulldown assay using S protein beads.

(**H-I**) PJA2 targets PLD1 protein expression but does not affect that of PLD2. Constructs encoding Flag-tagged PJA2 and HA-tagged PLD1 **(H)** or PLD2 **(I)** were co-transfected into HEK293T cells. Western blots were performed using the indicated antibodies.

(**J**) Overexpression of PJA2 but not its ring domain mutant (RM) induces the ubiquitination of PLD1. HEK293T cells were transfected with the constructs encoding the indicated proteins, treated with MG132 (20 μM) for 8 h, and subjected to pulldown assay using S protein beads. Western blots were performed using the indicated antibodies.

(**K**) PLD1 protein expression is increased in the PJA2 knockout (KO) HEK293A cells. Western blots were performed using the indicated antibodies.

(**L**) The ubiquitination of PLD1 is decreased in the PJA2 knockout (KO) HEK293A cells. The indicated HEK293A cells were transfected with the constructs encoding the indicated proteins,

treated with MG132 (20 μM) for 8 h, and subjected to pulldown assay using S protein beads. Western blots were performed using the indicated antibodies.

Figure 3.6

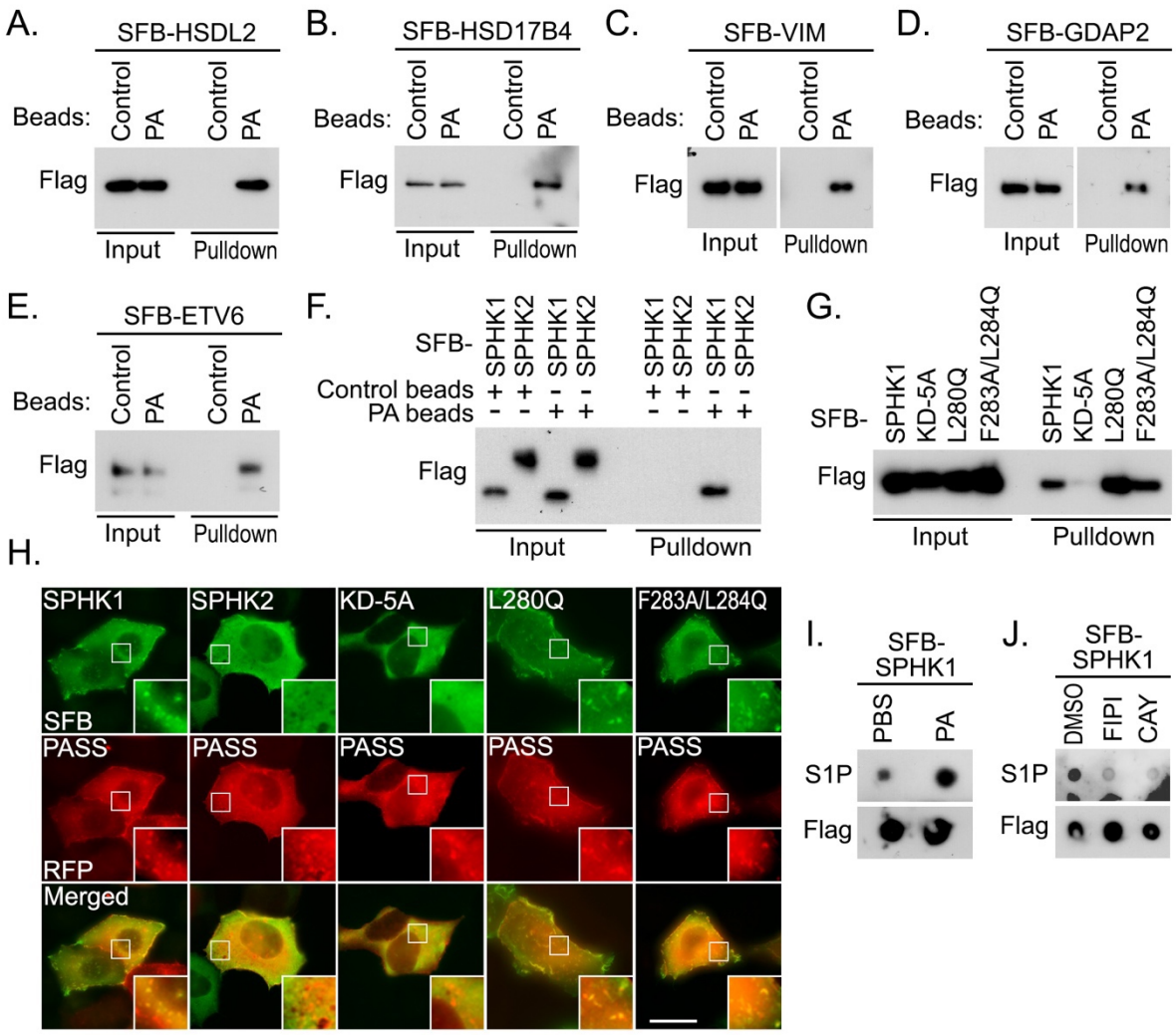


Figure 3.6. Analysis of the PA-mediated lipid-protein interaction network uncovers PA as a positive regulator of SPHK1.

(**A-E)** Validation of the interaction between PA and its selected HCIPs. Constructs encoding SFB-tagged HSDL2 (**A**), HSD17B4 (**B**), VIM (**C**), GDAP2 (**D**) and ETV6 (**E**) were transfected into HEK293T cells. Pulldown experiments were performed using PA beads.

(**F**) PA binds SPHK1 but not SPHK2. HEK293T cells were transfected with the constructs encoding the indicated proteins and subjected to pulldown assay using PA beads.

(**G**) PA dose not bind SPHK1 kinase dead mutant. HEK293T cells were transfected with the constructs encoding the indicated proteins and subjected to pulldown assay using PA beads. KD-5A, SPHK1 kinase dead mutant (S165A/G166A/D167A/G168A/L169A).

(**H**) PA is co-localized with SPHK1 and its two hydrophobic patch mutants but not with its kinase dead mutant or SPHK2. HEK293A cells were transfected with the constructs encoding the indicated proteins and subjected to immunofluorescent staining. PASS, PA sensor protein. Scale bar, 30 μm.

(**I**) Supplementing PA induces SPHK1 kinase activity *in vitro*. SFB-tagged SPHK1 was expressed in HEK293A cells, purified using S protein beads, and subjected to *in vitro* kinase assay in the presence of PA (300 μM), where sphingosine was used as a substrate. The extracted S1P was dotted on nitrocellulose paper and immunoblotted with the indicated antibodies.

(**J**) Inhibition of PA production suppresses SPHK1 kinase activity. SFB-tagged SPHK1 was expressed in HEK293A cells, treated with DMSO, FIPI (30 μM) or CAY10594 (20 μM) overnight, purified using S protein beads, and subjected to *in vitro* kinase assay, where sphingosine was used as a substrate. The extracted S1P was dotted on nitrocellulose paper and immunoblotted with the indicated antibodies.

3.14 Supplemental Figure

Figure S3.1

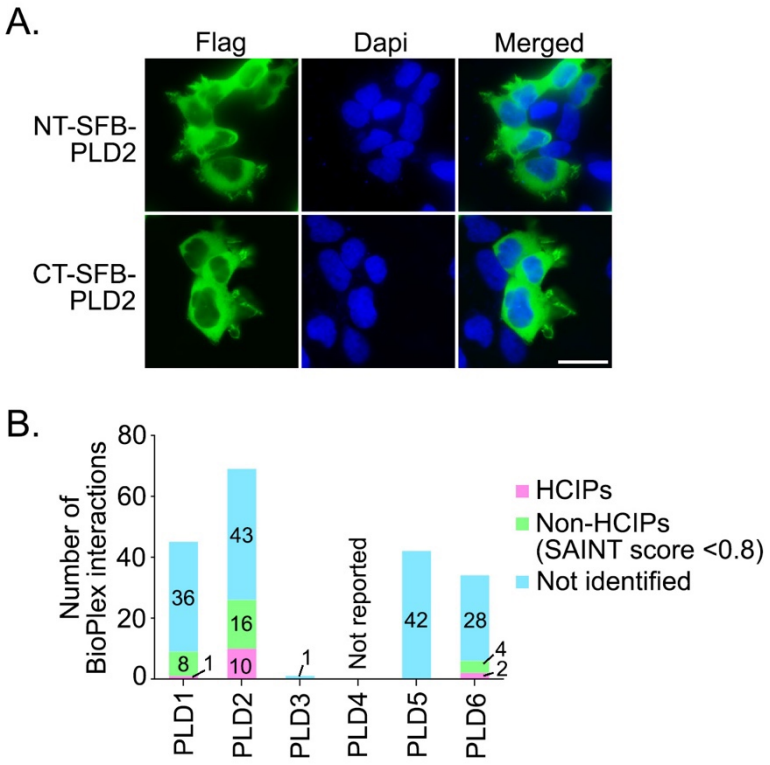


Figure S3.1. Characterization of PLDs-based protein-protein interaction network.

(**A**) SFB tag location does not affect PLD2 cellular localization. *N*-terminal SFB-tagged PLD2 and *C*-terminal SFB-tagged PLD2 constructs were expressed in HEK293A cells. Immunofluorescent staining was performed using Flag antibody. Scale bar, 30 μm.

(**B**) Comparison of the numbers of PLDs-interacting proteins identified in the BioPlex interactome database with the HCIPs of our data set.

Figure S3.2

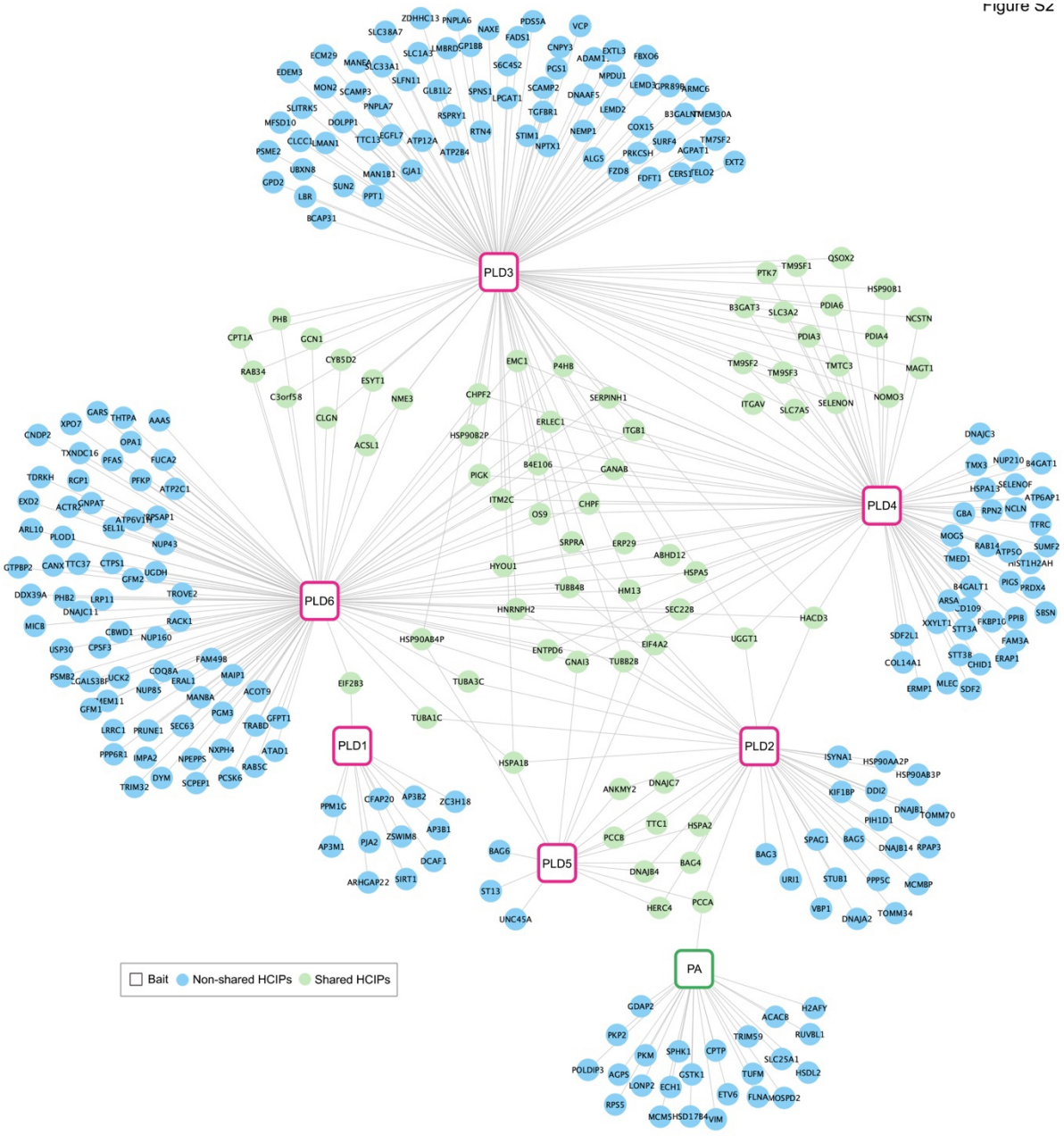


Figure S3.2. Interactome map of the PLDs and PA-associated proteins.

The cytoscape-generated merged interaction network for six PLD proteins and their lipid product PA showed a group of shared HCIPs (green) between PLD proteins and PA. Unique HCIPs for each PLD protein and PA are labelled in blue

3.15 Electronic Supplemental Tables

Table S3.1. Complete peptide identification list for human PLDs and PA-based MS experiments.

Table S3.2. Complete prey list for human PLDs and PA-based MS experiments.

Table S3.3. Complete HCIP list for human PLDs and PA-based MS experiments.

Table S3.4. Comparison of the PLDs HCIPs in BioPlex.

Table S3.5. GO analyses of human PLDs and PA-associated HCIPs.

Table S3.6: Sequence information of PJA2 sgRNAs.

CHAPTER 4

Summary and Conclusions

4.1 Summary of results

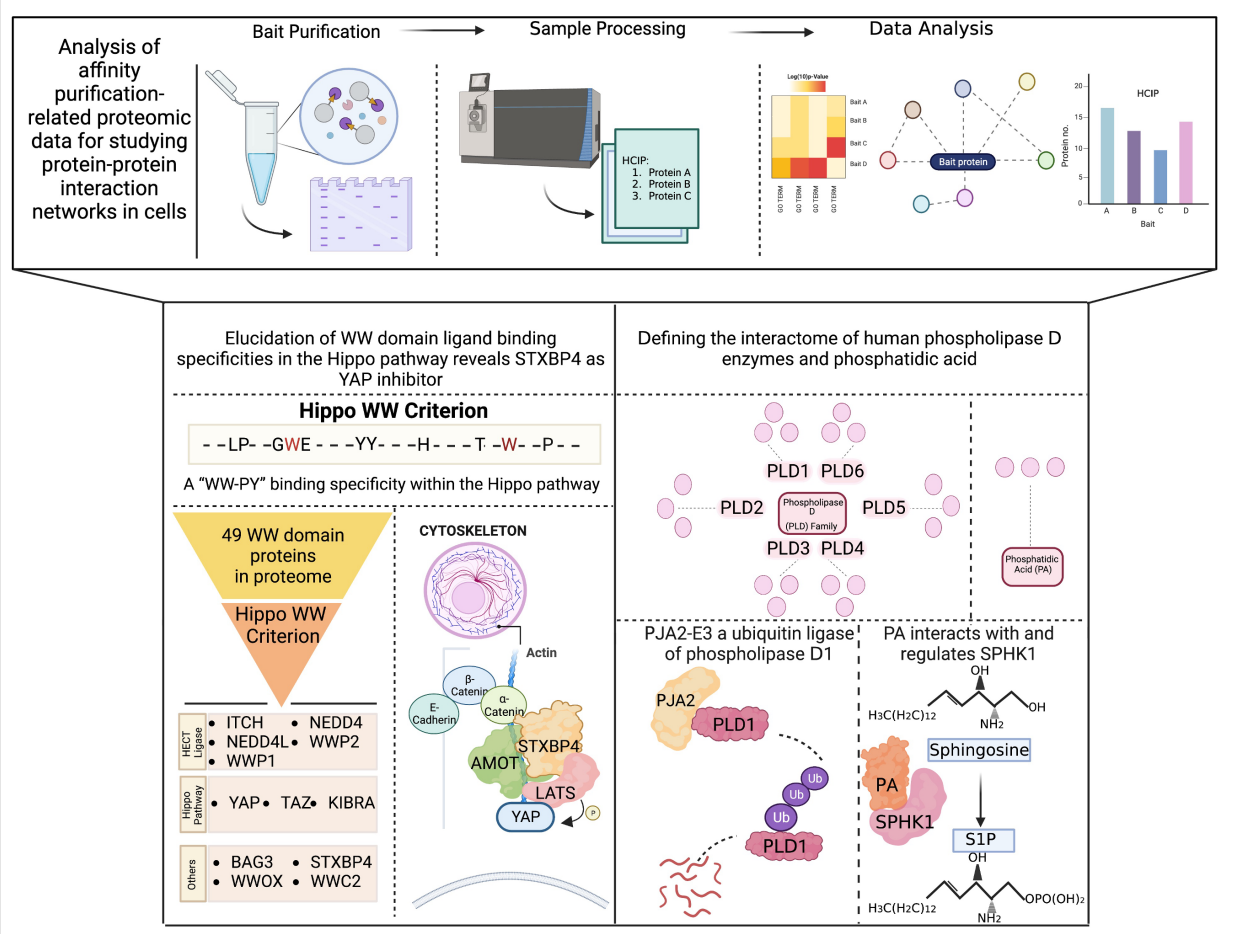


Figure 4.1: Overview of projects generated using the user-friendly proteomic pipeline.

The aim of this work was to generate a user-friendly pipeline for mass spectrometry data analysis to aid in the identification of biologically relevant interactions **(Figure 4.1).** While the proteomics field and mass spectrometry instruments has evolved over the decades, so does the amount of data produced creating a problem for users in the identification of *bona fide* interactions. We applied this generated pipeline for the identification of novel Hippo pathway regulators to further elucidate our understanding of this pathway's modulations. We

demonstrated using this generated pipeline, the identification of a unique 9-amino acid residue important for modulating interactions between WW-PxY motif containing proteins within the Hippo pathway. Additionally, a novel YAP inhibitor was discovered using this conserved 9 amino acid sequence, Syntaxin binding protein 4 (STXBP4). The application of this pipeline was also used to potentially identify novel upstream regulators governing the PLD-PA-Hippo axis. Though our further studies did not identify any Hippo pathway regulators amongst the PLD family, we did discover novel interactors that are Hippo independent and biologically relevant. Amongst the potential interactors of six PLD members, an E3 ubiquitin-protein ligase, PJA2 was identified to bind specifically to PLD1 (an oncogene). Additionally, we uncovered that PA binds, co-localizes, and regulates Sphingosine Kinase1 (SPHK1), which has been associated with regulating cell growth, proliferation, and survival.

In Chapter 1 (published in ((Kattan *et al.*, 2023)), we describe the MS pipeline for data analysis. Mass spectrometry (MS) is a powerful tool for studying small molecules in biological research. It can be used in conjunction with other techniques to identify interacting proteins in protein-protein interaction (PPI) studies. Common methods for isolating protein complexes include immunoprecipitation, affinity purification, and protein-proximity labelling approaches. Nevertheless, the growing size and intricacy of proteomic data associated with protein-protein interactions present difficulties when attempting to manually extract meaningful biological insights from the data. To address this issue, various bioinformatic tools have been developed to facilitate the analysis of PPI-related proteomic data. These tools include MS raw data filtration, gene ontology (GO) analysis, functional topology analysis, and PPI network visualization. These methods enable regular biological research labs to process, analyze, and present large-scale MS

data. Validating and characterizing the refined PPI network experimentally is also crucial to ensure the reliability of the interactome dataset. This review article discusses the current methods used for isolating and identifying interacting proteins using MS analysis. It provides an overview of how to process MS raw data, interpret their biological significance, construct a PPI network, and characterize the functions of newly identified interacting proteins. We aimed to offer guidance and highlight bioinformatics resources for researchers interested in using MS to investigate PPIs in their biological research. While this guideline provides current resources, there is room to incorporate other -omic resources to make the data more meaningful. Using this approach, we were able to identify novel regulators that were Hippo and non-Hippo related.

In Chapter 2, we put this generated pipeline into use for the discovery of novel hippo interactions(Vargas *et al*, 2020c). This study focused on WW domains, which are protein modules involved in protein-protein interactions. By using bioinformatic resources and tools, we screened many potential ligands, and identified the specific binding preferences of WW domains within the Hippo pathway. This led to the identification of a specific amino acid sequence governing this specificity. Upon further analysis, this specific sequence criterion was demonstrated to be specific for other proteins within the WW proteome, STXBP4 was discovered to contain this sequence. Mechanistically, studies demonstrated that STXBP4 recruits multiple Hippo PY motif-containing proteins, such as AMOT and LATS, to form a complex with α-catenin at adherens junctions. This complex also includes YAP/TAZ, which interact with AMOT and LATS. The findings of this study provide valuable insights into the regulatory mechanisms of the Hippo pathway and suggest a novel role for STXBP4 as an inhibitor of YAP. This research contributes to our understanding of the molecular interactions involved in the

Hippo signaling pathway and may have implications for developing therapeutic strategies targeting YAP in cancer and other diseases.

In Chapter 3, we aimed to identify novel regulators upstream of the PLD-PA-Hippo axis in hopes of finding an important player for this axis. The study aims to uncover the protein-protein interactions and functional relationships of PLD and PA-associated proteins in their relation to the Hippo pathway. Using a combination of experimental and bioinformatics approaches, we identified and validated a comprehensive network of proteins that interact with PLD and participate in PA signaling. The interactome analysis revealed several novel interactions and functional associations. While we did not find any Hippo related regulators, we found that PLD interacts with numerous proteins involved in various cellular pathways, including membrane trafficking, cytoskeleton organization, cell signaling, and lipid metabolism. Through our proteomic analysis, we discovered that PJA2 functions as an E3 ubiquitin ligase for PLD1. These findings emphasize the involvement of PJA2 in a complex downstream signaling network. Numerous studies, including our own, have identified PA as a significant signaling molecule due to its interaction with various proteins. This led us to employ a proteomic approach to explore the proteins associated with PA.

Through this analysis, we successfully identified a cluster of proteins as PA binding proteins. Among the identified proteins, we experimentally validated the lipid-protein interaction between PA and SPHK1, which has also been previously reported using different methods. We demonstrated that PA serves as a critical regulator of SPHK1, influencing its cellular localization and kinase activity. These findings shed light on the mechanisms underlying the modulation of

this essential lipid kinase. The findings of this study provide valuable insights into the complex network of interactions and functional roles of PLD and PA-associated proteins. This comprehensive analysis contributes to our understanding of the molecular mechanisms underlying PA signaling and its involvement in cellular processes. The identified protein-protein interactions and complexes serve as a resource for further investigations into the physiological and pathological roles of PLD and PA signaling in human biology and disease.

In conclusion, our understanding of complex datasets is still at the surface level. With so many new -omic data generation evolving overtime demands that bioinformatic resources stay up to date. With new data generating techniques being generated, convoluted data is generated, and the need arises for further establishments of guidelines on how to deconvolute and create a user-friendly guide on how to extract relevant data. The studies arising from our generated pipeline is just the beginning, further tools can be added in such as additional disease -omic resource to provide additional insight into these interactions.

4.2 Implications of generated proteomics pipeline for discovery of novel regulators

The analysis of affinity purification-related proteomic data for studying protein-protein interaction networks opens several exciting avenues for future research, including method development, network dynamics, functional annotation, validation, perturbation studies, and integration with other -omics data. These implications can contribute to a deeper understanding of cellular processes, disease mechanisms, and potential therapeutic targets(Hanash, 2003; Rolland *et al*, 2014). While this generated pipeline is a general guideline on how to handle a large dataset, there is room for further method development that can give provide further

biological information. Future modifications to this guideline can be incorporated once refinement on existing scoring methods (e.g., CRAPome, MUSE) enhance the accuracy, sensitivity, and throughput of proteomic data analysis in the context of studying protein-protein interaction (PPI) networks. While these scoring resources are useful, they have some limitations such as contaminant database coverage, lack of contextual information, and parameter determination that can lead a user to explore interactions that may not necessarily be true. More so, future implications include investigating the dynamic nature of generated PPI networks over time, capturing changes in interactions under different cellular conditions or in response to different treatments(Räschle *et al*, 2015). These types of studies could shed light on the temporal aspects and regulatory mechanisms governing protein-protein interactions.

Expanding on the analysis of proteomic data, future studies can focus on functional annotation of the identified protein-protein interactions. Integrating other -omics data, such as transcriptomics or metabolomics (Srivastava & Creek, 2019), can provide a more comprehensive understanding of the biological processes and pathways associated with the protein-protein interaction network (Huang *et al*, 2017; Reel *et al*, 2021). Future implications include developing computational approaches and tools to integrate multi-omic data and unravel the complex relationships between genetic variation, protein-protein interactions, and cellular phenotypes. Although affinity purification-based methods are powerful for identifying protein-protein interactions, experimental validation is crucial to establish biological function. Additionally, further characterization of specific interactions can be performed to understand their binding affinities and functional consequences.

In conclusion, the analysis of large datasets has proven to be a valuable approach for studying protein-protein interaction networks in cells. By utilizing affinity purification techniques coupled with mass spectrometry, researchers can identify and characterize protein complexes, elucidate protein interaction dynamics, and gain insights into the functional organization of cellular pathways. This powerful method has enabled the discovery of novel protein interactions, the identification of protein complexes involved in disease processes, and the exploration of signaling networks underlying cellular functions. As advances in proteomic technologies continue to refine and expand, the analysis of affinity purification-related proteomic data will undoubtedly remain a fundamental tool for unraveling the intricate web of protein interactions that drive cellular processes, ultimately leading to a deeper understanding of cell biology and the development of targeted therapeutics.

With the role of the mammalian Hippo pathway as a vital signaling pathway involved in organ size control and tumorigenesis, targeting this pathway is an attractive target. Understanding the specific binding interactions between WW domains and their ligands opens possibilities for developing targeted therapeutics. The generated proteomics analysis pipeline was utilized to initiate the study of the WW domain within the Hippo pathway which further elucidated important motif-driven interactions. Modulating the WW domain interactions within the Hippo pathway can lead to potential treatments of cancer. Downstream transcriptional coactivator, YAP, has been demonstrated to have a dramatic effect on organ size through its hyperactivation(Dong *et al*, 2007). The cytoplasmic-nuclear shuttling of YAP has been established to be critical for its regulation, therefore, it has been suggested that blocking nuclear signaling promoted by YAP could be an approach to inhibit its activity and restrict tumor

growth(Yu *et al*, 2014; Zavan *et al*, 2019). Through its high expression in many human cancers, YAP drives tumorigeneses by promoting cancer cell survival, metastasis, and cell proliferation, highlighting YAP as an attractive therapeutic cancer treatment. Research efforts have been geared towards generating inhibitors towards YAP(Dey *et al.*, 2020; Gibault *et al*, 2017; Wang *et al*, 2016b; Ye & Eisinger-Mathason, 2016), though no successful drug has been generated creating a need for a YAP targeting therapeutic.

To generate a YAP targeting therapy, a YAP WW domain containing small peptide can be constructed. As mentioned, the WW domain is defined by the presence of two tryptophan (W) residues separated apart by ~25 amino acids, which can bind its proline-rich peptide ligand, mostly "PPxY" motif (P, proline; Y, tyrosine; x, any amino acid; hereafter named as "PY" motif). Failure of their recognition is associated with many human diseases including cancer(Chang *et al.*, 2007b), giving the WW domain-mediated protein-protein a critical role in biological processes. With the elucidation of the Hippo WW domain binding specificity for Hippo pathway components, such a therapy can be created. YAP contains two WW domains, which are required to recruit nuclear PY motif-containing proteins such as ERBB4, ARID1A, WBP2 to facilitate YAP transcriptional activity and oncogenic functions. With this criteria, a small peptide containing a WW domain that can target and inhibit YAP in YAP dependent cancers. Given its central role in diverse pathological processes, targeting YAP with the use of a WW domain small peptide therapy with in the Hippo pathway emerges as an exciting therapeutic avenue for addressing a wide range of diseases, potentially offering novel and effective treatments

To address the role of phospholipase D (PLD) and phosphatidic acid (PA) on the Hippo pathway, a proteomics approach as the one mentioned above was taken. This study identified several proteins that interact with the PLD and PA. While there are numerous novel interactions that were identified, future investigations are needed to further establish biological relevance. Our functional investigations established connections between members of the PLD family and various cellular processes. While further biochemical assays are needed to validate the interactions between the specific PLD bait and their respective HCIP prey, proteomics related analysis revealed many potential biological pathways these PLD members could be participating in. PLD1 was found to be a substrate of PJA2, however, it is important to note that 3 additional HCIPs fall under the category of proteosome-mediated ubiquitin dependent protein processing. This highlights that PLD1 could be a substrate of other ubiquitin ligases, future studies could expand on the degradation mechanism of this phospholipid. PLD2 prey identified sixteen members playing a role in HSP90 chaperone cycle for steroid hormone receptor (SHR). SHRs are transcription factors that become activated upon sensing steroid hormones such as androgens and estrogens. Many unliganded hormone receptors are located mostly in the cytoplasm, and with PLD2 localizing to the cytoplasm it is possible that PLD2 plays a role in regulating novel SHR pathways.

Interestingly, when analyzing the proteomics data for PLD3 and PLD4 HCIPs, approximately 30 members were found to be involved in protein processing in the endoplasmic reticulum (ER). Although PLD3 and PLD4 are proteins primarily localized to the ER, their specific roles in ER-related functions have remained largely unknown. Notably, our research has uncovered several important participants involved in ER protein quality control as binding partners for PLD3 and

PLD4 indicating potential involvement of PLD3 and PLD4 in the regulation of the ER stress response. Additionally, we observed the formation of a complex between PLD3 and PLD4 with lysosome proteins TM9SFs and the overexpression of PLD3 induced contact formation between the ER and lysosomes. As both PLD3/4 and TM9SFs are transmembrane proteins located on the ER and lysosomes, respectively, future investigations will focus on determining whether their interaction is directly mediated by their cytosolic regions or influenced by external factors. Furthermore, the signaling contexts and functional significance underlying the formation of the PLD3/4–TM9SFs complex warrant further exploration. Remarkably, our data also unveiled the occurrence of heterodimerization and homodimerization between PLD3 and PLD4. It will be highly intriguing to elucidate the role of such dimer formation in regulating the ER-related functions of PLD3/4. With recent discovery of PLD3 and PLD4 function as redundant lysosomal 5'-3' exonuclease with major activity towards mtDNA(Gavin *et al.*, 2018b; Gonzalez *et al*, 2018a; Van Acker *et al*, 2023) , it is important to further characterize this ER finding.

PLD5 remains very loosely studied. While this study revealed some potential new partners through validation, further studies are warranted to establish a biological role. Through the proteomics analysis, many ER related prey were found. While it does remain possible that PLD5 could have some contact with the ER through other interacting partners, this finding could be due to the handling of sample during the purification process of PLD5. PLD6 identified fourteen preys to be involved within mitochondrial membrane organization and fifteen prey to be involved in protein processing in the ER. PLD6 was demonstrated to bind through interaction assays to mitochondrial proteins ACOT9, CPT1A, and FAM49B as well as ER proteins ERLEC1, HSPA5, and P4HB. These findings indicate that PLD6 could participate in ER-

mitochondria crosstalk or have some type of interaction at contact sites. Further studies are needed to show how PLD6 is involved in the ER and under what physiological conditions. Finally, the proteomics analysis of PA revealed to have many interesting new potential interaction partners including nine involved in peroxisomal protein import. It is possible that PA binds to proteins involved in peroxisomal protein import, potentially regulating their activity or localization, thereby affecting the import process.

This study also demonstrated that PJA2 interacts and ubiquitinates oncogene PLD1 (Chen *et al*, 2012; Henkels *et al*, 2013a), and facilitates its degradation, suggesting that PJA2 may function as tumor suppressor and promote tumorigenesis. However, we found that when abolishing PJA2 expression and looked for tumor growth via a cell-derived xenograft model, there was no tumor growth and PJA2 was highly expressed in breast carcinoma tissues (unpublished data). This not only suggests that PJA2 is also an oncogene but that PJA2 has other independent functions yet to be elucidated. With this in mind, we aimed to explore the possible mechanism behind its tumor suppressive functions and found in a colleagues MS dataset for LC3A, a protein involved in the elongation and closure of the autophagosome membrane, that PJA2 was a prey identified (unpublished data). Through further interactions studies, we saw that PJA2 interacted strongly with LC3A amongst a panel of autophagy related proteins leading to the notion that PJA2 plays a role in autophagy. Further investigation found that under CQ treatment, PJA2 colocalizes with LC3A punctate and its binding is abolished when deleting an LC3 interacting region identified within PJA2 (unpublished data). While autophagy remains to be a regulated and detailed biological process, E3 ligases have been seen to play a role with autophagy machinery(Sun *et al*,

2017). While this data is unpublished, further studies are required to further elucidate the role of PJA2.

In conclusion, the integration of the studies mentioned has provided a comprehensive understanding of different aspects of protein biology and interaction networks. The generation of a pipeline for the analysis of affinity purification-related proteomic data has allowed for a systematic exploration of protein-protein interactions within cells, offering valuable insights into the complex network of molecular associations. Furthermore, the elucidation of WW domain ligand binding specificities in the Hippo pathway has identified STXBP4 as a potent inhibitor of YAP, uncovering a novel regulatory mechanism. Additionally, the interactome analysis of human Phospholipase D and the associated phosphatidic acid protein network has shed light on the intricate cellular signaling pathways involved in lipid metabolism. By combining these studies, we have deepened our knowledge of protein-protein interactions, unveiled novel regulatory mechanisms, and gained insights into the complex orchestration of cellular processes. These findings contribute to the broader understanding of molecular biology and pave the way for further investigations aimed at deciphering the intricacies of cellular networks and developing targeted therapeutic approaches.

4.3 References

Chang NS, Hsu LJ, Lin YS, Lai FJ, Sheu HM, 2007. WW domain-containing oxidoreductase: a candidate tumor suppressor, Trends in Molecular Medicine. Elsevier Current Trends, pp. 12-22.

Chen Q, Hongu T, Sato T, Zhang Y, Ali W, Cavallo J-A, van der Velden A, Tian H, Di Paolo G, Nieswandt B (2012) Key roles for the lipid signaling enzyme phospholipase d1 in the tumor microenvironment during tumor angiogenesis and metastasis. *Science signaling* 5: ra79-ra79

Dey A, Varelas X, Guan KL (2020) Targeting the Hippo pathway in cancer, fibrosis, wound healing and regenerative medicine. *Nature Reviews Drug Discovery 2020 19:7* 19: 480-494

Dong J, Feldmann G, Huang J, Wu S, Zhang N, Comerford SA, Gayyed Mariana FF, Anders RA, Maitra A, Pan D (2007) Elucidation of a Universal Size-Control Mechanism in Drosophila and Mammals. *Cell* 130: 1120-1133

Gavin AL, Huang D, Huber C, Mårtensson A, Tardif V, Skog PD, Blane TR, Thinnes TC, Osborn K, Chong HS *et al*, 2018. PLD3 and PLD4 are single-stranded acid exonucleases that regulate endosomal nucleic-acid sensing, Nature Immunology 2018 19:9. Nature Publishing Group, pp. 942-953.

Gibault F, Bailly F, Corvaisier M, Coevoet M, Huet G, Melnyk P, Cotelle P, 2017. Molecular Features of the YAP Inhibitor Verteporfin: Synthesis of Hexasubstituted Dipyrrins as Potential Inhibitors of YAP/TAZ, the Downstream Effectors of the Hippo Pathway, ChemMedChem. John Wiley and Sons Ltd, pp. 954-961.

Gonzalez AC, Schweizer M, Jagdmann S, Bernreuther C, Reinheckel T, Saftig P, Damme Correspondence M, 2018. Unconventional Trafficking of Mammalian Phospholipase D3 to Lysosomes Yeast Mammalian cells Vacuole Late endosome / MVB Carboxypeptidase S (CPS) pathway Alkaline phosphatase (ALP) pathway Golgi Lysosome Late endosome / MVB Phospholipase D3 Lysosomal membr, CellReports. pp. 1040-1053.

Hanash S (2003) Disease proteomics. *Nature* 422: 226-232

Henkels KM, Boivin GP, Dudley ES, Berberich SJ, Gomez-Cambronero J, 2013. Phospholipase D (PLD) drives cell invasion, tumor growth and metastasis in a human breast cancer xenograph model, Oncogene 2013 32:49. Nature Publishing Group, pp. 5551-5562.

Huang S, Chaudhary K, Garmire LX (2017) More is better: recent progress in multi-omics data integration methods. *Frontiers in genetics* 8: 84

Kattan RE, Ayesh D, Wang W (2023) Analysis of affinity purification-related proteomic data for studying protein–protein interaction networks in cells. *Briefings in Bioinformatics* 24: bbad010

Reel PS, Reel S, Pearson E, Trucco E, Jefferson E (2021) Using machine learning approaches for multi-omics data analysis: A review. *Biotechnology Advances* 49: 107739

Rolland T, Taşan M, Charloteaux B, Pevzner SJ, Zhong Q, Sahni N, Yi S, Lemmens I, Fontanillo C, Mosca R (2014) A proteome-scale map of the human interactome network. *Cell* 159: 1212-1226

Räschle M, Smeenk G, Hansen RK, Temu T, Oka Y, Hein MY, Nagaraj N, Long DT, Walter JC, Hofmann K (2015) Proteomics reveals dynamic assembly of repair complexes during bypass of DNA cross-links. *Science* 348: 1253671

Srivastava A, Creek DJ (2019) Discovery and validation of clinical biomarkers of cancer: a review combining metabolomics and proteomics. *Proteomics* 19: 1700448

Sun A, Wei J, Childress C, Shaw IV JH, Peng K, Shao G, Yang W, Lin Q (2017) The E3 ubiquitin ligase NEDD4 is an LC3-interactive protein and regulates autophagy. *Autophagy* 13: 522-537

Van Acker ZP, Perdok A, Hellemans R, North K, Vorsters I, Cappel C, Dehairs J, Swinnen JV, Sannerud R, Bretou M *et al* (2023) Phospholipase D3 degrades mitochondrial DNA to regulate nucleotide signaling and APP metabolism. *Nature Communications* 14: 2847

Vargas RE, Duong VT, Han H, Ta AP, Chen Y, Zhao S, Yang B, Seo G, Chuc K, Oh S *et al* (2020) Elucidation of WW domain ligand binding specificities in the Hippo pathway reveals STXBP4 as YAP inhibitor. *EMBO Journal* 39

Wang C, Zhu X, Feng W, Yu Y, Jeong K, Guo W, Lu Y, Mills GB, 2016. Verteporfin inhibits YAP function through up-regulating 14-3-3σ sequestering YAP in the cytoplasm, American Journal of Cancer Research. E-Century Publishing Corporation, pp. 27-37.

Ye S, Eisinger-Mathason TSK, 2016. Targeting the Hippo pathway: Clinical implications and therapeutics, Pharmacological Research. Academic Press, pp. 270-278.

Yu FX, Luo J, Mo JS, Liu G, Kim YC, Meng Z, Zhao L, Peyman G, Ouyang H, Jiang W *et al* (2014) Mutant Gq/11 Promote Uveal Melanoma Tumorigenesis by Activating YAP. *Cancer Cell* 25: 822-830

Zavan B, Yuan H, Blank M, Sudol M, Chen Y-A, Chen H-F, Chang N-S, Lu C-Y, Cheng T-Y, Pan S-H, 2019. WW Domain-Containing Proteins YAP and TAZ in the Hippo Pathway as Key Regulators in Stemness Maintenance, Tissue Homeostasis, and Tumorigenesis, Frontiers in Oncology | wwwfrontiersinorg. p. 60.