Université de Montréal

Quantitative proteomics identifies substrates of SUMO E3 ligase PIAS proteins involved in cell growth and motility

Par

Chongyang Li

Institut de recherche en immunologie et cancérologie,

Programme de Biologie Moléculaire, option Biologie des Systèmes

Faculté de Médecine

Thèse présentée à la Faculté des études supérieures et postdoctorales

en vue de l'obtention du grade de Philosophiae Doctor (Ph.D.)

en Biologie Moléculaire, option Biologie des Systèmes

December 2020

© Chongyang Li, 2020

Université de Montréal

Unité académique : Institut de recherche en immunologie et cancérologie, Faculté de Médecine

---

Cette thèse intitulée

Quantitative proteomics identifies substrates of E3 SUMO ligase PIAS proteins involved in cancer cell growth and motility

Présenté par

Chongyang Li

A été évaluée par un jury composé des personnes suivantes

Katherine Borden

Président-rapporteur

Pierre Thibault

Directeur de recherche

James Omichinski

Membre du jury

Brian Raught

Examinateur externe

Résumé

La SUMOylation des protéines est une modification post-traductionnelle se produisant sur des lysines d'un large éventail de protéines cellulaires. Cette modification est dynamique et régit plusieurs évènement cellulaires essentiels, dont la translocation et la dégradation des protéines, la ségrégation chromosomique mitotique, la réparation de l'ADN, la progression du cycle cellulaire, la prolifération cellulaire et la migration. La conjugaison de la protéine SUMO sur son substrat se produit grâce à une triade enzymatique regroupant l'enzyme d'activation E1 SAE 1/2, la conjugase E2 UBC9 et dans la plupart des cas une ligase SUMO E3. Cette cascade enzymatique nécessite une source d'ATP pour son initiation. Parmi la famille des ligases SUMO E3, on retrouve un domaine spécifique nommé SP-RING présent chez une sous population de celles-ci. Parmi ces ligases on retrouve 7 protéines inhibitrices des protéines STAT activées regroupees sous le nom de PIAS. Les ligases PIAS ont été identifiées à l'origine comme des inhibiteurs spécifiques des protéines STAT responsable du signal de transduction et de l'activation de la transcription génique. Des études récentes ont montré que les protéines PIAS jouent également un rôle important sur la stabilité de leurs substrats et la transduction de leur signal. De plus, les substrats SUMOylés par les PIAS sont impliqués dans plusieurs processus cellulaires, notamment la réparation des dommages à l'ADN, la réponse immunitaire, la prolifération et la motilité cellulaire. Ces divers processus cellulaires peuvent être déréglés et entrainer le développement du cancer. Il s'avère que les protéines PIAS sont fortement exprimées dans divers types de cancer et sont impliquées dans la tumorigenèse. Plusieurs rapports suggèrent que les protéines PIAS pourraient favoriser la croissance et la progression des cellules cancéreuses en régulant le niveau de SUMOylation de plusieurs substrats. Initialement, les substrats des ligases PIAS ont été identifiés à partir de plusieurs études individuelles et plus récemment, des centaines de substrats spécifiques de la SUMO E3 ligase ont été identifiés à partir de criblage de micropuces à protéines interrogeant le protéome humain. Cependant, la manière dont ces substrats sont sélectionnés et quels sont les sites de SUMOylation ciblés par ces PIAS demeurent encore méconnus.

Afin d'aborder ces questions, j'ai commencé mon étude avec PIAS1, l'une des ligases SUMO E3 les plus étudiées. Pour ce faire, j'ai varié le niveau d'expression de PIAS1 dans des cellules

HeLa selon l'approche CRISPR/Cas9. Ainsi, deux modèles ont été construit, soit via une surexpression du gène ou via un knockout du gène. Ces mutants ont permis de constater que PIAS1 avait un impact physiologique sur la prolifération et la migration des cellules. J'ai tiré avantage d'une méthode protéomique précédemment développé sur les peptides SUMO pour déterminer les changements de SUMOylation lors de la surexpression de PIAS1. J'ai identifié 983 sites SUMO sur 544 protéines, dont 62 protéines ont été identifiées comme substrats potentiels de PIAS1. Parmi celles-ci, la vimentine (VIM), une protéine de la famille des filaments intermédiaire de type III impliquée dans l'organisation du cytosquelette et la motilité cellulaire, a été reconnu comme un substrat de PIAS1. Afin de valider le rôle de la SUMOylation des lysines Lys-439 et Lys-445 de VIM j'ai effectué des études fonctionelles de motilité cellulaire avec les mutants où ces sites ont été substitués en arginine. Ces expériences m'ont permis de constater que la SUMOylation de VIM aux sites Lys-439 et Lys-445 est nécessaire à l'assemblage et désassemblage dynamique des filaments intermédiaires de VIM, lesquels regulent la migration et la motilité cellulaire.

Dans la deuxième étude, j'ai élargi mon recherche sur toutes les ligases PIAS et avons découvert que ces dernières avaient toutes un impact sur la prolifération cellulaire et la migration des cellules du cancer du sein MDA-MB-231 suite à un knockout de ces gènes par CRISPR / Cas9. De plus, j'ai optimisé mon approche de protéomique quantitative SUMO via SILAC et l'avons complémenté d'une analyse transcriptomique. Cette combinaison a permis d'acquérir une compréhension des composants fonctionnels impliqués dans les réseaux de régulation PIAS. Il s'avère qu'un grand sous-ensemble de gènes / protéines impliqués dans la migration et la prolifération des cellules sont régulés par tous les membres de la famille PIAS, et suggère une certaine redondance fonctionnelle parmi ces ligases. De plus, chaque PIAS régule un ensemble unique de substrats / gènes impliqués dans plusieurs processus cellulaires différents, tels que la réparation des dommages de l'ADN, le remodelage de la chromatine et la formation de la chaîne SUMO. Ces résultats suggèrent que chacune des PIASs régule de façon spécifique les fonctions cellulaires. La combinaison des analyses protéomiques et transcriptomiques ont permi de dresser un portrait global des mécanismes de régulation régit par les protéines PIAS et ce au-delà de leur activité enzymatique directe.

Abstract

Protein SUMOylation is a highly dynamic and reversible post-translational modification that targets lysine residues on a wide range of proteins involved in several essential cellular events, including protein translocation and degradation, mitotic chromosome segregation, DNA damage response, cell cycle progression, cell proliferation, and migration. Protein SUMOylation is an ATP-dependent enzymatic process that involves an E1 activating enzyme SAE1/2, a E2 conjugase UBC9, and usually facilitated by SUMO E3 ligases. The SP-RING family is the largest family of SUMO E3 ligases, encompassing seven mammalian protein inhibitor of activated STAT (PIAS) proteins. PIAS family was originally identified as specific inhibitors for signal transducer and activator of transcription (STAT), which involves gene transcriptional regulation. Recent studies showed that PIAS proteins also play important roles in the regulation of protein stability and signal transduction through the SUMOylation of target substrates. In addition, PIAS-mediated protein SUMOylation is also involved in several cellular processes, including DNA damage repair, immune response, cellular proliferation, and motility. Most notably, PIAS proteins are highly expressed in different cancer types and have been implicated in tumorigenesis. Several reports suggest that PIAS proteins could promote cancer cell growth and progression by regulating the SUMOylation of different substrates. To date, a number of substrates of PIAS ligases have been identified from several individual studies, and hundreds of specific SUMO E3 ligase substrates were identified from a human proteome microarray-based activity screen. However, how these substrates are selected, and which SUMOylation sites are targeted by these PIAS are still unknown.

To answer these questions, I started my investigation with PIAS1, one of the most well studied SUMO E3 ligases. By changing the expression level of PIAS1 in HeLa cells using gene overexpression or CRISPR/Cas9 gene knockout, I found PIAS1 had a physiological impact on cell proliferation and migration. I took advantage of the previously developed SUMO proteomics workflow to quantitatively profile global SUMOylome changes upon PIAS1 overexpression in a site-specific manner. I identified 983 SUMO sites on 544 proteins, of which 62 proteins were assigned as putative PIAS1 substrates. In particular, Vimentin (VIM), a type III intermediate

filament protein involved in cytoskeleton organization and cell motility, was identified as PIAS1 substrates. Two SUMOylation sites mediated by PIAS1 at Lys-439 and Lys-445 residues were further evaluated and found to be necessary for dynamic disassembly and assembly of vimentin intermediate filaments, which further regulates cell migration and motility.

In the second study, I extended my investigation to all PIAS ligases and further found that all PIAS proteins impact cell proliferation and migration of breast cancer cell MDA-MB-231 after CRISPR/Cas9 gene knockout. I further optimized my SILAC-based quantitative SUMO proteomics approach and combined it with transcriptomics to gain a system-level understanding of the functional components involved in PIAS regulatory networks. A large subset of proteins/ genes involved in cell proliferation and migration were commonly regulated by all PIAS proteins, suggesting a redundancy of regulation within the PIAS family. In addition, each PIAS regulated a unique pool of substrates/genes involved in different cellular processes, such as DNA damage repair, chromatin remodeling, and SUMO chain formation, suggesting that each PIAS specifically regulates cellular functions. The trans-scale analyses between proteomics and transcriptomics shed light on the comprehensive pictures of the regulation networks by PIAS proteins beyond their direct enzymatic activity.

Overall, the quantitative SUMO proteomics approach provided a robust method for identifying substrates of PIAS SUMO E3 ligases. The combination of proteomic and transcriptomic analyzes made it possible to draw up a global portrait of the regulatory mechanisms governed by the PIAS proteins.

Key words: SUMOylation, SUMO E3 ligase, PIAS proteins, proteomics, mass spectrometry, cell proliferation, cell migration

Table of Contents

List of Figures

List of abbreviations

2D	Two dimensional
µM	Micromolar
ACN	Acetonitrile
AD	Acidic Domain
AGC	Automatic Gain Control
ATP	Adenosine Triphosphate
C18	Octadecyl carbon chain
CID	Collision induced dissociation
CTL	Control
Da	Dalton
DC	Direct Current
DAPI	4',6-diamidino-2-phenylindole
DDA	Data Dependent Acquisition
DDA	Data Dependent Acquisition
DIA	Data Independent Acquisition
DMEM	Dulbecco Modified Eagle's minimal Mssential Medium
DNA	Deoxyribonucleic Acid
DUBs	Deubiquitinating Enzymes
EGFR	Epidermal Growth Factor Receptor
EMT	Epithelial-Mesenchymal Transition
ESI	Electrospray Ionisation
FA	Formic Acid
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
GO	Gene Ontology
HCD	Higher-energy Collisional Dissociation
HeLa	Henrietta Lacks cervical cancer cell line
HLB	Hydrophilic-Lipophilic Balance
HPLC	High performance liquid chromatography
ISG15	Interferon Stimulated Gene 15
LC	Liquid Chromatography
LC-MS	Liquid Chromatography Mass Spectrometry
LFQ	Label Free Quantification
LIT	Linear Ion Trap
m/z	Mass-to-Charge
mM	millimolar
MALDI	Matrix-Assisted Laser Desorption/Ionization
mRNA	messenger Ribonucleic Acid
MS	Mass spectrometry
MS/MS	Tandem Mass Spectrometry
NTA	Nitriloacetic Acid
PAGE	Polyacrylamide Gel Electrophoresis
PBS	Phosphate Buffered Saline
PCNA	Proliferating Cell Nuclear Antigen

PSM	Peptide Spectrum Match
PIAS	Protein Inhibitor of Activated STAT
PINIT	Pro-Ile-Asn-Ile-Thr
PML	Promyelocytic Leukemia Protein
PPIs	Protein-Protein Interactions
PRM	Parrel Reaction monitoring
PTM	Posttranslational Modificaiton
RanGAP	Ran GTPase-Activating Protein
RLD	RING-finger-like zinc-binding domain
RNA	Ribonucleic Acid
RNF	RING Finger Protein
RP	Reverse Phase
RT	Room Temperature
SAE	SUMO Activating Enzyme
SAP	Scaffold Attachment factor A/B/acinus/PIAS
SCX	Strong Cation Exchange
SD	Standard Deviation
SDM	Site-directed mutagenesis
SDS	Sodium Dodecyl Sulfate
SENP	Sentrin-specific Protease
SILAC	Stable Isotope Labeling by Amino Acids in Cell Culture
SIM	SUMO Interacting Motif
SP-RING	Siz/Protein Inhibitor of Activated STAT-RING
sgRNA	Single guilde Ribonucleic Acid
S/T	Serine/Threonine-rich region
SUMO	Small Ubiquitin-like Modifier
STUbL	SUMO-targeted ubiquitin ligases
UBL	Ubiquitin-like
TBS	Tris Buffered Saline
TCE	Total Cell Extract
TCEP	Tris (2-Carboxyethyl) Phosphine Hydrochloride
TRIM	Tripartite Motif-Containing Protein
UBC9	E2 SUMO Conjugating Enzyme
VIM	Vimentin
WT	Wild-Type

Acknowledgements

Words are far less than enough to express my gratitude to those who have ever supported and helped me during my Ph.D. study and life in Montréal.

Firstly, I would like to sincerely say "Je vous remercie !" to my research director, Dr. **Pierre Thibault**, Time flies; it has been almost seven years since I arrived in Montreal. However, it was like yesterday that every moment was still so clearly remembered in my mind. We got in touch for the first time in January 6, 2014. Then the following several months, we discussed closely about my project and the study plan here through Skype and emails. Thanks for his acceptance of being my research director and his support for my application of FRQNT scholarship, which brought me here from another side of the earth to be his student. I greatly appreciate all the support from him during my research acitivies. His profound knowledge and broad vision in our field always lead me in an enlightened direction. I am so proud that I completed my thesis and published my research results in internationally renowned journals, not because of how smart I am, but because I always "Stand on the shoulders of giant"! His rigorous and meticulous attitude in research work has always been influencing me over the years. His careful guidance has created a scientist who is confident, independent and persistent. I believe that I am ready to meet all kinds of difficulties in my next research career, fight them and pursue the truth.

Secondly, I will also express my sincere thanks to Dr. **Francis McManus** (Le Gros et gym buddy). I remembered that he bought me a pasta salad the first day I was in the lab; I remembered that he said "Gung hei faat coi!" (Happy new year in Cantonese) to me for my first traditional Chinese new year in Canada. I remembered he came to fix MS around 6:00 am on a Saturday morning after receiving my text message. I remembered that he always came to brainstorm with me on my projects and helped me troubleshoot the issues I met. I cannot write down every moment he was here while I want to summarize that he is the one who took me on the track for the SUMO proteomics in practice. I felt safe when he was here. Finally, we published our paper on ***Nature Communications***.

« La victoire appartient aux plus persévérants. »

- Napoléon Bonaparte

CHAPTER ONE

1 Introduction

1.1 Proteome complexity

The central dogma of molecular biology elaborates the flow of genetic information encoded in the DNA sequence to its transcription into RNA and eventually translated into the amino acid sequences comprising the primary structure of a protein [1]. The term "proteome" was introduced at the 1st Siena meeting in 1994 by Marc Wilkins to describe the protein complement of a genome [2]. It represents the whole set of proteins expressed at a specific time in a particular cell or tissue in a species [3]. Over the last few decades, particularly due to advances in mass spectrometry (MS), scientists have gradually discovered that the human proteome is much more complex than the genome [4].
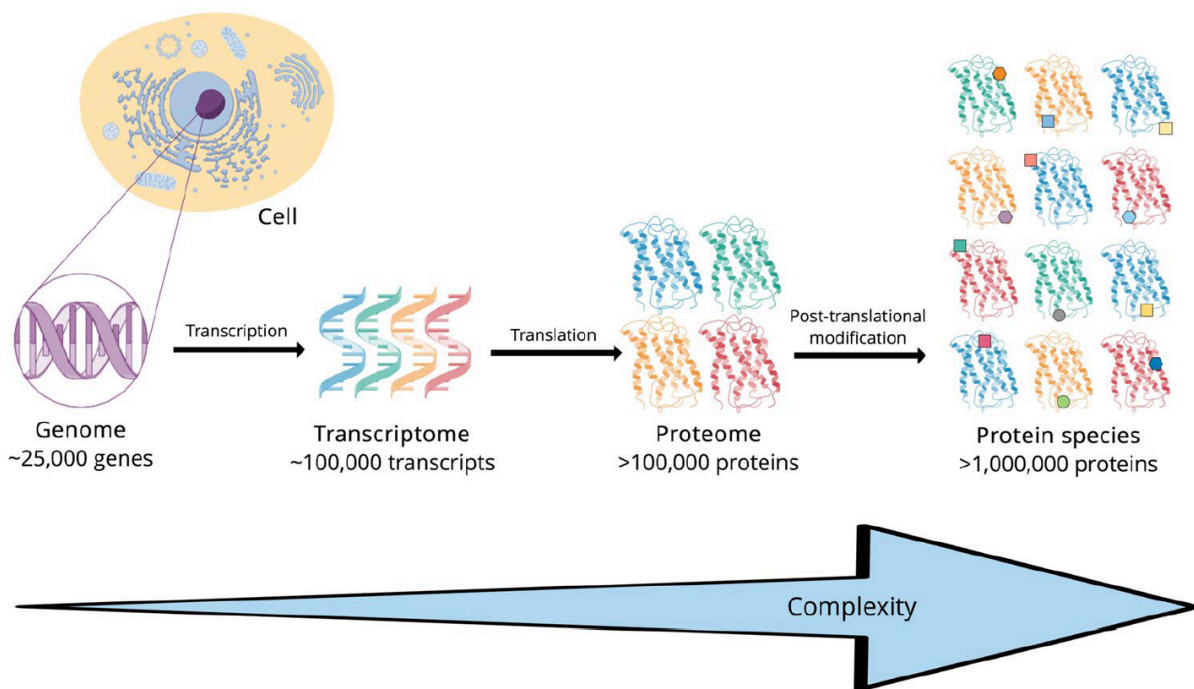
Figure 1-1 The increase in complexity from the genome to the proteome.Adapted with permission from Springer Nature: Chromatographia (Virág D. et al.) [5] © 2020.

Based on current discoveries, it is generally accepted in the scientific community that the human genome is composed of approximately 20,000 to 25,000 genes, whereas it is estimated

2

that there are more than one million variants [6]. Studies have shown that changes in mRNA levels during transcription, as well as alternative splicing, increase the size of the transcriptome, which reaches nearly 100,000 transcripts. These transcripts are then translated inside the cell to form corresponding peptides and eventually proteins with spatial structures. A growing number of studies have shown that the numerous chemical modifications involved in the subsequent synthesis and transformation of proteins lead to the completion of more than one million human proteins and exponentially increase the complexity of the human proteome relative to the human genome and transcriptome [7, 8].

1.2 Post-Translational Modifications

A post-translational modification (PTM) is a highly dynamic and reversible chemical modification that occurs on intracellular proteins and plays an essential role in the functioning of an organism's proteome [9]. After protein biosynthesis, the corresponding enzyme of a certain modification can covalently modify a specific amino acid site of the protein [10]. More than 200 post-translational modifications have been discovered to date and can be classified according to the properties of their chemical groups into the following categories [11, 12]:

- Based on **additional chemical groups:** phosphorylation, acetylation, methylation, etc.
- Based on **additional complex molecules:** glycosylation, AMPylation, Lipidation
- Based on **additional proteins/polypeptides:** ubiquitination modifications and various ubiquitin-liked modifications such as SUMOylation, Neddylation, ISGylation, etc.
- Based on **the protein cleavage:** Proteolysis
- And based on **amino acid modifications**: deamidation

A protein can be modified with a PTM at any stage during its "life cycle", e.g., many proteins are modified immediately after being translated in order to promote proper protein folding, or protein stability, or to be directed into distinct cell compartments such as the nucleus, cell membrane, different organelles, or microtubules. Subsequently, novel protein modifications may occur on the same protein to promote further activation or inactivation of the protein's catalytic activity, thus further influencing the biological activity of the protein [13-16]. In addition to simple modifications, proteins are often modified by a combination of different PTMs and the addition of functional groups through a progressive mechanism of protein maturation or activation [17, 18].

1.2.1 SUMOylation

Small ubiquitin-like modifier (SUMO) proteins are a group of small proteins ubiquitously expressed in all eukaryotic cells. They are highly conserved throughout evolution and dynamically reversible on the target proteins to increase the functional repertoire of the eukaryotic proteome

[19]. It was initially discovered by several independent research groups in a variety of contexts in the mid-1990s [20, 21]. Meluh and Koshland reported for the first time in 1995 that the deletion of the only SUMO isoform (SMT3) in *Saccharomyces cerevisiae* causes a loss of cell viability [22]. This was followed by three more studies in 1996, in which SUMO proteins were found to be binding partners of human DNA repair enzymes RAD51/52 [23], TNF receptor Fas/APO-1 [24] and PML [25]. Soon after, two studies showed that SUMO1 is a reversible protein modification covalently attached to the Ran GTPase-activating protein RanGAP1 and alters protein subcellular localization via the regulation of protein interactions [20, 21]. Through homology screening, other SUMO paralogs, designated SUMO2 and SUMO3, were hypothesized and subsequently shown to be functional ubiquitin-like proteins (UBLs) [26, 27].

SUMO proteins are approximately 12 kDa in size and composed of around 100 amino acids [19]. They have virtually superimposable three-dimensional (3D) structures of ubiquitin, featuring the ββαββαβ fold of the ubiquitin-protein family [28]. However, the sequence identity shared between SUMO and ubiquitin is less than 20%. Moreover, the overall surface charge distribution of SUMO differs significantly from that of ubiquitin, which creates distinct negative and positively charged patches [29]. In addition, SUMO proteins have a 10-25 amino acid unstructured N terminus, which is not found in any other ubiquitin-related proteins [30].

To date, all eukaryotes are found to express at least one SUMO protein. Some species only have a single gene for expressing SUMO protein, for instance, *SMT3* in *Saccharomyces cerevisiae*, *smo-1* in *Caenorhabditis elegans*, and *smt3* in *Drosophila melanogaster* [31-33]. Other organisms, such as plants and vertebrates, have several SUMO proteins expressed in the cells [34-36]. There are four distinct SUMO proteins confirmed in humans: SUMO1-SUMO4, and among them, SUMO1-3 are ubiquitously expressed [30]. The mature forms of human SUMO2 and SUMO3 proteins are 97% identical, only differing by three amino acids and presently indistinguishable by antibodies. Therefore, SUMO2 and SUMO3 are commonly referred to as SUMO2/3. However, SUMO1 is quite divergent and shares ~50% sequence identity with SUMO2/3 and can be distinguished by different antibodies [37]. The sequence alignment among human SUMO1, and SUMO2/3, and 3D structures are further elaborated in Chapter 2. In addition, SUMO4 protein is

quite similar to SUMO2/3 in sequence identity; however, instead of glutamine, there is a proline (Pro or P) residue located at position 90 that is unique to SUMO4 [38]. The maturation process of SUMO proteins involves the proteolytic cleavage of its C-terminus tail to expose a di-glycine necessary for conjugation to their substrates [3, 39]. It is shown that the maturation process is prevented by Pro-90 due to the inability of the precursor hydrolyzation. Thus, native SUMO4 cannot be activated into the mature form [40, 41]. Although the conjugation mechanism of SUMO4 is still unclear, the maturation of SUMO4 was found in cells stimulated with serum starvation, and this maturation capacity was shown to increase when the serum starvation treatment was prolonged [42]. Among these, human SUMO1-SUMO3 are widely expressed in various tissues, whereas the distribution of SUMO4 seems to be tissue- or organ-dependent, and it is predominantly found in tissues such as the kidney, lymph node, and spleen [43]. It is worth noting that the human genome also encodes a novel SUMO variant, SUMO5, which is reported to be involved in the formation of PML nuclear bodies (NBs), as well as their disruption [44]. With the development of proteomic techniques, thousands of novel substrates were found to be SUMOylated; although the majority of them were located in the nucleus, they were also reported to be distributed almost everywhere in the cell [45-49]. In cells, SUMO1 substrates were primarily found at the nuclear membrane and in nuclear bodies, whereas SUMO2 substrates are partially cytosolic and SUMO3 substrates are located mainly in nuclear bodies [50-52]. At the mitochondrial outer membrane, mitochondrial fission proteins are SUMOylated by all three SUMO isoforms [53]. In addition, all SUMO cascade enzymes undergo SUMOylation, including SUMO proteins themselves. SUMOylation plays crucial roles many important biological processes; within the nucleus, it regulates transcription regulation, chromatin organization, DNA damage repair, and the immune response [54]. In the cytoplasm, several other effects of SUMOylation have been reported, such as protein translocation, recruitment of proteins that possess a SUMO-interacting motif (SIM), steric interference with protein-protein interactions, and crosstalk with other PTMs. Like ubiquitylation and phosphorylation, in most cases, SUMOylation inhibits the activity of the target proteins [54-57]. A more detailed discussion on the functions of SUMOylation will be elaborated in Chapter two.
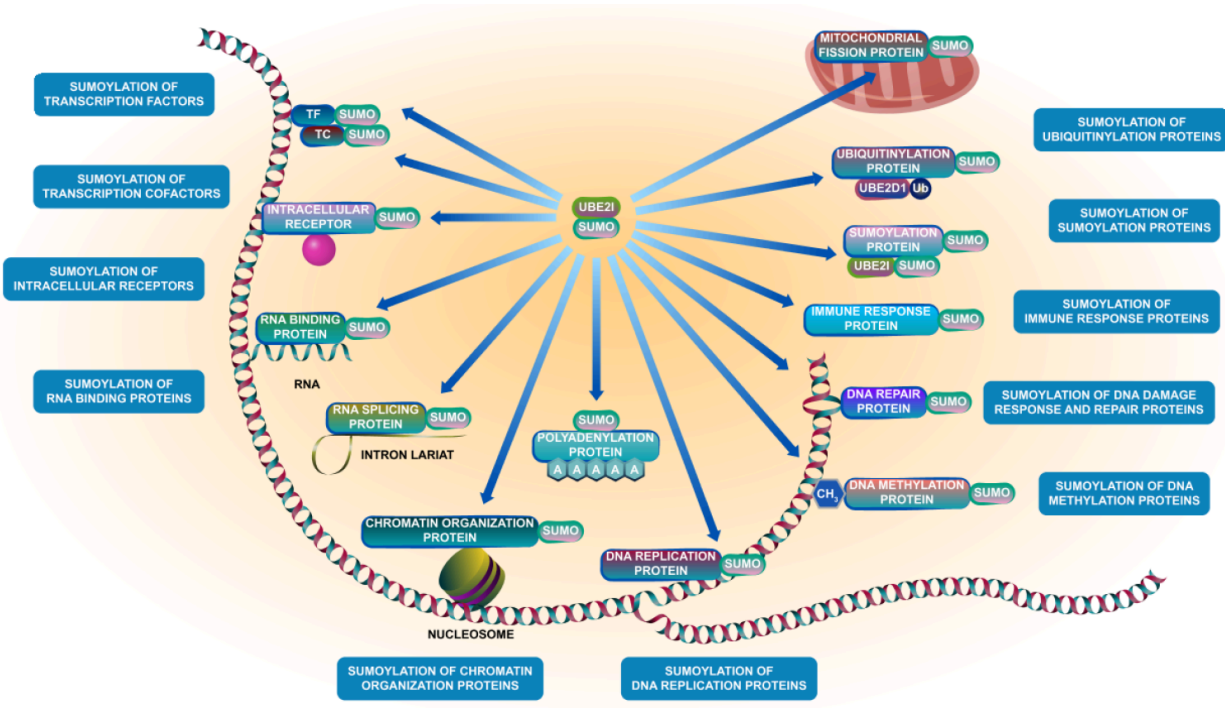
Figure 1-2 Different functions of SUMOylation.

Adapted with permission from Reactome: (May B.) [58] © 2013.

1.2.2 SUMO pathway

Similar to ubiquitin, the SUMO protein forms a covalent bond between the Gly residue at its C-terminus and the ε-NH$_2$ group on a substrate Lys residue [59, 60]. Prior to SUMO conjugation, the nascent SUMO needs to be C-terminally cleaved by SUMO-specific isopeptidases, sentrin-specific proteases (SENPs) to expose a Gly-Gly motif [61, 62]. This maturation process results in the removal of 4 amino acids from SUMO1 and 11 amino acids from SUMO2/3 [39]. Although all SUMO isoforms differ in the sequence identity, the conjugation processes share the same enzymatic cascade [30].
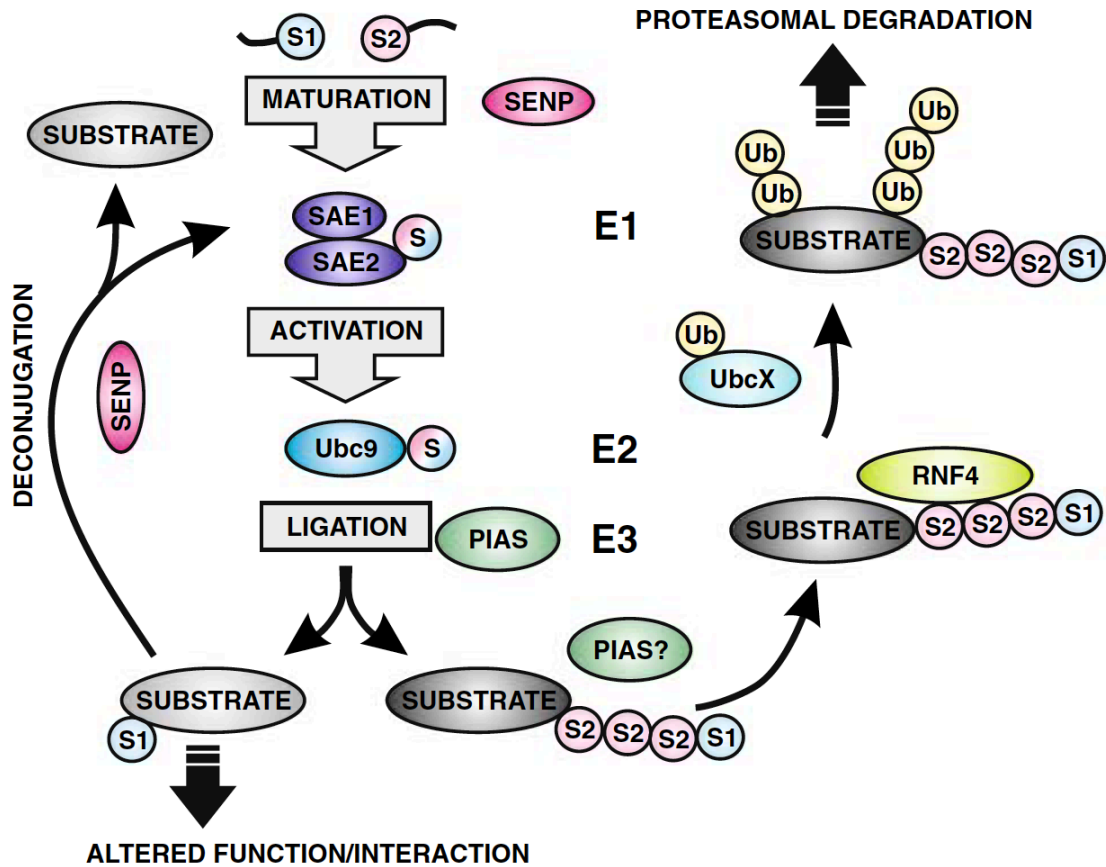
Figure 1-3 SUMO modification pathway.

Adapted with permission from Springer Nature: Cellular and Molecular Life Sciences (Miia M. Rytinki et al.) [63] © 2009.

For the reversible SUMOylation process, the first step is to activate mature SUMO by SUMO E1 activating enzyme heterodimer, which is composed of two subunits: E1-SUMO Activating Enzyme subunit 1 and 2 (SAE1/SAE2) [54, 64]. This activation process is an ATP-dependent reaction, which requires the formation of the SUMO adenylate, and the thioester bond between the C-terminal carboxyl group of SUMO and the catalytic Cys residue of SAE2 [65]. Next, SUMO is transferred to the E2 conjugating enzyme UBC9, which forms a new thioester linkage between the catalytic Cys residue of E2 and the C-terminal carboxyl group of SUMO [54]. Unlike the ubiquitin pathway, there is only one E2 conjugating enzyme that is providing activated SUMO and responsible for all the conjugation processes. Additionally, UBC9 directly interacts with target proteins which contains a consensus motif Ψ-K-x-D/E, where Ψ is a hydrophobic residue, K is the

acceptor lysine, x is any amino acid (aa), and D or E is an acidic residue [66]. Subsequently, SUMO is transferred to the substrate by UBC9 to form a covalent isopeptide bond between the Gly residue at the C-terminus of SUMO protein and the Lys residue on the substrate [67]. In most cases, this conjugation process can be completed solely by UBC9. However, SUMO ligation to target proteins is usually assisted by SUMO E3 ligases, which are enzymes that facilitate the transfer of SUMO from UBC9 to the substrate [68]. In addition, SUMO E3 ligases also facilitate the formation of SUMO polychains on the target proteins [69]. Next, the SUMOylated substrates will have two alternative fates depending on their SUMO status. When monoSUMOylation occurs on the substrates, the cellular localization and protein function of the substrates will be regulated accordingly, and the interaction with the other protein may also be altered. In contrast, substrate polySUMOylation often triggers SUMO-targeted ubiquitin ligase (STUbL) dependent proteasomal degradation. In this process, a SUMO-targeted ubiquitin ligase, such as RNF4, recognizes the polySUMO chain on the substrate and then promotes the ubiquitination of the substrates. Finally, polyubiquitination directs the substrates to the proteasome, resulting in protein degradation [70].

1.2.3 SUMO E3 Ligases

In the ubiquitination system, the estimated number of E3 ligases is ~1000, and E3 ligase plays a key role in substrates selection, therefore, the ubiquitination process is the result of the collaboration between E2 and E3 [71, 72]. However, unlike ubiquitin, less than 20 SUMO E3 ligases have been found to date. Moreover, only the E2 conjugase UBC9 seems to be able to complete SUMOylation *in vitro* without the aid of SUMO E3 ligase [73, 74]. Additionally, there is no covalent interaction between SUMO E3 ligases and their substrates during the conjugation/ligation process [75]. In 2001, Takahashi Y. et al. demonstrated for the first time in yeast that Siz1, a new member of the PIAS family, is necessary for the SUMO E3-dependent conjugation process both *in vitro* and *in vivo* [76]. After that, more and more evidence showed the important roles of SUMO E3 ligases in the SUMOylation of target proteins.

Based on the protein structure, SUMO E3 ligases have been grouped into two distinct groups: IR1-M-IR2 domains of the nucleoporin RanBP2/Nup358 and SP-RING E3 ligase family [77]. The polycomb group protein Pc2 was also described as a SUMO E3 ligase, the substrates of which are the transcriptional corepressors CtBP1 and CtBP2 both *in vitro* and in *vivo*; however, the mechanism remains unclear [78]. RanBP2 is a 358 kDa protein containing multi-domains that interact with nuclear transport receptors, the Ran GTPase-activating protein RanGAP1, and UBC9 [79]. Structural and biochemical evidence suggests that RanBP2 contains multiple E3 ligase regions that are unfolded under native conditions [80]. However, sequence similarity shows they are not obviously homologous to other proteins [81]. The IR1 domain of RanBP2 is responsible for binding Ubc9, but with a lower affinity comparing to IR2 binding [82, 83]. The RanBP2 M domain strongly enhances its SUMO E3 activity [84]. Through aligning the E2~SUMO thioester complex in an optimal orientation, RanBP2/Nup358 family ligases enhance the substrate interaction and facilitate the SUMO transfer [85]. Several studies show that RanBP2 promotes the SUMOylation of Sp100, HDAC, hnRNP C1, PML proteins, and a single IR domain is sufficient for binding to both UBC9 and SUMO [85-89].

The SP-RING family is the largest family of SUMO E3 ligases, encompassing seven mammalian PIAS proteins, two homologues identified in yeast (Siz1 and Siz2) and their homologues in other eukaryotes, such as *Drosophila melanogaster* and zebrafish [90-96]. In addition, two SP-RING E3 ligases, AtSiz1 and AtMms21, have been identified in *A. thaliana* [97]. Sequence similarity and evolutionary conservation analysis show that these proteins contain a highly conserved SP-RING domain that is homologous to RING domains in ubiquitin E3 ligases [77]. The SP-RING domain is required to form a complex by interacting with the backside of UBC9 through its UBC9 Binding Region (UBR) and flanking regions of SP-RING domain mediate its interaction with SUMO and specific targets [98]. Then, the flexible UBC9~SUMO thioester is locked in an orientation by SP-RING domain to favor the nucleophilic attack by the target lysine to further complete the SUMOylation process [99].

1.2.4 The protein inhibitor of activated STAT (PIAS) family

The protein inhibitor of activated STAT (PIAS) family was originally identified as specific inhibitors for signal transducer and activator of transcription (STAT), a gene transcription factor [100]. In 1997, PIAS3 was the first of its family to be identified using yeast two-hybrid assays by Chung C. D. et al. while studying the Janus kinase (JAK)/STAT signaling pathway [101]. PIAS1 and two isoforms of PIASx and PIASy were discovered by the same research group in 1998 [102]. Hereafter, an increasing number of reports have demonstrated that PIAS proteins play crucial roles in transcription regulation. They interact with the transcription factors, such as p53, androgen receptor, and p73 top, to form a complex and further activate or repress transcription [90, 103-107]. Since PIAS proteins were found to do far more than act as transcription coregulators; the PIAS acronym was also proposed as **P**leiotropic **I**nteractors **A**ssociated with **S**UMO based on their ability to regulate SUMO proteins as SUMO E3 ligase [63].

The mammalian PIAS family is composed of 7 proteins that are encoded by four genes: *PIAS1*, *PIAS2* (*PIASx*), *PIAS3*, and *PIAS4* (*PIASy*) [108]. Sequence alignment analysis indicates that approximately 40% of nucleotide sequences are conserved across all PIAS genes. With the exception of PIAS1, two isoforms are encoded by *PIAS2-4* genes respectively, due to alternative splicing [90-92]. Figure 1-4 shows the schematic structure of human PIAS proteins along with their homologues from *D. melanogaster*, *C. elegans,* and *S. cerevisiae*.
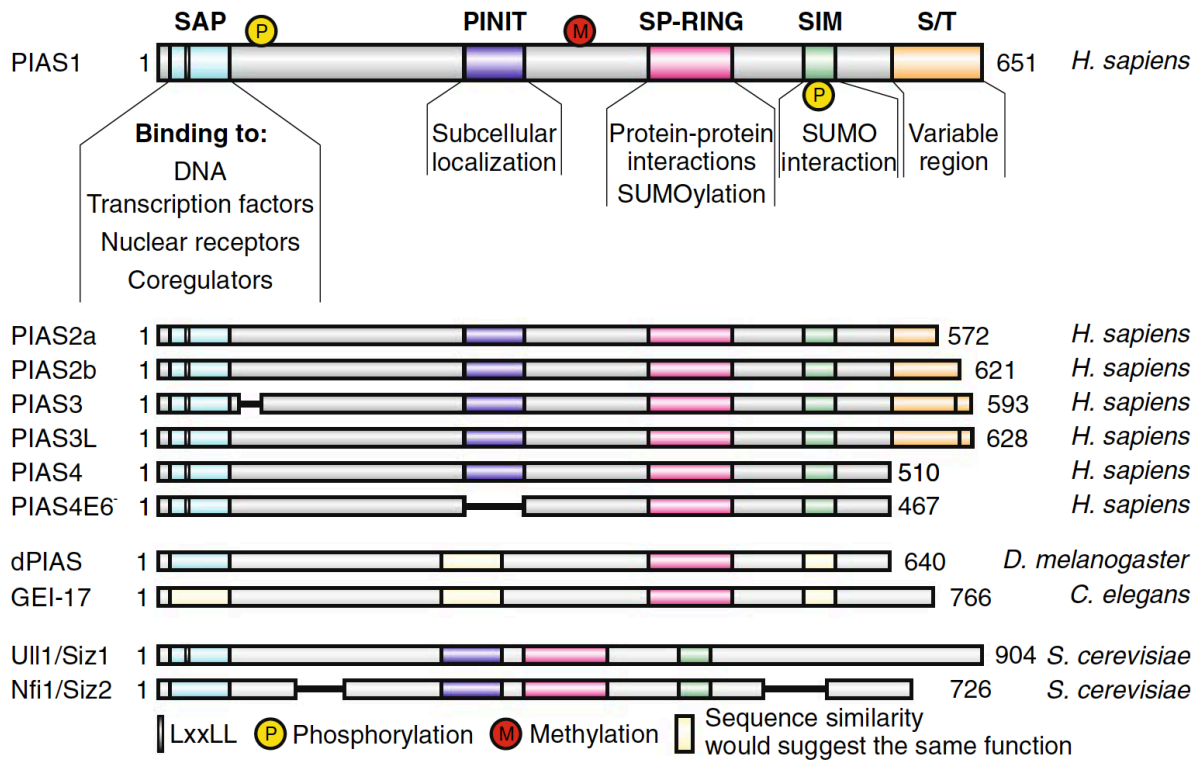
Figure 1-4 Schematic structures of PIAS proteins.

Identified conserved domain and motif along with PTMs on different human PIAS proteins and the orthologous PIAS proteins from *S. cerevisiae, C. elegans*, and *D. melanogaster* are illustrated. Adapted with permission from Springer Nature: Cellular and Molecular Life Sciences (Rytinki M. et al.) [63] © 2009.

Across the species listed in Figure 1-4, PIAS homologues share strong sequence homology and contain a relatively similar number of amino acid residues, ranging from 510 (PIAS4) to 651 (PIAS1) in humans, and from 640 (dPIAS) to 766 (GEI-17) in invertebrates [63]. Overall, three different functional domains and two motifs have been identified. From the N terminus, there are scaffold attachment factor-A/B, acinus and PIAS (SAP) domain, the Pro-Ile-Asn-Ile-Thr (PINIT) motif, the RING-finger-like zinc-binding domain (RLD), the SUMO-interacting motif (SIM), and the serine/threonine-rich region (S/T) at the C-terminus [63, 68, 109-111]. The SAP domain and the central RLD show a high degree of conservation through evolution and were found in most of the species, whereas the S/T-rich region at C-terminus is a poorly conserved region and is only present in human PIAS1, 2, and 3. Interestingly, among all human PIAS proteins, only PIAS4 lacks the S/T region, and the PINIT motif is also missing in the isoform PIAS4E6⁻.

The SAP domain is a motif comprising 35 amino acids and found in a variety of nuclear proteins involved in RNA processing, gene transcription, or DNA repair [112]. It has also been proposed as a DNA-binding motif due to the essential role in the structure-specific DNA binding activity in the process of chromosome organization [112]. NMR structural analysis of the PIAS1 N-terminal domain revealed the interaction between a unique four-helix bundle and A-T-rich DNA, as well proteins, such as p53 [113]. There is a LxxLL motif found in all PIAS proteins, which is located in the SAP domain. It has been reported that the LxxLL motif is involved in the complex assembly of nuclear receptors and their co-activators [114, 115]. The C-terminus of the SAP domain contains a highly conserved PINIT motif that is found in all PIAS proteins with the exception of PIAS4E6[-] [106]. In mouse embryonic stem cells, the PINIT motif was reported to be necessary for the retention of PIAS3L at the nucleus and its deletion affected PIAS3L subcellular localization [111]. In addition, the PINIT domain is also found to play an important role in dictating PIAS SUMO E3 ligase substrate specificity [77].

In the center of PIAS proteins, a cysteine-rich region forms a putative RING finger-type motif, termed the Siz/PIAS RING (SP-RING) domain. The RING domain found in all known ubiquitin E3 ligases contains a cysteine/histidine-rich sequence for zinc-binding [116]. These residues create a globular domain through an interleaved coordination of zinc ions, which is involved in the mediation of protein-protein interactions (PPIs) [117]. Although the SP-RING domain lacks two conserved cysteines, it is still found to be structurally similar to the classical E3 RING domain [68]. In yeast, structure-based mutational analysis and biochemical studies showed that the SP-RING domain interacts with the E2 conjugase UBC9 to form a SUMO E2:E3 complex, while the PINIT domain acts as a molecular scaffold to coordinate substrates [77]. Thereby the transfer of SUMO is facilitated from UBC9 to the lysine residues in both consensus and non-consensus regions of the substrates [99].

Between the SP-RING domain and C-terminal S/T-rich region, there is a conserved SUMO-interacting motif (SIM) in all PIAS proteins. The SIM is composed of hydrophobic and acidic amino acids which form a β-sheet that can bind to SUMO in a parallel or antiparallel fashion [118, 119]. The first SIM (hhXSXS/Taaa), where h stands for a hydrophobic amino acid and a is an acidic

amino acid, was first identified in a variety of proteins, such as PM-Scl75, PKY, and CHD3/ZFH in a yeast two-hybrid assay [109]. It is shown in the two-hybrid system that SIM interacts noncovalently with SUMO1, and the interaction can be abolished by its depletion. Later studies established more variable and less stringent consensus sequences such as ψxψψ or ψψxψ (where ψ is often V or I). To date, all SIMs are found to contain a hydrophobic core sequence that is surrounded by either glutamate or aspartate residues or phosphorylated serine or threonine residues. Several negatively charged amino acids have been found in the PIAS-SIM motif on the C-terminal side of the core region. These sites have shown the potential of being phosphorylated by the kinase CK2. The phosphorylation further enhances the interaction between the SIM and SUMO proteins [93, 120]. In addition to the main SIM described above, a recent study has shown there is a second SIM at the C-terminus of PIAS1-3, but not PIAS4 [121]. This new PIAS-SIM was shown to play a crucial role for UBC9-PIAS-SUMO1 complex formation.

PIAS proteins are overexpressed in various types of cancer and are frequently associated with tumorigenesis [69]. For instance, in prostate cancer, the overexpressed PIAS1 promotes cell proliferation through suppression of p21 [122, 123]. PIAS1 was found to regulate tumour initiating stem cells through modulating the epigenetic status of several clinically relevant genes, including cyclin D2 (*CCND2*), estrogen receptor (*ESR1*), and breast tumor suppressor WNT5A (*WNT5A*) in breast cancer cell line MDA-MB-231 [124]. In multiple myeloma, microRNA-21 activates the STAT3-dependent signal pathway through the repression of PIAS3 function, and down-regulation of PIAS3 contributes to the oncogenic function of microRNA-21 [125]. Several studies highlighted the relationship between hypoxic signaling and PIAS4 overexpression with respect to different cancers such as ovarian and pancreatic, implicating PIAS4 as part of the shift upon hypoxic signaling [126, 127].

As SUMO E3 ligases, PIAS1, PIAS2 and PIAS4 were reported to regulate the activity of the tumor suppressor p53 through SUMOylation [128]. The PIAS4 mediated SUMOylation of p53 enhances the transcriptional activity of wild-type (WT) p53, further promoting cellular senescence [128]. One clinical study reported that the 6-year survival rate after transplantation of multiple myeloma patients shows a negative correlation with PIAS1 overexpression [129]. The

transcription factor Myc has been reported to be associated with several cancers [130]. In B cell lymphoma, PIAS1-mediated Myc SUMOylation leads to its longer half-life and thus increases its oncogenic activity [131, 132]. In acute promyelocytic leukemia (APL), PIAS1-mediated promyelocytic leukemia (PML) SUMOylation regulates oncogenic signaling and further produces the fusion product with the retinoic acid receptor alpha (PML-RARα) [133]. In lung cancer, PIAS1 was reported to facilitate Focal adhesion kinase (FAK) autophosphorylation on tyrosine 397 via SUMOylation, which increased its kinase activity and promoted cancer survival and progression [134].

PIAS proteins also regulate cell migration and epithelial-mesenchymal transition (EMT) in various types of cancers through PIAS-depend SUMOylation. TGFβ is the crucial regulator of EMT in many cancer developments, including breast cancer and colorectal cancer [135]. PIAS1 was reported to mediate TGFβ SUMOylation and inhibit its activity, which resulted in the positive regulation of EMT and metastasis [136]. PIAS4 was reported to enhance epithelial cell migration through SUMOylation of C/EBPδ, leading to sequestration at the nuclear periphery and subsequent reduction of C/EBPδ transcriptional activity [137]. PIAS3-mediated GTP-Rac1 SUMOylation prolongs the Rac1-GTP activated state and further stimulates cell migration [138]. In MDA-MB-231 triple-negative breast cancer cells, PIAS1 and/or PIAS4 mediates PYK2 SUMOylation, subsequently triggering autophosphorylation, interaction with tyrosine kinase SRC, phosphorylation of paxillin, activation of ERK1/2, and promotion of cell migration [139]. These individual studies have shed light on the diverse roles that the PIAS family proteins have in cancer.

1.3 Mass spectrometry-based proteomic analysis

Proteomics is a systematic and experimental analysis of the entire set of proteins produced and modified within a defined biological system to determine their identity, quantity, and function [140, 141]. With the development of MS, especially for its advancement of high throughput, mass accuracy, and sensitivity, using LC-MS for proteomic analysis becomes the first choice. MS-based proteomic analysis is widely applied in the field of biology to answer questions related to biomarker identification, protein expression and isoforms, protein turnover, and PPIs in a large-scale fashion [142-145]. It can also be used to monitor the global proteome regulation following a change in gene expression or to study temporal dynamic changes following cell stimulation [146-148].
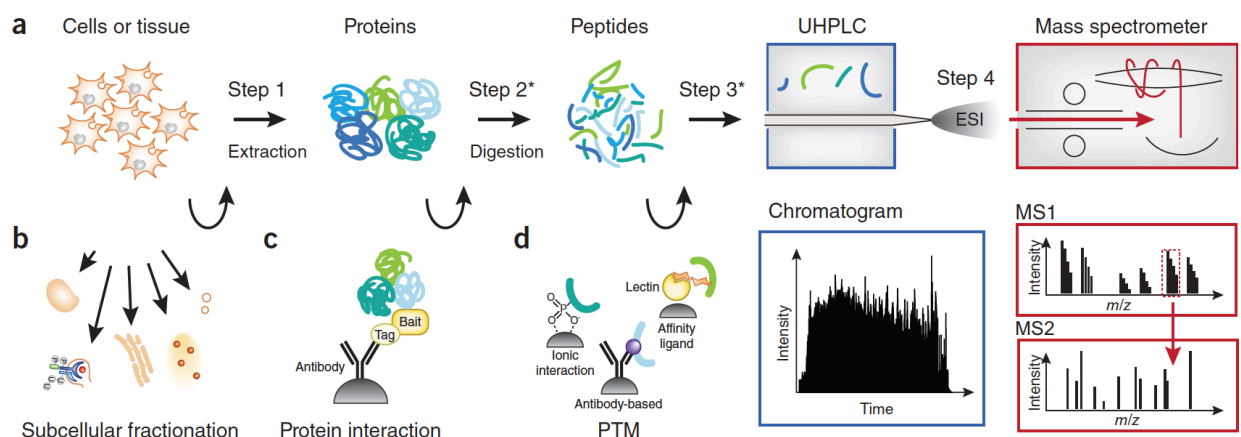
Figure 1-5 Shotgun proteomics workflow.

Adapted with permission from Springer Nature: Nature Immunology (Meissner F. et al.) [149] © 2014.

In shotgun proteomics, the generic workflow consists of four steps, as shown in Figure 1-5, a. In step 1, protein is extracted from a biological sample, such as cells or tissue. In step 2, proteins are digested into peptides using proteases. In step 3, peptides are separated by HPLC. In step 4, peptides are ionized and analyzed in a mass spectrometer. Based on different study objectives, a few other steps can be added to the workflow at steps 2 and 3 (*). As shown in Figure 1-5 (b–d), the add-on steps include subcellular fractionation, protein purification for interacting

proteins, immunoisolation for peptides with PTMs. Additionally, fractionation of proteins or peptides is optional.

1.3.1 Cell culture and treatment

As stated by a famous proverb: A good start is half the battle. Sample preparation is the first and most important step in the entire experimental process. Samples for proteomic analysis by MS can come from a variety of sources, such as different types of tissue or cultured cells that have been stimulated in various ways, for example, through drug treatment or salt stress [150-154]. In this thesis, I used human HEK293 SUMO3m cell line [155] to investigate the global changes in protein SUMOylation upon varying expression of SUMO E3 ligase gene using molecular biology techniques. In order to quantify the changes in protein SUMOylation, a stable isotope labeling of amino acids in cell culture (SILAC) based quantitation method was applied [156]. SILAC is known for its high performance in the identification and relative quantitation of differential changes from a complex sample [157]. The SILAC approach relies on the essential amino acids for mammals, such as arginine and lysine, and entails the *in vivo* metabolic incorporation of isotopically-labeled amino acids into proteins [158]. The more detailed principle and applications of SILAC-based quantitative proteomics will be explained in the quantification section below.

At the cell culture stage, the gene expression levels of a certain SUMO E3 ligase could be regulated through molecular biology techniques such as siRNA knockdown or gene overexpression [159, 160]. For the study of SUMO proteomics, one of the most commonly used treatments is proteosome inhibitor MG132 [161]. SUMO has a cooperative and complementary role to ubiquitin and is able to balance the ubiquitin-proteasome system for protein degradation [162, 163]. Thus, treating cells with MG132 could significantly prevent protein degradation and accumulate SUMO conjugates on the newly synthesized proteins [164]. MG132 helps lock the highly dynamic and transit SUMOylation on the lysine and allow mass spectrometry to identify the corresponding sites [155]. The cells can be collected through centrifugation, but for adherent cells, it is often required that they be dissociated with trypsin or scratched off from Petri dishes.

The collected cells are subsequently washed with cold isotonic buffers, such as phosphate buffered saline (PBS), to remove any extracellular proteins, dead cells, remaining media or trypsin that could contaminate the MS-based analysis. Here, it should be noted that for SILAC based experiments, an equal amount of cells from different SILAC channels are preferred to mix at the cell level to minimize errors at the different sample preparation steps.

Due to the dynamic nature of proteins and the complexity of cellular biochemistry, special care is required throughout the cell collection process. Biological material must be kept at low temperatures to minimize extraneous biochemical reactions. Furthermore, the duration of sample preparation should be minimal to avoid peptide degradation through hydrolysis [165]. After harvesting cells or extracting a tissue of interest from an animal model, these samples can be stored in -80 °C for future experiments.

1.3.2 Protein extraction and fractionation

Shotgun proteomics analysis is mostly conducted using total protein extract from harvested cells, tissues, or organs [166]. The three primary technical challenges are how to break down cell membranes, extract a maximum amount of protein in a relevant buffer, and inhibit the degradation caused by related proteases [167]. The most common methods used for protein extraction from cells are physical lysis, such as snap freezing using excess of cold (−80 °C) EtOH [168], homogenizer [169], or beadbeating [170]. The advantage of these methods is that no chemicals are introduced that might interfere with subsequent steps or MS analysis. However, some of the mechanical methods are time-consuming and not satisfactory to extract all proteins, especially membrane proteins. It could also destroy the molecule of interest or the weak interaction between it. In this case, it is best to use adequate buffers that weaken the cell membrane, such as hypertonic buffer [171], enzymatic digestion [172] or denaturing buffer [173]. Of note, proteasome inhibitor is always highly recommended to add in the lysis buffer prior to protein extraction. For PTMs analysis, the corresponding inhibitors, such as phosphatase inhibitor, deacetylase inhibitor, or deubiquitinating enzyme inhibitor are also required. Detergents such as SDS, NP-40, or Triton X-100 are very important ingredients of lysis buffer. The

use of a detergent could improve the protein solubilization, such as membrane proteins and hydrophobic proteins, and inactivate the protein functions. However, detergents are considered one of the biggest interferences for MS analysis since they will disrupt peptide ionization and must be removed for sensitive MS detection of peptides. A cell fractionation step is often required for analyzing organelle proteome or low abundant proteins. Membranes present different resistances to certain stress conditions, so it is also possible to perform successive lyses to fractionate proteins according to their cell compartment of origin. A fractionation often used for human cells consists of lysing the cytoplasmic membranes with a hypotonic buffer and then nuclear membranes with a hypertonic buffer [132].

Once the proteins have been extracted, it is common to fractionate the sample in order to identify proteins of interest. Sodium dodecyl sulphate–polyacrylamide gel electrophoresis (SDS-PAGE) is a commonly used method for separating proteins in the laboratory [174]. SDS is a strong protein-detergent, which covers the intrinsic charge of the proteins and provides them "charge-to-mass ratios" [175]. In the presence of SDS and a reducing agent, disulfide bonds are cleaved, and proteins are linearized with a negative charge. Under the influence of an applied electrical current, the proteins migrate through the gel at different rates and are thus separated based on their molecular weight (MW). This procedure allows researchers to separate the protein based on their mass and remove detergents in a simple and relatively accurate manner.

Protein fractionation can also be conducted in a solution using biochemical purification processes. Affinity chromatography is an in-solution separation method to purify biomolecules of interests from a mixture. This method is based on a highly specific binding interaction between an analyte of interest and another substrate, such as a binding partner or ligand. The Nickel-nitrilotriacetic acid (Ni-NTA) purification system [176] and Immunoprecipitation (IP) represent two commonly used purification procedures to purify proteins of interest under either denaturing conditions or native condition of the lysate.

The Ni-NTA purification system is designed to purify the fusion proteins, which contain an affinity tag of polyhistidine residues (6xHis) [177]. The chelating ligand nitrilotriacetic acid (NTA) with four pre-charged Ni^{2+} ions is coupled to a cross-linked agarose resin. These beads provide a

strong and specific interaction between the Ni^{2+} ion and consecutive histidine residues. When Ni-NTA beads are incubated with cell lysate, 6xHis-tagged proteins are efficiently retained on the beads, and the nonspecific proteins can subsequently be removed with beads washes. Protein purification using the Ni-NTA occurs independently of the 3D structure of the 6xHis tag or protein. Thus, the His-tagged protein can be purified under either native or denaturing conditions through a simple step and further be eluted from the beads using an imidazole buffer.

Immunoprecipitation (IP) is a technique of precipitating proteins of interest from cell lysate based on the principle of specific antibody-antigen interaction. In an IP experiment, the antibody is pre-coupled to agarose or magnetic beads and is preferentially immobilized on the beads to prevent the co-elution of the antibody with target proteins. Subsequently, the antibody-beads are incubated with a cell lysate containing the target proteins. During incubation, the target proteins are gradually captured by antibody and thus form an antibody-protein complex that precipitates with the beads. One should remember that immunoprecipitation relies highly on the functional antigen-binding sites; thus the native condition is necessary for this protein purification technique. The proteins purified through IP can be eluted using an acidic solution.

1.3.3 Protein digestion and desalting

Once the protein samples are extracted or purified, protein quantification is very crucial for the following steps. It determines the amount of proteolytic enzymes used, the antibody used for the immunoisolation at the peptide level, and label-free quantitation (LFQ) between different conditions, as well as the peptide amount loaded on the analytical column of the liquid chromatography (LC) system. In order to standardize all the protein quantitation procedures, the commercially available Bradford protein assay kit was used throughout the thesis.

Shotgun proteomics analysis relies on LC-MS to characterizes protein mixture by analyzing the peptides generated from the protein after proteolytic digestion [178]. The enzymes used to cleave the protein into peptides are proteases [179]. Trypsin is the most popularly used protease for protein digestion in shotgun proteomics [176]. It cleaves the peptide bond at the carboxyl end of the amino acids arginine and lysine with an exceptional cleavage specificity. After tryptic

digestion, the resulting peptides with positively charged arginine and lysine at their C-terminus are relatively short, on average 5 to 30 amino acids, making them ideal for MS analysis [180, 181]. Under basic conditions, trypsin has a high proteolytic activity with a relatively lower auto-proteolysis when it is chemically methylated. It is also very stable under a wide variety of conditions. Of note, as SUMOylation occurs on lysine residues, it is more common to have one more missed cleavage site on digested SUMO peptides than the unmodified tryptic peptides. Following trypsin, Lys-C, chymotrypsin, Glu-C, and pepsin are also commonly used enzymes in the proteomics laboratory depending on different study objectives [182]. For this thesis, trypsin allows the identification of most proteins and PTM sites within the proteome and was thus used to profile the SUMO proteome of cells upon different PIAS gene knockouts. Proteolytic digestion using trypsin is sensitive to the high concentration of guanidine and urea, so I dilute the protein lysate into a compatible concentration of salts that trypsin tolerates.

Peptide desalting before peptide level immunoisolation and analysis by LC-MS is quite important [183]. The salt may disturb the immunoaffinity of antibody binding and change the pH. When samples contain high salt concentrations, it may also disturb the ionization process in MS and potentially even block the LC flows. Off-Line reversed-phase-trapping (desalting) cartridges are often used for peptide mixture desalting. The Oasis HLB column is a commercially available cartridge that contains strongly hydrophilic and reversed-phase polymers. It has an excellent capacity with a high performance of peptide retention. Briefly, desalting steps comprise cartridge conditioning with an organic solvent such as acetonitrile (ACN) and methanol, equilibration with acidified water (pH ~2), then application of peptides on the solid phase followed with the washes to remove any hydrophilic compounds, and finally peptide elution with aqueous solutions containing up to 70% organic solvents as the conditioning step. When the peptides are all eluted from the cartridges, peptide lyophilization is a mandatory step, using vacuum systems to remove the solvents subjected to any further sample preparation or LC-MS analysis.

1.3.4 Immunoisolation of SUMO peptides

Over the past few decades, it has been a great challenge to identify endogenous SUMOylation sites by mass spectrometry on the large-scale proteome level. Due to the nature of SUMO modifiers, two major issues make it incredibly challenging to generate SUMO site information. First, unlike some PTMs, such as phosphorylation and ubiquitinoylation, SUMO is less abundant in the cell, and that leads to much lower levels of protein SUMOylation [184]. Therefore, additional enrichment at the SUMO peptide level is crucial to precisely target and obtain the SUMO site information. However, the sequences of mammalian SUMO isoforms lead to relatively long side chain remnants upon tryptic digestion, which leads to the second issue. As shown in Figure 1-6, these remnant peptides are 19 amino acids or 32-34 amino acids in length for SUMO1 and for SUMO2/3, respectively. These extremely large amino acid remnants result in complex MS/MS fragmentation, which make it increasingly challenging for bioinformatics tools to identify the corresponding peptide sequences. In Chapter two, I discussed a few popular strategies which were developed by different groups for the immnoisolation of site-specific SUMO proteomics.

HEK293 SUMO3m cell line used in this thesis was created based on the human HEK293 cell line. It expresses SUMO3m protein containing 6xHis-tag on the N-terminus for Ni-NTA purification. In addition, it also contains an Asn at the fifth residue and an Arg residue at the sixth residue from the C terminus. These two mutations result in a 5 amino acid peptide remnant left on the lysine residue of the target protein after tryptic digestion and facilitate SUMO peptide identification using remnant immunoaffinity purification. Briefly, the antibodies used in immunoisolation, specifically recognizing the NQTGG remnant are first coupled on the magnetic protein A beads, followed by an antibody-beads crosslink. Then the tryptic digests are incubated with antibody beads for an hour at 4°C to allow SUMO peptides to be immunoisolated. After incubation, the beads are washed using cold PBS buffer to remove the nonspecific bindings, and the SUMO peptides are eluted using a 0.2% FA solution.
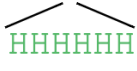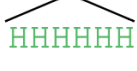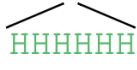
SMT3 (yeast)	N-ter... DSL**R**FLYDGI**R**IQADQTPEDLDMEDNDIIEAH**R**EQIGG[98]
SUMO1 (human)	N-ter... NSL**R**FLFEGQ**R**IADNHTP**K**ELGMEEEDVIEVYQEQTGG[97]
SUMO1 (mutant)	HHHHHH ↑ R
SUMO2 (human)	N-ter... RQI**R**F**R**FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG[93]
SUMO2 (mutant)	HHHHHH ↑ R
SUMO3 (human)	N-ter... RQI**R**F**R**FDGQPINETDTPAQLEMEDEDTIDVFQQQTGG[92]
SUMO3 (mutant)	HHHHHH ↑↑ RN

Figure 1-6 The sequence of SMT3 (*S. cerevisiae*) and three human SUMO paralogues.

The stable cell lines were generated using vectors expressing mutant SUMO (SUMOm) proteins. Each SUMOm contains an N terminal 6xHis and an Arg mutation at the sixth position from the C terminus. To distinguish SUMO3 mutant from SUMO2 mutant paralogue, the Q88N mutation was induced. Adapted with permission from Springer Nature: Nature Communications (Lamoliatte F. et al.) [155] © 2014

1.3.5 Offline strong cation exchange peptide fractionation

With growing interest in the proteomic study of PTMs, an extended proteome coverage is a key to ensures the success of confident identification and quantification by LC-MS. Many challenges remain in terms of how to further facilitate the detection of low-abundance peptides. For shotgun proteomics, where complex protein samples are digested, and peptides are resolved using different approaches prior to LC-MS analysis, more efforts need to be made to reduce sample complexity, although NiNTA purification and peptide immunoisolation have already been performed. Peptides contain different functional groups conferring them various charge states, polarity, hydrophobicity, and separation protocols can be devised according to these properties. The introduction of peptide separation techniques based on physical properties becomes extremely valuable for increasing the identification of low abundance peptides from highly complex peptide mixtures [185]. Accordingly, offline fractionation using ion-exchange chromatography or reversed-phase chromatography is very handy and improves proteome coverage. Since I am using a reversed-phase analytical column on the LC system, which will be introduced in the LC section, an offline strong cation exchange strategy was used in this thesis to

enhance the separation capacity of peptides prior to LC-MS analysis. SCX disks contain negative sulfonyl groups, which attract positively charged peptides under acidic pH conditions. Typically, the fractionation columns are made using pipette tips packed with SCX disks. After column equilibration with ACN, peptides dissolved in 0.2% FA solution are applied and bound to the SCX disks. Next, elution buffers are applied and then flushed through the column using centrifugation. Through stepwise elution with increasing salt concentration, separate fractions are generated, and peptides with the least positive net charge are resolved first, followed by neutral and then more positively charged peptides. Finally, individual fractions are collected and lyophilized using a SpeedVac Vacuum Concentrators.

1.3.6 RP-HPLC separation

When peptide samples are ready to be analyzed by MS, it is most common to use an autosampler coupled with LC to automatically perform all steps in a row [186]. In this thesis, the online separation technique used for all the peptide samples was reverse-phase high-performance liquid chromatography (RP-HPLC) . The technique uses a hydrophobic stationary phase to immobilize all molecules based on their hydrophobicity and uses mobile phases, a binary mixture of extremely volatile solvents, to separate and elute the molecules [187]. Briefly, the peptides are dissolved in low pH acidic solutions, such as 0.2% FA or 0.1% TFA to promote the ionization process before MS analysis. Subsequently, peptide samples are picked up and injected into the RP analytical column by the autosampler. Through increasing concentration of organic solvent in the mobile phase, for instance, acetonitrile (ACN) or methanol (MeOH), the peptides are gradually eluted and sprayed in the ion source. Depending on the sample complexity, the gradient time for peptide elution may vary from 70min to 250min for a shotgun proteomics analysis. The most commonly used materials for the RP analytical column are spherical silica particles, chemically modified with hydrophobic alkyl chains. The long *n*-octyldecyl (C18) hydrocarbon chain is most efficient for the analysis of standard tryptic peptides [188]. Thus an RP analytical column 15 to 20 cm in length and packed with these capillaries is highly recommended. In addition, shorter hydrocarbon chains, such as C4 or C8, labeled on a larger

particle with pore sizes (300 Å) is very suitable for the analysis of long peptides or intact proteins [188].

1.3.7 Mass spectrometry

Mass spectrometry (MS) is an analytical technique that is used to characterize the molecules by determining their m/z [189]. This is achieved by ionizing molecules and measuring their mass-to-charge ratio (m/z) [190]. In addition, MS can be used to identify unknown components and quantify known components present in a sample [191]. A recent study showed that crosslinking mass spectrometry (XL-MS) enabled the structural characterization of protein complexes and ligands bound to their target proteins [192]. All mass spectrometers consist of three major components: an ionization source, a mass analyzer, and an ion detection system [193]. In typical MS analysis, peptide precursors ions are first produced in the ionization chamber, then transferred to the mass analyzer and corresponding ions are sorted and separated based on their m/z ratio [194]. Modern MS instruments often comprise more than one analyzer in order to facilitate structural characterization of selected ions. In shotgun proteomics, to obtain the peptide sequence, tandem mass spectrometry (MS/MS) is required. MS/MS is an analysis technique where multiple mass analyzers are coupled together. Peptide precursor ions are selected by the first mass analyzer then dissociated in a collision cell, and the fragment ions are separated by a second mass analyzer. All fragment ions are transmitted to the ion detection system to measure their intensities. Finally, these MS-acquired data are transformed into spectra and further processed by bioinformatic tools to deduce the peptide sequence.
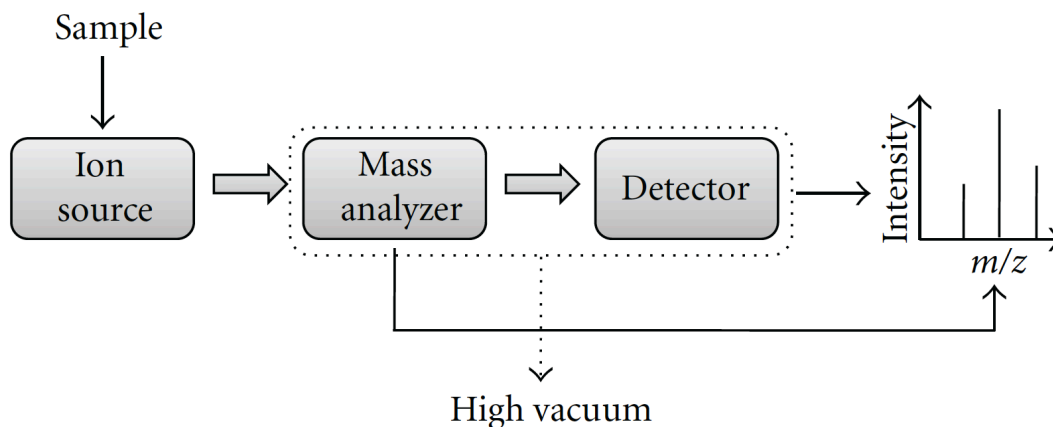
Figure 1-7 The basic components of the mass spectrometer.

Adapted with permission from Hindawi: International journal of Analytical Chemistry (Banerjee S. et al.) [193] © 2012

1.3.7.1 Sample ionization

As biological samples consist primarily of neutral molecules, these molecules need to be charged either negatively or positively prior to analysis by the MS analyzer [195]. The formation of ions is achieved in the ionization source, and different ionization techniques have been developed to accommodate a wide range of analytes. Based on the sample properties, such as inherent polarity, stability, and size, matrix-assisted laser desorption/ionization (MALDI) [118] and electrospray ionization (ESI) [196] were developed and frequently used for MS analysis. These two methods allow the ionization of the analyte with minimal fragmentation, also called "soft ionization", and are thus the most popular ionization techniques for the analysis of biological samples.

The matrix-assisted laser desorption/ionization (MALDI) technique creates ions from large molecules using laser energy absorbing matrix [197]. This technique was successfully used in laser desorption ionization of a large molecule in 1987 by Japanese researcher Koichi Tanaka for the first time [198]. MALDI enables the ionization of biomolecules as large as hundreds of thousands of Daltons when co-crystallized with the matrix. It also provides high accuracy and tolerance to salts or other impurities in the samples, thus is usefully applied in microbiological diagnostics,

histological imaging of clinical samples, epidemiologic studies, and taxonomical classification [199, 200]. However, MALDI results in a pulsed ionization, which is not suitable for on-line LC-MS/MS analyses [201].
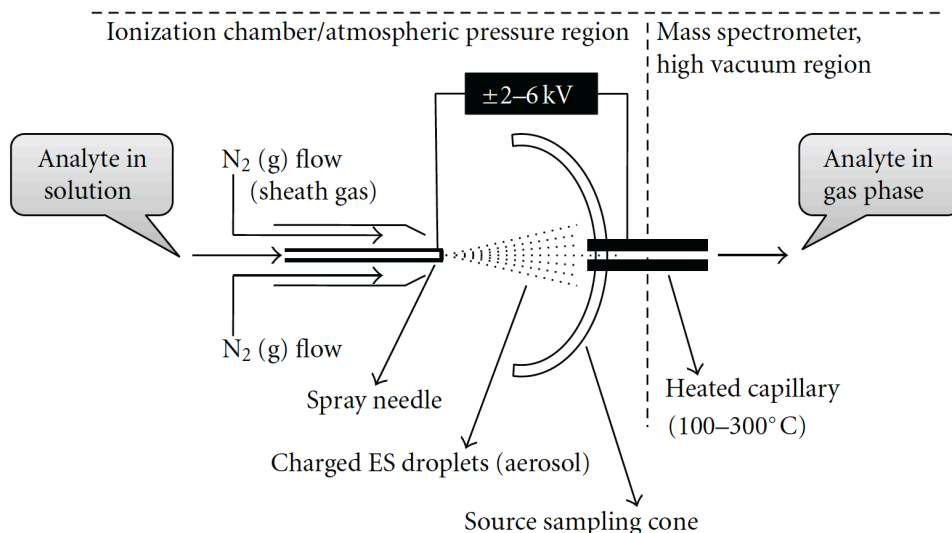
Figure 1-8 A schematic representation of the ESI-ion source.

Adapted with permission from Hindawi: International journal of Analytical Chemistry (Banerjee S. et al.) [193] © 2012

Alternatively, electrospray ionization (ESI) is popularly used in shotgun proteomics. It was first invented by John Fenn in the mid-1980s [202]. Developments in recent decades have made ESI the first choice in the shotgun proteomics analysis of biological samples [203]. In ESI, the sample is introduced through the tip of the spray needle into the source in the liquid state. Then a high "voltage offset" of 2-6 kV is applied between the tip and the inlet of the MS, which generates an aerosol of highly charged droplets from the liquid sample [204]. With the evaporation of the solvent, the preformed ions experience increasing electrostatic repulsion that eventually overcome the surface tension and favor droplet fission events ultimately leading to unsolvated gas phase ions. Ions formed by ESI are typically multi-charged following the addition or abstraction of protons from the neutral analytes. Occasionally, a concentric flow of sheath gas, such as dry N_2 is applied to facilitate the nebulization process especially for sprays introduced at flow rate >5 μL/min. Afterward, the charged analytes pass through the orifice of a heated capillary, which is typically kept at 100–300°C under vacuum. Finally, they enter into the mass

analyzer, which is maintained under high vacuum. For the contribution of soft ionization methods in MS analysis, John Fenn won one-half of the Nobel Prize in Chemistry jointly with Koichi Tanaka, the inventor of MALDI, in 2002 [205]. In this thesis, electrospray ionization was used to facilitate the ionization of peptide ions separated by LC in shotgun proteomics analysis. The applied spray voltage on the outer surface of the borosilicate glass spray capillary is also lower to 0.7-1.1 kV.

1.3.7.2 Mass analysis

Once peptides are ionized, the peptide ions are transferred into the heart of MS, a mass analyzer, to be further sorted according to their m/z. Currently, there are several types of mass analyzers available on the market. The most popular ones used in proteomic analysis include quadrupole (Q) mass filter, time-of-flight (TOF) analyzer, linear ion trap (LIT), orbital trap (Orbitrap), and Fourier transform–ion cyclotron resonance (FT-ICR) [206, 207]. Each of these analyzers has benefits and trade-offs relating to the performance based on acquisition speed reported as a frequency in *Hz*, dynamic range, and mass range [208, 209]. Among them, acquisition speed assesses the ability of the analyzer to acquire a certain number of spectra in a given time [210], while the dynamic range and mass range stand for the scan abilities of an MS for detectable signal and detectable m/z.

Additionally, mass resolving power (R) and mass accuracy are two important parameters for evaluating the characteristics of a mass analyzer. Here, R is the ability for a mass analyzer to separate two peaks and is defined by the following equation:

$$R = \frac{m}{\varDelta m}$$

Equation 1-1 mass resolving power calculation.

In the equation, *m* stands for the mass number of the peak of interest, *Δm* represents the mass difference between two adjacent peaks (peak width at a half-height is commonly used). Hence, the higher the R value, the better the ability of the mass analyzer to separate two adjacent peaks at a given mass.

Mass accuracy describes the difference between the observed and the theoretical mass of the analyte. It can be defined by the following equation:

$$\text{Mass accuracy} = \frac{Mm - Mt}{Mt}$$

Equation 1-2 Mass accuracy calculation.

In the equation, *Mm* is the mass of the analyte determined from the MS spectrum, and *Mt* is the theoretical value. The Mass accuracy usually express in expressed in Dalton or part per million (ppm).

In this thesis, all the samples were analyzed on two different Orbitrap-based mass spectrometers, the Q Exactive™ Hybrid Quadrupole-Orbitrap™ Mass Spectrometer or the Orbitrap Fusion™ Tribrid™ Mass Spectrometer.
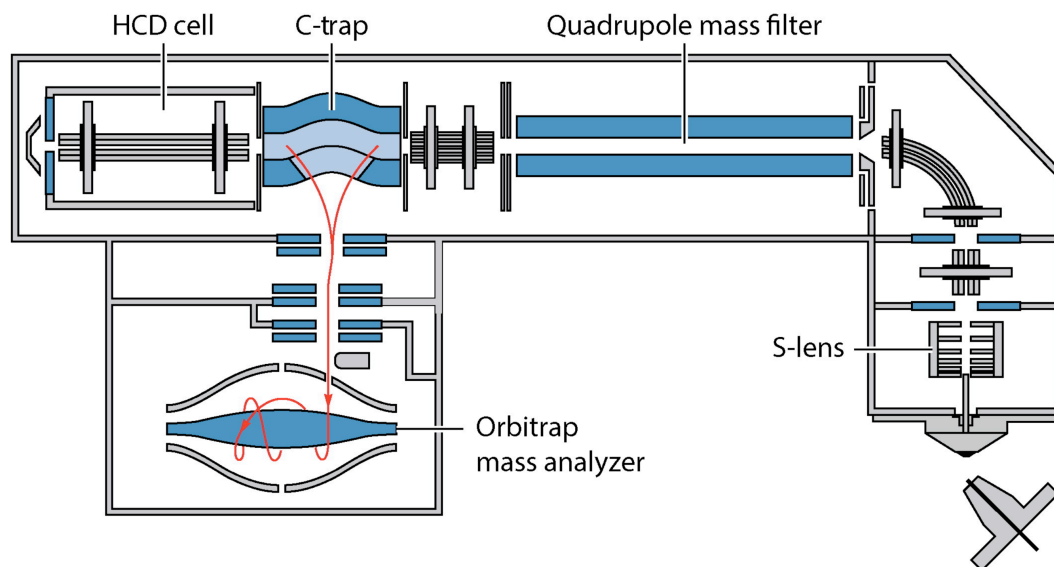
Figure 1-9 Schematic of the Q Exactive mass spectrometer.

Adapted with permission from Annual Reviews: Annual review of analytical chemistry (Eliuk S. et al.) [211] © 2015.

The Orbitrap Q Exactive mass spectrometer is a hybrid mass spectrometer that incorporates quadrupole and Orbitrap into one instrument as shown in Figure 1-9. Ionized molecules first

enter the S-lens, where they can be efficiently captured, focused, and transmitted to increase the sensitivity. Then all ions are passed through Active Beam Guide (ABG), where neutrals are prevented from entering the quadrupole, thus reducing the noise and improving operational robustness. Next, the ion precursor enters the first mass analyzer Quadrupole, which consists of two pairs of hyperbolic or circular electrode rods with opposite polarities. When both rod pairs are connected electrically, and a combination of alternating radio frequency (RF) voltage and direct current (DC) voltage are applied, an oscillating electric field in two dimensions (x and y) is created to separate ions based on their m/z [212]. Thus the quadrupole is considered as a low-resolution mass analyzer. However, it is known for its fast scan speed with minimal loss of ions, which facilitates the selection of specific ions or the scanning of m/z ranges without any delay or dead times. While the quadrupole mass filter can scan up to m/z 4000, in practice a full scan range that covers 1000 m/z units is sufficient for most shotgun proteomic applications. The quadrupole mass filter can also transmit selected m/z values for targeted analyses, and this mode of operation will be discussed in more detail in the next section.

In a Q-Exactive instrument, the filtered ions are collected in a storage device, called the C-trap, which provides an efficient accumulation of ions for subsequent analysis by the Orbitrap mass analyzer. The Orbitrap mass analyzer was first introduced by Makarov in 2000, based on the principle of orbital trapping described by Kingdon [213]. It became commercially available in 2005 after a significant improvement in the device performance. This analyzer consists of an outer electrode (barrel shape) and an central electrode (spindle shape), which enable its dual roles: analyzer and detector [214]. When a purely DC potential is applied, a logarithmic electrostatic field is generated between two electrodes that allows ion separation based on the motion of ions. After entering the Orbitrap, all ions are captured through "electrodynamic squeezing", then harmonically rotate around the central electrode at different frequencies based on the characteristics of their m/z values. Meanwhile, different ions also oscillate left and right along the z-axis of the electrode. These oscillation frequencies induced by the ion motions result in an image current on the outer electrode and are further measured and transformed into a mass spectrum. The Orbitrap provides high resolution up to 500000 at m/z 200 and scan rates of

up to 15 *Hz*. Additionally, a sub-1 ppm mass accuracy is routinely achieved using the Orbitrap for the analysis of peptide mixtures [211].

For MS/MS analysis in Q-Exactive instrument, the precursor ions of interest preselected by the quadrupole mass filter are injected into a higher energy collision induced dissociation (HCD) fragmentation-compatible ion routing multipole to be dissociated into fragment ions. Subsequently, these fragment ions are collected into ion packets, transferred to the C-trap for accumulation and finally to the Orbitrap for detection.
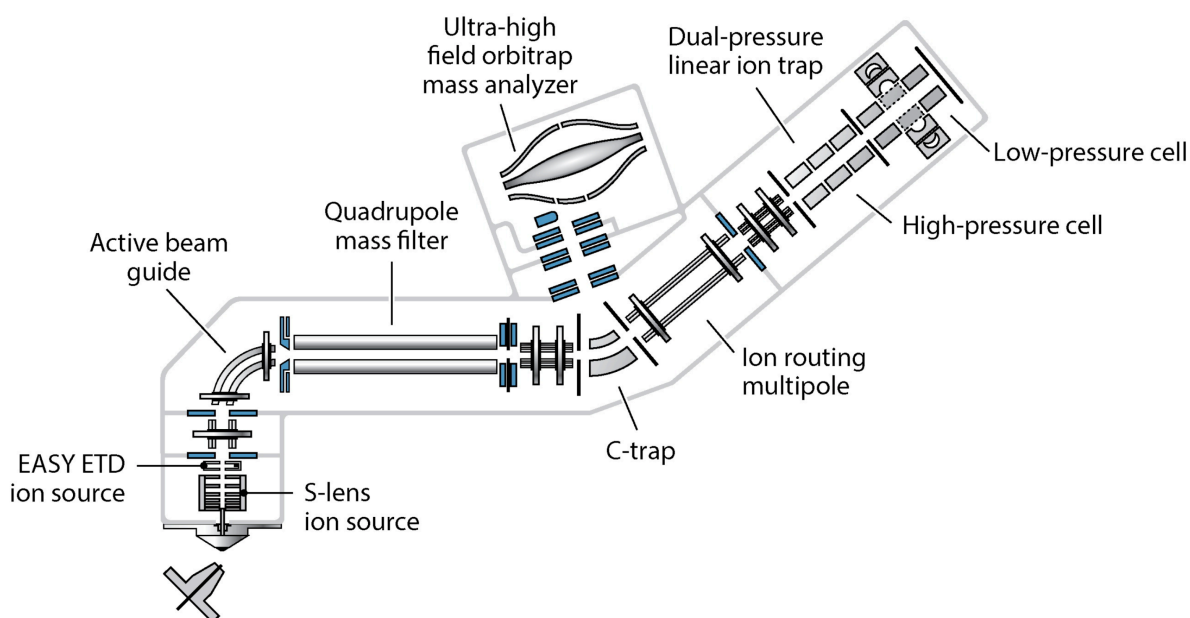
Figure 1-10 Schematic of the Orbitrap Fusion mass spectrometer.

Adapted with permission from Annual Reviews: Annual review of analytical chemistry (Eliuk S. et al.) [211] © 2015.

The Orbitrap Fusion mass spectrometer is a tribrid mass spectrometer that couples an extra dual-pressure LIT mass analyzer [215] to a Q-Exactive instrument. A LIT applies a stopping voltage at two ends of several quadrupole rods to confine ions by creating a potential well for the ions along the axis of the trap [216]. Based on the Mathieu stability diagram, the charged ions are trapped radially by a 2D RF field while confined axially by a static electric potential, which allows ions to be injected and accumulated. Through altering the RF frequency, ions can also be ejected via a slit along with one of the hyperbolic rods. In this dual-pressure LIT, the trapped ion can be

either stabilized or fragmented using CID or ETD in the high-pressure trap and then sent to the low-pressure trap for the detection or transferred back to the C-trap for eventual analysis in the Orbitrap.

1.3.8 Quantification of identified peptides and proteins

While PTMs occur at relatively low frequency in the proteome, they possess significant biological meaning. At present, it is possible to determine the absolute amount of a protein, a peptide, or even a PTM on a protein in a sample, described as the copy number per cell. The determination of the relative changes in protein abundance and modification thereof upon cell stimulation or changes in gene expression is the subject of many current studies. On one hand, how to accurately and precisely quantify those changes taking place in the biological system is one of the biggest challenges of a shotgun proteomics study. Conversely, how to reduce technical errors between multiple samples or temporal comparison also needs to be addressed. Depending on the sample type and research objectives, there are multiple methods available in a shotgun proteomics study. More detailed information on these three types of quantitation methods is illustrated in Chapter two.

For this thesis, the quantitation method used for proteomics study is based on SILAC metabolic labeling. SILAC labeling technology was developed by Mathias Mann's research group in 2002 and rapidly extended to many proteomic studies that use cultured cells [156]. It is known for the highest accuracy among MS quantitation methods to date and has been described as an ideal quantitation approach for cell-cultured samples [217, 218]. This method relies on ^{13}C- or ^{15}N-labeled amino acids supplemented in the medium to culture cells. For mammalian cells, amino acids that are essential for cell survival, such as lysine (Lys or K), should be used for labeling purposes to ensure they are provided as the only exogenous source for the culture medium. Arginine (Arg or R) has been shown as a conditionally essential amino acid in humans [219] and is necessary for many cultured human cell lines, such as HEK293 and HeLa [220]. The isotopically labeled Arg has also proven successful for the SILAC approach [220, 221]. Of note, the metabolic conversion of Arg to Pro could occur when there is a sufficient amount of supplemented Arg in

the medium. Thus, Pro should also be supplemented into the medium to prevent the conversion [222]. In addition, tryptic digested peptides contain at least one of these two amino acids. As such, isotopically labeled Arg and Lys become the best options for mammalian cells. Most SILAC experiments compare two conditions, while three SILAC channels are also commercially available. For a triple SILAC experiment, each condition is isotopically labeled with one type of amino acid as follows: L-arginine (Arg0) and L-lysine (Lys0) are used for light channel, $^{13}C_6^{14}N_4$ L-arginine (Arg6) and 4,4,5,5-D4-L-lysine (Lys4) are used for medium, and $^{13}C_6^{15}N_4$-L-arginine (Arg10) and $^{13}C_6^{15}N_2$-L-Lysine (Lys8) are used for heavy channel [223]. During cell division, those amino acids are *in vivo* metabolically incorporated into all proteins as they are synthesized. In order to ensure that sufficient incorporation of labeled amino acids is achieved, at least 7 passages of cell culture are required prior to any subsequent cell treatments [224]. Once the cells are fully labeled with medium or heavy version of Arg and Lys, different samples can be combined at either cell or protein level, which results in minimal experimental error or bias introduced during the sample preparation [225]. The data acquisition of SILAC samples on the LC-MS is not particular to most of the shotgun proteomics analysis. However, SILAC quantification is based on the measurements of peptide ion intensities reported in the MS^1-spectra [226]. SILAC is popularly used in the identification of differentially expressed proteins among biological samples [227, 228], protein turnover analysis [144] and interaction proteomics (interactomics) [143, 229]. SILAC is also wildly used in the analysis of global PTM changes. In this application, proteome is often analyzed in the meantime to ensure the changes at the PTM level are not attributed to protein changes. Although the SILAC approach is widely used in various proteomic analyses, the limitations of this approach are worth noting.

1. SILAC-based quantification analyzes the isotopomers of the same peptide which increases spectral complexity and redundancy of MS/MS sequencing.

2. SILAC labeling is only suitable for culturable samples, whereas the samples such as clinical samples or extracted tissues can only be labeled with other labeling methods.

3. In addition, the cost of isotopically labeled amino acids that are added to the cell culture medium is not very affordable to some labs.

1.3.9 Targeted analysis of SUMO peptides by parallel reaction monitoring mass spectrometry

Shotgun proteomics experiments by data-dependent acquisition provide an unbiased identification of proteins and PTMs during the discovery stage. However, due to the stochastic nature of data acquisition, it is less effective to consistently quantify a protein or peptide of interest that is of low abundance. In order to obtain appropriate quantitative information to determine the differences of protein abundance between samples in a statistical manner, it is imperative to establish a targeted experimental workflow. Parallel Reaction Monitoring (PRM) is a targeted approach for quantifying proteins or peptides of interest using high-resolution and high-precision mass spectrometry, such as quadrupole-Orbitrap (q-OT) [211, 230, 231]. Through the use of an inclusion list that contains m/z, charge state and retention time of precursor ions relevant to the analysis, the precise quantitative information of precursor ions of interests can be obtained.
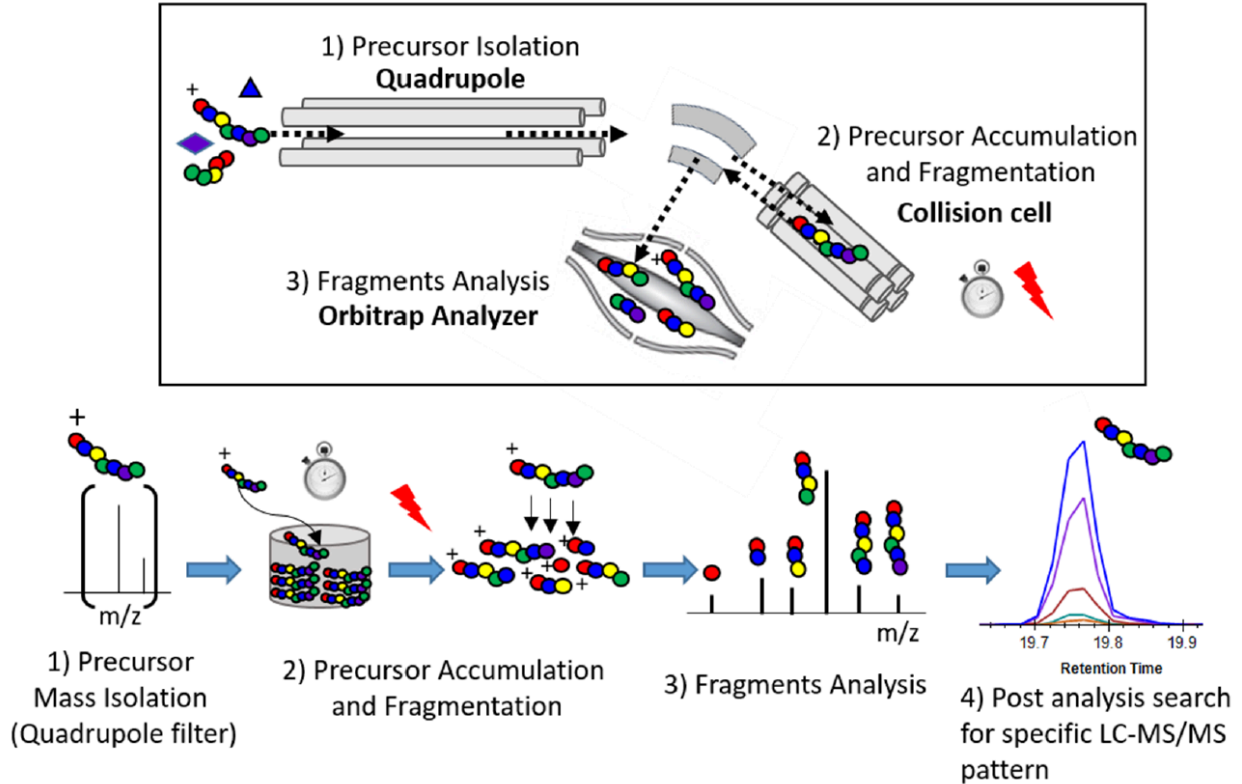
When performing a PRM experiment in q-OT, a predefined precursor is first selected in the quadrupole, and is then transferred to the HCD cell for fragmentation via the C-trap. To increase the signal-to-noise ratio of the ions measured in the Orbitrap, the ions can be accumulated in the C-trap before fragmentation. Then the fragmented ions from the HCD cell can be transferred back to the C-trap and finally injected into the Orbitrap mass analyzer to analyze. In this thesis, the PRM-MS approach was used to compare the regulation patterns of cytoskeletal protein SUMOylation by different PIAS ligases. It is very convenient and efficient to quantify multiple proteins in a complex sample using the PRM approach. It performs a full scan of each transition by a precursor ion and in parallel monitors all fragments from each precursor ion.

1.3.10 Data process

MS generates a tremendous amount of spectra, which makes it impossible to manually match all acquired spectra to the peptides from the database. Furthermore, assembling identified peptides into protein and calculating the quantitative information from corresponding peak intensity is also a big part of the data process for bottom-up proteomics. With the development of bioinformatics, using search algorithms to generate identification and quantification information from the acquired spectra significantly improves the data processing step. For shotgun proteomics, the data processing usually contains two different strategies: the *de novo* peptide sequencing [233] and the *in silico* database search [234]. In the *de novo* peptide sequencing strategy, no knowledge of the amino acid sequence is required. The developed algorithm computes and deduces the sequence of peptides directly from the experimental MS/MS spectra. In order to ensure the most accurate results, it is unquestionably essential to predict unknown information based on better mass accuracy and comprehensive peptide coverage. A commercially available software, PEAK studio [235], integrates scoring functions into a *de novo* sequencing algorithm that helps achieve the maximal identification with high confidence [236]. Since this strategy does not rely on pre-existing reference information of

peptide or protein sequences, it provides a huge advantage in identifying novel PTMs and endogenous peptides and can also be used for un-sequenced organisms or antibodies.

The *in silico* database search is the most popularly used strategy in bottom-up proteomics, and the method used in the *in silico* algorithm of this thesis requires a protein database as a reference to compute and deduce the peptide sequence from an MS/MS spectrum. Basically, the algorithm scans all experimental MS/MS spectra and compares them with the theoretical spectra generated from the reference database, which are typically translated and deduced from genomic data. The Uniprot database is one of the major free public resources that share billions of protein sequences derived from various species. There are many *in silico* algorithms available, such as Mascot [237] and SEQUEST [238] integrated into the Proteome Discoverer™ Software [239], and Andromeda integrated into free open-source software MaxQuant [240]. Among these software, two common features are normally integrated into the algorithms to help the peptide identification: Spectral Scoring [239] and false-discovery rates (FDR) [241]. Depending on the software used, the scoring functions may vary, however they all try to match the ion patterns generated from a spectrum to the database sequence and calculate the possibility of matched peptide sequence. The better matches, the higher scores it gives, and the more confident identification it will be. When all the matches and scoring are computed, a FDR is used to filter out the most significant identifications based on statistical calculation.

$$FDR = \frac{FP}{FP + TP}$$

Equation 1-3 FDR calculation with the number of false positive matches (FP) and the number of true positive matches (TP).

When algorithms perform peak and peptide match using a target database, they also match against a decoy database that contains reversed/shuffled sequences generated from the target database. As a result, true positive matches are obtained from the target database and the false positive matches are generated from the decoy database. Thus, the FDR is calculated using the equation above to reflect the most positive hits the data gives. In most shotgun proteomics studies, 1% FDR is used as a threshold to generate a 99% of the most positive hits.

1.4 Transcriptomic analysis by RNA-sequencing (RNA-Seq)

Transcriptome is the entire set of mRNA transcripts that contains all the transient biological information presented by the genome of a given organism under specific conditions and at a given time [242]. Transcriptomics is a quantitative technique using high-throughput methods to study the transcriptome of an organism at a system level [243]. Transcription profiling is one of the popular applications used in many research areas, such as clinical diagnosis, biomarker discovery, new drug assessment or environmental chemicals, to profile changes in the behavior of not just a single gene or a few genes but of a cell as a whole. In basic research, transcription profiling can be used for the identification of transcription factor mutants that are associated with observed phenotypes [244]. In addition, it has been also reported to be popularly used in the identification of gene functions and their corresponding pathways [245, 246].

Several different types of technologies have been developed for transcriptomic analysis, including hybridization-or sequence-based approaches [247]. RNA sequencing (RNA-seq) has proven its great advantages and is expected to revolutionize the analysis manner of eukaryotic transcriptomes [248]. RNA-Seq relies on the use of the high-throughput next generation sequencing (NGS) technology and the bioinformatic methods to identify and quantify the mRNA transcripts in the samples [249]. This tool is highly sensitive and accurate, which allows researchers to identify differential gene expression (DGE) and can be also used for the identification of disease-associated gene fusions and allele-specific expression [243, 250]. In general, the RNA-seq begins with the RNA extraction from a given biological sample. Then the total mRNA are converted to a cDNA library with adaptors attached to one or two ends [251]. Depending on the sequencing method used, the length of generated nucleotide sequences varies from 30 bp to over 10,000 bp, but is normally around 100 bp in length [252]. mRNA transcripts with different abundance can be quantified with a dynamic range of 5 orders of magnitude [253]. As a high-throughput technology, RNA-Seq experiments often generate raw data containing millions of reads which have to be processed and analyzed by bioinformatics tools to yield meaningful information [254]. RNA-seq data process usually includes several steps, such as quality control (QC), sequence mapping to reference genome, reads normalization and

quantification, and differential gene expression. These steps normally require a data processing pipeline composed of different bioinformatics software tools. In this thesis, QC and sequence alignment were performed using MultiQC v1.7 in a python environment [255]. Quantification was computed using RSEM [256] and differential expression analysis was performed using DESeq2 [257] packages in a R environment.
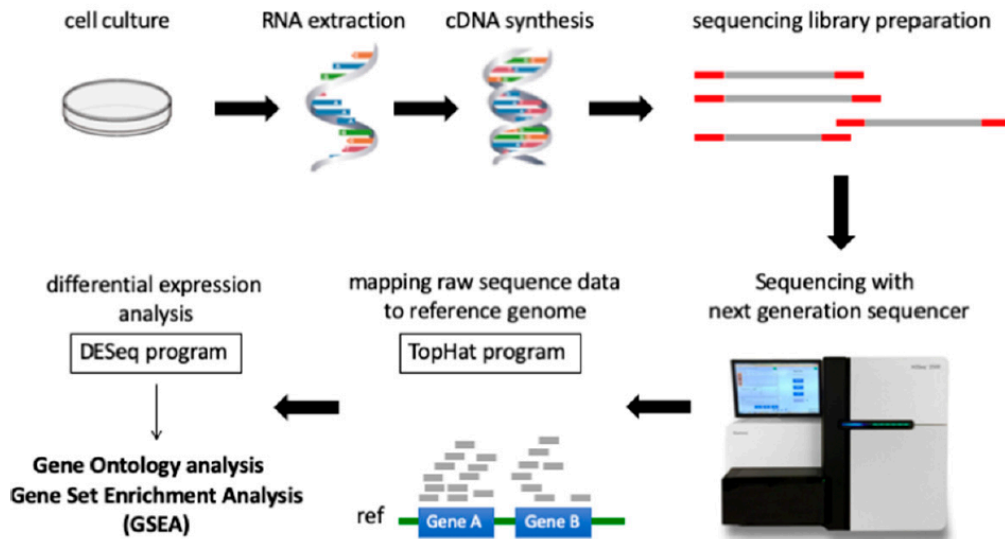
Figure 1-12 Workflow of RNA-seq based trnascriptomics analysis.

Adapted with permission from Multidisciplinary Digital Publishing Institute: Molecules (Adilijiang A. et al) [258] © 2019.

1.5 Biochemical and molecular biological approaches for functional study of protein SUMOylation

The robust SUMO proteomics approach based on MS allows the generation of a massive amount of data that depict the global SUMOylation changes regulated specifically by different PIASs.

1.5.1 CRISPR/Cas9-based genome editing technology

Genome editing technology has enabled the extension of gene function exploration by facilitating the creation of more accurate cellular and animal models [259]. In the last few decades, the development of genome editing made a big success in the research of human genome, which allowed researchers to better understand the gene function in an organism [259, 260]. With the discovery of new bacterial nucleases and the engineering improvement of enzymatic activity and specificity, genome editing technology in higher eukaryotic species has advanced at a rapid pace. Notably, in 2012, the identification of the CRISPR/Cas9 system by Jennifer Doudna and Emmanuelle Charpentier marked a huge breakthrough in precise DNA editing milestones. They received the Nobel Prize in Chemistry 2020 for the contribution in pioneering the revolutionary gene-editing technology.

Clustered regularly interspaced short palindromic repeats (CRISPR) is a technology in the field of molecular biology for changing the genome sequence of living organisms in a precise fashion [261, 262]. Based on the bacterial CRISPR/Cas9 antiviral defense system [261], the complex containing Cas9 nuclease and a synthetic guide RNA (sgRNA) is delivered into a cell, where it then searches and binds to the desired location on the genome to introduce a double-strand break (DSB) in the target area. The DSB may then be repaired through a non-homologous end joining (NHEJ) mechanism. As random deletions or insertions happen at the repair site, the gene functionality is disrupted or altered. This process is known as gene knockout [263]. In addition, the homology-directed repair (HDR) mechanism can also be triggered if DNA sequences similar to the repair template are employed as a donor. By incorporating the exogenous DNA

containing modified DNA sites into the original position, a nucleotide can be altered as desired, specifically referred to as gene knock-in [264]. In this thesis, I combined CRISPR/Cas9 gene-editing technology with Fluorescence-activated Cell Sorting (FACS) technology to knock out SUMO E3 ligase PIAS genes in human cells. In brief, the vectors co-express mCherry fluorescent protein, Cas9 nuclease, and sgRNAs that target the PIAS gene functional domain are transfected into human cells, HEK293 SUMO3m, HeLa and MDA-MB231. This approach allows flow cytometry to sort the cells which contain mCherry signals and most likely have PIAS sgRNA and Cas9 expressed. Experiment results showed the combination of these two techniques efficiently achieved a PIAS knockout cell line.

1.5.2 Site-directed mutagenesis (SDM)

Site-directed mutagenesis is a conventionally used molecular biology approach for investigating the structure and biological activity of protein molecules, as well as for protein engineering [264]. It allows researchers to make changes at the specific region of a gene and assess the functionality of a protein PTM site. In this thesis, point mutations are introduced in the proteins of interest using this technique, which allows the evaluation of the function of SUMOylation on cytoskeleton proteins regulated by SUMO E3 ligase PIAS. SUMO modified lysine on the substrates is generally mutated into arginine to abolish SUMOylation on the site of interest. When producing a point mutation, the coding region of the gene is inserted into an expression vector, which may contain a peptide tag or fluorescent protein. Subsequently, a pair of primers containing the point mutation sequence of nucleotides are used in the Polymerase chain reaction (PCR) to amplify the mutant gene. In this process, it is strongly recommended to use a High-Fidelity polymerase with a low error rate, which ensures a high degree of accuracy in the replication of the DNA. After molecular cloning and SDM, the WT or mutant proteins can be expressed in the cell system, followed by a phenotypic assay to evaluate the functionality of certain SUMOylation sites under a certain biological context [265]. Although this technique is popularly used in most biology labs today, the drawbacks of this method still need to be noted. First, SUMOylation is a highly dynamic and reversible modification, whereas mutating lysine into arginine permanently prevents the SUMOylation from happening on this site. Second, lysine can

be modified by various PTMs, such as other Ubi/Ubl modifications, acetylation, and methylation[266]; the functionality discovered may not attribute to SUMOylation only. Third, SUMO can form several types of poly chains (detailed information of SUMO poly chain is elaborated in Chapter two), and the SDM technique cannot determine if the functions observed are from mono SUMOylation or poly SUMO chains, which requires additional experiments to confirm. Last but not least, this technique is still very low throughput and time-consuming, and should be carefully used for the functional assay.

1.5.3 *In vitro* SUMO assay

Similar to ubiquitylation, SUMOylation assay could happen *in vitro* when SUMO E1 and E2 enzymes, SUMO modifier protein, and a protein of interest co-exist with ATP in a suitable buffer at RT or 37 °C [267-269]. Although all proteins need to be purified under a native condition to ensure protein activity, this biochemical reaction is still easy to set up and allows for the rapid *in vitro* identification and validation of SUMOylation sites on proteins of interest. In addition, Werner et al. have also generated a bacterial cell line producing the SUMOylation machinery, allowing us to carry out an *in vitro* assay by simple transformation and induction of the protein of interest in the bacteria [270]. The SUMOylated proteins produced in this way can then be identified by a Shotgun approach or by Western blot. It is worth noting that the biochemical reaction often forces SUMOylation on some sites, which are unlikely to be modified *in vivo*. Moreover, SUMO E3 ligases are not essential for *in vitro* assay due to ideal biochemical conditions. Thus, to assess the function of SUMO E3 ligase, the concentration used in the reaction needs to be carefully calculated. Finally, this *in vitro* method is modified and optimized by different groups to apply to large-scale SUMO proteomics identifications. Several examples based on MS and protein microarray are illustrated in Chapter two.

1.5.4 Fluorescence microscopy-based assays

Fluorescence microscopy is popularly used in the biology research field. It is an imaging technique that allows the fluorophore excitation and the fluorescence signal detection through

the same light [271]. It is highly sensitive, specific, reliable and extensively used by scientists in the cell biology domain to observe the localization of proteins within cells, monitor the dynamic changes of cellular structure, and thus evaluate cell physiology. In order to produce fluorescence and be detected under microscopy, immunofluorescence or engineered fluorescence fusion proteins are required.

Immunofluorescence relies on the use of antibodies covalently labeled with fluorescent dyes to visualize molecules under a light microscope. For endogenous proteins, cultured cells or tissue samples need to be fixed using paraformaldehyde or methanol, followed by primary and secondary antibody incubation. This limits the observation of bio-macromolecules into a transient situation, and many pictures need to be taken to generate a more comprehensive view. Alternatively, recombinant proteins fused with fluorescent tag or protein domains, such as red fluorescent protein (RFP), allow localizing proteins in live cells.

Fluorescence Recovery After Photobleaching (FRAP) is another technique used in this thesis to determine the kinetics of intermediate filament (IF) diffusion through living cells [272]. This technique is commonly used in conjunction with proteins fused with fluorescent compartments, such as a GFP. Before photobleaching a certain area, a background image of the living sample needs to be saved. Then, the targeted area of the sample is focused by the light source through either a microscope objective with higher magnification or using laser light with the appropriate wavelength. Next, the fluorophores of the targeted area are bleached by a laser source with high energy, which results in a uniformly fluorescent area with a noticeable dark spot in the image captured by the microscope. As Brownian motion proceeds, the beached molecules will exchange and diffuse with the still-fluorescing molecules from surrounding regions. Finally, fluorescent intensity recovers in the bleached region in an ordered fashion and can be quantitatively monitored using a microscope.

1.6 Cell phenotypic assays

1.6.1 Cell proliferation assay

Cell proliferation assays are one of the frequently often used phenotypic assays in cell biology to evaluate the function of a gene or PTMs on proteins during cell growth [273]. Cell counting by hemocytometer under a microscope is the most direct method to assess the growth rate of a cell line. It is simple and straightforward to perform for cells growing in suspension. There are many automated cell counters that are commercially available, which make suspension cell proliferation assays efficient and uncomplicated. However, for adherent cells, direct cell counting is not as simple as suspension cells because it usually requires cells to be trypsinized from the bottom of the cell culture dish and resuspend into a single cell solution. These steps could be time-consuming when multiple conditions need to be assessed, and technical errors could be introduced due to cell doublets or triplets in the case of insufficient cell digestion. An alternative assay was developed based on the use of chemiluminescence to measure the activity of mitochondrial dehydrogenases. In order to measure the cell numbers, the stable tetrazolium salt (WST-1) reagent is added into the media of cultured cells and incubated for one hour. During the incubation of this assay, WST-1 is cleaved by NAD(P)H-dependent mitochondrial dehydrogenases to produce formazan [274]. With the progress of bioreduction reaction in the live cells, the media color changes due to the increasing formazan dye present in the solution. Followed by measuring the absorbance at Optical Density at 450 nm (OD_{450}), the OD value positively correlates with the number of live cells in the culture [275]. Although it cannot show the absolute cell numbers, it simplifies the relative comparison of cell proliferation across several conditions.

1.6.2 Cell migration assay

Adhesion cell motility can normally be elevated using a wound-healing cell migration assay. It is another phenotypic assay wildly used in biology labs for the analysis of cell migration under different conditions [276]. The simplest assay is performed by making a scratch on a cell

monolayer using a pipette tip and taking pictures of wound closure at certain time points, thus it is also called a scratch assay. This assay allows for the evaluation of the cell motility rate in a quantitative manner and for the comparison of multiple conditions at the same time. In this thesis, commercially available 2 well culture inserts were used to generate the wound gaps and improve the reproducibility of the cell migration assay.

1.7 Research objectives

With the advancement in the technology of biochemical protein microarrays, analytical MS and fluorescence microscopy, there is an increasing number of high throughput methods developed for the identification and characterization of SUMO E3 ligase substrates. Massive data have been generated through different approaches that have improved my understanding of SUMO E3 ligase regulation. However, it is still challenging to systematically identify the SUMOylated lysine site regulated by SUMO E3 ligase and understand the associated biological mechanism.

Thus, this thesis focuses on the improvement of MS-based quantitative SUMO proteomics. This is achieved by integrating CRISPR/Cas9 gene-editing technology and RNA-seq technology into the proteomic analysis, characterizing PIAS regulatory networks in a site-specific manner, and studying the physiological effects of regulated SUMOylation sites using phenotypic assays. This thesis is centered on the following research objectives:

1. Evaluate the physiological functions of SUMO E3 ligase PIAS proteins in cell growth and cell motility.

2. Identify the PIAS substrates in a quantitative and site-specific manner.

3. Study the regulatory mechanism involved in intermediate filament dynamics.

4. Characterize the regulation networks of SUMO E3 ligase PIASs at the post-translational and transcriptional levels.

1.8 Thesis overview

The first chapter of my doctoral thesis presents the literature review of the current knowledge of cellular and molecular biology relevant to protein SUMOylation and cytoskeletal organization. It also provides a general overview of MS-based shotgun proteomics, including a summary of sample preparation, instrumental analysis, and data analysis. In addition, this chapter also illustrated the most relevant biological techniques used for studying the function of identified protein SUMOylation in a cell model. This introductory session covers the most relevant background knowledge and previous study that has very well guided the following chapters.

Chapter two is a review article entitled "Proteomic strategies for characterizing ubiquitin-like modifications", which was accepted in **Nature Reviews Methods Primers**. This chapter further extends the literature review of proteomic studies focusing on the application in the ubiquitin or other ubiquitin-like proteins (UBLs) field. It introduces a comprehensive overview of the method used in UBL proteomics studies. It also provides a synthesis of all information that researchers need to give the best practices for experimental design, data analysis, reproducibility and standardization. Finally, it highlights the biological questions that can be addressed by this method. SUMO proteomic studies account for a large proportion.

In the third chapter, my first research article, entitled "Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility", published in **Nature Communications**, is presented [132]. In this scientific work, I took advantage of the previously developed SUMO proteomic workflow, further combined with protein overexpression and CRISPR/Cas9 gene-editing technology to identify SUMO E3 ligase PIAS1 substrates in a site-specific manner. A new model involved in PIAS1-mediated intermediate filament dynamics through vimentin SUMOylation and promoting cell motility is proposed in this article.

Chapter four represents my second research article entitled "SUMO proteomics and transcriptomics analyses identify PIAS-mediated regulatory networks involved in cell proliferation and migration". In this chapter, I extend the study to the whole PIAS E3 family

proteins based on the findings in Chapter three. In the meanwhile, the transcriptomic analysis was introduced to complement the research view from another angle at the transcription level. The complementary data provided by proteomics and transcriptomics shed light on the regulation networks by different PIASs involved in cell proliferation and migration. This work is still in preparation.

Finally, Chapter five concludes the major findings from my entire Ph.D. study and discusses the results of each article. It also provides future perspectives and suggestions for future scientists in the field.

1.9 References

1.   Franklin, S. and T.M. Vondriska, *Genomes, proteomes, and the central dogma.* Circ Cardiovasc Genet, 2011. **4**(5): p. 576.

2.   Godovac-Zimmermann, J., *8th Siena meeting. From genome to proteome: integration and proteome completion.* Expert Rev Proteomics, 2008. **5**(6): p. 769-73.

3.   Graves, P.R. and T.A. Haystead, *Molecular biologist's guide to proteomics.* Microbiol Mol Biol Rev, 2002. **66**(1): p. 39-63; table of contents.

4.   Timp, W. and G. Timp, *Beyond mass spectrometry, the next step in proteomics.* Sci Adv, 2020. **6**(2): p. eaax8978.

5.   Virag, D., et al., *Current Trends in the Analysis of Post-translational Modifications.* Chromatographia, 2020. **83**(1): p. 1-10.

6.   International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.

7.   Liu, Y., et al., *Impact of Alternative Splicing on the Human Proteome.* Cell Rep, 2017. **20**(5): p. 1229-1241.

8.   Coate, J.E. and J.J. Doyle, *Variation in transcriptome size: are we getting the message?* Chromosoma, 2015. **124**(1): p. 27-43.

9.   Santos, A.L. and A.B. Lindner, *Protein Posttranslational Modifications: Roles in Aging and Age-Related Disease.* Oxid Med Cell Longev, 2017. **2017**: p. 5716409.

10.   Khoury, G.A., R.C. Baliban, and C.A. Floudas, *Proteome-wide post-translational modification statistics: frequency analysis and curation of the swiss-prot database.* Sci Rep, 2011. **1**.

11.   Reinders, J. and A. Sickmann, *Modificomics: posttranslational modifications beyond protein phosphorylation and glycosylation.* Biomol Eng, 2007. **24**(2): p. 169-77.

12.   Wang, Y.C., S.E. Peterson, and J.F. Loring, *Protein post-translational modifications and regulation of pluripotency in human stem cells.* Cell Res, 2014. **24**(2): p. 143-60.

13.   Muller, M.M., *Post-Translational Modifications of Protein Backbones: Unique Functions, Mechanisms, and Challenges.* Biochemistry, 2018. **57**(2): p. 177-185.

14. Chen, L., S. Liu, and Y. Tao, *Regulating tumor suppressor genes: post-translational modifications.* Signal Transduct Target Ther, 2020. **5**(1): p. 90.

15. Abdrabou, A. and Z. Wang, *Post-Translational Modification and Subcellular Distribution of Rac1: An Update.* Cells, 2018. **7**(12).

16. Prabakaran, S., et al., *Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding.* Wiley Interdiscip Rev Syst Biol Med, 2012. **4**(6): p. 565-83.

17. Knorre, D.G., N.V. Kudryashova, and T.S. Godovikova, *Chemical and functional aspects of posttranslational modification of proteins.* Acta Naturae, 2009. **1**(3): p. 29-51.

18. Millar, A.H., et al., *The Scope, Functions, and Dynamics of Posttranslational Protein Modifications.* Annu Rev Plant Biol, 2019. **70**: p. 119-151.

19. Hay, R.T., *SUMO: a history of modification.* Mol Cell, 2005. **18**(1): p. 1-12.

20. Matunis, M.J., E. Coutavas, and G. Blobel, *A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex.* J Cell Biol, 1996. **135**(6 Pt 1): p. 1457-70.

21. Mahajan, R., et al., *A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2.* Cell, 1997. **88**(1): p. 97-107.

22. Meluh, P.B. and D. Koshland, *Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C.* Mol Biol Cell, 1995. **6**(7): p. 793-807.

23. Shen, Z., et al., *UBL1, a human ubiquitin-like protein associating with human RAD51/RAD52 proteins.* Genomics, 1996. **36**(2): p. 271-9.

24. Okura, T., et al., *Protection against Fas/APO-1- and tumor necrosis factor-mediated cell death by a novel protein, sentrin.* J Immunol, 1996. **157**(10): p. 4277-81.

25. Boddy, M.N., et al., *PIC 1, a novel ubiquitin-like protein which interacts with the PML component of a multiprotein complex that is disrupted in acute promyelocytic leukaemia.* Oncogene, 1996. **13**(5): p. 971-82.

26. Kamitani, T., et al., *Characterization of a second member of the sentrin family of ubiquitin-like proteins.* J Biol Chem, 1998. **273**(18): p. 11349-53.

27. Kamitani, T., et al., *Covalent modification of PML by the sentrin family of ubiquitin-like proteins.* J Biol Chem, 1998. **273**(6): p. 3117-20.

28. Skilton, A., et al., *SUMO chain formation is required for response to replication arrest in S. pombe.* PLoS One, 2009. **4**(8): p. e6750.

29. Bayer, P., et al., *Structure determination of the small ubiquitin-related modifier SUMO-1.* J Mol Biol, 1998. **280**(2): p. 275-86.

30. Geiss-Friedlander, R. and F. Melchior, *Concepts in sumoylation: a decade on.* Nat Rev Mol Cell Biol, 2007. **8**(12): p. 947-56.

31. Broday, L., et al., *The small ubiquitin-like modifier (SUMO) is required for gonadal and uterine-vulval morphogenesis in Caenorhabditis elegans.* Genes Dev, 2004. **18**(19): p. 2380-91.

32. Gillies, J., et al., *SUMO Pathway Modulation of Regulatory Protein Binding at the Ribosomal DNA Locus in Saccharomyces cerevisiae.* Genetics, 2016. **202**(4): p. 1377-94.

33. Smith, M., W. Turki-Judeh, and A.J. Courey, *SUMOylation in Drosophila Development.* Biomolecules, 2012. **2**(3): p. 331-49.

34. Saracco, S.A., et al., *Genetic analysis of SUMOylation in Arabidopsis: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential.* Plant Physiol, 2007. **145**(1): p. 119-34.

35. Nacerddine, K., et al., *The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice.* Dev Cell, 2005. **9**(6): p. 769-79.

36. Melchior, F., *SUMO--nonclassical ubiquitin.* Annu Rev Cell Dev Biol, 2000. **16**: p. 591-626.

37. Eifler, K. and A.C. Vertegaal, *Mapping the SUMOylated landscape.* FEBS J, 2015. **282**(19): p. 3669-80.

38. Baczyk, D., et al., *SUMO-4: A novel functional candidate in the human placental protein SUMOylation machinery.* PLoS One, 2017. **12**(5): p. e0178056.

39. Nayak, A. and S. Muller, *SUMO-specific proteases/isopeptidases: SENPs and beyond.* Genome Biol, 2014. **15**(7): p. 422.

40. Guo, D., et al., *A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes.* Nat Genet, 2004. **36**(8): p. 837-41.

41.   Owerbach, D., et al., *A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation.* Biochem Biophys Res Commun, 2005. **337**(2): p. 517-20.

42.   Wei, W., et al., *A stress-dependent SUMO4 sumoylation of its substrate proteins.* Biochem Biophys Res Commun, 2008. **375**(3): p. 454-9.

43.   Tossidou, I., et al., *SUMOylation determines turnover and localization of nephrin at the plasma membrane.* Kidney Int, 2014. **86**(6): p. 1161-73.

44.   Liang, Y.C., et al., *SUMO5, a Novel Poly-SUMO Isoform, Regulates PML Nuclear Bodies.* Sci Rep, 2016. **6**: p. 26509.

45.   Bruderer, R., et al., *Purification and identification of endogenous polySUMO conjugates.* EMBO Rep, 2011. **12**(2): p. 142-8.

46.   Tatham, M.H., et al., *Comparative proteomic analysis identifies a role for SUMO in protein quality control.* Sci Signal, 2011. **4**(178): p. rs4.

47.   Da Silva-Ferrada, E., et al., *Analysis of SUMOylated proteins using SUMO-traps.* Sci Rep, 2013. **3**: p. 1690.

48.   Becker, J., et al., *Detecting endogenous SUMO targets in mammalian cells and tissues.* Nat Struct Mol Biol, 2013. **20**(4): p. 525-31.

49.   Vertegaal, A.C., et al., *Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics.* Mol Cell Proteomics, 2006. **5**(12): p. 2298-310.

50.   Fu, C., et al., *Stabilization of PML nuclear localization by conjugation and oligomerization of SUMO-3.* Oncogene, 2005. **24**(35): p. 5401-13.

51.   Mukhopadhyay, D., et al., *SUSP1 antagonizes formation of highly SUMO2/3-conjugated species.* J Cell Biol, 2006. **174**(7): p. 939-49.

52.   Matunis, M.J., J. Wu, and G. Blobel, *SUMO-1 modification and its role in targeting the Ran GTPase-activating protein, RanGAP1, to the nuclear pore complex.* J Cell Biol, 1998. **140**(3): p. 499-509.

53.   Figueroa-Romero, C., et al., *SUMOylation of the mitochondrial fission protein Drp1 occurs at multiple nonconsensus sites within the B domain and is linked to its activity cycle.* FASEB J, 2009. **23**(11): p. 3917-27.

54. Flotho, A. and F. Melchior, *Sumoylation: a regulatory protein modification in health and disease.* Annu Rev Biochem, 2013. **82**: p. 357-85.

55. Zhao, J., *Sumoylation regulates diverse biological processes.* Cell Mol Life Sci, 2007. **64**(23): p. 3017-33.

56. Jentsch, S. and I. Psakhye, *Control of nuclear activities by substrate-selective and protein-group SUMOylation.* Annu Rev Genet, 2013. **47**: p. 167-86.

57. Yang, X.J. and C.M. Chiang, *Sumoylation in gene regulation, human disease, and therapeutic action.* F1000Prime Rep, 2013. **5**: p. 45.

58. May, B. *SUMO E3 ligases SUMOylate target proteins*. 2013; Available from: https://reactome.org/content/detail/R-HSA-3108232.

59. Makhnevych, T., et al., *Global map of SUMO function revealed by protein-protein interaction and genetic networks.* Mol Cell, 2009. **33**(1): p. 124-35.

60. Balasubramaniyan, N., et al., *SUMOylation of the farnesoid X receptor (FXR) regulates the expression of FXR target genes.* J Biol Chem, 2013. **288**(19): p. 13850-62.

61. Hay, R.T., *SUMO-specific proteases: a twist in the tail.* Trends Cell Biol, 2007. **17**(8): p. 370-6.

62. Axelsen, L.N., et al., *Managing the complexity of communication: regulation of gap junctions by post-translational modification.* Front Pharmacol, 2013. **4**: p. 130.

63. Rytinki, M.M., et al., *PIAS proteins: pleiotropic interactors associated with SUMO.* Cell Mol Life Sci, 2009. **66**(18): p. 3029-41.

64. Kerscher, O., R. Felberbaum, and M. Hochstrasser, *Modification of proteins by ubiquitin and ubiquitin-like proteins.* Annu Rev Cell Dev Biol, 2006. **22**: p. 159-80.

65. Park, H.J. and D.-J. Yun, *Chapter Five - New Insights into the Role of the Small Ubiquitin-like Modifier (SUMO) in Plants*, in *International Review of Cell and Molecular Biology*, K.W. Jeon, Editor. 2013, Academic Press. p. 161-209.

66. Wilkinson, K.A., A. Nishimune, and J.M. Henley, *Analysis of SUMO-1 modification of neuronal proteins containing consensus SUMOylation motifs.* Neurosci Lett, 2008. **436**(2): p. 239-44.

67. Gareau, J.R. and C.D. Lima, *The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition.* Nat Rev Mol Cell Biol, 2010. **11**(12): p. 861-71.

68. Hochstrasser, M., *SP-RING for SUMO: new functions bloom for a ubiquitin-like protein.* Cell, 2001. **107**(1): p. 5-8.

69. Rabellino, A., C. Andreani, and P.P. Scaglioni, *The Role of PIAS SUMO E3-Ligases in Cancer.* Cancer Res, 2017. **77**(7): p. 1542-1547.

70. Hickey, C.M., N.R. Wilson, and M. Hochstrasser, *Function and regulation of SUMO proteases.* Nat Rev Mol Cell Biol, 2012. **13**(12): p. 755-66.

71. George, A.J., et al., *A Comprehensive Atlas of E3 Ubiquitin Ligase Mutations in Neurological Disorders.* Front Genet, 2018. **9**: p. 29.

72. Chen, L. and H. Hellmann, *Plant E3 ligases: flexible enzymes in a sessile world.* Mol Plant, 2013. **6**(5): p. 1388-404.

73. Cartier, E., et al., *The SUMO-Conjugase Ubc9 Prevents the Degradation of the Dopamine Transporter, Enhancing Its Cell Surface Level and Dopamine Uptake.* Front Cell Neurosci, 2019. **13**: p. 35.

74. Verma, D.K., et al., *The SUMO Conjugase Ubc9 Protects Dopaminergic Cells from Cytotoxicity and Enhances the Stability of alpha-Synuclein in Parkinson's Disease Models.* eNeuro, 2020. **7**(5).

75. Chu, Y. and X. Yang, *SUMO E3 ligase activity of TRIM proteins.* Oncogene, 2011. **30**(9): p. 1108-16.

76. Takahashi, Y., A. Toh-e, and Y. Kikuchi, *A novel factor required for the SUMO1/Smt3 conjugation of yeast septins.* Gene, 2001. **275**(2): p. 223-31.

77. Yunus, A.A. and C.D. Lima, *Structure of the Siz/PIAS SUMO E3 ligase Siz1 and determinants required for SUMO modification of PCNA.* Mol Cell, 2009. **35**(5): p. 669-82.

78. Kagey, M.H., T.A. Melhuish, and D. Wotton, *The polycomb protein Pc2 is a SUMO E3.* Cell, 2003. **113**(1): p. 127-37.

79. Zhu, S., et al., *Protection from isopeptidase-mediated deconjugation regulates paralog-selective sumoylation of RanGAP1.* Mol Cell, 2009. **33**(5): p. 570-80.

80.   Ritterhoff, T., et al., *The RanBP2/RanGAP1*SUMO1/Ubc9 SUMO E3 ligase is a disassembly machine for Crm1-dependent nuclear export complexes.* Nat Commun, 2016. **7**: p. 11482.

81.   Werner, A., A. Flotho, and F. Melchior, *The RanBP2/RanGAP1*SUMO1/Ubc9 complex is a multisubunit SUMO E3 ligase.* Mol Cell, 2012. **46**(3): p. 287-98.

82.   Pichler, A., et al., *The RanBP2 SUMO E3 ligase is neither HECT- nor RING-type.* Nat Struct Mol Biol, 2004. **11**(10): p. 984-91.

83.   Tatham, M.H., et al., *Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralog selection.* Nat Struct Mol Biol, 2005. **12**(1): p. 67-74.

84.   Reverter, D. and C.D. Lima, *Insights into E3 ligase activity revealed by a SUMO-RanGAP1-Ubc9-Nup358 complex.* Nature, 2005. **435**(7042): p. 687-92.

85.   Pichler, A., et al., *The nucleoporin RanBP2 has SUMO1 E3 ligase activity.* Cell, 2002. **108**(1): p. 109-20.

86.   Scognamiglio, A., et al., *HDAC-class II specific inhibition involves HDAC proteasome-dependent degradation mediated by RANBP2.* Biochim Biophys Acta, 2008. **1783**(10): p. 2030-8.

87.   Cho, K.I., et al., *Targeting the cyclophilin domain of Ran-binding protein 2 (Ranbp2) with novel small molecules to control the proteostasis of STAT3, hnRNPA2B1 and M-opsin.* ACS Chem Neurosci, 2015. **6**(8): p. 1476-85.

88.   Saitoh, H., M.D. Pizzi, and J. Wang, *Perturbation of SUMOlation enzyme Ubc9 by distinct domain within nucleoporin RanBP2/Nup358.* J Biol Chem, 2002. **277**(7): p. 4755-63.

89.   Saitoh, N., et al., *In situ SUMOylation analysis reveals a modulatory role of RanBP2 in the nuclear rim and PML bodies.* Exp Cell Res, 2006. **312**(8): p. 1418-30.

90.   Kahyo, T., T. Nishida, and H. Yasuda, *Involvement of PIAS1 in the sumoylation of tumor suppressor p53.* Mol Cell, 2001. **8**(3): p. 713-8.

91.   Sachdev, S., et al., *PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies.* Genes Dev, 2001. **15**(23): p. 3088-103.

92.   Schmidt, D. and S. Muller, *Members of the PIAS family act as SUMO ligases for c-Jun and p53 and repress p53 activity.* Proc Natl Acad Sci U S A, 2002. **99**(5): p. 2872-7.

93.   Johnson, E.S. and A.A. Gupta, *An E3-like factor that promotes SUMO conjugation to the yeast septins.* Cell, 2001. **106**(6): p. 735-44.

94.   Takahashi, Y., et al., *Yeast Ull1/Siz1 is a novel SUMO1/Smt3 ligase for septin components and functions as an adaptor between conjugating enzyme and substrates.* J Biol Chem, 2001. **276**(52): p. 48973-7.

95.   Zhao, X. and G. Blobel, *A SUMO ligase is part of a nuclear multiprotein complex that affects DNA repair and chromosomal organization.* Proc Natl Acad Sci U S A, 2005. **102**(13): p. 4777-82.

96.   Cheng, C.H., et al., *SUMO modifications control assembly of synaptonemal complex and polycomplex in meiosis of Saccharomyces cerevisiae.* Genes Dev, 2006. **20**(15): p. 2067-81.

97.   Zhang, S., Y. Qi, and C. Yang, *Arabidopsis SUMO E3 ligase AtMMS21 regulates root meristem development.* Plant Signal Behav, 2010. **5**(1): p. 53-5.

98.   Mascle, X.H., et al., *Identification of a non-covalent ternary complex formed by PIAS1, SUMO1, and UBC9 proteins involved in transcriptional regulation.* J Biol Chem, 2013. **288**(51): p. 36312-27.

99.   Streich, F.C., Jr. and C.D. Lima, *Capturing a substrate in an activated RING E3/E2-SUMO complex.* Nature, 2016. **536**(7616): p. 304-8.

100.   Niu, G.J., et al., *Protein Inhibitor of Activated STAT (PIAS) Negatively Regulates the JAK/STAT Pathway by Inhibiting STAT Phosphorylation and Translocation.* Front Immunol, 2018. **9**: p. 2392.

101.   Chung, C.D., et al., *Specific inhibition of Stat3 signal transduction by PIAS3.* Science, 1997. **278**(5344): p. 1803-5.

102.   Liu, B., et al., *Inhibition of Stat1-mediated gene activation by PIAS1.* Proc Natl Acad Sci U S A, 1998. **95**(18): p. 10626-31.

103.   Munarriz, E., et al., *PIAS-1 is a checkpoint regulator which affects exit from G1 and G2 by sumoylation of p73.* Mol Cell Biol, 2004. **24**(24): p. 10593-610.

104. Liao, J., Y. Fu, and K. Shuai, *Distinct roles of the NH2- and COOH-terminal domains of the protein inhibitor of activated signal transducer and activator of transcription (STAT) 1 (PIAS1) in cytokine-induced PIAS1-Stat1 interaction.* Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5267-72.

105. Gallagher, W.M., et al., *MBP1: a novel mutant p53-specific protein partner with oncogenic properties.* Oncogene, 1999. **18**(24): p. 3608-16.

106. Shuai, K., *Regulation of cytokine signaling pathways by PIAS proteins.* Cell Res, 2006. **16**(2): p. 196-202.

107. Sharrocks, A.D., *PIAS proteins and transcriptional regulation--more than just SUMO E3 ligases?* Genes Dev, 2006. **20**(7): p. 754-8.

108. Wang, R., et al., *The conserved ancient role of chordate PIAS as a multilevel repressor of the NF-kappaB pathway.* Sci Rep, 2017. **7**(1): p. 17063.

109. Minty, A., et al., *Covalent modification of p73alpha by SUMO-1. Two-hybrid screening with p73 identifies novel SUMO-1-interacting proteins and a SUMO-1 interaction motif.* J Biol Chem, 2000. **275**(46): p. 36316-23.

110. Tan, J.A., et al., *Protein inhibitors of activated STAT resemble scaffold attachment factors and function as interacting nuclear receptor coregulators.* J Biol Chem, 2002. **277**(19): p. 16993-7001.

111. Duval, D., et al., *The 'PINIT' motif, of a newly identified conserved domain of the PIAS protein family, is essential for nuclear retention of PIAS3L.* FEBS Lett, 2003. **554**(1-2): p. 111-8.

112. Aravind, L. and E.V. Koonin, *SAP - a putative DNA-binding motif involved in chromosomal organization.* Trends Biochem Sci, 2000. **25**(3): p. 112-4.

113. Okubo, S., et al., *NMR structure of the N-terminal domain of SUMO ligase PIAS1 and its interaction with tumor suppressor p53 and A/T-rich DNA oligomers.* J Biol Chem, 2004. **279**(30): p. 31455-61.

114. Moilanen, A.M., et al., *A testis-specific androgen receptor coregulator that belongs to a novel family of nuclear proteins.* J Biol Chem, 1999. **274**(6): p. 3700-4.

115.   Liu, B., et al., *A transcriptional corepressor of Stat1 with an essential LXXLL signature motif.* Proc Natl Acad Sci U S A, 2001. **98**(6): p. 3203-7.

116.   Eisenhaber, B., et al., *The ring between ring fingers (RBR) protein family.* Genome Biol, 2007. **8**(3): p. 209.

117.   Weissman, A.M., *Themes and variations on ubiquitylation.* Nat Rev Mol Cell Biol, 2001. **2**(3): p. 169-78.

118.   Hannich, J.T., et al., *Defining the SUMO-modified proteome by multiple approaches in Saccharomyces cerevisiae.* J Biol Chem, 2005. **280**(6): p. 4102-10.

119.   Song, J., et al., *Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation.* J Biol Chem, 2005. **280**(48): p. 40122-9.

120.   Stehmeier, P. and S. Muller, *Phospho-regulated SUMO interaction modules connect the SUMO system to CK2 signaling.* Mol Cell, 2009. **33**(3): p. 400-9.

121.   Lussier-Price, M., et al., *Characterization of a C-Terminal SUMO-Interacting Motif Present in Select PIAS-Family Proteins.* Structure, 2020. **28**(5): p. 573-585 e5.

122.   Puhr, M., et al., *PIAS1 is a determinant of poor survival and acts as a positive feedback regulator of AR signaling through enhanced AR stabilization in prostate cancer.* Oncogene, 2016. **35**(18): p. 2322-32.

123.   Hoefer, J., et al., *PIAS1 is increased in human prostate cancer and enhances proliferation through inhibition of p21.* Am J Pathol, 2012. **180**(5): p. 2097-107.

124.   Liu, B., et al., *PIAS1 regulates breast tumorigenesis through selective epigenetic gene silencing.* PLoS One, 2014. **9**(2): p. e89464.

125.   Xiong, Q., et al., *Identification of novel miR-21 target proteins in multiple myeloma cells by quantitative proteomics.* J Proteome Res, 2012. **11**(4): p. 2078-90.

126.   Sun, L., et al., *PIASy mediates hypoxia-induced SIRT1 transcriptional repression and epithelial-to-mesenchymal transition in ovarian cancer cells.* J Cell Sci, 2013. **126**(Pt 17): p. 3939-47.

127.   Chien, W., et al., *PIAS4 is an activator of hypoxia signalling via VHL suppression during growth of pancreatic cancer cells.* Br J Cancer, 2013. **109**(7): p. 1795-804.

128. Stehmeier, P. and S. Muller, *Regulation of p53 family members by the ubiquitin-like SUMO system.* DNA Repair (Amst), 2009. **8**(4): p. 491-8.

129. Driscoll, J.J., et al., *The sumoylation pathway is dysregulated in multiple myeloma and is associated with adverse patient outcome.* Blood, 2010. **115**(14): p. 2827-34.

130. Dang, C.V., *MYC on the path to cancer.* Cell, 2012. **149**(1): p. 22-35.

131. Rabellino, A., et al., *PIAS1 Promotes Lymphomagenesis through MYC Upregulation.* Cell Reports, 2016. **15**(10): p. 2266-2278.

132. Li, C., et al., *Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility.* Nat Commun, 2020. **11**(1): p. 834.

133. Rabellino, A., et al., *The SUMO E3-ligase PIAS1 Regulates the Tumor Suppressor PML and Its Oncogenic Counterpart PML-RARA.* Cancer Research, 2012. **72**(9): p. 2275-2284.

134. Constanzo, J.D., et al., *PIAS1-FAK Interaction Promotes the Survival and Progression of Non-Small Cell Lung Cancer.* Neoplasia, 2016. **18**(5): p. 282-293.

135. Heldin, C.H., M. Vanlandewijck, and A. Moustakas, *Regulation of EMT by TGFbeta in cancer.* FEBS Lett, 2012. **586**(14): p. 1959-70.

136. Lamouille, S., J. Xu, and R. Derynck, *Molecular mechanisms of epithelial-mesenchymal transition.* Nat Rev Mol Cell Biol, 2014. **15**(3): p. 178-96.

137. Zhou, S., et al., *PIASy represses CCAAT/enhancer-binding protein delta (C/EBPdelta) transcriptional activity by sequestering C/EBPdelta to the nuclear periphery.* J Biol Chem, 2008. **283**(29): p. 20137-48.

138. Castillo-Lluva, S., et al., *SUMOylation of the GTPase Rac1 is required for optimal cell migration.* Nat Cell Biol, 2010. **12**(11): p. 1078-85.

139. Uzoma, I., et al., *Global Identification of Small Ubiquitin-related Modifier (SUMO) Substrates Reveals Crosstalk between SUMOylation and Phosphorylation Promotes Cell Migration.* Mol Cell Proteomics, 2018. **17**(5): p. 871-888.

140. Anderson, N.L. and N.G. Anderson, *Proteome and proteomics: new technologies, new concepts, and new words.* Electrophoresis, 1998. **19**(11): p. 1853-61.

141.	Anastasi, E., et al., *Expression of Reg and cytokeratin 20 during ductal cell differentiation and proliferation in a mouse model of autoimmune diabetes.* European Journal of Endocrinology, 1999. **141**(6): p. 644-652.

142.	Zhang, W.H., et al., *Advances on diagnostic biomarkers of pancreatic ductal adenocarcinoma: A systems biology perspective.* Comput Struct Biotechnol J, 2020. **18**: p. 3606-3614.

143.	Chen, Y., et al., *Bcl2-associated athanogene 3 interactome analysis reveals a new role in modulating proteasome activity.* Mol Cell Proteomics, 2013. **12**(10): p. 2804-19.

144.	Ross, A.B., J.D. Langer, and M. Jovanovic, *Proteome Turnover in the Spotlight: Approaches, Applications & Perspectives.* Mol Cell Proteomics, 2020.

145.	Blakeley, P., et al., *Investigating protein isoforms via proteomics: a feasibility study.* Proteomics, 2010. **10**(6): p. 1127-40.

146.	Peng, Z., et al., *Integration of proteomic and transcriptomic profiles reveals multiple levels of genetic regulation of salt tolerance in cotton.* BMC Plant Biol, 2018. **18**(1): p. 128.

147.	Breen, M.S., et al., *Temporal proteomic profiling of postnatal human cortical development.* Transl Psychiatry, 2018. **8**(1): p. 267.

148.	Kondethimmanahalli, C., H. Liu, and R.R. Ganta, *Proteome Analysis Revealed Changes in Protein Expression Patterns Caused by Mutations in Ehrlichia chaffeensis.* Front Cell Infect Microbiol, 2019. **9**: p. 58.

149.	Meissner, F. and M. Mann, *Quantitative shotgun proteomics: considerations for a high-quality workflow in immunology.* Nat Immunol, 2014. **15**(2): p. 112-7.

150.	Dapic, I., et al., *Proteome analysis of tissues by mass spectrometry.* Mass Spectrom Rev, 2019. **38**(4-5): p. 403-441.

151.	Mo, R., et al., *Acetylome analysis reveals the involvement of lysine acetylation in photosynthesis and carbon metabolism in the model cyanobacterium Synechocystis sp. PCC 6803.* J Proteome Res, 2015. **14**(2): p. 1275-86.

152. Chen, Z., et al., *Phosphoproteomic analysis provides novel insights into stress responses in Phaeodactylum tricornutum, a model diatom.* J Proteome Res, 2014. **13**(5): p. 2511-23.

153. Rinfret Robert, C., et al., *Interplay of Ubiquitin-Like Modifiers Following Arsenic Trioxide Treatment.* J Proteome Res, 2020. **19**(5): p. 1999-2010.

154. Kanshin, E., et al., *A cell-signaling network temporally resolves specific versus promiscuous phosphorylation.* Cell Rep, 2015. **10**(7): p. 1202-14.

155. Lamoliatte, F., et al., *Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling.* Nat Commun, 2014. **5**: p. 5409.

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

157. Wang, X., et al., *SILAC-based quantitative MS approach for real-time recording protein-mediated cell-cell interactions.* Sci Rep, 2018. **8**(1): p. 8441.

158. Zhang, G. and T.A. Neubert, *Use of stable isotope labeling by amino acids in cell culture (SILAC) for phosphotyrosine protein identification and quantitation.* Methods Mol Biol, 2009. **527**: p. 79-92, xi.

159. Yan, Y.L., et al., *DPPA2/4 and SUMO E3 ligase PIAS4 opposingly regulate zygotic transcriptional program.* PLoS Biol, 2019. **17**(6): p. e3000324.

160. Varejao, N., et al., *DNA activates the Nse2/Mms21 SUMO E3 ligase in the Smc5/6 complex.* EMBO J, 2018. **37**(12).

161. Lamoliatte, F., et al., *Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification.* Nat Commun, 2017. **8**: p. 14109.

162. Liebelt, F. and A.C. Vertegaal, *Ubiquitin-dependent and independent roles of SUMO in proteostasis.* Am J Physiol Cell Physiol, 2016. **311**(2): p. C284-96.

163. Perry, J.J., J.A. Tainer, and M.N. Boddy, *A SIM-ultaneous role for SUMO and ubiquitin.* Trends Biochem Sci, 2008. **33**(5): p. 201-8.

164. Castoralova, M., et al., *SUMO-2/3 conjugates accumulating under heat shock or MG132 treatment result largely from new protein synthesis.* Biochim Biophys Acta, 2012. **1823**(4): p. 911-9.

165. Lygirou, V., M. Makridakis, and A. Vlahou, *Biological sample collection for clinical proteomics: existing SOPs.* Methods Mol Biol, 2015. **1243**: p. 3-27.

166. Lee, J., W.M. Garrett, and B. Cooper, *Shotgun proteomic analysis of Arabidopsis thaliana leaves.* J Sep Sci, 2007. **30**(14): p. 2225-30.

167. Bass, J.J., et al., *An overview of technical considerations for Western blotting applications to physiological research.* Scand J Med Sci Sports, 2017. **27**(1): p. 4-25.

168. Kanshin, E., M. Tyers, and P. Thibault, *Sample Collection Method Bias Effects in Quantitative Phosphoproteomics.* J Proteome Res, 2015. **14**(7): p. 2998-3004.

169. de Araujo, M.E., G. Lamberti, and L.A. Huber, *Homogenization of Mammalian Cells.* Cold Spring Harb Protoc, 2015. **2015**(11): p. 1009-12.

170. Roberts, A.V., *The use of bead beating to prepare suspensions of nuclei for flow cytometry from fresh leaves, herbarium leaves, petals and pollen.* Cytometry A, 2007. **71**(12): p. 1039-44.

171. Glatter, T., E. Ahrne, and A. Schmidt, *Comparison of Different Sample Preparation Protocols Reveals Lysis Buffer-Specific Extraction Biases in Gram-Negative Bacteria and Human Cells.* J Proteome Res, 2015. **14**(11): p. 4472-85.

172. Kohn, A., *Lysis of frozen and thawed cells of Escherichia coli by lysozyme and their conversion into spheroplasts.* J Bacteriol, 1960. **79**: p. 697-706.

173. Weston, L.A., K.M. Bauer, and A.B. Hummon, *Comparison of bottom-up proteomic approaches for LC-MS analysis of complex proteomes.* Anal Methods, 2013. **5**(18).

174. Laemmli, U.K., *Cleavage of structural proteins during the assembly of the head of bacteriophage T4.* Nature, 1970. **227**(5259): p. 680-5.

175. Smith, B.J., *SDS Polyacrylamide Gel Electrophoresis of Proteins.* Methods Mol Biol, 1984. **1**: p. 41-55.

176. Olsen, J.V., S.E. Ong, and M. Mann, *Trypsin cleaves exclusively C-terminal to arginine and lysine residues.* Mol Cell Proteomics, 2004. **3**(6): p. 608-14.

177.   Bornhorst, J.A. and J.J. Falke, *Purification of proteins using polyhistidine affinity tags.* Methods Enzymol, 2000. **326**: p. 245-54.

178.   Zhang, Y., et al., *Protein analysis by shotgun/bottom-up proteomics.* Chem Rev, 2013. **113**(4): p. 2343-94.

179.   Walker, J.M., *The protein protocols handbook*. 3rd ed. 2009, New York, N.Y.: Humana Press. xxxv, 1991 p.

180.   Laskay, U.A., et al., *Proteome digestion specificity analysis for rational design of extended bottom-up and middle-down proteomics experiments.* J Proteome Res, 2013. **12**(12): p. 5558-69.

181.   Huesgen, P.F., et al., *LysargiNase mirrors trypsin for protein C-terminal and methylation-site identification.* Nat Methods, 2015. **12**(1): p. 55-8.

182.   Tsiatsiani, L. and A.J. Heck, *Proteomics beyond trypsin.* FEBS J, 2015. **282**(14): p. 2612-26.

183.   Link, A.J. and J. LaBaer, *Off-line desalting of peptide mixtures.* Cold Spring Harb Protoc, 2011. **2011**(3): p. prot5591.

184.   Thibault, F.P.M.a.P., *Identification of SUMOylated and Ubiquitylated Substrates by Mass Spectrometry* in *SUMOylation and Ubiquitination: Current and Emerging Concepts*, V.G. Wilson, Editor. 2019, Caister Academic Press. p. 71-94.

185.   Manadas, B., et al., *Peptide fractionation in proteomics approaches.* Expert Rev Proteomics, 2010. **7**(5): p. 655-63.

186.   Price, P., *Standard definitions of terms relating to mass spectrometry : A report from the committee on measurements and standards of the American society for mass spectrometry.* J Am Soc Mass Spectrom, 1991. **2**(4): p. 336-48.

187.   Pitt, J.J., *Principles and applications of liquid chromatography-mass spectrometry in clinical biochemistry.* Clin Biochem Rev, 2009. **30**(1): p. 19-34.

188.   Pearson, J.D. and F.E. Regnier, *The Influence of Reversed-Phase n-Alkyl Chain Length on Protein Retention, Resolution and Recovery: Implications for Preparative HPLC.* Journal of Liquid Chromatography, 1983. **6**(3): p. 497-510.

189.   Urban, P.L., *Quantitative mass spectrometry: an overview.* Philos Trans A Math Phys Eng Sci, 2016. **374**(2079).

190. Dang, H.V. and F. Marini, *Chemometrics-based Spectroscopy for Pharmaceutical and Biomedical Analysis.* Frontiers in Chemistry, 2019. **7**.

191. Little, J.L., C.D. Cleven, and S.D. Brown, *Identification of "Known Unknowns" Utilizing Accurate Mass Data and Chemical Abstracts Service Databases.* Journal of The American Society for Mass Spectrometry, 2011. **22**(2): p. 348-359.

192. Schneider, M., A. Belsom, and J. Rappsilber, *Protein Tertiary Structure by Crosslinking/Mass Spectrometry.* Trends Biochem Sci, 2018. **43**(3): p. 157-169.

193. Banerjee, S. and S. Mazumdar, *Electrospray Ionization Mass Spectrometry: A Technique to Access the Information beyond the Molecular Weight of the Analyte.* International Journal of Analytical Chemistry, 2012. **2012**.

194. Maier, M., *The effect of acrolein and biodiesel exhaust exposure on the endocannabinoid levels in human plasma and method development for future analysis of oxylipins* in *Chemistry*. 2013, UMEA university.

195. Banerjee, S. and S. Mazumdar, *Electrospray ionization mass spectrometry: a technique to access the information beyond the molecular weight of the analyte.* Int J Anal Chem, 2012. **2012**: p. 282574.

196. Fenn, J.B., et al., *Electrospray ionization for mass spectrometry of large biomolecules.* Science, 1989. **246**(4926): p. 64-71.

197. Hillenkamp, F., et al., *Matrix-assisted laser desorption/ionization mass spectrometry of biopolymers.* Anal Chem, 1991. **63**(24): p. 1193A-1203A.

198. Karas, M., D. Bachmann, and F. Hillenkamp, *Influence of the wavelength in high-irradiance ultraviolet laser desorption mass spectrometry of organic molecules.* Analytical Chemistry, 1985. **57**(14): p. 2935-2939.

199. Dreisewerd, K., *Recent methodological advances in MALDI mass spectrometry.* Anal Bioanal Chem, 2014. **406**(9-10): p. 2261-78.

200. Feenstra, A.D., M.E. Duenas, and Y.J. Lee, *Five Micron High Resolution MALDI Mass Spectrometry Imaging with Simple, Interchangeable, Multi-Resolution Optical System.* J Am Soc Mass Spectrom, 2017. **28**(3): p. 434-442.

201. Clark, A.E., et al., *Matrix-assisted laser desorption ionization-time of flight mass spectrometry: a fundamental shift in the routine practice of clinical microbiology.* Clin Microbiol Rev, 2013. **26**(3): p. 547-603.

202. Yamashita, M. and J.B. Fenn, *Electrospray Ion-Source - Another Variation on the Free-Jet Theme.* Journal of Physical Chemistry, 1984. **88**(20): p. 4451-4459.

203. Ho, C.S., et al., *Electrospray ionisation mass spectrometry: principles and clinical applications.* Clin Biochem Rev, 2003. **24**(1): p. 3-12.

204. Wilm, M., *Principles of electrospray ionization.* Mol Cell Proteomics, 2011. **10**(7): p. M111 009407.

205. Markides, K.G., A, *Advanced information on the Nobel Prize in Chemistry 2002.* 2002.

206. McLuckey, S.A. and J.M. Wells, *Mass analysis at the advent of the 21st century.* Chemical Reviews, 2001. **101**(2): p. 571-606.

207. Scigelova, M. and A. Makarov, *Orbitrap mass analyzer - Overview and applications in proteomics.* Proteomics, 2006: p. 16-21.

208. Domon, B. and R. Aebersold, *Review - Mass spectrometry and protein analysis.* Science, 2006. **312**(5771): p. 212-217.

209. Mallick, P. and B. Kuster, *Proteomics: a pragmatic perspective.* Nature Biotechnology, 2010. **28**(7): p. 695-709.

210. Pawel Lukasz Urban, Y.C.C., Yi-Sheng Wang, *Balancing Acquisition Speed and Analytical Performance of Mass Spectrometry*, in *Time-Resolved Mass Spectrometry: From Concept to Applications*. 2016.

211. Eliuk, S. and A. Makarov, *Evolution of Orbitrap Mass Spectrometry Instrumentation.* Annu Rev Anal Chem (Palo Alto Calif), 2015. **8**: p. 61-80.

212. Geib, T., et al., *Triple Quadrupole Versus High Resolution Quadrupole-Time-of-Flight Mass Spectrometry for Quantitative LC-MS/MS Analysis of 25-Hydroxyvitamin D in Human Serum.* J Am Soc Mass Spectrom, 2016. **27**(8): p. 1404-10.

213. Kingdon, K.H., *A Method for the Neutralization of Electron Space Charge by Positive Ionization at Very Low Gas Pressures.* Physical Review, 1923. **21**(4): p. 408-418.

214. Michalski, A., et al., *Ultra high resolution linear ion trap Orbitrap mass spectrometer (Orbitrap Elite) facilitates top down LC MS/MS and versatile peptide fragmentation modes.* Mol Cell Proteomics, 2012. **11**(3): p. O111 013698.

215. Second, T.P., et al., *Dual-pressure linear ion trap mass spectrometer improving the analysis of complex protein mixtures.* Anal Chem, 2009. **81**(18): p. 7757-65.

216. Douglas, D.J., A.J. Frank, and D. Mao, *Linear ion traps in mass spectrometry.* Mass Spectrom Rev, 2005. **24**(1): p. 1-29.

217. Bantscheff, M., et al., *Quantitative mass spectrometry in proteomics: a critical review.* Anal Bioanal Chem, 2007. **389**(4): p. 1017-31.

218. Zhang, G., D. Fenyo, and T.A. Neubert, *Evaluation of the variation in sample preparation for comparative proteomics using stable isotope labeling by amino acids in cell culture.* J Proteome Res, 2009. **8**(3): p. 1285-92.

219. Morris, S.M., Jr., *Arginine: beyond protein.* Am J Clin Nutr, 2006. **83**(2): p. 508S-512S.

220. Wheatley, D.N., et al., *Single amino acid (arginine) restriction: growth and death of cultured HeLa and human diploid fibroblasts.* Cell Physiol Biochem, 2000. **10**(1-2): p. 37-55.

221. Ong, S.E., I. Kratchmarova, and M. Mann, *Properties of 13C-substituted arginine in stable isotope labeling by amino acids in cell culture (SILAC).* J Proteome Res, 2003. **2**(2): p. 173-81.

222. Bendall, S.C., et al., *Prevention of amino acid conversion in SILAC experiments with embryonic stem cells.* Mol Cell Proteomics, 2008. **7**(9): p. 1587-97.

223. Hilger, M. and M. Mann, *Triple SILAC to determine stimulus specific interactions in the Wnt pathway.* J Proteome Res, 2012. **11**(2): p. 982-94.

224. Ong, S.E. and M. Mann, *A practical recipe for stable isotope labeling by amino acids in cell culture (SILAC).* Nat Protoc, 2006. **1**(6): p. 2650-60.

225. Chen, X., et al., *Quantitative proteomics using SILAC: Principles, applications, and developments.* Proteomics, 2015. **15**(18): p. 3175-92.

226. Pappireddi, N., L. Martin, and M. Wuhr, *A Review on Quantitative Multiplexed Proteomics.* Chembiochem, 2019. **20**(10): p. 1210-1224.

227. Cox, J. and M. Mann, *Quantitative, high-resolution proteomics for data-driven systems biology.* Annu Rev Biochem, 2011. **80**: p. 273-99.

228. Roepstorff, P., *Mass spectrometry based proteomics, background, status and future needs.* Protein Cell, 2012. **3**(9): p. 641-7.

229. Zheng, P., et al., *QUICK identification and SPR validation of signal transducers and activators of transcription 3 (Stat3) interacting proteins.* J Proteomics, 2012. **75**(3): p. 1055-66.

230. Gallien, S., et al., *Targeted proteomic quantification on quadrupole-orbitrap mass spectrometer.* Mol Cell Proteomics, 2012. **11**(12): p. 1709-23.

231. Kim, H.J., et al., *Quantitative Profiling of Protein Tyrosine Kinases in Human Cancer Cell Lines by Multiplexed Parallel Reaction Monitoring Assays.* Mol Cell Proteomics, 2016. **15**(2): p. 682-91.

232. Barthelemy, N.R., et al., *Tau Phosphorylation Rates Measured by Mass Spectrometry Differ in the Intracellular Brain vs. Extracellular Cerebrospinal Fluid Compartments and Are Differentially Affected by Alzheimer's Disease.* Front Aging Neurosci, 2019. **11**: p. 121.

233. Medzihradszky, K.F. and R.J. Chalkley, *Lessons in de novo peptide sequencing by tandem mass spectrometry.* Mass Spectrom Rev, 2015. **34**(1): p. 43-63.

234. Moskovets, E., et al., *Limitation of predictive 2-D liquid chromatography in reducing the database search space in shotgun proteomics: in silico studies.* J Sep Sci, 2012. **35**(14): p. 1771-8.

235. Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry.* Rapid Commun Mass Spectrom, 2003. **17**(20): p. 2337-42.

236. Zhang, J., et al., *PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification.* Mol Cell Proteomics, 2012. **11**(4): p. M111 010587.

237. Hirosawa, M., et al., *MASCOT: multiple alignment system for protein sequences based on three-way dynamic programming.* Comput Appl Biosci, 1993. **9**(2): p. 161-7.

238. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-89.

239. Paulo, J.A., *Practical and Efficient Searching in Proteomics: A Cross Engine Comparison.* Webmedcentral, 2013. **4**(10).

240. Tyanova, S., T. Temu, and J. Cox, *The MaxQuant computational platform for mass spectrometry-based shotgun proteomics.* Nat Protoc, 2016. **11**(12): p. 2301-2319.

241. Elias, J.E. and S.P. Gygi, *Target-decoy search strategy for mass spectrometry-based proteomics.* Methods Mol Biol, 2010. **604**: p. 55-71.

242. LaRossa, R.A., *Transcriptome*, in *Brenner's Encyclopedia of Genetics (Second Edition)*, S. Maloy and K. Hughes, Editors. 2013, Academic Press: San Diego. p. 101-103.

243. Blumenberg, M., *Transcriptome Analysis*, in *Transcriptome Analysis*. 2019, IntechOpen.

244. Viola, I.L. and D.H. Gonzalez, *Chapter 2 - Methods to Study Transcription Factor Structure and Function*, in *Plant Transcription Factors*, D.H. Gonzalez, Editor. 2016, Academic Press: Boston. p. 13-33.

245. Garg, R., et al., *Transcriptome analyses reveal genotype- and developmental stage-specific molecular responses to drought and salinity stresses in chickpea.* Sci Rep, 2016. **6**: p. 19228.

246. Garcia-Sanchez, S., et al., *Candida albicans biofilms: a developmental state associated with specific and stable gene expression patterns.* Eukaryot Cell, 2004. **3**(2): p. 536-45.

247. Liu, F., et al., *Comparison of hybridization-based and sequencing-based gene expression technologies on biological replicates.* BMC Genomics, 2007. **8**: p. 153.

248. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.

249. Kukurba, K.R. and S.B. Montgomery, *RNA Sequencing and Analysis.* Cold Spring Harb Protoc, 2015. **2015**(11): p. 951-69.

250. Costa, V., et al., *RNA-Seq and human complex diseases: recent accomplishments and future perspectives.* Eur J Hum Genet, 2013. **21**(2): p. 134-42.

251.	Chatterjee, A., et al., *A Guide for Designing and Analyzing RNA-Seq Data.* Methods Mol Biol, 2018. **1783**: p. 35-80.

252.	Henson, J., G. Tischler, and Z. Ning, *Next-generation sequencing and large genome assemblies.* Pharmacogenomics, 2012. **13**(8): p. 901-15.

253.	McGettigan, P.A., *Transcriptomics in the RNA-seq era.* Curr Opin Chem Biol, 2013. **17**(1): p. 4-11.

254.	Conesa, A., et al., *A survey of best practices for RNA-seq data analysis.* Genome Biol, 2016. **17**: p. 13.

255.	Ewels, P., et al., *MultiQC: summarize analysis results for multiple tools and samples in a single report.* Bioinformatics, 2016. **32**(19): p. 3047-8.

256.	Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC Bioinformatics, 2011. **12**: p. 323.

257.	Varet, H., et al., *SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data.* PLoS One, 2016. **11**(6): p. e0157022.

258.	Adilijiang, A., et al., *Next Generation Sequencing-Based Transcriptome Predicts Bevacizumab Efficacy in Combination with Temozolomide in Glioblastoma.* Molecules, 2019. **24**(17).

259.	Li, H., et al., *Applications of genome editing technology in the targeted therapy of human diseases: mechanisms, advances and prospects.* Signal Transduct Target Ther, 2020. **5**(1): p. 1.

260.	in *Human Genome Editing: Science, Ethics, and Governance*. 2017: Washington (DC).

261.	Barrangou, R., *The roles of CRISPR-Cas systems in adaptive immunity and beyond.* Curr Opin Immunol, 2015. **32**: p. 36-41.

262.	Verheyden, J.P. and J.G. Moffatt, *Halo sugar nucleosides. IV. Synthesis of some 4',5'-unsaturated pyrimidine nucleosides.* J Org Chem, 1974. **39**(24): p. 3573-9.

263.	Jiang, F. and J.A. Doudna, *CRISPR-Cas9 Structures and Mechanisms.* Annu Rev Biophys, 2017. **46**: p. 505-529.

264. Devkota, S., *The road less traveled: strategies to enhance the frequency of homology-directed repair (HDR) for increased efficiency of CRISPR/Cas-mediated transgenesis.* BMB Rep, 2018. **51**(9): p. 437-443.

265. Lundblad, R.L. and F. Macdonald, *Handbook of biochemistry and molecular biology*. Fifth edition. ed. 2018, Boca Raton: CRC Press, Taylor & Francis Group. xv, 1001 pages.

266. Zee, B.M. and B.A. Garcia, *Discovery of lysine post-translational modifications through mass spectrometric detection.* Essays Biochem, 2012. **52**: p. 147-63.

267. Knipscheer, P., et al., *Ubc9 sumoylation regulates SUMO target discrimination.* Mol Cell, 2008. **31**(3): p. 371-82.

268. Sarge, K.D. and O.K. Park-Sarge, *Detection of proteins sumoylated in vivo and in vitro.* Methods Mol Biol, 2009. **590**: p. 265-77.

269. McManus, F.P., C.D. Altamirano, and P. Thibault, *In vitro assay to determine SUMOylation sites on protein substrates.* Nat Protoc, 2016. **11**(2): p. 387-97.

270. Werner, A., et al., *Performing in vitro sumoylation reactions using recombinant enzymes.* Methods Mol Biol, 2009. **497**: p. 187-99.

271. Dunst, S. and P. Tomancak, *Imaging Flies by Fluorescence Microscopy: Principles, Technologies, and Applications.* Genetics, 2019. **211**(1): p. 15-34.

272. Loren, N., et al., *Fluorescence recovery after photobleaching in material and life sciences: putting theory into practice.* Quarterly Reviews of Biophysics, 2015. **48**(3): p. 323-387.

273. Riss, T.L., et al., *Cell Viability Assays*, in *Assay Guidance Manual*, S. Markossian, et al., Editors. 2004: Bethesda (MD).

274. Huet, O., et al., *NADH-dependent dehydrogenase activity estimation by flow cytometric analysis of 3-(4,5-dimethylthiazolyl-2-yl)-2,5-diphenyltetrazolium bromide (MTT) reduction.* Cytometry, 1992. **13**(5): p. 532-9.

275. Leung, E.Y., et al., *Hormone Resistance in Two MCF-7 Breast Cancer Cell Lines is Associated with Reduced mTOR Signaling, Decreased Glycolysis, and Increased Sensitivity to Cytotoxic Drugs.* Front Oncol, 2014. **4**: p. 221.

276. Pijuan, J., et al., *In vitro Cell Migration, Invasion, and Adhesion Assays: From Cell Imaging to Data Analysis.* Front Cell Dev Biol, 2019. **7**: p. 107.

CHAPTER TWO

2 Proteomic strategies for characterizing ubiquitin-like modifications

Chongyang Li [1,2], Trent Nelson [1,2], Alfred C.O. Vertegaal [3*], Pierre Thibault [1,4*]

[1]Institute for Research in Immunology and Cancer, Université de Montréal, P.O. Box 6128, Station Centreville, Montréal, Québec H3C 3J7, Canada

[2]Molecular biology program, Université de Montréal, Montréal, Québec H3C 3J7, Canada

[3]Cell and Chemical Biology, Leiden University Medical Center, Einthovenweg 20, 2333 ZC, Leiden, The Netherlands

[4]Department of Chemistry, Université de Montréal, Montréal, Québec H3C 3J7, Canada

Authors Contribution:

Introduction (P.T.); Methods (C.L., T.N., P.T.); Results (C.L., A.C.O.V.), Applications (C.L., T.N., P.T.); Reproducibility and data deposition (A.C.O.V.), Limitations and Optimization, Outlook (A.C.O.V.). All authors reviewed and edited the final manuscript. C.L. and T.N. contributed equally and are co- first authors. A.C.O.V. and P.T. contributed equally and are co- last authors.

2.1 Abstract

The modification of proteins by the addition of ubiquitin and ubiquitin-like proteins (UBLs) is involved in a wide range of cellular processes including cell cycle progression, the DNA damage response, endocytosis, cell signaling, autophagy and protein quality control. The UBL family comprises over a dozen structurally related members, with ubiquitin, small ubiquitin-like modifier (SUMO) proteins, NEDD8, ISG15, and FAT10 being the most commonly known. Each UBL is associated with a distinct set of enzymes that alter the architecture and fate of their cognate proteins. UBL-conjugating enzymes add one or more UBLs to lysine and non-lysine acceptor sites on their target proteins, forming a complex distribution of monomeric and polymeric modifications. Different approaches and strategies are available to identify the sites of UBL modification, the types of modification and their dynamics upon various cellular stimuli; these techniques can decipher the complex architecture of UBL substrates and expand our understanding of UBL functions and their importance in cellular homeostasis and human diseases. This Primer will cover the current methods for identifying UBL substrates, their modification sites and UBL chain linkages, and will describe where the application of these methods can be used to gain biological insights into UBL functions.

2.2 Introduction

Post-translational modification of proteins with the small protein ubiquitin or other ubiquitin-like proteins (UBLs) provides a mechanism for the diversification of protein structure and function. Ubiquitin is a highly evolutionarily conserved, 76-amino-acid protein that can alter the turnover, localisation and interactions of its conjugated substrates [1, 2]. Other UBLs include the SUMO proteins (SUMO1, SUMO2 and SUMO3), NEDD8, ISG15, FAT10, ATG12 and URM1. UBLs, including ubiquitin itself, share a common conjugation mechanism whereby an E1 enzyme activates the mature UBL before transferring it to an E2 enzyme, which conjugates the UBL to its target protein with the aid of an E3 ligase [3] (Figure 2-1). The number of proteins involved in conjugation varies significantly depending on the specific UBL; the ubiquitin machinery, for example, comprises two E1 enzymes, ~40 E2 enzymes and >600 E3 ligases, whereas other UBLs typically have a single E1 and E2 enzyme and no more than a dozen known E3 ligases [4]. Protein modification by UBLs is highly dynamic and reversible, and deconjugation is carried out by UBL-specific proteases, which in the case of ubiquitin consist of a family of ~100 Deubiquitylases (DUBs) [5]. With the exception of ATG12, FAT10, and URM1, which are genetically encoded and translated in their mature forms, all UBLs must also be processed by these proteases to generate the C-terminal diglycine motif required for conjugation. UBL proteins have minimal amino-acid sequence similarity to ubiquitin, although they commonly share a five-stranded β-sheet that partially wraps around a central α-helix, known as the β-grasp fold [6] and conjugate to proteins through the formation of an isopeptide bond between the ε-atom of a lysine residue on the target protein and a diglycine motif at the C-terminus of the mature UBL [4] (Figure 2-2). Modifications can also take place at non-lysine sites such as the free amino group of the protein N-terminus and through the formation of thioester bonds with cysteine, or hydroxyester serine and threonine residues [7].
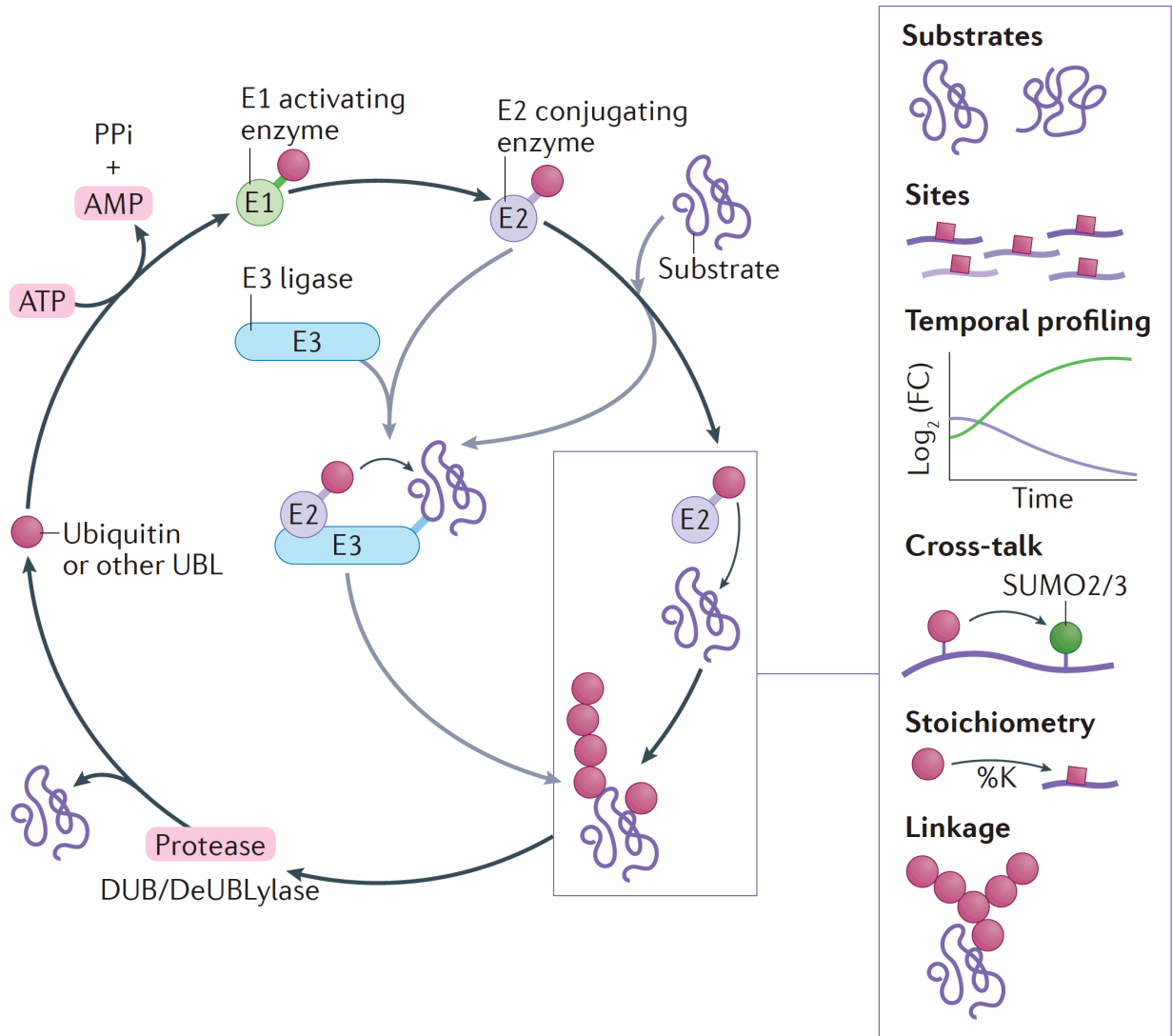
Figure 2-1 Overview of the UBL conjugation machinery and information available from UBL proteomic experiments.

Free ubiquitin-like proteins (UBLs) are activated by an E1 activating enzyme and transferred to an E2 conjugating enzyme, which mediates the conjugation of UBLs to an acceptor lysine of a target substrate via an isopeptide bond, with the aid of E3 ligases. In some cases, E2s can directly modify the substrate without the assistance of an E3 ligase. Cleavage of the isopeptide bond is mediated by deubiquitylases (DUBs), leading to deconjugation of ubiquitin or UBL from the substrate. UBL proteomics approaches can identify modified substrate proteins and their specific modified residues, profile modifications over time to understand the dynamics of modifications, profile crosstalk amongst UBLs and between other modifications such as phosphorylation and acetylation, can determine the percentage occupancy at a given site, and determine the types of polyUBL linkages.

UBLs can form polymeric chains at modification sites. Each monomer may be connected to the next at the same residues (homotypic polymeric chains) or linkage sites can vary between different monomers (heterotypic polymeric chains). The extent of these chains varies considerably between UBLs and the type and structure of these modifications influence protein function, the basis of the ubiquitin code [2]. For ubiquitin, chains can be initiated from the N-terminal moiety of a ubiquitin residue and many different internal lysine residues on the subsequent monomer, including K6, K11, K27, K29, K33, K48 and K63. Chain diversity is further complicated by the fact that heterotypic chains can branch at two or more sites within a single ubiquitin molecule. Polymeric chains have been reported for members of the small ubiquitin-like modifier (SUMO) subfamily [8] and the ubiquitin-like protein NEDD8 [9, 10], and several UBLs can form mixed hybrid chains of different UBLs [11].

Over the past few years, key technological advances in molecular biology, antibody arrays, affinity chromatography and mass spectrometry (MS) have extended our understanding of the structural diversity and function of UBL conjugates. Excellent reviews have previously described the conjugation and deconjugation machinery of UBLs [4, 6], the structure of various UBLs and their functions [3, 12, 13] and the diversity of UBL chain architecture [8, 11, 14]. This Primer describes the current analytical approaches and methods for identifying UBL target proteins, their acceptor sites, the type of UBL modification, the degree of UBL branching, and strategies to quantify changes in the abundance of UBL targets in a site-specific manner (Figure 2-1). Although ubiquitin and SUMO represent the most studied UBLs, this Primer also presents current methods for the analysis of other UBLs and outlines strategies to distinguish substrates modified with different types of UBLs and determine crosstalk between modifications. The application of proteome-wide identification methods is also described for the profiling of UBL targets in response to different cell stimuli. Methods enabling the detailed analysis of UBL chain architecture such as ubiquitin clipping have thus far been described only for ubiquitin and readers

are referred to a recent report for additional information on this approach [15].

Ubiquitin	1	MQI--------------------FVKTLTGKTIT---LE-VEPSD-------TIENV-KAK
SUMO1	1	MSDQEAKPSTEDLGDKKE-GEYIKLKVIGQDSSEIH---FK-VKMTT-------HLKKL-KES
SUMO2	1	MADEKPKEG-----VKTENNDHINLKVAGQDGSVVQ---FK-IKRHT-------PLSKL-MKA
SUMO3	1	MSEEKPKEG-----VKTE-NDHINLKVAGQDGSVVQ---FK-IKRHT-------PLSKL-MKA
NEDD8	1	MLI--------------------KVKTLTGKEIE---ID-IEPTD-------KVERI-KER
ISG15	1	-------------------EPLSILVRNNKGRSST---YE-VRLTQ-------TVAHL-KQQ
FAT10	1	VVKPSDE-----------ELPLFLVESGDEAKRHL---LQ-VRRSS-------SVAQV-KAM
UFM1	1	MSKVS-----------------FKITLTSDPRLPYKVLS-VPEST-------PFTAV-LKF
URM1	1	MAAP----------------LSVEVEFGGGAELL---FDGIKKHRVTLPGQEEPWDIRNLL

Ubiquitin	27	IQDKEGIP-PDQQRLIFAGKQLED------------GRTLSDYNIQKESTLHLVLRLRGG
SUMO1	48	YCQRQGVP-MNSLRFLFEGQRIAD------------NHTPKELGMEEEDVIEVYQEQTGG
SUMO2	44	YCERQGLS-MRQIRFRFDGQPINE------------TDTPAQLEMEDEDTIDVFQQQTGG
SUMO3	43	YCERQGLS-MRQIRFRFDGQPINE------------TDTPAQLEMEDEDTIDVFQQQTGG
NEDD8	27	VEEKEGIP-PQQQRLIYSGKQMND------------EKTAADYKILGGSVLHLVLALRGG
ISG15	29	VSGLEGVQ-DDLFWLTFEGKPLED------------QLPLGEYGLKPLSTVFMNLRLRGG
FAT10	37	IETKTGII-PETQIVTCNGKRIED------------GKMMADYGIRKGNLLFLACYCIGG
UFM1	33	AAFEFKVPAATSAIITNDGIGINP------------AQTAGNVFLKHGSELRIIPRDRVG
URM1	40	IWIKKNLL-KERPELFIQGDSVRPGILVLINDADWELLGELDYQLQDQDSVLFISTLHGG

Figure 2-2 Sequence identity among UBLs.

In mammals, ubiquitin is encoded by four genes (UBB,UBC,UBA52 and RPS27A) translated as head-to-tail concatemers or as fusions with ribosome proteins. Other UBLs, such as UFM1, ISG15, NEDD8, SUMO and ATG8, are expressed as single preproteins. Whole sequence identity among UBLs is low; however, the C-terminal diglycine motif (highlighted in red) is highly conserved, indicating the importance of these residues. In contrast to their modest sequence identity, the structural conservation of these proteins is evident as the β-grasp fold is observed in all UBLs.

2.3 Experimentation

The general workflow of UBL analysis can be divided into different steps that typically comprise sample preparation, enrichment of UBL-modified proteins and/or peptides, mass spectrometry (MS) analysis — usually using liquid chromatography-tandem mass spectrometry (LC-MS/MS) techniques — and computational analyses. These steps are further described below.

2.3.1 Sample preparation

2.3.1.1 Treatments for enhence of UBL modifications

Cells or tissues can serve as a starting point for UBL identification protocols and can be treated with external stimuli to cause the accumulation of UBL modifications that would otherwise be present at levels below the detection limit of MS. For example, treatment of cells with interferon (IFN) or tumor necrosis factor alpha (TNFα) is essential for the enrichment of proteins modified by ISG15 and FAT10 [1, 2]. Further, several studies have employed heat shock or proteasome inhibitors such as MG132 and bortezomib to increase the number of modified substrates and obtain a wider range of UBLs in stress conditions [3, 4]. Whether a treatment is used for a given experiment ultimately depends on the type of UBL and the overall goal of the study.

2.3.1.2 Choice of lysis buffer

Cells or homogenized tissue must be lysed to extract cellular proteins, which can be performed through addition of a lysis buffer or sonication to dissociate and break open cells. The choice of lysis buffer primarily depends on the enrichment steps performed later in the protocol. Protocols that rely on the enrichment of peptide fragments, rather than proteins, or that use affinity tags can use denaturing lysis buffers, such as those containing urea or guanidine. The main advantage of these buffers is that their denaturing properties greatly reduce the activity of cellular proteases, which can remove UBL modifications during sample preparation [16-18]. Strategies that enrich for proteins modified with endogenous UBLs or that enable UBL-interacting

protein identification must use non-denaturing buffers, such as RIPA-based buffers, supplemented with protease inhibitors. Commonly used protease inhibitors include the serine protease inhibitors Pefabloc and Aprotinin; the aspartyl protease inhibitor Pepstatin; the serine/cysteine protease Leupeptin; and N-ethylmaleimide (NEM) — a cysteine protease inhibitor that also inhibits endogenous DUBs [19]. Other, metal-based DUB inhibitors can be used, including copper pyrithione (CuPT), auranofin and gold-based compounds that target the DUBs UCHL5 and USP14 [20-22]. If enriching for NEDD8-modified substrates, 2-orthophenanthroline (OPT) is commonly added to the lysis buffer to inhibit deneddylation. After cell lysis, organelles can be separated using isopycnic centrifugation to enrich cell extracts in specific proteins that may not be detected in total cell lysates. For example, this technique can be used to separate proteins from the rough endoplasmic reticulum (density: 1.20 g/mL) from those of Golgi vesicles (density: 1.14 g/mL) and plasma membrane (density: 1.12 g/mL) [23].

2.3.1.3 Digestion of purified proteins

According to the purification method used, protein extracts are first treated with reducing and alkylating agents (such as chloroacetamide) to cleave disulfide bonds and derivatize cysteine residues to prevent their recombination. Extracts are then digested either in solution or on beads by proteases to produce a pool of peptides for subsequent MS analyses. Trypsin or Lys-C are generally the most popular proteases for UBL identification as these proteases cleave UBL modifications to generate distinctive remnant peptides (Figure 2-3), which can then be targeted by antibodies to allow the immunopurification of peptides with UBL modification sites. This site-specific approach differs to global approaches that do not use any UBL enrichment at the peptide level. For example, cleavage of ubiquitin by trypsin produces a diglycine remnant (εGG), which can be targeted by an anti-εGG antibody to purify peptides corresponding to the ubiquitin modification site. At least two monoclonal antibodies are available for ubiquitin remnant immunoprecipitation; each has been shown to enrich a distinct subset of amino acids flanking the ubiquitylation site and therefore a combination of antibodies should be used to ensure comprehensive analysis of the ubiquitylome [24]. Remnant-specific antibodies must be chemically cross-linked to agarose or magnetic beads using dimethyl pimelimidate (DMP) and

prepared prior to immunopurification; cross-linking efficiency can be determined by comparing pre-crosslinked and post-crosslinked antibody using SDS-PAGE [18, 25]. Peptides can be lyophilized at this stage prior to LC-MS/MS analyses [17, 18]. Before analysing peptides by LC-MS/MS or related methods, offline fractionation using strong cation exchange (SCX) chromatography or high-pH reversed-phase (bRP) fractionation can be performed to increase the depth of UBL identification [17, 18]; this allows for the identification of low-abundance peptides that would remain undetected in unfractionated samples.
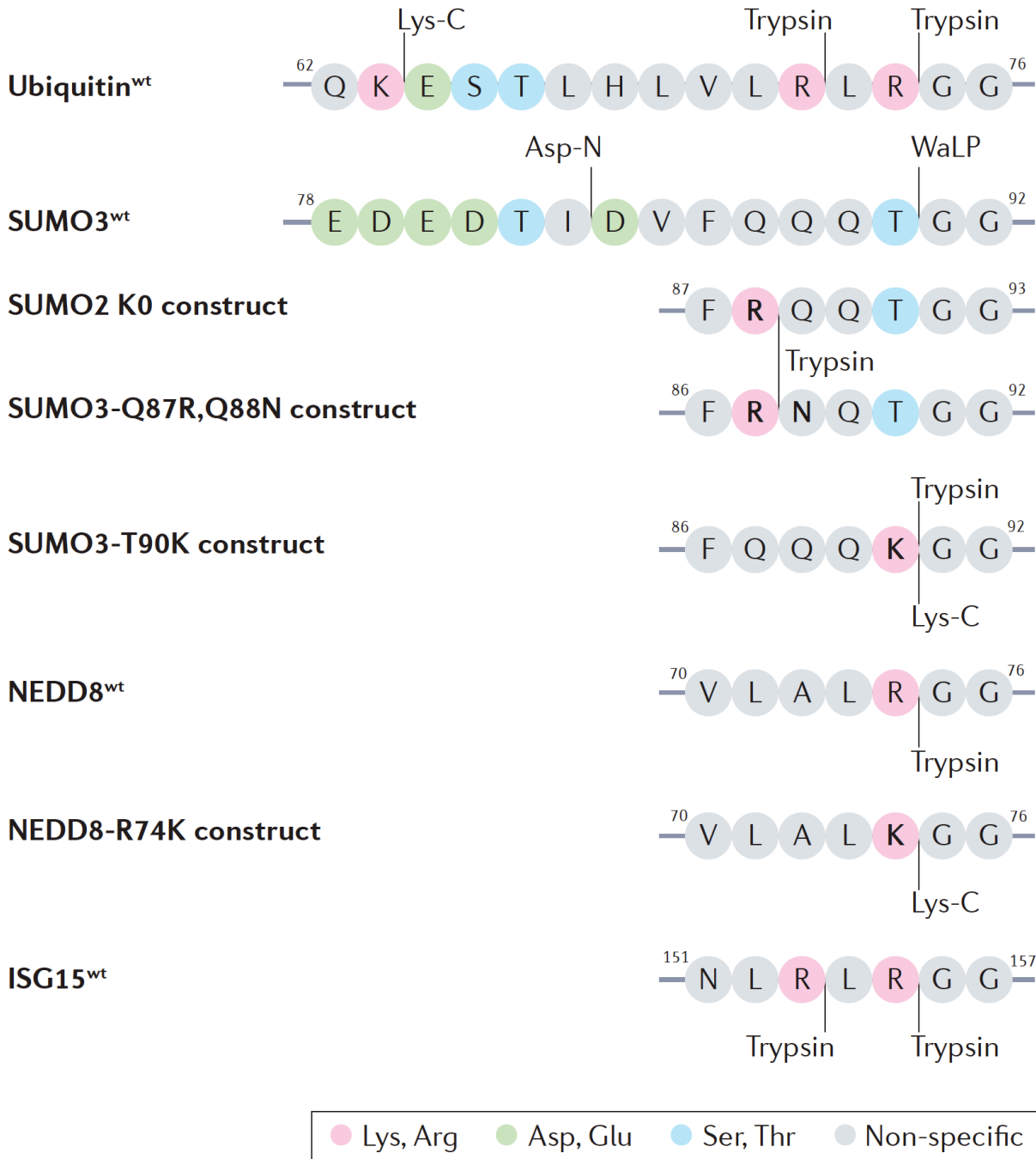
Figure 2-3 Protease specificity for UBL proteomics.

The use of protease enzymes is primarily based on the C-terminal amino acid sequence of the UBL, producing a remnant peptide for further enrichment through immunoprecipitation strategies. Trypsin is often used owing to its specificity and the release of peptides with K/R residues at the C-terminus, facilitating MS/MS sequencing. Exogenous constructs have been generated for several UBLs to produce distinct remnant peptides upon digestion of a protease of choice, as seen with the SUMO2/3 constructs and the NEDD8 R74K construct.

To facilitate the identification of Ubls at the proteome level, an enrichment step is essential, and often requires exogenous gene expression in cells and affinity chromatography. Enrichment of Ubl modified targets typically uses affinity tags to purify Ubl proteins covalently linked to their substrates where the tag of choice (6xHis, 10xHis, Strep or Flag) is appended to the N-terminal of ubiquitin or Ubls by cloning the tag upstream of the ubiquitin/ubl gene into a plasmid [26-28]. An alternative to affinity tags is to use proximity-dependent biotin labeling (BioID) [29]. Cells expressing a Ubl E2-conjugating enzyme [30] or E3 ligase [31] fused to the BirA* tag can be induced to biotinylate the corresponding UBL substrates. These biotinylated proteins are subsequently captured on a streptavidin affinity matrix and identified by MS. To identify the modification site, an extra immunoisolation step is required to enrich the Ubl amino acid remnant tag on the side-chain of the target lysine. For ubiquitin, NEDD8 and ISG15, a diglycine remnant is left on the side-chain lysine after trypsin digestion, and antibodies have been raised against the diglycine motif (αK-GG) [32, 33]. However, this method does not discriminate between these three Ubls since the same diglycine remnant is produced when enriched extracts are digested with trypsin. To overcome this issue, the UbiSite strategy used Lys-C digestion to reveal a 13 amino acid remnant unique to ubiquitin that can be immunopurified with a specific antibody [34]. A recent strategy has been proposed to analyze protein NEDDylation using serial NEDD8-ubiquitin substrate profiling (sNUSP) by replacing arginine 74 of endogenous NEDD8 with lysine via CRISPR-Cas9 [35]. This approach enabled the analysis of NEDDylated peptides by remnant immunoaffinity enrichment of peptides containing the diglycine motif revealed upon Lys-C.

2.3.2 Protein enrichment strategies

The low abundance of UBLs and their target proteins and the dynamic nature of modifications [36] mean that enrichment of modified proteins is essential. Enrichment strategies can be divided into two groups of methods that enable the identification of exogenously expressed or endogenous UBLs or UBL enzymes, described below and summarized in Table 1.

Table 1 Proteomic strategies applied to the identification of ubiquitin and UBL modified targets.

Methods	UBL	Advantages[1]	Limitations[2]	Risk of artefacts[3]	False positives[4]
Exogenous systems					
Affinity purification	Ubiquitin [26-28, 37, 38], SUMO [39-44], NEDD8 [30, 45-49], ISG15 [50], UFM1 [51], FAT10 [52], URM1 [53]	One step purification	Lower throughput than other systems	Yes	Low-Medium
BioID	Ubiquitin [31], SUMO [54], NEDD8 [30]	Identification of weak and/or transient interactions	Cannot distinguish between substrates and interactors	Yes	Low-Medium
Site-directed mutagenesis and peptide immunoprecipitation	SUMO [16, 17, 55, 56], NEDD8 [35]	Site-specific identification	Specific peptide/antibody required	Yes	Low
Endogenous systems					
In vitro UBL assay	Ubiquitin [57], SUMO [25, 57, 58], ISG15 [58], URM1 [58]	Can be either targeted or high thoughput	Requires recombinant cascade enzymes and ATP	Yes	Medium
UBL trap	Ubiquitin [59-64], SUMO [65-68]	Endogenous UBL protein identification	Low enrichment rate	Yes	Medium-High
Proteolytic digestion of UBL modification	Ubiquitin [34, 69], SUMO [69-71], NEDD8 [69], ISG15 [72]	Endogenous UBL identification in a site-specific manner	Low specificity, cannot distinguish between other UBL proteins	No	Medium
Immunoprecipitation of endogenous UBL proteins or UBL peptides after proteolytic digestion	Ubiquitin [15, 17, 18, 24, 26, 32-35, 39, 59, 73-77], SUMO [19, 78-82], NEDD8 [9], ISG15 [83, 84], UFM1 [85, 86]	Site-specific identification	Requires large amount of antibody	No	Low
Protein microarray approaches					
Protein microarrays	Ubiquitin [85, 86], SUMO [85, 87, 88], NEDD8 [85], ISG15 [85], UFM1 [85], FAT10 [85]	Easy sample preparation prior to instrumental identification	Requires recombinant cascade enzymes and ATP	Yes	Medium

[1]All exogenous systems decomplexify samples for LC-MS/MS identification, and all endogenous systems can be applied to any sample and tend to be more time-efficient. [2]All exogenous systems are time consuming and dependent on sample culture, and all endogenous systems are associated with greater interference from non-modified proteins and have lower sequence coverage. [3]'Risk of artefacts' indicates whether artificial UBL modifications on target proteins may be introduced by cell engineering, such as overexpression of a UBL modifier or relative enzymes, or *in vitro* biochemical reaction. [4] 'False positives' indicates the possibility of

identifying proteins without UBL modifications or noncovalent UBL-binding proteins. Levels of nonspecific binding proteins are rated from low to high. Approaches with site-specific identification typically have low false positive rates, owing to the identification of the UBL remnant peptides in addition to the substrate peptides.

2.3.2.1 Enrichment of proteins through exogenous expression

Enrichment of UBL-modified target proteins can be achieved using strategies that exogenously express UBL genes and allow downstream purification of UBL targets by affinity purification. These strategies are required when target UBL substrates are of low abundance. A common approach involves appending a 6xHis-tag, 10xHis-tag or Strep-tag to the N-terminal of the UBL to enable affinity purification of proteins modified by the recombinant UBL. A FLAG-tag can also be used and subsequently immunoisolated using high-affinity monoclonal antibodies. In these cases, tagging is achieved by cloning the tag upstream of the UBL gene in a plasmid [26-28] (Figure 2-4).
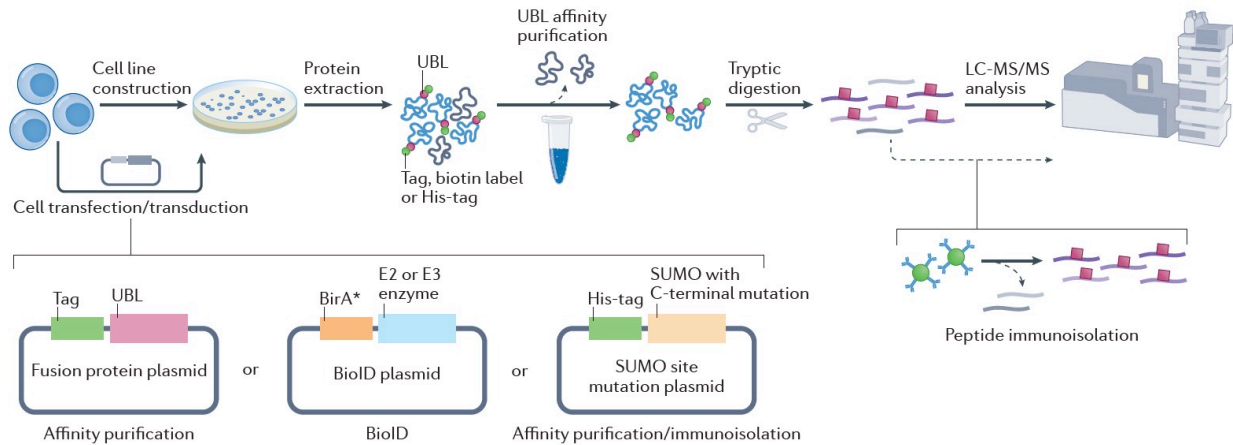
Figure 2-4 Identification of targets of ubiquitin-like proteins (UBLs) using strategies that exogenously express UBLs or related enzymes.

Stable cell lines are generated using plasmids for subsequent affinity purification of conjugated substrates. Plasmids incorporating UBL mutants can aid the immunopurification of modified peptides as they can give distinct remnant peptides following proteolytic digestion. Cellsexogenously expressing tagged ubiquitin/UBL modifiers, or that are tagged with biotin, in the case of BioID, are lysed prior to protein extraction and UBL-modified proteins are subsequently purified and digested with trypsin. UBL remnant immunopurification with one or more specific antibodies can be used to identify modified peptides in a site-specific manner. Peptide fractionation following this last enrichment step can also be applied to enhance the depth of the UBL proteome

analysis. Finally, peptides are analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS). The optional UBL peptide IP step is essential if site-specific information is desired. Made in ©Biorender (biorender.com).

Proximity labelling strategies can be used to identify the substrates of specific UBL-conjugating enzymes. In the proximity-dependent biotin identification (BioID) approach [29], a variant of the *E. coli* biotin ligase BirA (BirA*), is fused to an E2-conjugating enzyme [30] or E3 ligase [31], allowing biotinylation of their corresponding UBL substrates and subsequent capture of biotinylated proteins on a streptavidin affinity matrix before identification by LC-MS/MS (Figure 2-4). Recently, engineered versions of BirA* such as TurboID have been used and markedly reduced labelling time [89]. It should be noted that when using proximity labelling strategies, proximal non-substrate proteins will likely be biotinylated and included in the data, and further validation steps using orthogonal techniques (such as reciprocal tagging of interacting proteins) should be performed to reduce false positives.

For the above protocols, the amount of protein loaded onto affinity or immunopurification columns should be optimized to maximize recovery [16-19]. Binding capacity varies markedly with the type of resin used for affinity purification; for example, the Ni-NTA agarose resin used for purifying His-tag proteins has a high binding capacity (up to 50 mg of protein per mL of resin), whereas the capacity of the FLAG-antibody gel is ~0.6 mg/mL.

Exogenous expression of UBLs can also aid the enrichment of tryptic peptides of interest. For example, determination of SUMOylation sites in the proteome challenging because the tryptic digestion of endogenous SUMO paralogs produces large (26 or 32 amino acid) remnants on the lysine residue of the protein target, which are not amenable to peptide-level immunoprecipitation methods and complicate peptide sequencing. Various SUMO2/3 constructs have been developed to overcome these issues, in which arginine or lysine residues are inserted on the C-terminus of SUMO to create smaller remnants upon tryptic digestion and facilitate the enrichment of SUMO peptides by immunoprecipitation [55, 90-92] (Figure 2-4). Both Lys-C and trypsin are used in the K0 method for identification of SUMO sites, in which a lysine-deficient exogenous His_{10}-SUMO2 construct containing an N-terminal polyHis tag is used to enrich

SUMOylated proteins at a protein and peptide level. The lack of lysine residues in the exogenous SUMO2 construct prevents cleavage in an initial Lys-C digestion, and enables an additional round of SUMOylated protein purification before tryptic digestion to generate SUMOylated peptides that can be analysed by LC-MS/MS [16].

2.3.2.2 Endogenous protein enrichment strategies

A number of methods exist for enriching proteins modified with endogenous UBLs (Figure 2-5). The expression of UBL traps can enrich for modified protein *in vivo*; in these methods, fusion proteins containing UBL-binding domains are expressed in cells to bind UBL conjugates and allow their affinity purification following cell lysis (Figure 2-5). Ubiquitin-specific traps include ubiquitin-activated interaction traps (UBAITs) [93] and targets of ubiquitin ligases identified by proteomics (TULIP) [94]. These traps are linear fusions of E3 ligases and ubiquitin [93, 94]; substrates conjugated to the fused ubiquitin can be co-purified with the E3 ligase of interest and subsequently identified by LC-MS/MS. Traps such as tandem ubiquitin-binding entities (TUBEs) [64]'[62] and tandem hybrid ubiquitin-binding domains (ThUBDs) [95] bind to specific types of ubiquitin chains and can be used for enrichment of modified substrates by affinity purification. Traps for other UBLs include SUMO-traps, which use SUMO-interacting motifs (SIMs) arranged in tandem to capture endogenous SUMOylated proteins [66, 67], and a molecular trapping unit known as the NEDDylator [96]. The NEDDylator approach relies on the high specificity of endogenous neddylation, which mainly targets Cullin proteins [97]; fusing the NEDD8-conjugating E2 enzyme UBC12 to an E3 ubiquitin ligase of interest therefore enables non-physiological neddylation of its substrates. Subsequent purification and LC-MS/MS analysis of neddylated proteins from cells expressing the fusion construct facilitates the identification of E3 ubiquitin ligase substrates. Trap-based methods for ISG15, FAT10, URM1 and UFM1 are currently unavailable and strategies for their characterization rely on the expression of engineered fusion proteins composed of several units of domains from various UBL binding proteins [51-53]. These fusion proteins act as affinity media to purify UBL substrates in a selective manner before proteolytic digestion and MS analysis to identify UBL substrates.
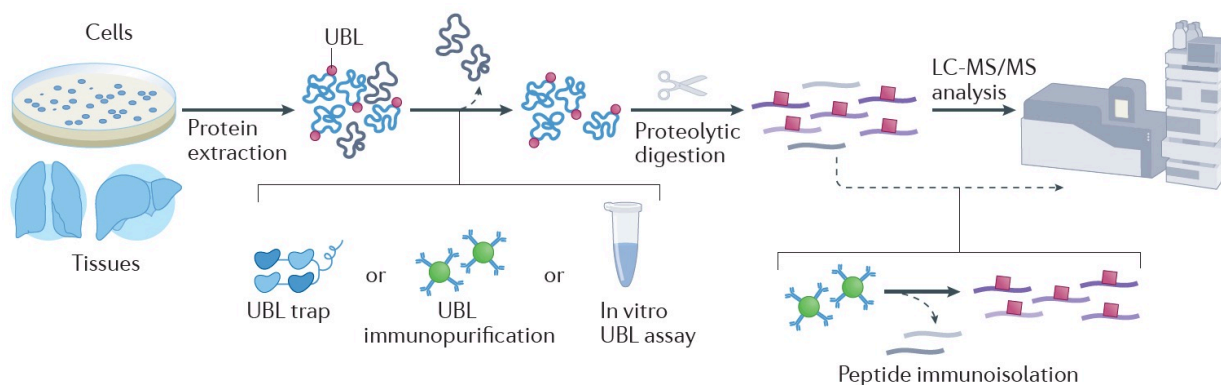
Figure 2-5 Identification of ubiquitin or ubiquitin-like proteins (UBLs) using strategies that enrich endogenous UBL proteins.

In UBL trap and UBL immunopurification strategies, the endogenous ubiquitin/UBL-modified proteins are pulled down following binding to a UBL-specific interaction domain (UBL trap) or using a specific ubiquitin/UBL antibody, respectively, prior to tryptic digestion. In vitro UBL assays are performed on total cell extract using recombinant UBL E1, E2, and E3 enzymes in the presence of ATP to determine modified residues on putative substrates. As with exogenous enrichment strategies, purified proteins are trypsinated and peptides can be purified using remnant-specific antibodies and analyzed by liquid-chromatography-tandem mass spectrometry (LC-MS/MS) to identify sites of modification. A peptide fractionation step can be applied after immunoprecipitation of to increase MS coverage. The optional UBL peptide IP step is essential if site-specific information is desired. Made in ©Biorender (biorender.com).

To facilitate the identification of endogenous UBLs in a global or site-specific manner, proteins can be modified by UBLs *in vitro* by incubating them with a relevant UBL protein, ATP and corresponding E1, E2 or E3 enzymes (Figure 2-5). Performing *in vitro* ubiquitylation [57] or SUMOylation [25] reactions on whole-cell lysates before tryptic digestion and MS identification is a valid strategy to study the UBL proteome in a site-specific manner. The use of protein microarrays and *in vitro* assays have enabled the study of substrates of almost all UBL modifiers on a protein level [85, 87, 88]; however, the results of in vitro assays do not reflect the spatial restrictions of enzymes and substrates in the cell.

MS provides insight into the relative frequency of different ubiquitin-ubiquitin, UBL-UBL and ubiquitin-UBL linkages, but cannot give details on the structure of entire polymers. Recently, a ubiquitin clipping strategy that uses the viral protease Lb[pro] enabled the identification of multiple

branching. In this approach, a trackable diglycine remnant enables the identification of ubiquitin polymer branching points and acceptor lysines on target proteins [15].

2.3.3 Enrichment of modified peptides

Immunoprecipitation of remnant peptides following trypsinization, as mentioned above, can be applied following enrichment of modified proteins to further enrich modified peptide sequences. For ubiquitin and the UBLs NEDD8 and ISG15, a diglycine remnant is left on the side-chain lysine after trypsin digestion, and the anti-εGG antibody can be used to purify the corresponding peptides. However, this method does not discriminate between these three UBLs. The UbiSite strategy can be used to identity ubiquitin remnants, specifically. It uses Lys-C digestion to reveal a 13-amino-acid ubiquitin remnant that can be immunoprecipitated with a specific antibody [34]. The corresponding remnant is left on substrates modified by monoubiquitin (except where the ubiquitin residue has been linked through K63, in which case Lys-C digestion gives rise to a 28-amino-acid remnant). Alternatively, the characterization of polyubiquitin chain structure on modified substrates can be performed using partial trypsinization of protein samples followed by LC-MS/MS analysis [98]. This approach can potentially facilitate the identification of chain linkages on specific acceptor lysine of the substrate. Further, neddylation can be identified using serial NEDD8-ubiquitin substrate profiling (sNUSP), which replaces Arg74 of NEDD8 with a lysine residue using CRISPR-Cas9 [35]; neddylated peptides can then be purified after trypsination using the anti-εGG antibody.

Sequential peptide purifications can be conducted to perform large-scale analyses of crosstalk between UBL modifications and phosphorylation [17, 78, 99]. For example, crosstalk between ubiquitin and SUMO in the context of protein degradation was studied using a two-step immunoprecipitation where ubiquitylated peptides are first purified using an anti-εGG antibody while SUMOylated peptides are isolated from the unretained eluate in a second immunoprecipitation using an εNQTGG antibody [17, 39]. This strategy can be used with any combination of UBLs for which UBL remnant antibodies are available, as long as the remnant peptides of the modifications are not identical.

The use of specific endoproteases and improvements in immunoprecipitation strategies can allow for the identification of SUMO modification sites in endogenous UBL proteomic analysis workflows [100]. The first method developed to identify SUMOylation sites without exogenous expression of a modified SUMO protein used the wild-type alpha-lytic protease (WaLP) [70] (Figure 2-3). This enzyme cleaves preferentially after threonine, alanine, serine and valine residues, resulting in a diglycine remnant motif following cleavage of SUMO modifications. Sequential digestion with different enzymes can also produce relatively small remnants of endogenous SUMO peptides; combinations include trypsin/Lys-C and Glu-C for SUMO1 [71] or Lys-C and Asp-N for SUMO2/3 [78] (Figure 2-3).

2.3.4 LC-MS/MS systems for UBL proteomics

Most large-scale UBL analyses use high-throughput LC-MS/MS to enable protein identification and localization of modification sites, although some studies have used protein microarrays with some success. MS instruments for UBL proteomic studies commonly comprise a quadrupole as a first mass analyzer to select precursor peptide ions (MS1 scan), a collision cell to facilitate ion fragmentation, and either an Orbitrap or a time-of-flight as a second mass analyzer to separate the resulting fragment ions (MS2 scan). Fragmentation through collisional activation is a key step that provides peptide sequence information, facilitating the identification of protein substrates and the site of modification using database search engines.

LC-MS/MS systems can use data-dependent acquisition (DDA) or data-independent acquisition (DIA) modes [101]. In DDA, the most abundant peptide ions detected in MS1 are selected for fragmentation and MS2 analysis. In DIA, peptide ions are sequentially transmitted to MS2, based on a moving window 10–20 m/z wide in MS1, yielding chimeric MS/MS spectra representing multiple peptides. These spectra can then be deconvoluted to correlate individual peptides with their corresponding fragments before peptide sequencing using database search engines. DDA is most commonly used, although recent technological advances in acquisition speed and MS sensitivity make DIA a viable alternative for comprehensive UBL analysis. An approach using the anti-εGG antibody and an optimized DIA method enabled the identification

of 35,000 ubiquitylated peptides following single LC-MS/MS analysis of proteasome inhibitor-treated cells [102]. Although DIA has the potential to identify and quantify protein digests over a large dynamic range in a reproducible manner, further development of software tools for DIA analysis is needed to deconvolute complex spectra and determine false discovery rates.

LC-MS/MS analysis can be carried out following the use of metabolic or isobaric mass labels. We discuss label-free and labelling techniques briefly below, but refer readers to recent reviews on quantitative proteomics for more information [103-106].

2.3.4.1 Label-free quantification

Label-free quantification (LFQ) mass spectrometry is widely used to profile protein abundance and have been successfully applied to different UBL identification strategies [59, 107]. LFQ enables relative protein quantification based on peptide ion abundance, using peptide ion intensities, areas or spectral counts to compare changes across different biological conditions and replicates (Figure 2-6) [108]. LFQ does not require special growth media or chemical derivatization and can be applied to various types of samples including primary material and biological tissues. It is easy to integrate into most experimental workflows and requires that each sample be measured individually. However, the stochastic sampling of peptide ions for MS/MS sequencing can result in undersampling, leading to missing quantification values for low-abundance ions [109].
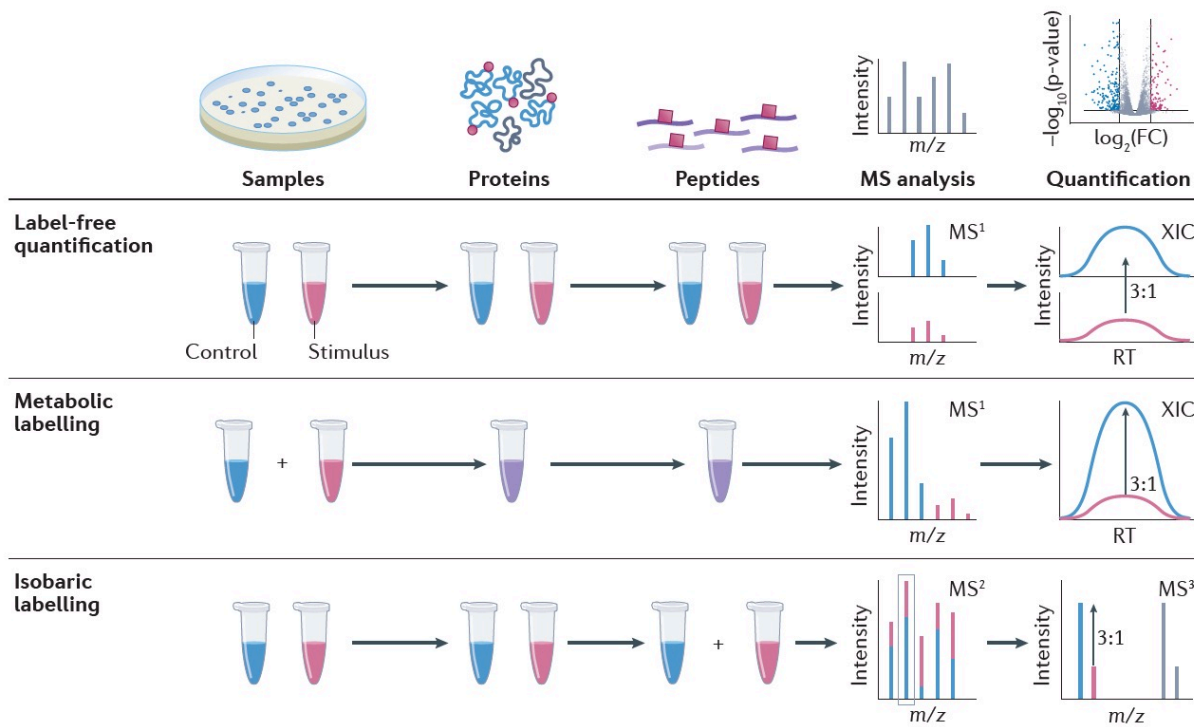
Figure 2-6 Quantitative proteomic strategies applied to UBL proteomics.

Quantification of the UBL proteome can be achieved via three main approaches: label-free quantification, metabolic labelling and isobaric labelling. In label-free quantification, samples are processed and analyzed separately without the use of any labels for the proteins or peptides. Relative quantification of different samples is based on the intensity of the peptide signals in each sample (a 3:1 ratio of control:stimulus is shown as an example). Metabolic labelling approaches such as stable isotope labeling by amino acids in cell culture (SILAC) involves the incorporation of isotopically labelled amino acids into the proteins during cell culture. The inherent difference in mass of the isotopically labelled proteins allows for the mixing of samples early in the sample processing workflow. Quantification is performed within the same sample by comparing the abundance of isotopically labeled peptides. Isobaric labelling strategues such as isobaric tags for relative and absolute quantitation (iTRAQ) and tandem mass tagging (TMT) involve the chemical labelling of free amino groups, resulting in peptide isotopomers of identical nominal mass that yield distinct reporter fragment ions in MS/MS spectra. Up to 16 samples can be compared together using isobaric peptide labeling. XIC, extracted-ion chromatogram. Made in ©Biorender (biorender.com).

2.3.4.2 Stable-isotope labeling by amino acids in cell culture

Stable-isotope labeling by amino acids in cell culture (SILAC) is a simple, reliable and robust MS-based quantitative proteomic approach that can be used in any cell culture system [110] and

has been applied in a pulse chase format to identify ubiquitin CRL4 E3 ligase targets [111]. In the most common variant of SILAC, cells are cultured in light (^{0}Lys, ^{0}Arg), medium (^{4}Lys, ^{6}Arg) or heavy (^{8}Lys, ^{10}Arg) isotopic forms of lysine and arginine (Figure 2-6). Once more than 95% of proteins have incorporated the isotopic labels, cells can be pooled and lysed at an early stage of sample preparation. The quantification of SILAC is based on calculating the ratio of isotope-labeled peptide pairs to light-labeled peptides at the MS1 level; this technique can therefore measure the relative changes in protein abundance between up to three samples within a single MS run. In a technique known as super SILAC, a pooled extract of heavy-labeled cells can be added as a spike-in to an endogenous tissue extract, and MS analysis is performed on the combined sample to minimize sample handling variability and experimental biases [112]. Although SILAC is an accurate quantification technique, it is limited to cell lines and can only compare up to three conditions; further, SILAC approaches give higher spectral complexity than LFQ methods, leading to over-sequencing of peptide isotope variants.

2.3.4.3. Isobaric labeling quantification

In isobaric peptide labeling approaches such as isobaric tag for relative and absolute quantitation (iTRAQ) [113] or tandem mass tag (TMT) systems [114], peptides are derivatized with isotopic reagents that confer identical isobaric masses to the corresponding peptides, but yield distinct reporter ions once they are fragmented by MS/MS (Figure 2-6). The labeling reagents comprise a mass reporter group, a mass normalizer group and an amine-reactive group, which reacts with NH_2-terminal and side-chain lysine amino groups. TMT systems enable the simultaneous comparison of up to 16 samples in a single MS/MS run, facilitating a reduction in the amount of starting material needed for UBL enrichment and significantly minimizing the number of missing quantification values across all experimental conditions [115]. TMTs can also be used to perform absolute quantification by adding an internal standard peptide; however, this approach usually yields lower quantification accuracy compared to SILAC owing to ratio compression. This is caused by co-selection of precursor ions, which is especially pronounced when measuring large fold changes. Several techniques have been reported to resolve ratio compression, such as extensive peptide fractionation during sample preparation [116], using the complement tandem mass tag (TMT^c) approach at the MS2 level [117], and MS3-based approaches [118]. For UBL modified peptides, an additional derivatization site on the remnant from the lysine side-chain can complicate the distribution of charge

state, peptide hydrophobicity and thus reporter ion quantification. To overcome these issues, derivatization is conducted directly following immunoprecipitation of the remnants on the beads with the antibody-bound peptides, to protect the primary amine of the UBL remnant from being labeled by the TMT reagents [119].

2.3.4.4 Analysis software

A number of software applications with integrated database-searching engines are available for identifying and quantifying UBL peptides from MS data, including Mascot [120], Proteome discoverer/SEQUEST [121], MaxQuant/Andromeda [122] and PEAKS studio [123]. MaxQuant is a freely available software package that can match fragment ions from UBL remnant chains after specifying the residue composition and can also calculate site localization probability.

2.3.5 Computational approaches

Computational approaches can be used to predict UBL sites and corresponding motifs. UbiBrowser provides an integrated bioinformatics platform for predicting ubiquitin E3-substrate interaction network in the human proteome [124]. GPS-SUMO (previously SUMOsp) [125, 126] integrates the GPS [127] and MotifX [128] tools to provide an *in silico* prediction of SUMOylation sites and SIMs [129]. JASSA also predicts SUMOylation sites and SIMs, based on a position weight matrix [130]. SumSec [131] and HseSUM [132] are both machine learning-based methods that use the predicted secondary structure of amino acids and half-sphere exposure to improve prediction capability. SUMO-Forest [133] and mUSP [134] are two of the most recent SUMO prediction software based on deep learning. Similar tools are available for ubiquitylation [135, 136] and neddylation [137]. Finally, PhosphoSitePlus [138] integrates both low-throughput and high-throughput experimental data and provides a host of online tools to study PTMs including ubiquitylation, SUMOylation and neddylation. There are currently limited site-prediction softwares for other UBLs such as ISG15, FAT10, URM, and UFM1. Most of these approaches use algorithms to predict UBL sites or motifs, except for PhosphoSitePlus which provides experimental data on UBL sites. All approaches can complement experimental approaches in order to identify novel sites that are of low abundance and not identified in high throughput

proteomics studies alone. Confidence in the prediction of UBL proteins and sites can vary depending on the parameter settings used in the different computational approaches.

2.4 Results

Proteome-wide analyses of UBL networks in a site-specific and site-independent manner result in large sets of raw MS data that require thorough analysis to identify modified lysines, determine quantitative changes in modifications and identify which proteins and related cell process are regulated by these PTMs. We discuss these approaches below.

2.4.1 Identifying modified lysines

The immunopurification of diglycine-containing peptides, an approach originally developed for ubiquitin-modified substrates, has now been adapted to study other UBLs such as SUMO and NEDD8 [9, 56, 92]. UBL-modified lysines are identified from their corresponding diagnostic mass remnants in the MS2 scan and must be specified in the search parameters of the analysis software [34, 78, 91]. The MS/MS spectrum of a tryptic peptide from the promyelocytic leukemia protein (PML) with an ubiquitylated site at K490 is shown in Figure 2-7; in this specific case, the modification can be deduced from a mass difference of 242.1374 Da between the y_8 and y_7 fragment ions, corresponding to the combined mass of a lysine (128.0950 Da) and a diglycine remnant of ubiquitin (114.0429 Da). Detailed protocols are available for the identification of UBLs such as ubiquitin and SUMO using different search engines [17, 18] It is key to consider fragments that have resulted from missed cleavages or chemical conversion of remnants; for example, pyroQ — a cyclic derivative of glutamate — is present in some SUMO remnants, including pyro-QQTGG (pQQTGG), which has a monoisotopic mass of 454.18121 [90]. Different free software packages are available to analyse raw MS data including MaxQuant [122, 139, 140] and PEAKS [141] as well as commercial software from MS manufacturers.
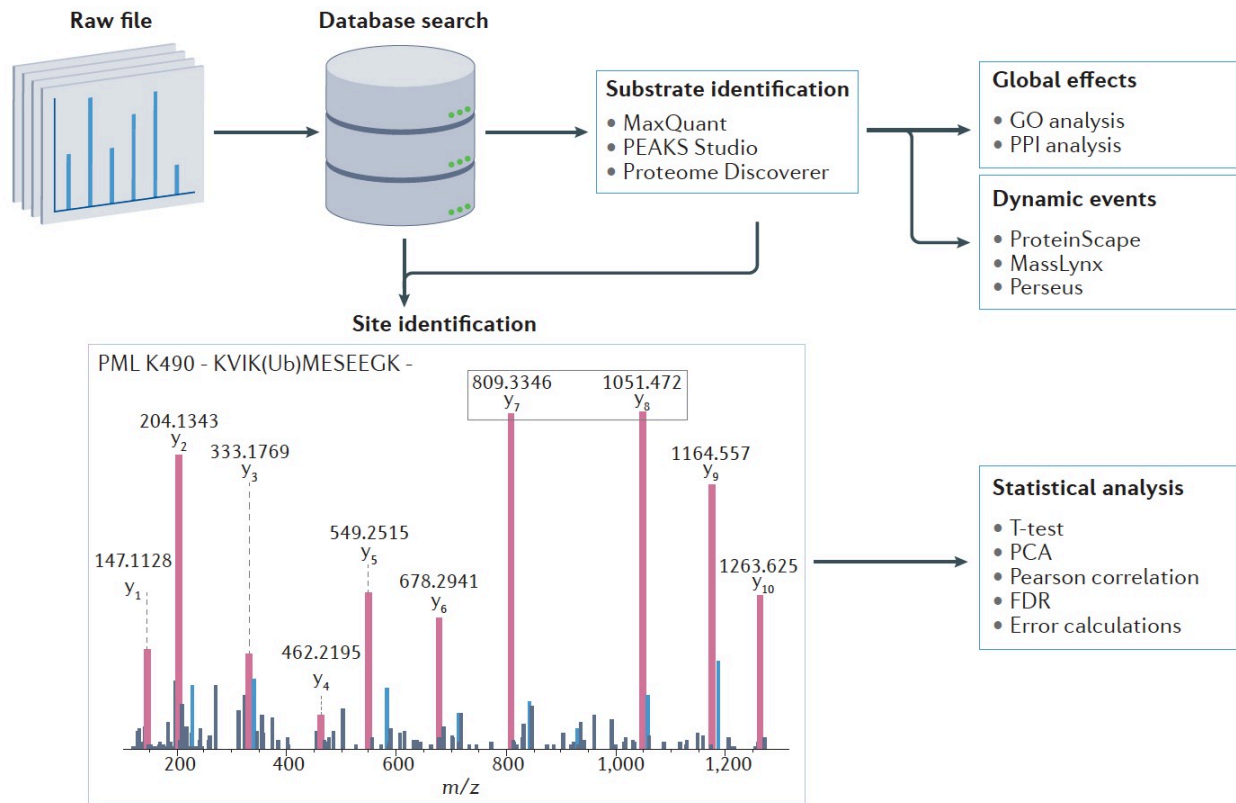
Figure 2-7 Data analysis workflow and example of MS/MS spectra.

MS data is first processed through a background proteome database search to identify proteins and their modifications. Several analyses can be performed to better understand important aspects of the data. Gene Ontology (GO) and protein-protein interaction (PPI) analyses can provide insight into the impact of site modifications by identifying interacting proteins associated with specific cellular functions. MS/MS spectra can enable peptide sequencing and identify the site of modification. The example spectrum shows an MS/MS spectra for human promyelocytic leukaemia protein (PML K490); the fragment ion harboring the UBL-specific remnant can be inferred from the large mass difference of 242.1374 Da between the y8 and y7 fragment ions, corresponding to a lysine (128.0950) and diglycine remnant (114.0429) of ubiquitin. Statistical analysis of MS data is imperative to confidently assign modification sites. Global analysis of dynamic events where stimulation promotes or removes modifications can be readily performed using software such as MassLynx and Proteinscape.

2.4.2 Identifying dynamic events

Dynamic alterations in UBL-modified proteomes in response to stimuli can be addressed in a site-specific manner. However, site-specific approaches can miss out on relevant targets owing to the generation of tryptic peptides with a single modified lysine residue that are too small or too large for MS analysis. The sequence of small peptides can match to large sets of reference proteins, confounding the identification of the protein of origin from the peptide sequence. Larger peptides are less efficiently ionized during MS/MS analysis and can remain undetected; in these cases, it is optimal to use several different proteases for cleavage [142].

Site-independent enrichments enrich full-size intact target proteins of UBLs and therefore yield multiple different peptides per target protein and can therefore give more accurate estimates of modification dynamics than site-specific approaches that only enrich peptides that contain the modification site and therefore provide lower number of peptides for quantification. In datasets comprising large numbers of MS runs, dynamic events can be studied by combining MS/MS data with MS data to allow matching of peptides across runs. This site-independent methodology was recently used to identify a decrease in the SUMOylation of the transcription factor TFAP2A in response to hypoxia, enhancing the transcriptional activity of HIF-1α [143]. Software packages that enable the identification of dynamic events include MaxQuant [140], Perseus [144], Peaks [141] and commercial packages from manufacturers of mass spectrometers including Proteome Discoverer, MassLynx and ProteinScape (Figure 2-7). Machine and deep learning techniques have become powerful approaches for predicting UBL modification sites and motifs, with a prediction accuracy of over 85% for ubiquitylated motifs [145, 146].

2.4.3 Identifying global effects of UBLs

Global analysis of UBL signalling networks frequently yield large datasets, which can be analyzed using Gene Ontology (GO) analysis (Figure 2-7) [147]. Sets of ubiquitin or UBL-modified targets are matched to known networks of physically and functionally related proteins; this enables the identification of cellular processes regulated by UBL modifications. For example, knockdown of E1 SUMO activating enzymes leads to the disruption of SUMO signalling relating

to nuclear processes, resulting in a delay in mitosis and defects in mitotic chromosome separation that was correlated through GO terms analysis [148].

The specific enrichment of protein domains can be found using the PFAM resource [149] and members of larger protein complexes can be identified and visualized using STRING analysis [150]. These analyses can be applied to any UBL dataset as the input is dependent on the proteins themselves and not the enriched UBL modification. Softwares such as PFAM and STRING are particularly useful given that PTMs frequently co-regulate larger sets of proteins, a concept known as group modification [151]. Indeed, modifications might have little effect on a single protein where co-regulation of a larger protein group is of biological importance.

2.4.4 Finding UBL ligase/protease interactions

Understanding the organization of the ubiquitin enzyme-substrate network is complicated by the huge number of different ubiquitin ligases and proteases. One approach to uncover novel interactions between UBL enzymes and target proteins is to compare ubiquitin target proteomes following knockdown or overexpression of specific UBL enzymes or using pharmacological UBL enzyme inhibitors, although achieving full specificity for inhibitors is rare. These approaches do not necessarily reveal direct substrates owing to the presence of enzymatic cascades. For example, E3 ligases can regulate other E3 ligases; this cascade has been extensively studied in the context of the DNA damage response and includes the E3 ligases RNF8, RNF168, HERC2, BRCA1/BARD1 [152, 153]. Further, multiple E3 ligases can regulate the same target protein and therefore the above approaches can yield false negative results owing to redundancy.

The identification of UBL protease substrates is challenging. These proteases can themselves be modified by ubiquitin and UBLs; therefore, simply knocking them down or inhibiting them does not necessarily result in enhanced modification of their target proteins and could affect other modified UBL and UBL proteases, leading to indirect effects. Multiple complementary approaches are required to establish enzyme-substrate relationships — for example, one can combine an approach that determines proteome-wide changes in UBL-modified targets following inhibition, knockdown or knockout of a protease with a proximity labelling approach such as

BioID [29] that enables the identification of proteins directly interacting with the protease of interest; proteins identified using both approaches will most likely be direct substrates of the ubiquitin or UBL protease of interest. Proximity labelling alone can also be used to identify interacting proteins of the UBL network, as demonstrated for SENP2, where BioID revealed that this SUMO protease interacts with proteins associated with the endoplasmic reticulum, Golgi, and inner nuclear membrane [54].

2.4.5 Determining ubiquitin targets and polymers

The development of site-specific enrichment methodologies was a major advance in studying the ubiquitin-modified proteome [154] and enabled the identification of ~17,000 modification sites [32]. Prior to this, a landmark study on ubiquitin chain linkages revealed that all linkages accumulate upon inhibition of the proteasome with the exception of the K63 linkage, indicating that all other linkages play a role in proteasomal protein degradation [33]. The UbiSite technology was developed to specifically identify ubiquitin diglycine remnants [34]. UbiSite datasets are searched to identify diglycine-modified lysines that represent ubiquitylation sites; corresponding MS/MS spectra show a size-shift of the lysine caused by the extra mass derived from the diglycine modification. This technology is particularly suited for conditions where large amounts of starting material are available. Analogous approaches have been developed to enrich and identify conjugation sites for SUMO using antibodies that recognize C-terminal proteolytic remnants of SUMO, and identified nearly 1,000 SUMOylation sites on 500 target proteins [39]. The methodology was used to reveal dynamic changes in protein SUMOylation in a site-specific manner upon proteasome inhibition. Moreover, recent studies have used genetic models in combination with GlyGly peptidomics to describe ISGylomes in the context of listeria infection in mice lacking ISG15 [83] and in cancer cells lacking USP18 (the main deISGylating enzyme *in vivo*) [155]. Further, ubiquitin clipping technology is particularly useful for the analysis of homotypic and heterotypic Ub polymers and provides information on the relative frequency of Ub polymer branching as it can detect multiple diglycine-modified lysines on a single ubiquitin molecule [15].

2.4.6 Statistical analysis

Statistical analysis must be performed to minimize false assignments of ubiquitin and UBL modifications. Individual software packages including Mascot [120], Proteome discoverer/SEQUEST [121], MaxQuant/Andromeda [122] and PEAKS studio [123], employ package-specific scoring methods to ensure that spectra are reliably annotated (Figure 2-7). Scoring methods usually correlate the fragment ions observed in the MS/MS spectrum with those predicted from *in silico* fragmentation of database sequence candidates that match the peptide mass. Optimal fragmentation of peptides is required to obtain sets of fragments that correspond to b-type and y-type fragment ion series, respectively from the N-terminal and C-terminal of the peptide [156]. A score is then calculated and used to rank all sequence candidates. Error probabilities can be computed for peptides based on the spectrum score as described above, the length of the peptide, the number of missed cleavages, the number of modifications and the charge state. Spectrum scores for individual peptides can be combined to obtain scores at the protein level. Further, the presence of diagnostic mass remnants in MS/MS spectra, such as diglycine remnants for ubiquitin modifications, can be used to limit false positives. To further limit false positives, a database containing both forward and reversed sequences of all proteins is used to determine the score cutoff, at which a preset false discovery rate is obtained. UniProt is typically used as a reference database as it provides a high-quality and freely accessible resource of protein sequences. False discovery rates should be set to ≤1 % according to common practice in the MS field and manual validation is helpful to further filter the data; we refer readers to a detailed technical overview for reducing false positive identifications of ubiquitin sites [157].

Peptide intensities should be normalized to compare peptide and protein abundances across samples in all forms of quantitative proteomic experiments. MaxQuant will automatically normalize all peak areas during analysis. Other normalization approaches include normalizing for total intensities of chromatograms, normalizing based on the median peptide intensities of each sample, normalizing based on reference signals or normalizing based on spiked internal standards [103, 158, 159]. Internal peptides of UBLs can be employed for normalization when equal amounts of total UBL conjugates are compared in different samples.

2.5 Applications

The system-level identification of UBL substrates and their modifications has provided valuable biological insights into UBL functions, which influence nearly all aspects of eukaryotic biology. The ability to define acceptor sites, the types of branching and the domains on which these modifications are located on target substrates in response to cell stimuli has greatly extended our understanding of UBL regulation, and some examples of applications are outlined below.

2.5.1 Interrogating UBL crosstalk

UBLs have distinct enzymatic machinery, including specific E1, E2 and E3 enzymes and unique proteases, and cannot compensate for each other's absence. Despite their distinct functions, UBLs act together in large networks by co-modifiying target proteins, forming heterotypic polymers and modifying each other's enzymatic machinery (Figure 2-8) [160]. Crosstalk can be interrogated with ubiquitin proteomics techniques; for example, the affinity purification of HA-tagged FAT10 was used to show that this UBL can bind to the SUMO E1 and block SUMOylation [161]. Further, remnant immunopurification was recently used to show that ISGylation and SUMOylation are increased in response to interferon stimulation, a process that requires activation of the ISG15 E3 ligase TRIM25 [162].
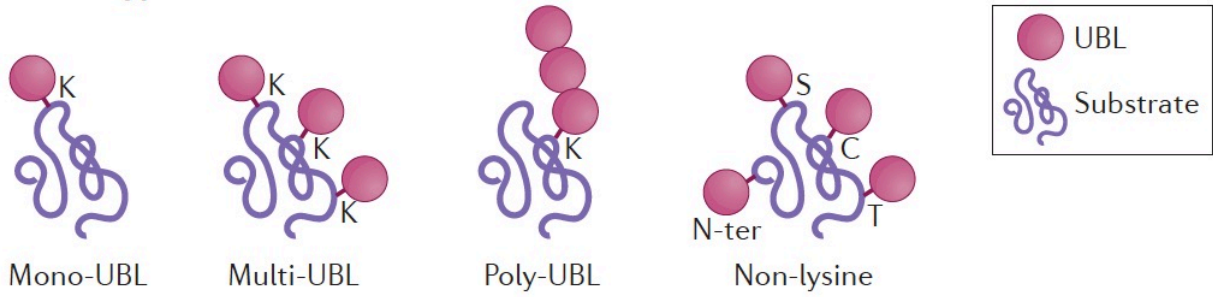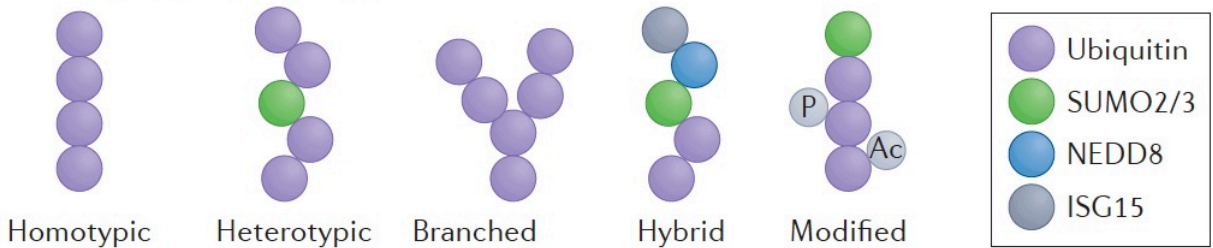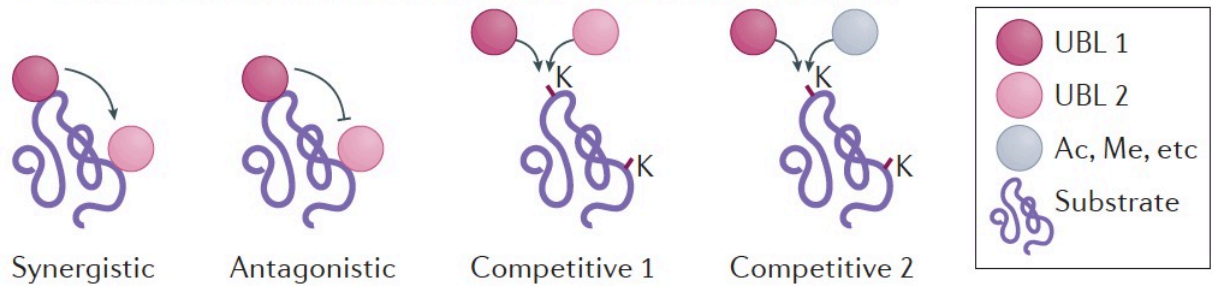
Figure 2-8 Crosstalk between ubiquitin and ubiquitin-like proteins (UBLs).

(a) UBL modifications can occur in a variety of forms, including mono-UBL, multi-UBL, poly-UBL and N-terminal modifications, and each can lead to a unique functional outcome. (b) Complex Ubl chain topology allows for a variety of outputs. Polymers of UBL chains vary in structure (homotypic, heterotypic, branched) and mixed chains can exist, which contain UBLs of different types (hybrid). Further, UBLs can themselves be modified by post-translational modifications such as phosphorylation and acetylation (modified). The mixing and modification of UBLs highlights the extensive crosstalk between PTMs. (c) Crosstalk between UBL modifications can occur between two sites by either facilitating (synergistic) or impeding (antagonistic) the modification of a nearby site. Further, crosstalk can occur on the same site between UBLs (Competitive 1) or other lysine modifications such as acetylation or methylation (Competitive 2).

The type of UBL chains appended to a substrate triggers a distinct outcome. Homotypic and heterotypic ubiquitin polymers have different capacities for mediating protein degradation, changes in substrate activity, localization and other non-proteolytic events [11, 163, 164]. Further, UBL chains can be phosphorylated, acetylated, and modified by other types of UBL to form complex hybrid chains (Figure 2-8b). Ubiquitin is extensively modified by other PTMs through polyubiquitylation, SUMOylation, neddylation, ISGylation, phosphorylation and acetylation. Phosphorylation of ubiquitin Ser65 by the kinase PINK1, for example, has been shown to interact with the ubiquitin E3 ligase Parkin and ultimately generates a feed-forward mechanism for activating Parkin ubiquitin E3 ligase activity at the mitochondria [165-169].

Site-directed ubiquitin proteomics approaches have shown different UBLs can modify the same receptor lysine or different lysines of the same substrates, leading to competitive, synergistic or antagonistic interactions (Figure 2-8c) [170-172]. A two-step enrichment protocol for purifying target proteins modified by both ubiquitin and UBLs [39, 173] showed that the overwhelming majority of proteins modified by SUMO and ubiquitin are subsequently degraded by the proteasome. Different affinity purification methods have further been used to identify heterotypic polymers such as ubiquitylated SUMO, SUMOylated ubiquitin, neddylated ubiquitin and ubiquitylated NEDD8 [11].

One of the first examples of crosstalk between UBL modifications and ubiquitin/UBL enzymatic components was the SUMOylation of the ubiquitin-conjugating enzyme E2 K (UBE2K). SUMOylation was shown to inhibit enzymatic activity by blocking its interaction with the ubiquitin-activating E1 enzyme (UBA1) [174]. Studying crosstalk is important as it has been implicated in mediating the expression of oncogenes; for example, SUMO and ubiquitin modifications are both required for efficient proteasomal degradation of the p63 isoform ΔNp63α, showing cooperation between these UBLs [175]. In the context of genome stability and tumor development, SUMOylation of the BRCA1/BARD1 dimer by the E3 SUMO-protein ligase PIAS1 has been shown to strongly enhance its E3 ubiquitin-protein ligase activity [94]. Another interesting functional example is the targeting of SUMOylated proteins by the ubiquitin protease

USP7 to control replication [176]. Proteomics approaches have unearthed extensive lists of this type of crosstalk between ubiquitylation and SUMOylation [34, 177].

SUMO-targeted ubiquitin ligases (STUbLs) are a well-known example of crosstalk between ubiquitin and UBL proteins. These ubiquitin E3 ligases contain SUMO-Interaction Motifs (SIMs) that enable their targeting to SUMOylated proteins, resulting in the formation of heterotypic chains or modification of separate acceptor lysines on the target protein. The E3 ubiquitin-protein ligase and STUbL RNF4 has been identified as a major regulator of the ubiquitylation and proteolysis of SUMOylated substrates [172, 178, 179]. TULIP has been used to identify novel STUbl and RNF4 target proteins on a system level[94]; intriguingly, many of the identified targets were enzymes of the SUMO conjugation pathway, including the SUMO-conjugating enzyme UBC9 and the E3 SUMO-protein ligases PIAS1, PIAS2, PIAS3, ZNF451, and NSMCE2. STUbLs other than RNF4 have been identified, such as RNF111/Arkadia and RNF216/TRIAD3, suggesting possible substrate specificity among these and potentially other unidentified STUbLs [180, 181]. Using GFP-HA tagged SUMO1/2, a recent report indicated that RNF111/Arkadia preferentially selects substrates modified by SUMO1-capped SUMO2/3 chains [182]. Both RNF4 and RNF111 are important for maintaining genome stability; RNF4 regulates BRCA1/BARD1 [94], replication protein A (RPA) and the DNA repair protein Rad51[183], whereas RNF111 regulates SUMOylated XPC [180, 184]. In addition to STUbls, NEDD8 activates a large number of E3 ubiquitin ligases, especially Cullins [185].

2.5.2 Characterizing UBL modification sites

UbiSite technology has enabled the identification of over 63,000 ubiquitin sites on 9,200 proteins [34] whereas SUMO site enrichment enabled the identification of over 40,000 SUMO sites on over 6,700 proteins [142]. Several studies have identified different types of ubiquitin binding domains (UBDs) [186, 187] and enrichment of acidic residues surrounding sites of diglycine modification has been observed using specific monoclonal antibodies [24]; however, studies have not shown clear sequence biases or structural preferences around acceptor lysines for ubiquitin. The lack of a clearly defined ubiquitylation motif could reflect the large diversity of

ubiquitin-modifying enzymes, each recognizing different structural determinants. Examination of lysine ubiquitylation by affinity purification of Strep-HA-tagged ubiquitin and MS revealed that modified sites show low evolutionary conservation across eukaryotic species and that significant overlap exists with other modifications such as acetylation [27]. Interestingly, an inverse relationship is observed between protein N-terminal ubiquitylation and acetylation upon proteasome inhibition, consistent with a role for acetylation in increasing protein half lives [188]. Although protein N-terminal ubiquitylation was suggested to target selected proteins for degradation [189], no correlation has been observed between changes in protein abundance and ubiquitylation sites [32, 34].

The analysis of residues surrounding the acceptor lysine from large-scale SUMO proteome datasets has revealed several SUMO-dependent motifs. These include the canonical UBC9 consensus motif ψKxE (where ψ is an aliphatic residue and x is any amino acid), a hydrophobic motif, a phospho-dependent sequence, and reverse consensus and non-consensus regions [177]. The correlation of all SUMO receptor sites with other types of modification revealed a 29% overlap with sites of ubiquitin, acetylation and methylation. However, this overlap was reduced significantly when sequence motifs KXE and [IVL]KXE were selected, implying that these consensus motifs are specific for SUMO whereas other motifs show a larger overlap with ubiquitylation sites. Distinct modification site motifs for other UBLs such as ISG15, FAT10, URM1 and UFM1 have yet to be determined.

Synergistic, antagonistic and competitive crosstalk exists between SUMO and ubiquitin. For example, competition at K164 of proliferating cell nuclear antigen (PCNA) can lead to the inhibition of homologous recombination when this site is SUMOylated, whereas its ubiquitylation favors translesion synthesis during the DNA damage response (Figure 2-9a) [190]. By contrast, synergistic interactions are observed between SUMOylated PML and RNF4 to promote ubiquitylation of PML [191], whereas the SUMOylation of the Hepatitis C viral protein NS5A inhibits its ubiquitylation and degradation by the host [192]. Remnant immunoprecipitation has been used to show that SUMOylation can mediate protein interactions between PML nuclear bodies and UBC9 [193] or the proteasome [78]. during cellular senescence or proteasome

inhibition. Alternatively, SUMOylation can impede protein interactions by shielding functional domains; this has been observed for poly(ADP-ribose) polymerase 1 (PARP1) where SUMOylation at K203 delays its proteolytic cleavage by caspase 3 and subsequently delays cell apoptosis [99]. Similarly, SUMOylation of the DUB USP37 at K452 affects its deconjugation activity, leading to a stabilization in the ubiquitylation of its substrate c-Myc [39]. Other examples of UBL-mediated changes in enzymatic activity have been reported for ISG15, where conjugation of ISG15 to sites next to the catalytic sites of UBC13 and UBCH6 suppressed their ubiquitin E2 enzyme activities (Figure 2-9b) [194, 195]. Changes in enzymatic activity can also be observed through crosstalk with other UBLs; for example, FAT10 causes the proteasomal degradation of the ubiquitin E1 activating enzyme UBE1 upon binding, suggesting a putative regulatory role of FAT10 in the ubiquitin conjugation pathway (Figure 2-9c) [196].
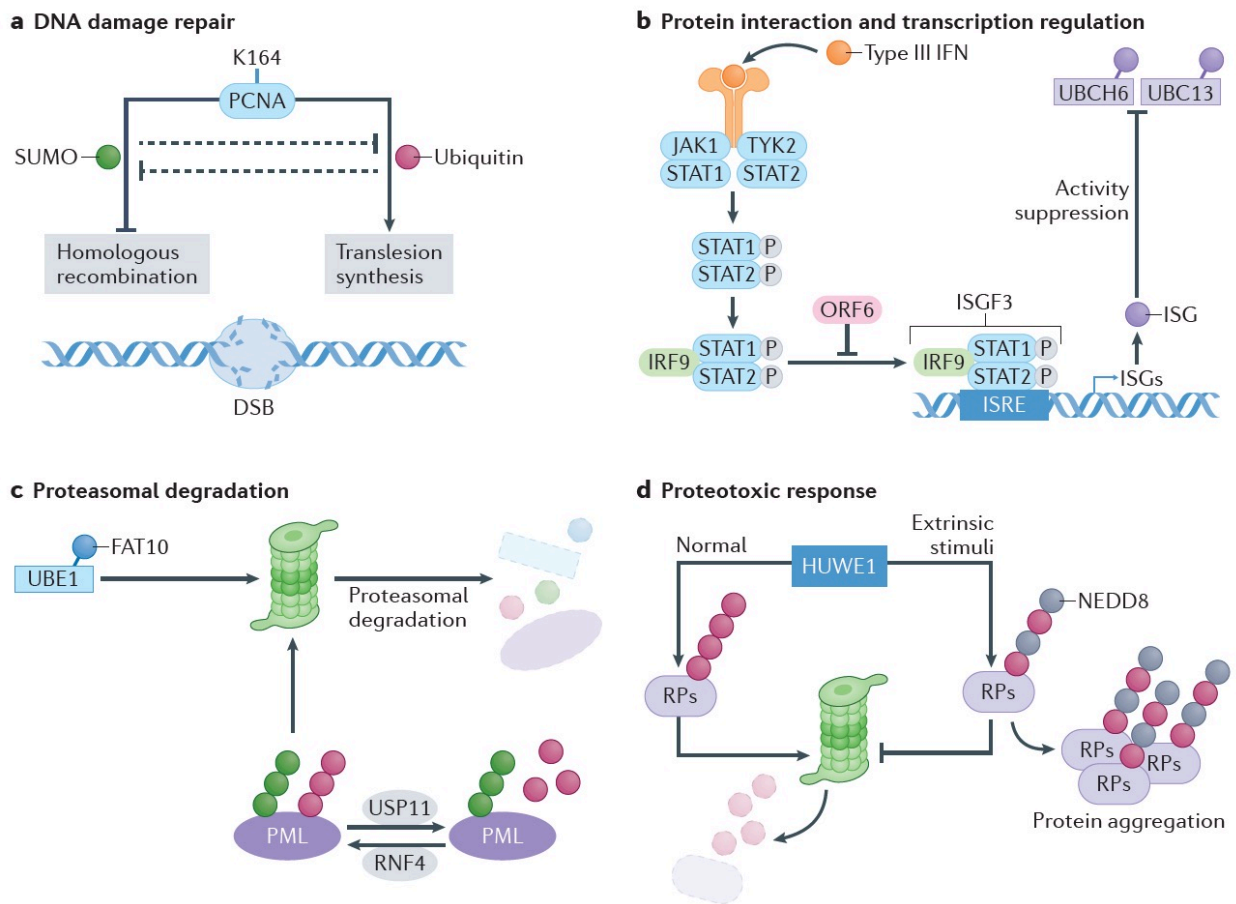
Figure 2-9 Roles of UBL modifications in the cell.

(a) Competition of UBL modifications at Lys164 of the DNA damage response protein PCNA impacts its downstream function. SUMOylation of Lys164 leads to the inhibition of homologous recombination at double-strand breaks (DSBs), wheras ubiquitylation leads to the recruitment of translesion synthesis polymerases. (b) UBL modifications can modify the activity of UBL enzymes. Protein ISGylation is induced by type III interferons (Type III IFN) through a signal transduction pathway leading to the expression of ISG15. ISG15 can be conjugated to the ubiquitin E2 conjugating enzymes UBC13 and UBCH6 on residues next to their catalytic sites, suppressing their ubiquitin E2 enzyme activity by impeding the thioester bond formation with ubiquitin. (c) Modification of UBL enzymes can lead to their degradation. For example, synergistic UBL modification of promyelocytic leukemia protein (PML) leads to the proteasomal degradation of the protein. Once SUMOylated, PML is ubiquitylated by the SUMO-targeted Ubiquitin Ligase (STUbL) RNF4 and subsequently degraded. This process is antagonized by the deubiquitylation of PML by USP11. Modification of the ubiquitin E1 activating enzyme UBE1 by FAT10 also leads to its proteasomal degradation. (d) The ubiquitin E3 ligase HUWE1 mediates the polyubiquitylation of ribosomal proteins (RPs), leading to their proteasomal degradation. Following exposure of cells to extrinsic stimuli such as proteotoxic stress, HUWE1 mediates the formation of ubiquitin-NEDD8 mixed chains on RPs, leading to protein aggregation that protects them from stress-induced toxicity. Made in ©Biorender (biorender.com).

Studying the effect of different environmental cues and cellular perturbations such as growth factors and pharmacological inhibitors is important as these can modulate UBL modifications. For example, the epidermal growth factor receptor (EGFR) is ubiquitylated upon stimulation, along with several downstream components of the endocytic machinery [197]. The profiling of ubiquitylated proteins following EGF stimulation using affinity chromatography and metabolic labeling using SILAC identified many ubiquitinated proteins involved in intracellular signaling, including cell-cell adhesion and actin remodeling, as evidenced from GO term analysis [26]. 23% of proteins in both this EGF ubiquitylome data set and a published data set of EGF-induced phosphotyrosine proteomes contained both modifications, suggesting interplay between kinase/phosphatase and E3 ligase/DUB networks. This is indeed the case for the tyrosine-protein kinase receptor ECK, which is phosphorylated by EGFR upon stimulation and subsequently monoubiquitylated by the Cbl E3 ubiquitin-protein ligase [26].

2.5.3 Understanding the roles of UBL linkages

Novel affinity purification methods have provided valuable insights into the structure and function of previously poorly understood ubiquitin signals. The combination of a Lys-less ubiquitin with affinity purification and MS identified novel linear polyubiquitin targets including TRAF6, a novel linear ubiquitin assembly complex (LUBAC) substrate essential for proper interleukin-1β-dependent NFκB signaling [37]. Phage-display-derived affimers corresponding to rare ubiquitin chains such as K6 and K33/K11 linkages have been used as affinity media to identify proteins with these rare chains [61]; this approach, in combination with MS analysis, identified the E3 ubiquitin ligase HUWE1 as a major source of cellular K6 chains, along with several substrates that were modified in a Parkin-dependent manner upon mitochondrial depolarization.

Ubiquitin clipping approaches have determined that 10–20% of ubiquitin polymers exist as chains with one, two or three branching points, and that PINK1/Parkin-mediated mitophagy predominantly exploits monoubiquitin and short-chain polyubiquitin chains with terminal phospho-ubiquitin moieties [15, 198]. Use of K11/K48 bispecific antibodies to immunopurify ubiquitylated proteins before MS analysis identified several substrates including mitotic regulators and newly synthesized and misfolded proteins, suggesting a function for this heterotypic chain in proteasomal degradation [75].

The modification of ubiquitin with phosphorylation, acetylation, and hybrid polymer chains of other UBLs can lead to different functional outcomes. These modifications can have a positive or negative impact on the activity of enzymes from the ubiquitin cascade pathway that recognize these structural variants. For example, in the context of mitochondrial turnover, ubiquitin chains containing S65 phosphorylation are regulated by several DUBs including USP30 [169]. Similarly, polyubiquitin chains can be modified by other UBLs, the best-known example being SUMO-ubiquitin mixed chains. It is conceivable that specialized DUBs can recognize and hydrolyze mixed chains; for example, USP11 can reverse ubiquitylation on hybrid SUMO-ubiquitin chains to counteract the enzymatic activity of RNF4 during the DNA damage response [199]. Further insights into these events could be revealed by a combination of ubiquitin and SUMO remnant

immunoprecipitation followed by MS analysis. Although mixed SUMO-ubiquitin chains are commonly detected upon proteasome inhibition, they play an important role in maintaining genome stability through the recruitment of RAP80 and its subsequent interaction with the BRCA1 complex [200].

Ubiquitin-NEDD8 mixed chains have been described primarily in response to proteotoxic stress to protect the nuclear ubiquitin proteasome system from stress-induced dysfunction (Figure 2-8d) [201]. These ubiquitin-NEDD8 chains involve the E3 ligase HUWE1, which targets ribosomal proteins to induce their aggregation and protects them from stress-induced toxicity. Although several non-Cullin NEDD8 substrates have been identified, experiments should avoid overexpression of exogenous NEDD8 as this can result in an imbalance of the free ubiquitin:NEDD8 ratio, which may favor atypical conjugation of NEDD8 by ubiquitin E1-activating enzyme UBE1 instead of the NEDD8 E1-activating enzyme [202]. The use of serial NEDD8-ubiquitin substrate profiling has proved useful in identifying NEDD8 substrates under physiological conditions [35]. This technique showed that polyneddylation at K6, K11, K48 and K54 is regulated by the deneddylation enzyme NEDP1, whereas other sites, such as K22, K27 and the NEDD8–ubiquitin K48 appear to be deneddylated by other NEDD8 proteases such as the COP9 signalosome. Interestingly, this approach identified cofilin-1, as well as other cytoskeletal proteins such as stathmin, profilin-1, actin and myosins as non-Cullin neddylation substrates. Cofilin-1 is neddylated at K112, a site located within its α4-helix actin binding domain, suggesting that neddylation could regulate its access to actin filaments.

Protein ISGylation is induced by type I and type III interferons and viral and bacterial infections, and appears to be restricted to higher eukaryotes as it is not found in yeast, nematode or drosophila cells [203]. ISG15 polymers have not been observed on protein conjugates and no ISG15-interacting motif has been reported thus far. It should be noted that the architecture of ISG15 includes two UBL domains, so it functions essentially as a stable UBL dimer. Further, ubiquitin forms hybrid chains with ISG15 primarily through ubiquitin K29, highlighting potential crosstalk between ISG15 and ubiquitin conjugation pathways [204]. Although ISG15 substrates have a role in regulating transcription and pre-mRNA splicing during the interferon response [50],

the biological function of these hybrid ubiquitin-ISG15 chains is still unknown, though they appear to play a role in reducing the turnover of cellular ubiquitylated proteins [204]. Affinity purification combined with MS analysis has revealed crosstalk between ISG15 and SUMO, where interferon type I increased the levels of both UBLs [162]. SUMOylation was found to stabilize several ISG15 restriction factors and this stabilization was dependent on the E3 ISG15 ligase TRIM25 but not HERC5 or the E2 conjugating enzyme UBE2L6.

2.5.4 Profiling the ubiquitylome

The availability of methods enabling the large-scale profiling of ubiquitin conjugates in a site-specific manner opens up interesting opportunities to uncover unsuspected pleiotropic effects of pharmacological inhibitors, substrates of specific E3 ligases/hydrolases and changes in the ubiquitylome landscape upon different cell perturbations. Developments in affinity purification and isobaric labeling have enhanced the sensitivity and scalability of ubiquitin proteomic experiments. For example, the recently developed UbiFast approach, in which diglycine tryptic peptides are directly labeled while immobilized on antibody beads, was able to quantify changes in ubiquitylation at ~10,000 sites in 10 samples, using as little as 500 µg material per sample [119]. This approach was used to map ubiquitylated proteins in models of basal and luminal human breast cancer xenografts and enabled the identification of the Ikaros transcription factors (e.g. IKZF1 and IKZF3) and CSNK1A1 as targets for the ubiquitin E3 ligase cereblon.

Although several reports have documented global changes in protein ubiquitylation upon proteotoxic stress, proteasome inhibition, or DNA damage using whole-cell lysates, subcellular fractionation or enrichment of individual proteins can be necessary to identify low abundance conjugates. Tagging and isolation of membrane proteins including ErbB-2 (HER2), dishevelled-2 (DVL2), and T cell receptor α (TCRα) before diglycine immunoprecipitation provided far greater site coverage than traditional affinity purification MS experiments [205]. Organelle fractionation has been used for more efficient identification of target proteins that may be underrepresented in whole-cell extracts; for example, isopycnic centrifugation combined with diglycine immunoprecipitation has been successfully applied to reveal previously unknown functions of

the mitochondrial degradation and import pathways [206]. By combining pharmacological inhibitors and genetic ablation with quantitative ubiquitin proteomics, E3 ubiquitin-protein ligase MARCHF5 and the deubiquitylase USP30 were shown to modify import substrates at the translocase of the outer membrane complex [198, 206]. These studies identified a role for USP30 in regulating mitochondrial import of ubiquitylated proteins, possibly to recycle ubiquitin at the membrane entry, whereas MARCH5F was found to direct the degradation of import intermediates by the proteasome. A valosin containing protein (VCP/p97) inhibitor substrate trapping approach was used to identify improperly folded proteins that are cleared from the early secretory pathway via endoplasmic reticulum-associated degradation (ERAD) [207]. In combination with quantitative ubiquitin proteomics, this method uncovered novel and previously known endogenous ERAD substrates in HepG2 liver cells. Although these examples specifically relate to ubiquitin substrates, the use of subcellular fractionation could also be applied to study proteins modified by other UBLs present in specific organelles that might be underrepresented in total cell lysates.

2.6 Reproducibility and data deposition

2.6.1 Field standards on data deposition

Specific repositories are available for the deposition of MS data, including the PRoteomics IDEntifications Database (PRIDE) [208], the Mass spectrometry Interactive Virtual Environment (MassIVE) [209], the Japan ProteOme Standard repository (jPOSTrepo) [210] and the integrated Proteome resource iProX [211]. Raw files and processed protein and peptide output files from ubiquitin proteomics experiments should be stored along with metadata on enrichment protocols and the biological material used in experiments, including details regarding animal tissues and the identity of any cell lines employed. Information regarding chromatography conditions and MS acquisition and software parameters for processing of raw data files should also be stored. Raw files constitute the unprocessed, non-curated output files containing full sequential sets of MS and MS/MS files. Processed files contain the results of proteome-database searches carried out using the raw mass spectrometry files to indentify peptides and proteins. Submission of complete datasets is recommended as a minimum reporting standards for the field, to enable the parsing, integration and visualisation of the datasets in the repositories and to connect the processed data to the corresponding mass spectra. Processed identification results can be deposited in a Proteomics Standard Initiative (PSI) open standard format such as mzIdentML or mzTab. Targeted proteomics data can be uploaded to repositories such as PeptideAtlas [212] and Panorama Public [213].

2.6.2 Reproducibility issues

Results can be influenced by seemingly small issues such as using different batches of serum to culture cells or whether incubators containing cells are frequently opened or not, both leading to small changes in cellular proteomes. Further, abundant peptides can be carried over in subsequent LC-MS/MS runs. To limit these issues, it is best to perform experiments in large batches to limit differences between biological and technical replicates; however, this requires considerable MS runtime. Alternatively, metabolic labelling with SILAC or isobaric peptide

labelling can improve reproducibility between runs. Further, the use of different software packages for the same dataset can lead to partially different results. Making raw and processed MS data available to the scientific community is therefore important to enable researchers to query the same set of raw data using different software packages.

2.7 Limitations and optimizations

Proteomics studies have revealed tens of thousands of ubiquitin-modified and UBL-modified sites; however, our understanding of the functional consequences of these modifications on individual target proteins is limited. Further, we have limited knowledge of ubiquitin/UBL chain compositions — owing to the proteolytic approaches used for MS — and the specific identity of the targets for different E3 ligases and proteases. We discuss these limitations further below.

2.7.1 Functional relevance of UBL modifications

Although proteomic techniques can reveal many modified proteins and modification sites, our understanding of the functional relevance of modifications is limited because functional investigations are difficult to achieve in a similarly high-throughput fashion. The comparison of wild-type target proteins with mutants with non-functioning UBL modification sites can reveal the functional consequences of modifications. However, as SUMOs are thought to co-regulate large numbers of substrates in a group-like manner [151, 214], and this mode of action is likely prevalent for ubiquitin and other UBLs as well as for other PTMs, modification-deficient mutants may not show a phenotype as other substrates may compensate for the absence of the modification on the selected target protein. Alternative experimental approaches must be developed to enable simultaneous mutation of large sets of target proteins in cells and the development of cell lines expressing many modification-deficient mutant proteins.

Activity-based probes can be used to study the cellular activity of UBL-conjugating and UBL-deconjugating enzymes, particularly for enzymes that employ catalytic cysteines [215]. Coupling these probes to enrichment strategies and MS allows identification of the UBL enzymes involved in biological responses to stimuli [216]. These activity-based probes have particularly demonstrated the key roles of DUBs in remodelling the ubiquitome [217].

2.7.2 Targets modified by specific chains

Significant progress has been made in the field regarding the characterization of the composition and heterogeneity of ubiquitin polymers, including homotypic and heterotypic chain linkages [218]. However, ubiquitin proteomics methodologies do not enable the simultaneous identification of substrate proteins together with the characterization of the composition of their conjugated ubiquitin polymers, owing to the proteolytic digests that are employed to study these modifications. Alternate approaches that allow for the enrichment of specific targets with limited proteolysis must be developed to identify the modification site and type of modification in a single experiment. The ubiquitin clipping technology can be used for this purpose but can only be applied to ubiquitin; developing clipping technologies for other UBLs would therefore be beneficial to enable investigation of UBL signaling. These approaches would require top-down proteomics approaches, which start with intact modified proteins [219].

2.7.3 Importance of PTM crosstalk

Improved methodology, such as the development of PTM-specific antibodies and increased MS sensitivity have facilitated the characterization of individual PTMs [220]. However, our understanding of crosstalk between ubiquitin/UBLs and other PTMs at a proteome-wide scale is limited. Methods that can probe the substrate selectivity towards UBLs and other types of modifications in the context of specific cell stimulation are needed to further understand their interplay and crosstalk.

2.7.4 Unexpected outcomes and workaround

An important caveat when mapping ubiquitylation sites following tryptic digestion is the erroneous identification of diglycine remnants associated with artefactual alkylation caused by iodoacetamide treatment [221]. Iodoacetamide is typically and frequently used to alkylate free cysteine but can also form covalent acetamide adducts (C_2H_3NO) on lysine residues. The addition of two iodoacetamide molecules to free amino groups, such as the N-terminus or one lysine residue, produces an adduct ($C_4H_6N_2O_2$) of an identical mass to that of diglycine remnant, which

cannot be distinguished by conventional search engines. This pitfall can be overcome by using chloroacetamide instead of iodoacetamide. Alternatively, iodoacetamide can be used at room temperature instead of at higher temperature to prevent adduct formation [76]. The anti-εGG antibody frequently used to enrich ubiquitylated tryptic peptides is not expected to enrich iodoacetamide adducts.

2.8 Outlook

The methodology described in this review is expected to enhance our functional understanding of UBL signalling in biological processes, pathological conditions and following treatment with novel drugs such as proteolysis targeting chimeras (PROTACs). However, high reproducibility and high throughput at limited costs is essential for clinical applications and the sensitivity of the methodology needs to be improved to enable comprehensive analysis of UBLs in situations of limited sample availability, for example when using needle biopsies. These issues have been adequately dealt with in the phosphorylation field [222] and need to be further improved with UBLs.

The assignment of different homotypic and heterotypic UBL chains to specific substrates and modification sites still represents an important challenge with current proteomic methods. Connecting protein targets to the different types of chains that modify them is possible using top-down analysis of modified proteins [223], although improvements in sensitivity are required to reduce the amounts of starting material required for successful assignment [224]. Ubiquitin clipping following enrichment of ubiquitin-modified proteins can provide a meaningful strategy to study the architecture of polyubiquitin chains. Further, whereas traps exist for enriching specific types of ubiquitin chains, more are needed to cover cover the many different types of ubiquitin linkages. Traps could include novel ubiquitin binding domains [225] or antibodies that can bind to specific ubiquitin linkages [61].

PROTACs exploit the great potential of the ubiquitin system for the tailored degradation of any protein, even for proteins that have so far been considered undruggable [226, 227]. PROTACs connects a protein of interest to a ubiquitin ligase, enabling its ubiquitylation and subsequent proteasomal degradation. Cell-wide proteomics approaches are required to establish the specificity of PROTACs, enabling identification and quantification of all cellular proteins in order to identify proteins that are degraded in response to PROTACs. In human cells, it is still challenging to identify proteins that are expressed at very low levels, given the large dynamic

range of proteins [228] and MS methods using data independent acquisition (DIA) will be vital to reach the required depth of analysis for characterizing responses to PROTACs [101].

One of the most challenging issues in the field is uncovering how the host of cellular E3 ligases, proteases and targets interact. Effective approaches to uncover this complex interacting network so far have used substrate traps, such as UBAIT and TULIP approaches. These are most useful for profiling HECT and RBR E3 ligases, as ubiquitin is normally transferred to these E3s prior to substrate modification; however, they have also been applied to RING-type ligases. Whether the fused ubiquitin can be employed for substrate modification by the E3 ligase of interest must be established experimentally. Key negative controls are a ligase-dead E3 mutant and a non-conjugatable ubiquitin mutant lacking the diglycine motif. The main difference between UBAIT and TULIP is that the latter approach allows the purification of substrates under denaturing conditions to remove non-covalent interactors. It remains to be established whether the NEDDylator approach can be applied to all E3 ligases. Ubiquitin Ligase Substrate Traps employ a fusion construct of an ubiquitin-binding domain coupled to an E3 ligase of interest [229]. Ubiquitylated substrates can subsequently be purified and identified by MS. Ligase-dead mutants are also important negative controls for this technique. Ligase traps of eight different F box proteins can be used to identify known and novel substrates, demonstrating the specificity of these F box proteins. These methods can be employed to link E3 ligases of interest to their substrates.

The complexity of the ubiquitin system is overwhelming, with over 600 putative E3s and substrate-recognition subunits of E3 complexes and around 100 proteases regulating thousands of substrates at tens of thousands of sites. The wiring of this complex system is largely unclear [230, 231] and. connecting enzymes and substrates is a daunting task, owing to the redundancy of enzymes and E3 ligase cascades [232]. Knockdown, knockout or inhibition of a protease enables the identification of potential substrates using UBL proteomics techniques [233]; however, whether the identified proteins represent *bona fide* substrates requires subsequent validation. We expect significant progress in the next decade to link ubiquitin and UBL enzymes to their substrates. Further, system-wide studies at the interface of multiple PTMs are needed to

uncover PTM crosstalk. Sequential enrichment of different PTMs in tryptic digests of samples can be employed for this purpose to provide further insights into the interrelationship between protein modifications [17, 39, 173, 234-237]. A major challenge here is the processing of raw MS files due to the steep increase in computing time required to handle data for multiple different PTMs.

Upon the clinical success of proteasome inhibitors for the treatment of multiple myeloma[238], pharmaceutical companies started to develop other inhibitors to target ubiquitin and UBL signaling [226, 239]. These include a set of inhibitors for targeting ubiquitin and UBL E1 enzymes. The most advanced compound clinically is Pevonedistat (also known as MLN4924), an inhibitor of the NEDD8 activating enzyme (NAE) that is currently in phase III clinical trials in a combination therapy for the treatment of acute and chronic myeloid leukemia [240]. Specific inhibitors for the ubiquitin activating enzyme UBE1 [241] and the SUMO E1 enzyme [242] have subsequently been developed. Any novel inhibitors should specifically block one type of ubiquitin or UBL signal transduction cascade, which may require extensive fine-tuning of the compounds targeting these related enzymes. Other clinical opportunities for ubiquitin proteomics include evaluating the specificity of PROTACs as described above, and evaluating the specificity of novel VCP/p97 inhibitors [243]. The clinical development of any of the above compounds will require detailed analysis of system-wide ubiquitin and UBL signaling in clinical samples, which could pose specific challenges including small amounts of biopsy samples, and the need for rapid high-throughput analysis. Similar challenges have been addressed in the phosphorylation field due to the large sets of kinase inhibitors that are being developed [244]. Solutions include parallel processing of samples in a robust and user-friendly manner, for example using UbiFast [119] and improvements in MS sensitivity and resolution [245, 246] that will likely benefit the ubiquitin and UBL field [247].

Technology to investigate ubiquitin and UBL signalling has matured over the last decade and can now be used efficiently to chart complex cellular UBL signalling networks, affecting thousands of substrates. We expect our limited knowledge of enzyme-substrate connections to rapidly improve over the coming years. Deep learning and machine learning techniques are emerging as

analysis techniques and are expected to make strong contributions to the field. The vast datasets generated by ubiquitin proteomics approaches will suit machine learning approaches, enabling reliable predictions of UBL modifications.

2.9 Related links

JASSA: http://www.jassa.fr

PhosphoSitePlus: https://www.phosphosite.org

UniProt: https:www.uniprot.org

Proteomics IDEntifications Database (PRIDE): https://www.ebi.ac.uk/pride

Mass Spectrometry Interactive Virtual Environment (MassIVE):

https://massive.ucsd.edu/ProteoSAFe/static/massive.jsp

Japan Proteome Standard Repository/Database: https://jpostdb.org

iProX: https://www.iprox.org

Peptide Atlas: http://www.peptideatlas.org

Panorama Public: https://panoramaweb.org

2.10 Glossary

UBL traps	Affinity purification methods where fusion proteins containing units of UBL binding domains are expressed in cells to trap UBL conjugates
Pulse chase	An experiment where cells are exposed to a labeled compound incorporated into proteins, that is later replaced with an unlabeled form to determine the time of exchange
Ubiquitin clipping	A technique that uses an engineered viral protease, Lbpro*, to cleave ubiquitin conjugates and leave a tracetable diglycine remnant on the modified substrates
Isopycnic centrifugation	A fractionation method where cell components can be separated based on a density gradient pre-formed or formed during high speed centrifugation
Translesion synthesis	A process whereby the DNA replication machinery can bypass the blocked replication fork caused by DNA damage
Affimers	Non-antibody binding proteins that mimic the molecular recognition features of antibodies
Proteolysis targeting chimeras	Heterobifunctional small molecules that consist of an ubiquitin E3 ligase binding domain linked to a domain that binds specifically to a protein targeted for degradation
Ubiquitin code	The concept that distinct conformation of ubiquitin chains and modifications lead to different cellular outcomes
Deneddylation	Removal of NEDD8 from modified substrates
Remnant peptides	Amino acids left over on modified lysine residues after proteolytic digestion
Ubiquitylome	Repertoire of ubiquitylated proteins
Offline fractionation	Fractionation of peptide extracts by methods such as ion exchange or high pH reverse phase chromatography; fractions are subsequently analyzed by mass spectrometry
Ratio compression	Underestimated fold-change ratio of peptide/protein abundance in isobaric peptide labeling owing to co-selection of different peptides during MS/MS
Top-down proteomics	Protein identification method that relies on the selection of protein ions as precursors for MS/MS fragmentation
Head-to-tail concatemers	A long, continuous DNA molecule containing multiple copies of the same gene assembled head to tail

2.11 Acknowledgements

2.12 References

1.   Rajalingam, K. and I. Dikic, *SnapShot: Expanding the Ubiquitin Code.* Cell, 2016. **164**(5): p. 1074-1074 e1.

2.   Yau, R. and M. Rape, *The increasing complexity of the ubiquitin code.* Nat Cell Biol, 2016. **18**(6): p. 579-86.

3.   Kerscher, O., R. Felberbaum, and M. Hochstrasser, *Modification of proteins by ubiquitin and ubiquitin-like proteins.* Annu Rev Cell Dev Biol, 2006. **22**: p. 159-80.

4.   Cappadocia, L. and C.D. Lima, *Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism.* Chem Rev, 2018. **118**(3): p. 889-918.

5.   Clague, M.J., S. Urbe, and D. Komander, *Breaking the chains: deubiquitylating enzyme specificity begets function.* Nat Rev Mol Cell Biol, 2019. **20**(6): p. 338-352.

6.   Ronau, J.A., J.F. Beckmann, and M. Hochstrasser, *Substrate specificity of the ubiquitin and Ubl proteases.* Cell Res, 2016. **26**(4): p. 441-56.

7.   McClellan, A.J., S.H. Laugesen, and L. Ellgaard, *Cellular functions and molecular mechanisms of non-lysine ubiquitination.* Open Biol, 2019. **9**(9): p. 190147.

8.   Jansen, N.S. and A.C.O. Vertegaal, *A Chain of Events: Regulating Target Proteins by SUMO Polymers.* Trends Biochem Sci, 2020.

9.   Bailly, A.P., et al., *The Balance between Mono- and NEDD8-Chains Controlled by NEDP1 upon DNA Damage Is a Regulatory Module of the HSP70 ATPase Activity.* Cell Rep, 2019. **29**(1): p. 212-224 e8.

10.   Keuss, M.J., et al., *Unanchored tri-NEDD8 inhibits PARP-1 to protect from oxidative stress-induced cell death.* EMBO J, 2019. **38**(6).

11.   Perez Berrocal, D.A., et al., *Hybrid Chains: A Collaboration of Ubiquitin and Ubiquitin-Like Modifiers Introducing Cross-Functionality to the Ubiquitin Code.* Front Chem, 2019. **7**: p. 931.

12.   Streich, F.C., Jr. and C.D. Lima, *Structural and functional insights to ubiquitin-like protein conjugation.* Annu Rev Biophys, 2014. **43**: p. 357-79.

13.   van der Veen, A.G. and H.L. Ploegh, *Ubiquitin-like proteins.* Annu Rev Biochem, 2012. **81**: p. 323-57.

14.   Akutsu, M., I. Dikic, and A. Bremm, *Ubiquitin chain diversity at a glance.* J Cell Sci, 2016. **129**(5): p. 875-80.

15.   Swatek, K.N., et al., *Insights into ubiquitin chain architecture using Ub-clipping.* Nature, 2019. **572**(7770): p. 533-537.

16.   Hendriks, I.A. and A.C. Vertegaal, *A high-yield double-purification proteomics strategy for the identification of SUMO sites.* Nat Protoc, 2016. **11**(9): p. 1630-49.

17.   McManus, F.P., F. Lamoliatte, and P. Thibault, *Identification of cross talk between SUMOylation and ubiquitylation using a sequential peptide immunopurification approach.* Nat Protoc, 2017. **12**(11): p. 2342-2358.

18.   Udeshi, N.D., et al., *Large-scale identification of ubiquitination sites by mass spectrometry.* Nat Protoc, 2013. **8**(10): p. 1950-60.

19.   Barysch, S.V., et al., *Identification and analysis of endogenous SUMO1 and SUMO2/3 targets in mammalian cells and tissues using monoclonal antibodies.* Nat Protoc, 2014. **9**(4): p. 896-909.

20.   Liu, N., et al., *Clinically used antirheumatic agent auranofin is a proteasomal deubiquitinase inhibitor and inhibits tumor growth.* Oncotarget, 2014. **5**(14): p. 5453-71.

21.   Liu, N., et al., *A novel proteasome inhibitor suppresses tumor growth via targeting both 19S proteasome deubiquitinases and 20S proteolytic peptidases.* Sci Rep, 2014. **4**: p. 5240.

22.   Zhang, J.J., et al., *Deubiquitinases as potential anti-cancer targets for gold(III) complexes.* Chem Commun (Camb), 2013. **49**(45): p. 5153-5.

23.   Graham, J.M., *Fractionation of Subcellular Organelles.* Curr Protoc Cell Biol, 2015. **69**: p. 3 1 1-3 1 22.

24.   Wagner, S.A., et al., *Proteomic analyses reveal divergent ubiquitylation site patterns in murine tissues.* Mol Cell Proteomics, 2012. **11**(12): p. 1578-85.

25.   McManus, F.P., C.D. Altamirano, and P. Thibault, *In vitro assay to determine SUMOylation sites on protein substrates.* Nat Protoc, 2016. **11**(2): p. 387-97.

26.   Argenzio, E., et al., *Proteomic snapshot of the EGF-induced ubiquitin network.* Mol Syst Biol, 2011. **7**: p. 462.

27.   Danielsen, J.M., et al., *Mass spectrometric analysis of lysine ubiquitylation reveals promiscuity at site level.* Mol Cell Proteomics, 2011. **10**(3): p. M110 003590.

28.   Peng, J., et al., *A proteomics approach to understanding protein ubiquitination.* Nat Biotechnol, 2003. **21**(8): p. 921-6.

29.   Roux, K.J., et al., *BioID: A Screen for Protein-Protein Interactions.* Curr Protoc Protein Sci, 2018. **91**: p. 19 23 1-19 23 15.

30.   Hill, Z.B., et al., *Direct Proximity Tagging of Small Molecule Protein Targets Using an Engineered NEDD8 Ligase.* J Am Chem Soc, 2016. **138**(40): p. 13123-13126.

31.   Coyaud, E., et al., *BioID-based Identification of Skp Cullin F-box (SCF)beta-TrCP1/2 E3 Ligase Substrates.* Mol Cell Proteomics, 2015. **14**(7): p. 1781-95.

32.   Kim, W., et al., *Systematic and quantitative assessment of the ubiquitin-modified proteome.* Mol Cell, 2011. **44**(2): p. 325-40.

33.   Xu, G., J.S. Paige, and S.R. Jaffrey, *Global analysis of lysine ubiquitination by ubiquitin remnant immunoaffinity profiling.* Nat Biotechnol, 2010. **28**(8): p. 868-73.

34.   Akimov, V., et al., *UbiSite approach for comprehensive mapping of lysine and N-terminal ubiquitination sites.* Nat Struct Mol Biol, 2018. **25**(7): p. 631-640.

35.   Vogl, A.M., et al., *Global site-specific neddylation profiling reveals that NEDDylated cofilin regulates actin dynamics.* Nat Struct Mol Biol, 2020. **27**(2): p. 210-220.

36.   Jeram, S.M., et al., *Using mass spectrometry to identify ubiquitin and ubiquitin-like protein conjugation sites.* Proteomics, 2009. **9**(4): p. 922-34.

37.   Kliza, K., et al., *Internally tagged ubiquitin: a tool to identify linear polyubiquitin-modified proteins by mass spectrometry.* Nat Methods, 2017. **14**(5): p. 504-512.

38.   Meierhofer, D., et al., *Quantitative analysis of global ubiquitination in HeLa cells by mass spectrometry.* J Proteome Res, 2008. **7**(10): p. 4566-76.

39.   Lamoliatte, F., et al., *Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification.* Nat Commun, 2017. **8**: p. 14109.

40.   Li, C., et al., *Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility.* Nat Commun, 2020. **11**(1): p. 834.

41.   Schimmel, J., et al., *Uncovering SUMOylation dynamics during cell-cycle progression reveals FoxM1 as a key mitotic SUMO target protein.* Mol Cell, 2014. **53**(6): p. 1053-66.

42.   Schimmel, J., et al., *The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle.* Mol Cell Proteomics, 2008. **7**(11): p. 2107-22.

43.   Tatham, M.H., et al., *Comparative proteomic analysis identifies a role for SUMO in protein quality control.* Sci Signal, 2011. **4**(178): p. rs4.

44.   Vertegaal, A.C., et al., *Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics.* Mol Cell Proteomics, 2006. **5**(12): p. 2298-310.

45.   Li, Z., et al., *Functions and substrates of NEDDylation during cell cycle in the silkworm, Bombyx mori.* Insect Biochem Mol Biol, 2017. **90**: p. 101-112.

46.   Liao, S., et al., *The Protein Neddylation Pathway in Trypanosoma brucei: Functional characterization and substrate identification.* J Biol Chem, 2017. **292**(3): p. 1081-1091.

47.   Bennett, E.J., et al., *Dynamics of cullin-RING ubiquitin ligase network revealed by systematic quantitative proteomics.* Cell, 2010. **143**(6): p. 951-65.

48.   Xirodimas, D.P., et al., *Ribosomal proteins are targets for the NEDD8 pathway.* EMBO Rep, 2008. **9**(3): p. 280-6.

49.   Jones, J., et al., *A targeted proteomic analysis of the ubiquitin-like modifier nedd8 and associated proteins.* J Proteome Res, 2008. **7**(3): p. 1274-87.

50.   Zhao, C., et al., *Human ISG15 conjugation targets both IFN-induced and constitutively expressed proteins functioning in diverse cellular pathways.* Proc Natl Acad Sci U S A, 2005. **102**(29): p. 10200-5.

51.   Yoo, H.M., et al., *Modification of ASC1 by UFM1 is crucial for ERalpha transactivation and breast cancer development.* Mol Cell, 2014. **56**(2): p. 261-274.

52.   Leng, L., et al., *A proteomics strategy for the identification of FAT10-modified sites by mass spectrometry.* J Proteome Res, 2014. **13**(1): p. 268-76.

53.   Khoshnood, B., I. Dacklin, and C. Grabbe, *A proteomics approach to identify targets of the ubiquitin-like molecule Urm1 in Drosophila melanogaster.* PLoS One, 2017. **12**(9): p. e0185611.

54.   Odeh, H.M., et al., *The SUMO-specific isopeptidase SENP2 is targeted to intracellular membranes via a predicted N-terminal amphipathic alpha-helix.* Mol Biol Cell, 2018. **29**(15): p. 1878-1890.

55.   Impens, F., et al., *Mapping of SUMO sites and analysis of SUMOylation changes induced by external stimuli.* Proc Natl Acad Sci U S A, 2014. **111**(34): p. 12432-7.

56.   Tammsalu, T., et al., *Proteome-wide identification of SUMO modification sites by mass spectrometry.* Nat. Protoc, 2015. **10**(9): p. 1374-1388.

57.   Bakos, G., et al., *An E2-ubiquitin thioester-driven approach to identify substrates modified with ubiquitin and ubiquitin-like molecules.* Nat Commun, 2018. **9**(1): p. 4776.

58.   Hemelaar, J., et al., *Specific and covalent targeting of conjugating and deconjugating enzymes of ubiquitin-like proteins.* Mol Cell Biol, 2004. **24**(1): p. 84-95.

59.   Lear, T.B., et al., *Kelch-like protein 42 is a profibrotic ubiquitin E3 ligase involved in systemic sclerosis.* J Biol Chem, 2020. **295**(13): p. 4171-4180.

60.   O'Connor, H.F., et al., *Ubiquitin-Activated Interaction Traps (UBAITs): Tools for Capturing Protein-Protein Interactions.* Methods Mol Biol, 2018. **1844**: p. 85-100.

61.   Michel, M.A., et al., *Ubiquitin Linkage-Specific Affimers Reveal Insights into K6-Linked Ubiquitin Signaling.* Mol Cell, 2017. **68**(1): p. 233-246 e5.

62.   Yoshida, Y., et al., *A comprehensive method for detecting ubiquitinated substrates using TR-TUBE.* Proc Natl Acad Sci U S A, 2015. **112**(15): p. 4630-5.

63.   Shi, Y., et al., *A data set of human endogenous protein ubiquitination sites.* Mol Cell Proteomics, 2011. **10**(5): p. M110 002089.

64.   Hjerpe, R., et al., *Efficient protection and isolation of ubiquitylated proteins using tandem ubiquitin-binding entities.* EMBO Rep, 2009. **10**(11): p. 1250-8.

65.   Lopitz-Otsoa, F., et al., *SUMO-Binding Entities (SUBEs) as Tools for the Enrichment, Isolation, Identification, and Characterization of the SUMO Proteome in Liver Cancer.* J Vis Exp, 2019(153).

66.   Lang, V., et al., *Using Biotinylated SUMO-Traps to Analyze SUMOylated Proteins.* Methods Mol Biol, 2016. **1475**: p. 109-21.

67.   Da Silva-Ferrada, E., et al., *Analysis of SUMOylated proteins using SUMO-traps.* Sci Rep, 2013. **3**: p. 1690.

68.   Bruderer, R., et al., *Purification and identification of endogenous polySUMO conjugates.* EMBO Rep, 2011. **12**(2): p. 142-8.

69.   Jeram, S.M., et al., *An improved SUMmOn-based methodology for the identification of ubiquitin and ubiquitin-like protein conjugation sites identifies novel ubiquitin-like protein chain linkages.* Proteomics, 2010. **10**(2): p. 254-65.

70.   Lumpkin, R.J., et al., *Site-specific identification and quantitation of endogenous SUMO modifications under native conditions.* Nat Commun, 2017. **8**(1): p. 1171.

71.   Cai, L., et al., *Proteome-wide Mapping of Endogenous SUMOylation Sites in Mouse Testis.* Mol Cell Proteomics, 2017. **16**(5): p. 717-727.

72.   Baldanta, S., et al., *ISG15 governs mitochondrial function in macrophages following vaccinia virus infection.* PLoS Pathog, 2017. **13**(10): p. e1006651.

73.   Kozuka-Hata, H., et al., *System-Wide Analysis of Protein Acetylation and Ubiquitination Reveals a Diversified Regulation in Human Cancer Cells.* Biomolecules, 2020. **10**(3).

74.   Akimov, V., et al., *StUbEx PLUS-A Modified Stable Tagged Ubiquitin Exchange System for Peptide Level Purification and In-Depth Mapping of Ubiquitination Sites.* J Proteome Res, 2018. **17**(1): p. 296-304.

75.   Yau, R.G., et al., *Assembly and Function of Heterotypic Ubiquitin Chains in Cell-Cycle and Protein Quality Control.* Cell, 2017. **171**(4): p. 918-933 e20.

76.   Udeshi, N.D., et al., *Methods for quantification of in vivo changes in protein ubiquitination following proteasome and deubiquitinase inhibition.* Mol Cell Proteomics, 2012. **11**(5): p. 148-59.

77.   Wagner, S.A., et al., *A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles.* Mol Cell Proteomics, 2011. **10**(10): p. M111 013284.

78.   Hendriks, I.A., et al., *Site-specific characterization of endogenous SUMOylation across species and organs.* Nat Commun, 2018. **9**(1): p. 2456.

79.   Hendriks, I.A., et al., *System-wide identification of wild-type SUMO-2 conjugation sites.* Nat Commun, 2015. **6**: p. 7289.

80.   Cubenas-Potts, C., et al., *Identification of SUMO-2/3-modified proteins associated with mitotic chromosomes.* Proteomics, 2015. **15**(4): p. 763-72.

81.   Becker, J., et al., *Detecting endogenous SUMO targets in mammalian cells and tissues.* Nat Struct Mol Biol, 2013. **20**(4): p. 525-31.

82.   Matafora, V., et al., *Proteomics analysis of nucleolar SUMO-1 target proteins upon proteasome inhibition.* Mol Cell Proteomics, 2009. **8**(10): p. 2243-55.

83.   Zhang, Y., et al., *The in vivo ISGylome links ISG15 to metabolic pathways and autophagy upon Listeria monocytogenes infection.* Nat Commun, 2019. **10**(1): p. 5383.

84.   Giannakopoulos, N.V., et al., *Proteomic identification of proteins conjugated to ISG15 in mouse and human cells.* Biochem Biophys Res Commun, 2005. **336**(2): p. 496-506.

85.   Merbl, Y., et al., *Profiling of ubiquitin-like modifications reveals features of mitotic control.* Cell, 2013. **152**(5): p. 1160-72.

86.   Gupta, R., et al., *Ubiquitination screen using protein microarrays for comprehensive identification of Rsp5 substrates in yeast.* Mol Syst Biol, 2007. **3**: p. 116.

87.   Uzoma, I., et al., *Global Identification of Small Ubiquitin-related Modifier (SUMO) Substrates Reveals Crosstalk between SUMOylation and Phosphorylation Promotes Cell Migration.* Mol Cell Proteomics, 2018. **17**(5): p. 871-888.

88.   Oh, Y.H., et al., *Chip-based analysis of SUMO (small ubiquitin-like modifier) conjugation to a target protein.* Biosens Bioelectron, 2007. **22**(7): p. 1260-7.

89.   Branon, T.C., et al., *Efficient proximity labeling in living cells and organisms with TurboID.* Nat Biotechnol, 2018. **36**(9): p. 880-887.

90.   Hendriks, I.A., et al., *Uncovering global SUMOylation signaling networks in a site-specific manner.* Nat Struct Mol Biol, 2014. **21**(10): p. 927-36.

91.   Lamoliatte, F., et al., *Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling.* Nat Commun, 2014. **5**: p. 5409.

92. Tammsalu, T., et al., *Proteome-Wide Identification of SUMO2 Modification Sites.* Sci. Signal, 2014. **7**(323): p. rs2.

93. O'Connor, H.F., et al., *Ubiquitin-Activated Interaction Traps (UBAITs) identify E3 ligase binding partners.* EMBO Rep, 2015. **16**(12): p. 1699-1712.

94. Kumar, R., et al., *The STUbL RNF4 regulates protein group SUMOylation by targeting the SUMO conjugation machinery.* Nat Commun, 2017. **8**(1): p. 1809.

95. Gao, Y., et al., *Enhanced Purification of Ubiquitinated Proteins by Engineered Tandem Hybrid Ubiquitin-binding Domains (ThUBDs).* Mol Cell Proteomics, 2016. **15**(4): p. 1381-96.

96. Zhuang, M., et al., *Substrates of IAP ubiquitin ligases identified with a designed orthogonal E3 ligase, the NEDDylator.* Mol Cell, 2013. **49**(2): p. 273-82.

97. Salas-Lloret, D., G. Agabitini, and R. Gonzalez-Prieto, *TULIP2: An Improved Method for the Identification of Ubiquitin E3-Specific Targets.* Front Chem, 2019. **7**: p. 802.

98. Xu, P. and J. Peng, *Characterization of polyubiquitin chain structure by middle-down mass spectrometry.* Anal Chem, 2008. **80**(9): p. 3438-44.

99. Rinfret Robert, C., et al., *Interplay of Ubiquitin-Like Modifiers Following Arsenic Trioxide Treatment.* J Proteome Res, 2020. **19**(5): p. 1999-2010.

100. Fulzele, A. and E.J. Bennett, *Ubiquitin diGLY Proteomics as an Approach to Identify and Quantify the Ubiquitin-Modified Proteome.* Methods Mol Biol, 2018. **1844**: p. 363-384.

101. Ludwig, C., et al., *Data-independent acquisition-based SWATH-MS for quantitative proteomics: a tutorial.* Mol Syst Biol, 2018. **14**(8): p. e8126.

102. Hansen, F.M., et al., *Data-independent acquisition method for ubiquitinome analysis reveals regulation of circadian biology.* Nat Commun, 2021. **12**(1): p. 254.

103. Calderon-Celis, F., J.R. Encinar, and A. Sanz-Medel, *Standardization approaches in absolute quantitative proteomics with mass spectrometry.* Mass Spectrom Rev, 2018. **37**(6): p. 715-737.

104. Cox, J. and M. Mann, *Quantitative, high-resolution proteomics for data-driven systems biology.* Annu Rev Biochem, 2011. **80**: p. 273-99.

105. Ong, S.E. and M. Mann, *Mass spectrometry-based proteomics turns quantitative.* Nat Chem Biol, 2005. **1**(5): p. 252-62.

106. Rodriguez-Suarez, E. and A.D. Whetton, *The application of quantification techniques in proteomics for biomedical research.* Mass Spectrom Rev, 2013. **32**(1): p. 1-26.

107. Cox, J., et al., *Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ.* Mol Cell Proteomics, 2014. **13**(9): p. 2513-26.

108. Zhu, W., J.W. Smith, and C.M. Huang, *Mass spectrometry-based label-free quantitative proteomics.* J Biomed Biotechnol, 2010. **2010**: p. 840518.

109. Hogrebe, A., et al., *Benchmarking common quantification strategies for large-scale phosphoproteomics.* Nat Commun, 2018. **9**(1): p. 1045.

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

111. An, J., et al., *pSILAC mass spectrometry reveals ZFP91 as IMiD-dependent substrate of the CRL4(CRBN) ubiquitin ligase.* Nat Commun, 2017. **8**: p. 15398.

112. Geiger, T., et al., *Super-SILAC mix for quantitative proteomics of human tumor tissue.* Nat Methods, 2010. **7**(5): p. 383-5.

113. Wiese, S., et al., *Protein labeling by iTRAQ: a new tool for quantitative mass spectrometry in proteome research.* Proteomics, 2007. **7**(3): p. 340-50.

114. Thompson, A., et al., *Tandem mass tags: a novel quantification strategy for comparative analysis of complex protein mixtures by MS/MS.* Anal Chem, 2003. **75**(8): p. 1895-904.

115. Rose, C.M., et al., *Highly Multiplexed Quantitative Mass Spectrometry Analysis of Ubiquitylomes.* Cell Syst, 2016. **3**(4): p. 395-403 e4.

116. Niu, M., et al., *Extensive Peptide Fractionation and y1 Ion-Based Interference Detection Method for Enabling Accurate Quantification by Isobaric Labeling and Mass Spectrometry.* Anal Chem, 2017. **89**(5): p. 2956-2963.

117. Wuhr, M., et al., *Accurate multiplexed proteomics at the MS2 level using the complement reporter ion cluster.* Anal Chem, 2012. **84**(21): p. 9214-21.

118. Ting, L., et al., *MS3 eliminates ratio distortion in isobaric multiplexed quantitative proteomics.* Nat Methods, 2011. **8**(11): p. 937-40.

119. Udeshi, N.D., et al., *Rapid and deep-scale ubiquitylation profiling for biology and translational research.* Nat Commun, 2020. **11**(1): p. 359.

120. Perkins, D.N., et al., *Probability-based protein identification by searching sequence databases using mass spectrometry data.* Electrophoresis, 1999. **20**(18): p. 3551-67.

121. Eng, J.K., A.L. McCormack, and J.R. Yates, *An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database.* J Am Soc Mass Spectrom, 1994. **5**(11): p. 976-89.

122. Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.

123. Zhang, J., et al., *PEAKS DB: de novo sequencing assisted database search for sensitive and accurate peptide identification.* Mol Cell Proteomics, 2012. **11**(4): p. M111 010587.

124. Li, Y., et al., *An integrated bioinformatics platform for investigating the human E3 ubiquitin ligase-substrate interaction network.* Nat Commun, 2017. **8**(1): p. 347.

125. Xue, Y., et al., *SUMOsp: a web server for sumoylation site prediction.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W254-7.

126. Ren, J., et al., *Systematic study of protein sumoylation: Development of a site-specific predictor of SUMOsp 2.0.* Proteomics, 2009. **9**(12): p. 3409-3412.

127. Xue, Y., et al., *GPS: a comprehensive www server for phosphorylation sites prediction.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W184-7.

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

129. Zhao, Q., et al., *GPS-SUMO: a tool for the prediction of sumoylation sites and SUMO-interaction motifs.* Nucleic Acids Res, 2014. **42**(Web Server issue): p. W325-30.

130. Beauclair, G., et al., *JASSA: a comprehensive tool for prediction of SUMOylation sites and SIMs.* Bioinformatics, 2015. **31**(21): p. 3483-91.

131.   Dehzangi, A., et al., *SumSec: Accurate Prediction of Sumoylation Sites Using Predicted Secondary Structure.* Molecules, 2018. **23**(12).

132.   Sharma, A., et al., *HseSUMO: Sumoylation site prediction using half-sphere exposures of amino acids residues.* BMC Genomics, 2019. **19**(Suppl 9): p. 982.

133.   Qian, Y., et al., *SUMO-Forest: A Cascade Forest based method for the prediction of SUMOylation sites on imbalanced data.* Gene, 2020. **741**: p. 144536.

134.   Xu, H.D., et al., *mUSP: a high-accuracy map of the in situ crosstalk of ubiquitylation and SUMOylation proteome predicted via the feature enhancement approach.* Brief Bioinform, 2020.

135.   Qiu, W., et al., *Computational Prediction of Ubiquitination Proteins Using Evolutionary Profiles and Functional Domain Annotation.* Curr Genomics, 2019. **20**(5): p. 389-399.

136.   He, F., et al., *Large-scale prediction of protein ubiquitination sites using a multimodal deep architecture.* BMC Syst Biol, 2018. **12**(Suppl 6): p. 109.

137.   Yavuz, A.S., N.B. Sozer, and O.U. Sezerman, *Prediction of neddylation sites from protein sequences and sequence-derived properties.* BMC Bioinformatics, 2015. **16 Suppl 18**: p. S9.

138.   Hornbeck, P.V., et al., *PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse.* Nucleic Acids Res, 2012. **40**(Database issue): p. D261-70.

139.   Cox, J., et al., *A practical guide to the MaxQuant computational platform for SILAC-based quantitative proteomics.* Nat.Protoc., 2009. **4**(5): p. 698-705.

140.   Tyanova, S., T. Temu, and J. Cox, *The MaxQuant computational platform for mass spectrometry-based shotgun proteomics.* Nat Protoc, 2016. **11**(12): p. 2301-2319.

141.   Ma, B., et al., *PEAKS: powerful software for peptide de novo sequencing by tandem mass spectrometry.* Rapid Commun Mass Spectrom, 2003. **17**(20): p. 2337-42.

142.   Hendriks, I.A., et al., *Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation.* Nat Struct Mol Biol, 2017. **24**(3): p. 325-336.

143.   Chachami, G., et al., *Hypoxia-induced Changes in SUMO Conjugation Affect Transcriptional Regulation Under Low Oxygen.* Mol Cell Proteomics, 2019. **18**(6): p. 1197-1209.

144.   Tyanova, S., et al., *The Perseus computational platform for comprehensive analysis of (prote)omics data.* Nat. Methods, 2016.

145.   Fu, H., et al., *DeepUbi: a deep learning framework for prediction of ubiquitination sites in proteins.* BMC Bioinformatics, 2019. **20**(1): p. 86.

146.   Liu, Y., et al., *DeepTL-Ubi: A novel deep transfer learning method for effectively predicting ubiquitination sites of multiple species.* Methods, 2020.

147.   The Gene Ontology, C., *The Gene Ontology Resource: 20 years and still GOing strong.* Nucleic Acids Res, 2019. **47**(D1): p. D330-D338.

148.   Eifler, K., et al., *SUMO targets the APC/C to regulate transition from metaphase to anaphase.* Nat Commun, 2018. **9**(1): p. 1119.

149.   El-Gebali, S., et al., *The Pfam protein families database in 2019.* Nucleic Acids Res, 2019. **47**(D1): p. D427-D432.

150.   Szklarczyk, D., et al., *STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets.* Nucleic Acids Res, 2019. **47**(D1): p. D607-D613.

151.   Psakhye, I. and S. Jentsch, *Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair.* Cell, 2012. **151**(4): p. 807-820.

152.   Schwertman, P., S. Bekker-Jensen, and N. Mailand, *Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers.* Nat Rev Mol Cell Biol, 2016. **17**(6): p. 379-94.

153.   Jackson, S.P. and D. Durocher, *Regulation of DNA damage responses by ubiquitin and SUMO.* Mol.Cell, 2013. **49**(5): p. 795-807.

154.   Xu, P., et al., *Quantitative proteomics reveals the function of unconventional ubiquitin chains in proteasomal degradation.* Cell, 2009. **137**(1): p. 133-145.

155.   Pinto-Fernandez, A., et al., *Deletion of the deISGylating enzyme USP18 enhances tumour cell antigenicity and radiosensitivity.* Br J Cancer, 2021. **124**(4): p. 817-830.

156. Steen, H. and M. Mann, *The ABC's (and XYZ's) of peptide sequencing.* Nat.Rev.Mol.Cell Biol., 2004. **5**(9): p. 699-711.

157. Na, C.H., et al., *Synaptic protein ubiquitination in rat brain revealed by antibody-based ubiquitome analysis.* J Proteome Res, 2012. **11**(9): p. 4722-32.

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

159. Murie, C., et al., *Normalization of mass spectrometry data (NOMAD).* Adv Biol Regul, 2018. **67**: p. 128-133.

160. Vertegaal, A.C., *Uncovering ubiquitin and ubiquitin-like signaling networks.* Chem.Rev., 2011. **111**(12): p. 7923-7940.

161. Aichem, A., et al., *The ubiquitin-like modifier FAT10 interferes with SUMO activation.* Nat Commun, 2019. **10**(1): p. 4452.

162. El-Asmi, F., et al., *Cross-talk between SUMOylation and ISGylation in response to interferon.* Cytokine, 2020. **129**: p. 155025.

163. French, M.E., C.F. Koehler, and T. Hunter, *Emerging functions of branched ubiquitin chains.* Cell Discov, 2021. **7**(1): p. 6.

164. Haakonsen, D.L. and M. Rape, *Branching Out: Improved Signaling by Heterotypic Ubiquitin Chains.* Trends Cell Biol, 2019. **29**(9): p. 704-716.

165. Kane, L.A., et al., *PINK1 phosphorylates ubiquitin to activate Parkin E3 ubiquitin ligase activity.* J Cell Biol, 2014. **205**(2): p. 143-53.

166. Kazlauskaite, A., et al., *Parkin is activated by PINK1-dependent phosphorylation of ubiquitin at Ser65.* Biochem J, 2014. **460**(1): p. 127-39.

167. Koyano, F., et al., *Ubiquitin is phosphorylated by PINK1 to activate parkin.* Nature, 2014. **510**(7503): p. 162-6.

168. Ordureau, A., et al., *Dynamics of PARKIN-Dependent Mitochondrial Ubiquitylation in Induced Neurons and Model Systems Revealed by Digital Snapshot Proteomics.* Mol Cell, 2018. **70**(2): p. 211-227 e8.

169. Wauer, T., et al., *Ubiquitin Ser65 phosphorylation affects ubiquitin structure, chain assembly and hydrolysis.* EMBO J, 2015. **34**(3): p. 307-25.

170.   Baek, K., et al., *NEDD8 nucleates a multivalent cullin-RING-UBE2D ubiquitin ligation assembly.* Nature, 2020. **578**(7795): p. 461-466.

171.   Da Costa, I.C. and C.K. Schmidt, *Ubiquitin-like proteins in the DNA damage response: the next generation.* Essays Biochem, 2020. **64**(5): p. 737-752.

172.   Uzunova, K., et al., *Ubiquitin-dependent proteolytic control of SUMO conjugates.* J Biol Chem, 2007. **282**(47): p. 34167-75.

173.   Cuijpers, S.A.G., E. Willemstein, and A.C.O. Vertegaal, *Converging Small Ubiquitin-like Modifier (SUMO) and Ubiquitin Signaling: Improved Methodology Identifies Co-modified Target Proteins.* Mol Cell Proteomics, 2017. **16**(12): p. 2281-2295.

174.   Pichler, A., et al., *SUMO modification of the ubiquitin-conjugating enzyme E2-25K.* Nat Struct Mol Biol, 2005. **12**(3): p. 264-9.

175.   Ranieri, M., et al., *Sumoylation and ubiquitylation crosstalk in the control of DeltaNp63alpha protein stability.* Gene, 2018. **645**: p. 34-40.

176.   Lecona, E., et al., *USP7 is a SUMO deubiquitinase essential for DNA replication.* Nat. Struct. Mol. Biol, 2016.

177.   Hendriks, I.A. and A.C. Vertegaal, *A comprehensive compilation of SUMO proteomics.* Nat Rev Mol Cell Biol, 2016. **17**(9): p. 581-95.

178.   Prudden, J., et al., *SUMO-targeted ubiquitin ligases in genome stability.* EMBO J, 2007. **26**(18): p. 4089-101.

179.   Sun, H., J.D. Leverson, and T. Hunter, *Conserved function of RNF4 family proteins in eukaryotes: targeting a ubiquitin ligase to SUMOylated proteins.* EMBO J, 2007. **26**(18): p. 4102-12.

180.   Poulsen, S.L., et al., *RNF111/Arkadia is a SUMO-targeted ubiquitin ligase that facilitates the DNA damage response.* J.Cell Biol., 2013. **201**(6): p. 797-807.

181.   Seenivasan, R., et al., *Mechanism and chain specificity of RNF216/TRIAD3, the ubiquitin ligase mutated in Gordon Holmes syndrome.* Hum Mol Genet, 2019. **28**(17): p. 2862-2873.

182. Sriramachandran, A.M., et al., *Arkadia/RNF111 is a SUMO-targeted ubiquitin ligase with preference for substrates marked with SUMO1-capped SUMO2/3 chain.* Nat Commun, 2019. **10**(1): p. 3678.

183. Yin, Y., et al., *SUMO-targeted ubiquitin E3 ligase RNF4 is required for the response of human cells to DNA damage.* Genes Dev, 2012. **26**(11): p. 1196-1208.

184. van Cuijk, L., et al., *SUMO and ubiquitin-dependent XPC exchange drives nucleotide excision repair.* Nat Commun, 2015. **6**: p. 7499.

185. Enchev, R.I., B.A. Schulman, and M. Peter, *Protein neddylation: beyond cullin-RING ligases.* Nat Rev Mol Cell Biol, 2015. **16**(1): p. 30-44.

186. Hicke, L., H.L. Schubert, and C.P. Hill, *Ubiquitin-binding domains.* Nat Rev Mol Cell Biol, 2005. **6**(8): p. 610-21.

187. Husnjak, K. and I. Dikic, *Ubiquitin-binding proteins: decoders of ubiquitin-mediated cellular functions.* Annu Rev Biochem, 2012. **81**: p. 291-322.

188. Aksnes, H., R. Ree, and T. Arnesen, *Co-translational, Post-translational, and Non-catalytic Roles of N-Terminal Acetyltransferases.* Mol Cell, 2019. **73**(6): p. 1097-1114.

189. Bloom, J., et al., *Proteasome-mediated degradation of p21 via N-terminal ubiquitinylation.* Cell, 2003. **115**(1): p. 71-82.

190. Choe, K.N. and G.L. Moldovan, *Forging Ahead through Darkness: PCNA, Still the Principal Conductor at the Replication Fork.* Mol Cell, 2017. **65**(3): p. 380-392.

191. Weisshaar, S.R., et al., *Arsenic trioxide stimulates SUMO-2/3 modification leading to RNF4-dependent proteolytic targeting of PML.* FEBS Lett, 2008. **582**(21-22): p. 3174-8.

192. Lee, H.S., et al., *SUMOylation of nonstructural 5A protein regulates hepatitis C virus replication.* J Viral Hepat, 2014. **21**(10): p. e108-17.

193. McManus, F.P., et al., *Quantitative SUMO proteomics reveals the modulation of several PML nuclear body associated proteins and an anti-senescence function of UBC9.* Sci Rep, 2018. **8**(1): p. 7754.

194. Takeuchi, T., et al., *Link between the ubiquitin conjugation system and the ISG15 conjugation system: ISG15 conjugation to the UbcH6 ubiquitin E2 enzyme.* J Biochem, 2005. **138**(6): p. 711-9.

195. Takeuchi, T. and H. Yokosawa, *ISG15 modification of Ubc13 suppresses its ubiquitin-conjugating activity.* Biochem Biophys Res Commun, 2005. **336**(1): p. 9-13.

196. Bialas, J., M. Groettrup, and A. Aichem, *Conjugation of the ubiquitin activating enzyme UBE1 with the ubiquitin-like modifier FAT10 targets it for proteasomal degradation.* PLoS One, 2015. **10**(3): p. e0120329.

197. Mukhopadhyay, D. and H. Riezman, *Proteasome-independent functions of ubiquitin in endocytosis and signaling.* Science, 2007. **315**(5809): p. 201-5.

198. Ordureau, A., et al., *Global Landscape and Dynamics of Parkin and USP30-Dependent Ubiquitylomes in iNeurons during Mitophagic Signaling.* Mol Cell, 2020. **77**(5): p. 1124-1142 e10.

199. Hendriks, I.A., et al., *Ubiquitin-specific Protease 11 (USP11) Deubiquitinates Hybrid Small Ubiquitin-like Modifier (SUMO)-Ubiquitin Chains to Counteract RING Finger Protein 4 (RNF4).* J Biol Chem, 2015. **290**(25): p. 15526-37.

200. Guzzo, C.M., et al., *RNF4-dependent hybrid SUMO-ubiquitin chains are signals for RAP80 and thereby mediate the recruitment of BRCA1 to sites of DNA damage.* Sci Signal, 2012. **5**(253): p. ra88.

201. Maghames, C.M., et al., *NEDDylation promotes nuclear protein aggregation and protects the Ubiquitin Proteasome System upon proteotoxic stress.* Nat Commun, 2018. **9**(1): p. 4376.

202. Leidecker, O., et al., *The ubiquitin E1 enzyme Ube1 mediates NEDD8 activation under diverse stress conditions.* Cell Cycle, 2012. **11**(6): p. 1142-50.

203. Skaug, B. and Z.J. Chen, *Emerging role of ISG15 in antiviral immunity.* Cell, 2010. **143**(2): p. 187-90.

204. Fan, J.B., et al., *Identification and characterization of a novel ISG15-ubiquitin mixed chain and its role in regulating protein homeostasis.* Sci Rep, 2015. **5**: p. 12704.

205. Anania, V.G., et al., *Peptide level immunoaffinity enrichment enhances ubiquitination site identification on individual proteins.* Mol Cell Proteomics, 2014. **13**(1): p. 145-56.

206. Phu, L., et al., *Dynamic Regulation of Mitochondrial Import by the Ubiquitin System.* Mol Cell, 2020. **77**(5): p. 1107-1123 e10.

207.	Huang, E.Y., et al., *A VCP inhibitor substrate trapping approach (VISTA) enables proteomic profiling of endogenous ERAD substrates.* Mol Biol Cell, 2018. **29**(9): p. 1021-1030.

208.	Perez-Riverol, Y., et al., *The PRIDE database and related tools and resources in 2019: improving support for quantification data.* Nucleic Acids Res, 2019. **47**(D1): p. D442-D450.

209.	Wang, M., et al., *Assembling the Community-Scale Discoverable Human Proteome.* Cell Syst, 2018. **7**(4): p. 412-421 e5.

210.	Okuda, S., et al., *jPOSTrepo: an international standard data repository for proteomes.* Nucleic Acids Res, 2017. **45**(D1): p. D1107-D1111.

211.	Ma, J., et al., *iProX: an integrated proteome resource.* Nucleic Acids Res, 2019. **47**(D1): p. D1211-D1217.

212.	Deutsch, E.W., H. Lam, and R. Aebersold, *PeptideAtlas: a resource for target selection for emerging targeted proteomics workflows.* EMBO Rep, 2008. **9**(5): p. 429-34.

213.	Sharma, V., et al., *Panorama Public: A Public Repository for Quantitative Data Sets Processed in Skyline.* Mol Cell Proteomics, 2018. **17**(6): p. 1239-1244.

214.	Jentsch, S. and I. Psakhye, *Control of Nuclear Activities by Substrate-Selective and Protein-Group SUMOylation.* Annu.Rev.Genet., 2013.

215.	Hewings, D.S., et al., *Activity-based probes for the ubiquitin conjugation-deconjugation machinery: new chemistries, new tools, and new insights.* FEBS J, 2017. **284**(10): p. 1555-1576.

216.	Ovaa, H. and A.C.O. Vertegaal, *Probing ubiquitin and SUMO conjugation and deconjugation.* Biochem Soc Trans, 2018. **46**(2): p. 423-436.

217.	Gui, W., et al., *Cell-Permeable Activity-Based Ubiquitin Probes Enable Intracellular Profiling of Human Deubiquitinases.* J Am Chem Soc, 2018. **140**(39): p. 12424-12433.

218.	Swatek, K.N. and D. Komander, *Ubiquitin modifications.* Cell Res, 2016. **26**(4): p. 399-422.

219.	Gomes, F., et al., *Top-down analysis of novel synthetic branched proteins.* J Mass Spectrom, 2019. **54**(1): p. 19-25.

220.   Olsen, J.V. and M. Mann, *Status of large-scale analysis of post-translational modifications by mass spectrometry.* Mol Cell Proteomics, 2013. **12**(12): p. 3444-52.

221.   Nielsen, M.L., et al., *Iodoacetamide-induced artifact mimics ubiquitination in mass spectrometry.* Nat Methods, 2008. **5**(6): p. 459-60.

222.   Humphrey, S.J., et al., *High-throughput and high-sensitivity phosphoproteomics with the EasyPhos platform.* Nat Protoc, 2018. **13**(9): p. 1897-1916.

223.   Tran, J.C., et al., *Mapping intact protein isoforms in discovery mode using top-down proteomics.* Nature, 2011. **480**(7376): p. 254-8.

224.   Lee, A.E., et al., *Preparing to read the ubiquitin code: top-down analysis of unanchored ubiquitin tetramers.* J Mass Spectrom, 2016. **51**(8): p. 629-637.

225.   Mattern, M., et al., *Using Ubiquitin Binders to Decipher the Ubiquitin Code.* Trends Biochem Sci, 2019. **44**(7): p. 599-615.

226.   Huang, X. and V.M. Dixit, *Drugging the undruggables: exploring the ubiquitin system for drug development.* Cell Res, 2016. **26**(4): p. 484-98.

227.   Verma, R., D. Mohl, and R.J. Deshaies, *Harnessing the Power of Proteolysis for Targeted Protein Inactivation.* Mol Cell, 2020. **77**(3): p. 446-460.

228.   Karayel, O., et al., *DIA-based systems biology approach unveils E3 ubiquitin ligase-dependent responses to a metabolic shift.* Proc Natl Acad Sci U S A, 2020.

229.   Mark, K.G., T.B. Loveless, and D.P. Toczyski, *Isolation of ubiquitinated substrates by tandem affinity purification of E3 ligase-polyubiquitin-binding domain fusions (ligase traps).* Nat Protoc, 2016. **11**(2): p. 291-301.

230.   Kliza, K. and K. Husnjak, *Resolving the Complexity of Ubiquitin Networks.* Front Mol Biosci, 2020. **7**: p. 21.

231.   Li, W., et al., *Genome-wide and functional annotation of human E3 ubiquitin ligases identifies MULAN, a mitochondrial E3 that regulates the organelle's dynamics and signaling.* PLoS One, 2008. **3**(1): p. e1487.

232.   O'Connor, H.F. and J.M. Huibregtse, *Enzyme-substrate relationships in the ubiquitin system: approaches for identifying substrates of ubiquitin ligases.* Cell Mol Life Sci, 2017. **74**(18): p. 3363-3375.

233. Liebelt, F., et al., *The poly-SUMO2/3 protease SENP6 enables assembly of the constitutive centromere-associated network by group deSUMOylation.* Nat Commun, 2019. **10**(1): p. 3987.

234. Andaluz Aguilar, H., et al., *Sequential phosphoproteomics and N-glycoproteomics of plasma-derived extracellular vesicles.* Nat Protoc, 2020. **15**(1): p. 161-180.

235. Melo-Braga, M.N., et al., *Comprehensive protocol to simultaneously study protein phosphorylation, acetylation, and N-linked sialylated glycosylation.* Methods Mol Biol, 2015. **1295**: p. 275-92.

236. Zhang, Y., H. Wang, and H. Lu, *Sequential selective enrichment of phosphopeptides and glycopeptides using amine-functionalized magnetic nanoparticles.* Mol Biosyst, 2013. **9**(3): p. 492-500.

237. Young, N.L., M.D. Plazas-Mayorca, and B.A. Garcia, *Systems-wide proteomic characterization of combinatorial post-translational modification patterns.* Expert Rev Proteomics, 2010. **7**(1): p. 79-92.

238. Manasanch, E.E. and R.Z. Orlowski, *Proteasome inhibitors in cancer therapy.* Nat Rev Clin Oncol, 2017. **14**(7): p. 417-433.

239. Harrigan, J.A., et al., *Deubiquitylating enzymes and drug discovery: emerging opportunities.* Nat Rev Drug Discov, 2018. **17**(1): p. 57-78.

240. Soucy, T.A., et al., *An inhibitor of NEDD8-activating enzyme as a new approach to treat cancer.* Nature, 2009. **458**(7239): p. 732-6.

241. Hyer, M.L., et al., *A small-molecule inhibitor of the ubiquitin activating enzyme for cancer treatment.* Nat Med, 2018. **24**(2): p. 186-193.

242. He, X., et al., *Probing the roles of SUMOylation in cancer cell biology by using a selective SAE inhibitor.* Nat Chem Biol, 2017. **13**(11): p. 1164-1171.

243. Lan, B., et al., *VCP/p97/Cdc48, A Linking of Protein Homeostasis and Cancer Therapy.* Curr Mol Med, 2017. **17**(9): p. 608-618.

244. Ferguson, F.M. and N.S. Gray, *Kinase inhibitors: the road ahead.* Nat Rev Drug Discov, 2018. **17**(5): p. 353-377.

245.   Bekker-Jensen, D.B., et al., *A Compact Quadrupole-Orbitrap Mass Spectrometer with FAIMS Interface Improves Proteome Coverage in Short LC Gradients.* Mol Cell Proteomics, 2020. **19**(4): p. 716-729.

246.   Bache, N., et al., *A Novel LC System Embeds Analytes in Pre-formed Gradients for Rapid, Ultra-robust Proteomics.* Mol Cell Proteomics, 2018. **17**(11): p. 2284-2296.

247.   Doll, S., F. Gnad, and M. Mann, *The Case for Proteomics and Phospho-Proteomics in Personalized Cancer Medicine.* Proteomics Clin Appl, 2019. **13**(2): p. e1800113.

CHAPTER THREE

3 Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility

Chongyang Li[1,2], Francis P. McManus[1], Cédric Plutoni[1], Cristina Mirela Pascariu[1], Trent Nelson[1,2], Lara Elis Alberici Delsin[1,2], Gregory Emery[1,3] & Pierre Thibault[1,4,5*]

[1] Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Québec, Canada.

[2] Molecular Biology Program, Université de Montréal, Montréal, Canada.

[3] Department of Pathology and Cell Biology, Université de Montréal, Montréal, Québec, Canada.

[4] Department of Chemistry, Université de Montréal, Montréal, Québec, Canada.

[5] Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, Québec, Canada.

Authors Contribution:

C.L. generated overexpression and knockout cell lines, and performed phenotypic assay, proteomics experiments and analyzed data. F.P.M. performed vimentin IP assay, C.P. performed IF and FRAP assays. C.M.P performed WB validation of substrates. C.L., F.P.M. and P.T. wrote the manuscript. G.E. and P.T. developed the concept and managed the project. All authors reviewed and edited the final manuscript.

3.1 Abstract

The Protein Inhibitor of Activated STAT 1 (PIAS1) is an E3 SUMO ligase that plays important roles in various cellular pathways. Increasing evidence shows that PIAS1 is overexpressed in various human malignancies, including prostate and lung cancers. Here, we used quantitative SUMO proteomics to identify potential substrates of PIAS1 in a system-wide manner. We identified 983 SUMO sites on 544 proteins, of which 62 proteins were assigned as putative PIAS1 substrates. In particular, Vimentin (VIM), a type III intermediate filament protein involved in cytoskeleton organization and cell motility, was SUMOylated by PIAS1 at Lys-439 and Lys-445 residues. VIM SUMOylation was necessary for its dynamic disassembly, and cells expressing a non-SUMOylatable VIM mutant showed a reduced level of migration. Our approach not only enables the identification of E3 SUMO ligase substrates, but also yields valuable biological insights into the unsuspected role of PIAS1 and VIM SUMOylation on cell motility.

3.2 Introduction

The small ubiquitin-like modifier (SUMO) protein is an ubiquitin-like (UBL) protein that is highly dynamic and can reversibly target lysine residues on a wide range of proteins involved in several essential cellular events, including protein translocation and degradation, mitotic chromosome segregation, DNA damage response, cell cycle progression, cell differentiation and apoptosis [2]. SUMO proteins are highly conserved through evolution, and the human genome encodes 4 SUMO genes, of which 3 genes (*SUMO1*, *SUMO2* and *SUMO3*) are ubiquitously expressed in all cells [2, 3]. Prior to conjugation, the immature SUMO proteins are C-terminally processed by sentrin-specific proteases (SENPs) [4]. These proteases also cleave the isopeptide bond formed between the ε-amino group of the acceptor lysine residues and the C-terminus residue of the conjugated SUMO proteins. The conjugation of SUMO to target proteins requires an E1 activating enzyme (SAE1/2), an E2 conjugating enzyme (UBC9) and one of several E3 SUMO ligases [5]. Unlike ubiquitination, *in vitro* SUMOylation can occur without E3 SUMO ligases, although enhanced substrate specificity is conferred by E3 SUMO ligases [6]. It is believed that SUMOylation events occurring without the aid of E3 SUMO ligases arise primarily on the consensus motif composed of ψKxE, where ψ represents a large hydrophobic residue and x, any amino acid [7]. To date, several structurally unrelated classes of proteins appear to act as E3 SUMO ligases in mammalian cells, such as the protein inhibitor of activated STAT (PIAS) family of proteins, Ran-binding protein 2 (RanBP2), the polycomb group protein (Pc2), and topoisomerase I- and p53-binding protein (TOPORS) [8, 9].

PIAS orthologs can be found throughout eukaryote cells, and comprise four PIAS proteins (PIAS1, PIASx (PIAS2), PIAS3, and PIASy (PIAS4)) that share a high degree of sequence homology [10]. Overall, five different domains or motifs on PIAS-family proteins recognize distinct sequences or conformations on target proteins, unique DNA structures, or specific "bridging" molecules to mediate their various functions [11]. An example of this is the PIAS Scaffold attachment factor (SAP) domain which has a strong affinity towards A–T rich DNA [12] and binds to Matrix attachment regions in DNA [13], in addition to having an important role in substrate recognition [14]. The PINIT motif affects subcellular localization and contributes to substrate

selectivity [15, 16]. The Siz/PIAS RING (SP-RING) domain interacts with UBC9 and facilitates the transfer of SUMO to the substrate [17]. The PIAS SIM (SUMO interaction motif) recognizes SUMO moieties of modified substrates and alters subnuclear targeting and/or assembly of transcription complex [17-19]. While several functions have been attributed to these domains, relatively little is known about the role of the poorly conserved C-terminus serine/threonine-rich region.

PIAS1 is one of the most well studied E3 SUMO ligases, and was initially reported as the inhibitor of signal transducers and activators of transcription 1 (STAT1) [20]. Previous studies indicated that PIAS1 interacts with activated STAT1 and suppresses its binding to DNA [9]. PIAS1 overexpression was reported in several cancers, including prostate cancer, multiple myeloma, and B cell lymphomas [21-24]. PIAS1 can SUMOylate the Focal adhesion kinase (FAK) at Lys-152, a modification that dramatically increases its ability to autophosphorylate Thr-397, activate FAK, and promotes the recruitment of several enzymes including Src family kinases [25]. In yeast, Lys-164 SUMOylation on Proliferating Cell Nuclear Antigen (PCNA) is strictly dependent on the PIAS1 ortholog Siz1, and is recruited to the anti-recombinogenic helicase Srs2 during S-phase [26]. PIAS1 can also regulate oncogenic signaling through the SUMOylation of promyelocytic leukemia protein (PML) and its fusion product with the retinoic acid receptor alpha (PML-RARα) as observed in acute promyelocytic leukemia (APL) [27]. In addition to its regulatory role in PML/PML-RARα oncogenic signaling, PIAS1 has been shown to be involved in the cancer therapeutic mechanism of arsenic trioxide (ATO). This is accomplished by ATO promoting the hyper-SUMOylation of PML-RARα in a PIAS1-dependent fashion, resulting in the ubiquitin-dependent proteasomal degradation of PML-RARα and APL remission [27]. In B cell lymphoma, PIAS1 has been reported as a mediator in lymphomagenesis through SUMOylation of MYC, a proto-oncogene transcription factor associated with several cancers. SUMOylation of MYC leads to a longer half-life and therefore an increase in oncogenic activity [24]. Altogether, these reports suggest that PIAS1 could promote cancer cell growth and progression by regulating the SUMOylation level on a pool of different substrates.

In this study, we first evaluate the effects of PIAS1 overexpression in HeLa cells. PIAS1 overexpression has a significant influence on cell proliferation, cell migration and motility. To

identify putative PIAS1 substrates, we develop a system level approach based on quantitative SUMO proteomic analysis [28] to profile changes in protein SUMOylation in cells overexpressing this E3 SUMO ligase. Our findings reveal that 91 SUMO sites on 62 proteins were regulated by PIAS1. Bioinformatic analysis indicates that many PIAS1 substrates are involved in transcription regulation pathways and cytoskeleton organization. Interestingly, several PIAS1 substrates, including cytoskeletal proteins (Actin filaments, Intermediate filaments and Microtubules), are SUMOylated at lysine residues located in non-consensus motif. We confirm the SUMOylation of several PIAS1 substrates using both a reconstituted *in vitro* and cell based *in vitro* SUMOylation assays. Further functional studies reveal that PIAS1 mediates the SUMOylation of vimentin (VIM) at two conserved sites on its C-terminus that affect the dynamic disassembly of this intermediate filament protein.

3.3 Methods

3.3.1 Cell Culture, Vector Construction and Gene Knockout

Human cervical cancer cell line (HeLa) (ATCC® CCL-2™, Cedarlane) and HEK293 stably expressing the 6xHis-SUMO3-Q87R/Q88N mutant (HEK293-SUMO3m) [28] were cultured in Dulbecco's modified Eagle's medium (DMEM) (HyClone) supplemented with 10% fetal bovine serum (Wisent), 1% L-glutamine (Thermo Fisher Scientific), 1% penicillin/streptomycin (Invitrogen) in 5% CO_2 at 37 °C.

The mammalian expression vector for Myc-PIAS1 was constructed by inserting the full-length cDNAs into pcDNA3.0-Myc. The mammalian expression vector for PIAS1-GFP-WT was constructed by cloning full-length PIAS1 cDNA into pcDNA3.1-c-GFP10. The mammalian expression vectors pReceiver-M11 (Flag-NSMCE2, Flag-PFDN2, Flag-VIM and Flag-empty control) were purchased from Genecopoeia, Inc. (Rockville, MD). The mammalian expression vector Emerald-Vimentin was purchased from addgene (#54300). The PIAS1-GFP-K137R, PIAS1-GFP-K238R, PIAS1-GFP-K315R and PIAS1-GFP-3XKR, Flag-VIM^{mt} and Emerald-VIM^{mt} plasmids were generated by site-directed mutagenesis using the GeneArt Site-Directed Mutagenesis System according to the instructions of the manufacturer (InvitrogenTM). The PIAS CRISPR/Cas9-based gene knockout vectors pCRISPR were also purchased from Genecopoeia, Inc. (Rockville, MD). Primer sequences used for vector construction, site-directed mutagenesis and sgRNA sequences used for CRISPR/Cas9 gene knockout are listed below: PIAS1_F: 5'-GGGTACCATGGCGGACAGTGCGGAAC-3', PIAS1_R: 5'-GGAATTCTCAGTCCAATGAAATAATGTCTGG-3', PIAS1_K137R_F: 5'-GTCCATCCGGATATAAGACTTCAAAAATTACCA-3', PIAS1_K137R_R: 5'-TGGTAATTTTTGAAGTCTTATATCCGGATGGAC-3', PIAS1_K238R_F: 5'-TACCTTCCACCTACAAGAAATGGCGTGGAACCA-3', PIAS1_K238R_R: 5'-TGGTTCCACGCCATTTCTTGTAGGTGGAAGGTA-3', PIAS1_K315R_F: 5'-GCTTTAATTAAAGAGAGGTTGACTGCGGATCCA-3', PIAS1_K315R_R: 5'-CGGATCCGCAGTCAACCTCTCTTTAATTAAAGC-3', VIM_K439R_F: 5'-GTTGATACCCACTCAAGAAGGACACTTCTGATT-3', VIM_K439R_R: 5'-

AATCAGAAGTGTCCTTCTTGAGTGGGTATCAAC-3', VIM_K445R_F: 5'-AGGACACTTCTGATTAGGACGGTTGAAACTAGA-3', VIM_K445R_R: 5'-TCTAGTTTCAACCGTCCTAATCAGAAGTGTCCT-3', PIAS1-sgRNA: 5'-TTCTGAACTCCAAGTACTGT-3', PIAS2-sgRNA: 5'-CAAGTATTACTAGGCTTTGC-3', PIAS3-sgRNA: 5'-GCCCTTCTATGAAGTCTATG-3', PIAS4-sgRNA: 5'-GGCTTCGCGCCGTAGTCTTAG-3' and scrambled sgRNA: 5'-GGCTTCGCGCCGTAGTCTTA-3'.

For transient transfection, cells were transfected with 1 µg plasmid per million cells using JetPrime Reagent (Polyplus-transfection) according to the manufacturer's protocol. Cells were harvested 36h or 48h after transfection for further experiments and protein overexpression was confirmed by western blot.

For PIAS gene knockout, cells were transfected with 1 µg plasmid per million cells using JetPrime Reagent (Polyplus-transfection) according to the manufacturer's protocol. Cells were sorted by FACS based on Red Fluorescent Proteins-mCherry signals 48h after transfection. mCherry positive cells were cultured and expended for another week (Supplementary Figure 3-1). PIAS knockout efficiency was confirmed by western blot.

3.3.2 Cell Proliferation Assay

The cell proliferation assay was carried out using WST-1 Cell Proliferation Assay Kit (Roche). 48h post-transfected cells or knockout (KO) cells were seeded into 96-well plates at a density of 1000 cells/well. WST-1 reagents were added to each well at time point day 0, day 2 and day 4, and cells were further incubated at 37 °C for 1h. The absorbance was measured on a microplate Reader Infinite® M1000 PRO (TECAN) with a test wavelength at 450 nm and a reference wavelength at 630 nm The relative numbers of viable cells were estimated by subtracting the 630 nm background absorbance from 450 nm measurements.

3.3.3 Cell Migration Assays

Cell motility was evaluated by using a wound-healing assay as follows: 36h post-transfection, cells were harvested by a brief trypsinization and were seeded in Ibidi wound healing 2 well-

Culture-Inserts (Ibidi) into 24-well plates. Cells were grown to confluence in DMEM containing 10% FBS for another 12h before the Ibidi wound healing 2 well-Culture-Inserts were removed. The cells were washed twice with PBS to remove the cell debris and grown in DMEM containing 1% FBS. The cell migration into the gap area was observed and photographed at time 0h, 24h and 48h. Closure of the gap was measured using a phase-contrast microscope. Wound healing was analyzed using "MRI Wound Healing Tool" plugin in ImageJ and estimated as percentage of the starting wound area.

3.3.4 SILAC Labeling and Protein Extraction

HEK293-SUMO3m cells were grown in DMEM (Thermo Fisher Scientific) containing light (0Lys, 0Arg), medium (4Lys, 6Arg) or heavy (8Lys, 10Arg) isotopic forms of lysine and arginine (Silantes) for at least 6 passages to ensure full labeling. For each triple SILAC experiment, the control channel was transfected with an empty-pcDNA3.0-Myc vector while the other two channels were transfected with the Myc-PIAS1 plasmid. Similar conditions were used for the culture of PIAS KO cells using the combination scheme described in Supplementary Figure 3-9. After 48h transfection, an equal amount of cells from each SILAC channel were combined and washed twice with ice-cold PBS, lysed in NiNTA denaturing incubation buffer (6 M Guanidinium HCl, 100 mM NaH$_2$PO$_4$, 20 mM 2-Chloroacetamide, 5 mM 2-Mercaptoethanol, 10 mM Tris-HCl pH=8) and sonicated. Protein concentration was determined using micro Bradford assay (Bio-Rad).

3.3.5 Protein Purification, Digestion and Desalting

For each replicate, 16 mg of total cell extract (TCE) were incubated with 320 μL of NiNTA beads (Qiagen) at 4 °C. After 16h incubation, NiNTA beads were washed once with 10 mL of NiNTA denaturing incubation buffer, 5 times with 10 mL of NiNTA denaturing washing buffer (8 M urea, 100 mM NaH$_2$PO$_4$, 20 mM imidazole, 5 mM 2-Mercaptoethanol, 20 mM Chloroacetamide, 10 mM Tris-HCl pH=6.3) and twice with 10 mL of 100 mM ammonium bicarbonate. Protein concentration was determined using micro Bradford assay (Bio-Rad). Protein digestion on beads was carried out using a ratio 1:50 sequencing grade modified trypsin (Promega): protein extract in 100 mM ammonium bicarbonate at 37 °C overnight. Proteins from

the flowthrough of the NiNTA purification were trypsinized using a ratio 1:50 sequencing grade modified trypsin (Promega) at 37 °C overnight for further proteome analyses. To quench the reaction, 0.1% trifluoroacetic acid (TFA) was added. The solution was desalted on hydrophilic-lipophilic balance (HLB) cartridges (1cc, 30 mg) (Waters) and eluted in LoBind tubes (Eppendorf) before being dried down by Speed Vac.

3.3.6 SUMO Peptide Enrichment

PureProteome protein A/G magnetic beads (Millipore) were equilibrated with anti-K (NQTGG) antibody (UMO-1-7-7, Abcam) at a ratio of 1:2 (v/w) for 1h at 4 °C in PBS. Saturated beads were washed 3 times with 200 mM triethanolamine pH=8.3. For crosslinking, 10 µl of 5 mM DMP in 200 mM triethanolamine pH=8.3 was added per µl of slurry and incubated for 1h at room temperature. The reaction was quenched for 30min by adding 1 M Tris-HCl pH=8 to a final concentration 5 mM. Cross-linked beads were washed 3 times with ice-cold PBS and once with PBS containing 50% glycerol. The tryptic digests were resuspended in 500 µl PBS containing 50% glycerol and supplemented with cross-linked anti-K-(NQTGG) at a ratio of 1:2 (w/w). After 1h incubation at 4 °C, anti-K-(NQTGG) antibody bound beads were washed three times with 1 ml of 1 × PBS, twice with 1 ml of 0.1 × PBS and once with double-distilled water. SUMO peptides were eluted three times with 200 µl of 0.2% formic acid in water and dried down by Speed Vac.

3.3.7 SCX Fractionation

Enriched SUMO peptides were reconstituted in water containing 15% acetonitrile and 0.2% formic acid and loaded on conditioned strong cation exchange (SCX) StageTips (Thermo Fisher Scientific). Peptides were eluted with ammonium formate pulses at 50, 75, 100, 300, 600 and 1,500 mM in 15% acetonitrile, pH=3. Eluted fractions were dried down by Speed Vac and stored at −80 °C for MS analysis.

3.3.8 Sample Fractionation and In-gel Digestion

Cell pellets were resuspended in 5 pellet volumes of ice cold RIPA buffer. The lysate was spun at 13000 g for 10min to separate the extract into RIPA soluble and insoluble fractions. The soluble and insoluble VIM samples were separated on a 4-12% SDS-PAGE (Bio-rad), and the

proteins were visualized by coomassie staining. The gel lane around vimentin corresponding position (54 kDa) was cut and then diced into ~1 mm 3 cubes. During the process of in-gel digestion, the gel pieces were first destained completely using destaining solution (50% H_2O, 40% methanol, and 10% acetic acid). Then, the gel pieces were dehydrated by washing several times in 50% acetonitrile (ACN) until the gel pieces shriveled and looked completely white. The proteins were reduced in 10 mM DTT at 56 °C for 30min, alkylated in 55 mM chloroacetamide at room temperature (RT) in the dark for 30min, and digested overnight with 300 ng of sequencing grade modified trypsin in 50 mM ammonium bicarbonate. The supernatants were transferred into Eppendorf tubes, and the gel pieces were sonicated twice in extraction buffer (67% ACN and 2.5% TFA). Finally, the peptide extraction and the initial digest solution supernatant were combined and then dried down using a Speed Vac and stored at -80 °C for MS analysis.

3.3.9 Mass Spectrometry Analysis

For PIAS1 substrates identification and proteome analysis, peptides were reconstituted in water containing 0.2% formic acid and analyzed by nanoflow-LC-MS/MS using an Orbitrap Fusion Mass spectrometer (Thermo Fisher Scientific) coupled to a Proxeon Easy-nLC 1000. Samples were injected on a 300 μm ID × 5 mm trap and separated on a 150 μm × 20 cm nano-LC column (Jupiter C18, 3 μm, 300 Å, Phenomenex). The separation was performed on a linear gradient from 7 to 30% acetonitrile, 0.2% formic acid over 105min at 600 nL/min. Full MS scans were acquired from m/z 350 to m/z 1,500 at resolution 120,000 at m/z 200, with a target AGC of 1E6 and a maximum injection time of 200ms. MS/MS scans were acquired in HCD mode with a normalized collision energy of 25 and resolution of 30,000 using a Top 3s method, with a target AGC of 5E3 and a maximum injection time of 3,000ms. The MS/MS triggering threshold was set at 1E5 and the dynamic exclusion of previously acquired precursor was enabled for 20s within a mass range of ±0.8 Da.

For targeted LC-MS/MS analyses of PIAS KO cells generated through CRISPR/Cas9 gene knockout technology, SUMO peptides obtained from SUMO peptide enrichment were analyzed using an inclusion list to detect and identify each isotopologue of the VIM SUMOylated peptides at K439 (e.g. ETNLDSLPLVDTHSK*R) and K445 (e.g. TLLIK*TVETR). The SUMO peptides were

155

reconstituted in water containing 0.2% formic acid and analyzed on an Orbitrap Q Exactive HF system (Thermo Fisher Scientific) coupled to a Proxeon Easy-nLC 1000. Samples were injected on a 300 μm ID × 5 mm trap and separated on a 150 μm × 20 cm nano-LC column (Jupiter C18, 3 μm, 300 Å, Phenomenex). The separation was performed on a linear gradient from 7 to 30% ACN, 0.2% formic acid over 105min at 600 nL/min. The mass spectrometer was operated in a targeted-MS^2 acquisition mode with a maximum injection time of 1000ms, 1 microscan, 30 000 resolution, 2E5 AGC target, 1.6 m/z isolation window, and 25% normalized collision energy.

For in-gel digested sample analyses, peptides were reconstituted in water containing 0.2% formic acid and analyzed on an Orbitrap Q Exactive HF system (Thermo Fisher Scientific) coupled to a Proxeon Easy-nLC 1000. Samples were injected on a 300 μm ID × 5 mm trap and separated on a 150 μm × 20 cm nano-LC column (Jupiter C18, 3 μm, 300 Å, Phenomenex). The separation was performed on a linear gradient from 7 to 30% ACN, 0.2% formic acid over 105 min at 600 nL/min. Full MS scans were acquired from m/z 350 to m/z 1,200 at resolution 120,000 at m/z 200, with a target AGC of 5E6 and a maximum injection time of 50ms. The precursor isolation window is set to 1.6 m/z with an offset of 0.3 m/z. MS/MS scans were acquired in HCD mode with a normalized collision energy of 25 and resolution of 30,000 using a TopN = 5 method, with a target AGC of 2E4 and a maximum injection time of 1,000ms.

3.3.10 Data Processing

The SUMO proteome MS data were analyzed using MaxQuant (version 1.5.3.8) [29, 30]. MS/MS spectra were searched against UniProt/SwissProt database (http://www.uniprot.org/) including Isoforms (released on 10 March 2015). The maximum missed cleavage sites for trypsin was set to 2. Carbamidomethylation (C) was set as a fixed modification and acetylation (Protein N term), oxidation (M), deamination (NQ) and NQTGG (K) were set as variable modifications. The option match between runs was enabled to correlate identification and quantitation results across different runs. The false discovery rate for peptide, protein, and site identification was set to 1%. SUMO sites with a localization probability of >0.75 were retained.

The flow through proteome MS data were searched with PEAKS X engine (Bioinformatics Solutions, Inc.) against the UniProt/SwissProt database (http://www.uniprot.org/) released on June 05, 2019. The precursor tolerance was set to 10 ppm and fragment ion tolerance to 0.01Da. The maximum missed cleavage sites for trypsin was set to 2. Carbamidomethylation (C) was set as a fixed modification and oxidation (M), deamination (NQ) and NQTGG (K), $^2H_4$-lysine and $^{13}C_6$-arginine (SILAC medium) and $^{15}N_2^{13}C_6$-lysine and $^{15}N_4^{13}C_6$-arginine (SILAC heavy) were set as variable modifications with a maximum of five modifications per peptide. The false discovery rate for peptides was set to 1.0% with decoy removal. Proteins were quantified with ≥2 unique peptides. The relative change in protein abundance across samples was determined using the PEAKS X software. The targeted MS data were analyzed using Skyline version 19.1, MacCoss Lab Software, Seattle, WA; (https://skyline.ms/wiki/home/software/Skyline/page.view?name=default), fragment ions for each targeted mass were extracted, and peak areas were integrated. Fold change ratios between the CTL and KO samples were calculated based on peak areas after normalizing peak intensities using the normalization factor determined from the proteome analysis.

The in-gel digested MS data were analyzed using MaxQuant (version 1.5.3.8) [29, 30]. MS/MS spectra were searched against UniProt/SwissProt database (http://www.uniprot.org/) including Isoforms (released on 10 March 2015). The maximum missed cleavage sites for trypsin was set to 2. Carbamidomethylation (C) was set as a fixed modification and acetylation (Protein N term), oxidation (M), deamination (NQ) and phosphorylation (STY) were set as variable modifications. The false discovery rate for peptide, protein, and site identification was set to 1%. Phospho sites with a localization probability of >0.75 were retained.

3.3.11 Bioinformatics Analysis

Classification of identified PIAS1 substrates was performed using PANTHER (Protein Analysis Through Evolutionary Relationships) (http://www.pantherdb.org), which classifies genes and proteins by their functions [31, 32]. The identified PIAS1 substrates were grouped into the biological process, molecular function and cellular component classes against the background of quantified SUMOylome using DAVID Bioinformatics Resources 6.7 [33]. The aligned peptide

sequences with ±6 amino acids sounding the modified lysine residue obtained in Andromeda were submitted to IceLogo [34]. For peptide sequences corresponding to multiple proteins, only the leading sequence was submitted. The secondary structures surrounding the PIAS1-regulated SUMO sites were investigated using NetSurfP-1.1 [35]. INTERPRO Protein Domains Analysis and Protein-Protein Interaction (PPI) network of identified PIAS1 substrates were built by searching against the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database version 9.1 [36, 37]. All predictions were based on experimental evidence with the minimal confidence score of 0.4, which is considered as the highest confidence filter in STRING. PPI networks were then visualized by Cytoscape v3.5.1 [38, 39].

3.3.12 Recombinant Protein Purification and in vitro SUMO Assay

The bacterial expression vectors pReceiver-B11 for His-NSMCE2 and His-PFDN2 recombinant protein expression were purchased from Genecopoeia, Inc (Rockville, MD). pReceiver-B11 was individually transformed into ArcticExpress Competent Cells (#230191) which were derived from *E. coli* B strains (Agilent Technologies). Transformed bacteria were cultured in LB medium until an Optical Density (OD) 600 of 0.5. Isopropyl β-D-1-thiogalactopyranoside (IPTG) (Bioshop) was added to the culture at a final concentration of 1 mM and the desired expression of the protein was inducted for 24h at 13 °C. Harvested bacterial pellets were lysed and sonicated in a solution containing 500 mM NaCl, 5 mM imidazole, 5 mM β-mercaptoethanol, 50 mM Tris-HCl pH=7.5. His-tagged proteins were purified on NiNTA beads (Qiagen). Purified proteins were eluted with 500 mM NaCl, 250 mM imidazole, 5 mM β- mercaptoethanol , 50 mM Tris-HCl pH=7.5 and concentrated on 3 kDa Ultra centrifugal filters (Amicon). The *in vitro* SUMO assay was carried out in a buffer containing 5 mM MgCl$_2$, 5 mM ATP, 50 mM Tris-HCl pH=7.5. The enzymes responsible for the SUMOylation reaction were added at the following concentrations: 0.1 µM SAE1/2, 1 µM UBC9, 10 µM SUMO-3 and 60 nM PIAS1. The reactions were incubated at 37 °C for 4h and analyzed by western blot.

3.3.13 Cell based in vitro SUMO Assay and Immunoprecipitation

HEK293-SUMO3m cells or HeLa cells were co-transfected with Myc-PIAS1, Flag-NSMCE2, Flag-PFDN2 or Flag-VIM. At 48h post-transfection, cells were harvested and lysed with 1 ml of

Triton lysis buffer (150 mM NaCl, 0.2% Triton X-100, 1 mM EDTA, 10% Glycerol, 50 mM Tris-HCl pH 7.5) supplemented with protease inhibitor cocktail (Sigma Aldrich) and phosphatase Inhibitor Cocktail (Sigma Aldrich) at 4 °C for 15 min with gentle rocking. TCE was incubated with anti-Flag M2 Affinity Agarose Gel (Sigma Aldrich) with gentle rocking at 4 °C for overnight. Immunoprecipitates were then washed three times with cold Triton lysis buffer and were analyzed by western blot.

3.3.14 Western Blot

TCE prepared in Triton lysis buffer were diluted in Laemmli buffer (10% (w/v) glycerol, 2% SDS, 10% (v/v) 2-mercaptoethanol and 62.5 mM Tris-HCl, pH=6.8), boiled for 10min and separated on a 4–12% SDS-PAGE (Bio-rad) followed by transfer onto nitrocellulose membranes. Before blocking the membrane for 1h with 5% non-fat milk in TBST (Tris-buffered saline with Tween 20), membranes were briefly stained with 0.1% Ponceau-S in 5% acetic acid to represent total protein content. Membranes were subsequently incubated overnight with a 1:1000 dilution of antibodies at 4 °C. Membranes were then incubated with peroxidase-conjugated Rabbit-anti-mouse IgG (Light Chain Specific) (58802S, Cell Signaling Technology) or Goat-anti-rabbit IgG (7074S, Cell Signaling Technology) for 1h at room temperature at a 1:5000 dilution. Membranes were washed three times with TBST for 10min each and revealed using ECL (GE Healthcare) as per the manufacturer's instructions. Chemiluminescence was captured on Blue Ray film (VWR). The following antibodies were used for western blot analyses: rabbit anti-Flag Antibody (F7425, Sigma Aldrich), mouse anti-Flag Antibody (F3165, Sigma Aldrich), rabbit anti-PIAS1 Antibody (3550S, Cell Signaling Technology), rabbit anti-PIAS2 Antibody (ab126601, Abcam), rabbit anti-PIAS3 Antibody (9042S, Cell Signaling Technology), rabbit anti-PIAS4 Antibody (4392S, Cell Signaling Technology), rabbit anti-Myc-Tag Antibody (2278S, Cell Signaling Technology), rabbit anti-His-Tag Antibody (SAB4301134, Sigma Aldrich), rabbit anti-α-Tubulin Antibody (2144S, Cell Signaling Technology), rabbit anti-β-Actin Antibody (4970S, Cell Signaling Technology), rabbit anti-Histone H3 Antibody (4499S, Cell Signaling Technology).

3.3.15 Fluorescence Imaging and Co-localization Analysis

HEK-SUMO3m cells were plated on 12-mm-diameter coverslips until they reached the desired density level and then transfected with the desired plasmids for 48h. Cells were fixed in 4% paraformaldehyde in PBS for 15min, followed by a 2-min permeabilization with 0.1% Triton X-100 in PBS and saturation with 2% BSA in PBS for 15min. Cells were incubated with the rabbit anti-PML antibody (sc-5621, santa cruz) with a 1:100 dilution for 1 h at 37 °C, rinsed and incubated with secondary antibodies conjugated to Alexa Fluor 555 Phalloidin (8953S, Cell Signaling Technology) with a 1:250 dilution and DAPI (D9542, Sigma Aldrich) with a 1:10,000 dilution for 1h at room temperature. Both primary and secondary antibodies are diluted in PBS / BSA 2%. To reach sub-diffraction resolution, images from fixed samples were acquired with the Airyscan detector of a Zeiss LSM880 confocal equipped with a 63X/1.43 Plan Apochromat oil immersion objective. The number of PIAS1 and PML positive structures were automatically detected and assessed by using the "particle analysis" tool in ImageJ. The Pearson's correlation coefficient was analyzed using ImageJ.

3.3.16 Fluorescence Recovery After Photobleaching (FRAP) assay

MCF-7 cells (ATCC® HTB-22™, Cedarlane) were plated in μ-Dish 35 mm dishes (Ibidi) until they reached the desired density level and then transfected with the Emerald-VIM^{wt} or Emerald-VIM^{mt} plasmids for 48h. The FRAP assays were conducted on a LSM 880 confocal microscope equipped with a thermostatized chamber at 37 °C. The Vimentin Emerald expressing cells were detected using a GaAsp detector. Bleaching was done by combining "Time", "Bleach" and "Region" modes on Zen software from Zeiss. Briefly, 5 pre-bleach images were taken every 5sec, after which five pulses of a 488 nm laser were applied to bleach an area of 25 × 2 μm. Post-bleach images were acquired every 5 sec for 5min. For fluorescence recovery analysis, the intensity in the bleached region was measured varying time points with "Frap profiler" plugin in ImageJ. Bleach data were normalized to unbleached regions for all the time points and expressed in arbitrary units in the recovery graphs.

3.3.17 Quantification of Vimentin Organization

Fixed images of cells expressing Emerald-VIM^{wt} or Emerald-VIM^{mt} were taken using a LSM 880 confocal microscope. The "title" mode in the Zen software from Zeiss was used to cover a

large area of cells (2,13mm x 2,13mm). To avoid localization and conformation artefacts due to expression levels, only the cells expressing Emerald-Vimentin at an intermediate level were evaluated using the threshold module in ImageJ. The same manual threshold was used for all conditions. Cells were tabulated using the "Cell counter" plugin in ImageJ into 4 categories: ULF, VIFs, cytosolic and total cells. Category specifications were performed manually and subjectively. A representative image of each category is shown in Supplementary Figure 12.

3.3.18 Statistical Analysis

Statistical analysis was carried out to assess differences between experimental groups. Data were presented as the means ± S.D.. Statistical significance was analyzed by the Student's t-tests. $p<0.05$ was considered to be statistically significant. One asterisk and two asterisks indicate $p<0.05$ and $p<0.01$, respectively.

3.3.19 Data Availability

The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository with the dataset identifier PXD011932 (http://proteomecentral.proteomexchange.org/cgi/GetDataset?ID=PXD011932).

3.4 Results

3.4.1 PIAS1 Regulates HeLa Cell Proliferation and Motility

To investigate the physiological function of PIAS1 in HeLa cells, we overexpressed PIAS1 (Figure 3-1a) and generated a PIAS1 knockout cell line (Supplementary Figure 3-1 and Figure 3-1b). For the PIAS1 overexpression, the abundance of PIAS1 in HeLa cells was increased by 6-fold at 48h post-transfection (Figure 3-1a). PIAS1 overexpression promotes HeLa cell proliferation by ~50% (Figure 3-1c), whereas the knockout cell reduced the rate of proliferation by ~50% (Figure 3-1d). We further examined the phenotypic effects of PIAS1 expression on cell migration using a wound-healing assay. The ability of Hela cells to migrate was increased after PIAS1 overexpression (Figure 3-1e), whereas the knockout cell displayed a reduced rate of migration (Figure 3-1g). Taken together, these results highlight the role that PIAS1 plays in regulating cell growth and cell migration in HeLa cells.
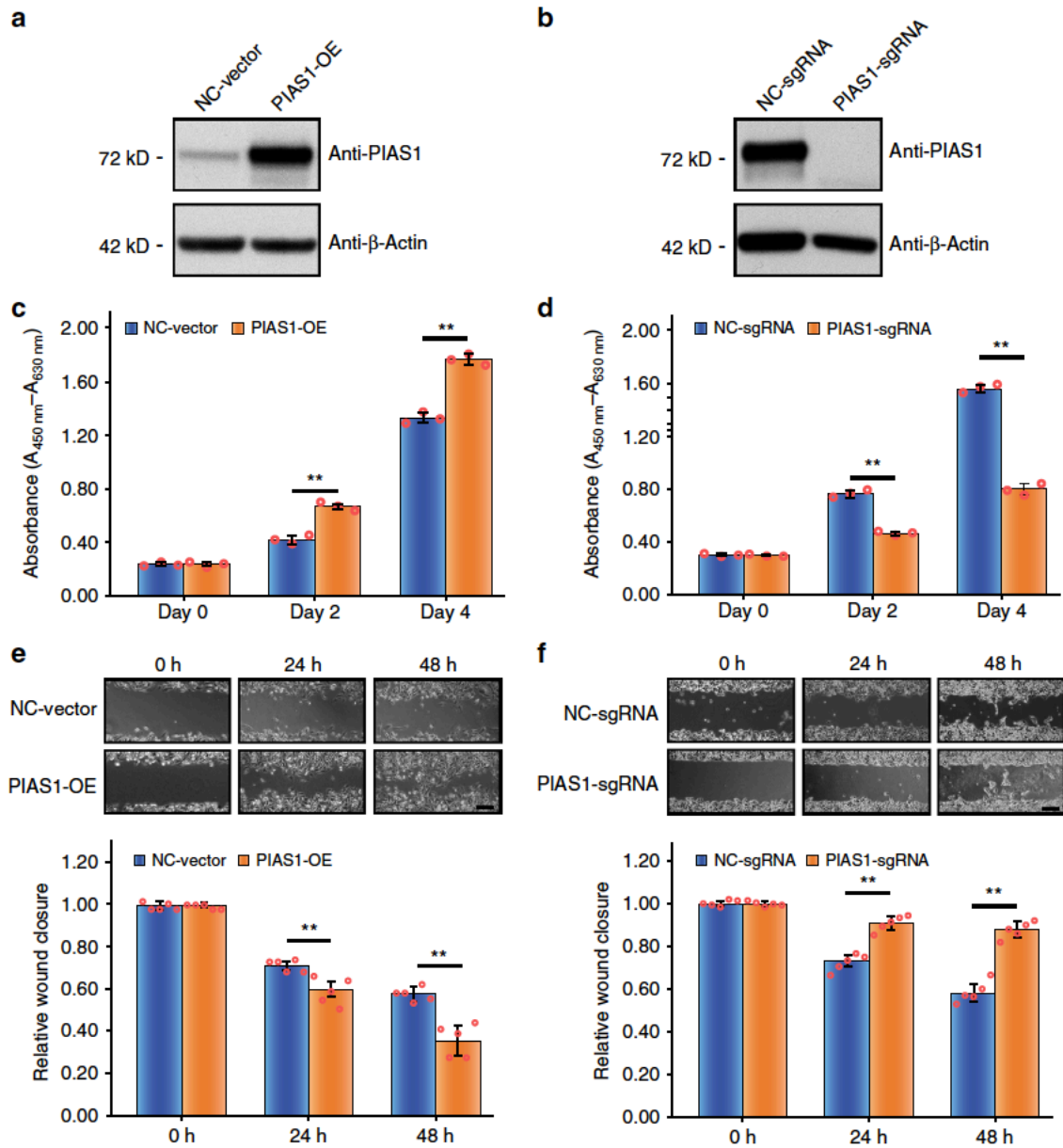
Figure 3-1 Functional effects of PIAS1 expression on HeLa cells.

(a) HeLa cells were transfected with Myc-PIAS1 (PIAS1-OE) or Empty vector (NC-vector) for 48 h. PIAS1 overexpression efficiency was determined by western blot. Actin was used as a loading control. (b) PIAS1 KO HeLa cells were generated by CRISPR/Cas9. PIAS1 KO was determined by western blot. Actin was used as a loading control. (c) PIAS1 overexpression in HeLa cells significantly promotes cell growth, n=3 biologically independent samples. (d) PIAS1 KO HeLa cells (PIAS1-sgRNA) showed impeded cell growth compared to the negative control cells (NC-sgRNA), n=3 biologically independent samples. (e) PIAS1 overexpression increases cell migration as determined by a wound-healing assay. Scale bar: 200 μm, n=5 biologically independent samples. (f) PIAS1 KO cells displayed decreased cell migration compared to the sgRNA negative control cells,

as determined by a wound-healing assay. Scale bar: 200 µm, n=5 biologically independent samples. Data represent the mean ± S.D., error bars represent S.D., **p<0.01, Student's t-test. Source data are provided as a Source Data file.

3.4.2 Identification of PIAS1 Substrates by SUMO Proteomics

To gain a better understanding of the role that PIAS1 plays in cell proliferation, migration and motility, we modified our previously published large-scale SUMO proteomic approach to identify PIAS1 substrates in a site-specific manner (Figure 3-2) [40]. We combined a SUMO remnant immunoaffinity strategy [28] with metabolic labeling (stable isotope labeling of amino acid in cell culture, SILAC) to study the global changes in protein SUMOylation upon PIAS1 overexpression. HEK293 cells stably expressing SUMO3m (Supplementary Figure 3-2) were grown at 37 °C in media containing light (^{0}Lys, ^{0}Arg), medium (^{4}Lys, ^{6}Arg), or heavy (^{8}Lys, ^{10}Arg) isotopic forms of lysine and arginine. Three biological replicates were performed, and for each replicate, one SILAC channel was transfected with an empty vector while the other two were transfected with Myc-PIAS1 vectors (Figure 3-2a). At 48 h post-transfection, an equal amount of cells from each SILAC channel were harvested and combined before lysis in a highly denaturant buffer. PIAS1 overexpression efficiency in HEK293 SUMO3m cells was evaluated by western blot (Figure 3-2b). Protein extracts were first purified by NiNTA beads to enrich SUMO-modified proteins and digested on beads with trypsin (Figure 3-2c). Following tryptic digestion, SUMO-modified peptides were immunopurified using an antibody directed against the NQTGG remnant that is revealed on the SUMOylated lysine residue. Next, peptides were fractionated by offline strong cation exchange (SCX) STAGE tips and analyzed by LC-MS/MS on a Tribrid Fusion instrument. To determine that abundance changes were attributed to SUMOylation and not to change in protein expression, we also performed quantitative proteomic analyses on the total cell extracts from PIAS1 overexpression (Figure 3-3a and Supplementary Figure 3-3). PIAS1 overexpression caused a global increase in protein SUMOylation with negligible changes on protein abundance (Figure 3-3b). In total, 12080 peptides on 1756 proteins (Figure 3-3a, Supplementary Table 3-1) and 983 SUMO peptides on 544 SUMO proteins (Figure 3-3b, Supplementary Table 3-2) were quantified for the proteome and SUMO proteome analyses,

respectively. A total of 91 SUMOylation sites on 62 proteins were found to be upregulated by PIAS1 overexpression including its known substrate promyelocytic leukemia (PML) protein. A summary of these analyses is shown in Figure 3-3c.
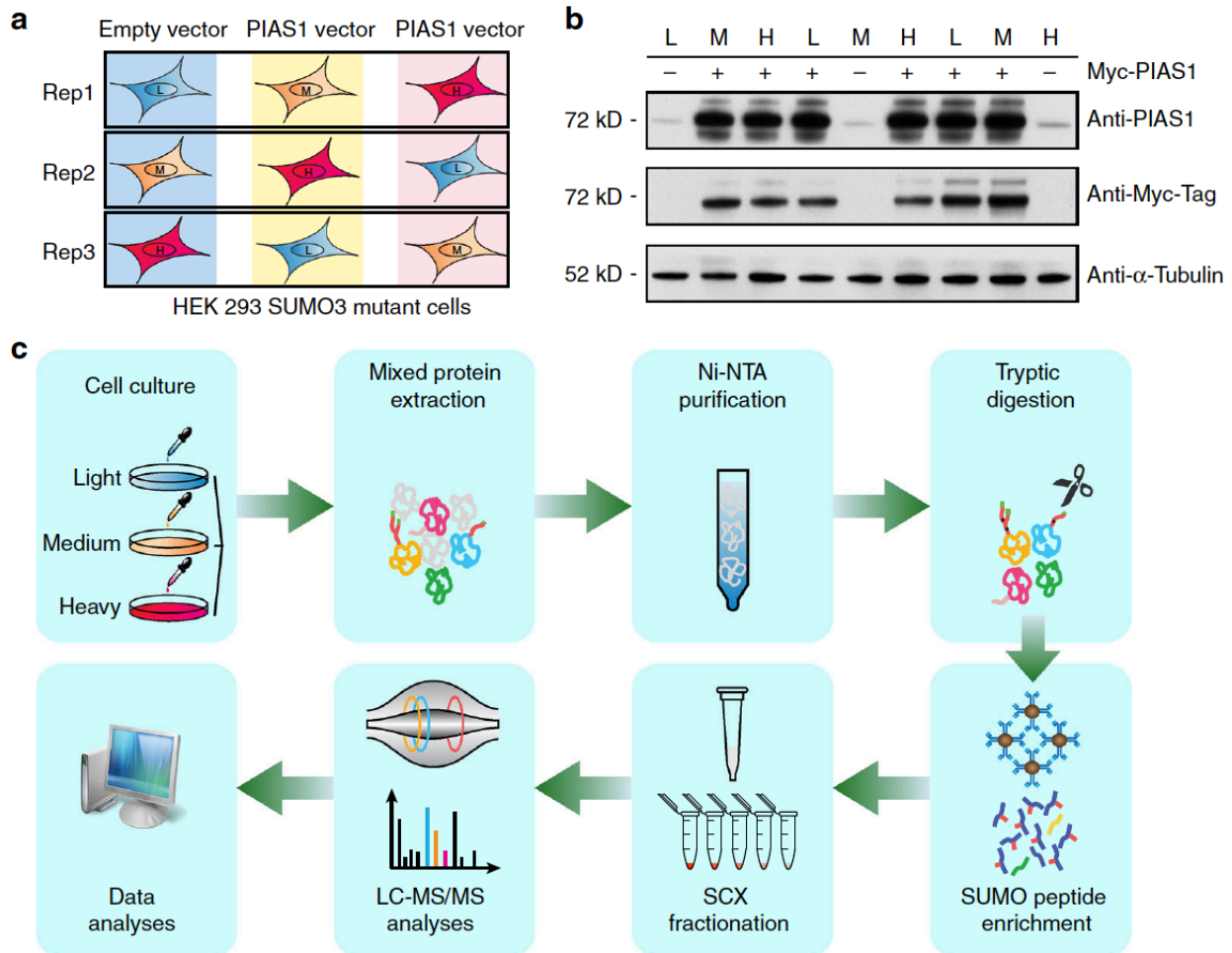

Figure 3-2 Workflow for the identification of PIAS1 substrates.

(a) HEK293 SUMO3m cells were cultured in SILAC medium with reverse labeling in biological triplicates. For replicate 1, PIAS1 overexpression was performed in medium and heavy channels with Myc-PIAS1 vector, while the cells cultured in light media were transfected with the pcDNA3.0 (Empty vector). For replicates 2 and 3 the empty vector was transduced in the medium and heavy labelled cells, respectively. (b) Western blot showing the level of overexpression of PIAS1 in the transfected cells. Detection of PIAS1 overexpression by both Anti-PIAS1 antibody and Anti-Myc antibody. β-Tubulin is used as loading control. (c) SILAC labelled cells were lysed and combined in a 1:1:1 ratio based on protein content. SUMOylated proteins were enriched from the cell extract on an IMAC column prior to their tryptic digestion. After desalting and drying, peptides containing the SUMO3 remnant were enriched using a custom anti-K-ε-NQTGG antibody that was crosslinked on magnetic

beads. Enriched peptides were further fractionated on SCX columns and injected on a Tribrid Fusion mass spectrometer. Peptide identification and quantification were performed using MaxQuant. Source data are provided as a Source Data file.

Protein classification ontology analysis of the PIAS1 substrates using PANTHER clustered the targets into 11 groups (Figure 3-3d). PIAS1 mediated SUMOylation predominantly occurred on nucleic acid binding proteins, transcription factors, cytoskeletal proteins, chaperone proteins, enzyme modulators and ligases. We next classified putative PIAS1 substrates by their gene ontology (GO) molecular function, biological process and cellular components (Figure 3-3e) using the whole identified SUMOylome as background. GO cellular component classification revealed that PIAS1 substrates were enriched in PML body, plasma membrane and microtubule compared to the global SUMOylome (Figure 3-3e). GO biological process analysis revealed that identified PIAS1 substrates are involved in a variety of biological processes, including protein stabilization, protein sumoylation and protein folding (Figure 3-3e). GO molecular function analysis indicated that PIAS1 substrates are associated with ubiquitin protein ligase binding, structural molecule activity, protein tag, SUMO transferase activity and unfolded protein binding (Figure 3-3e). Indeed, PIAS1 regulates the SUMOylation of several proteins whose roles in the cell are diverse. Much like global SUMOylation, PIAS1 mediated SUMOylation may play a role in several biological processes that are independent from each other.
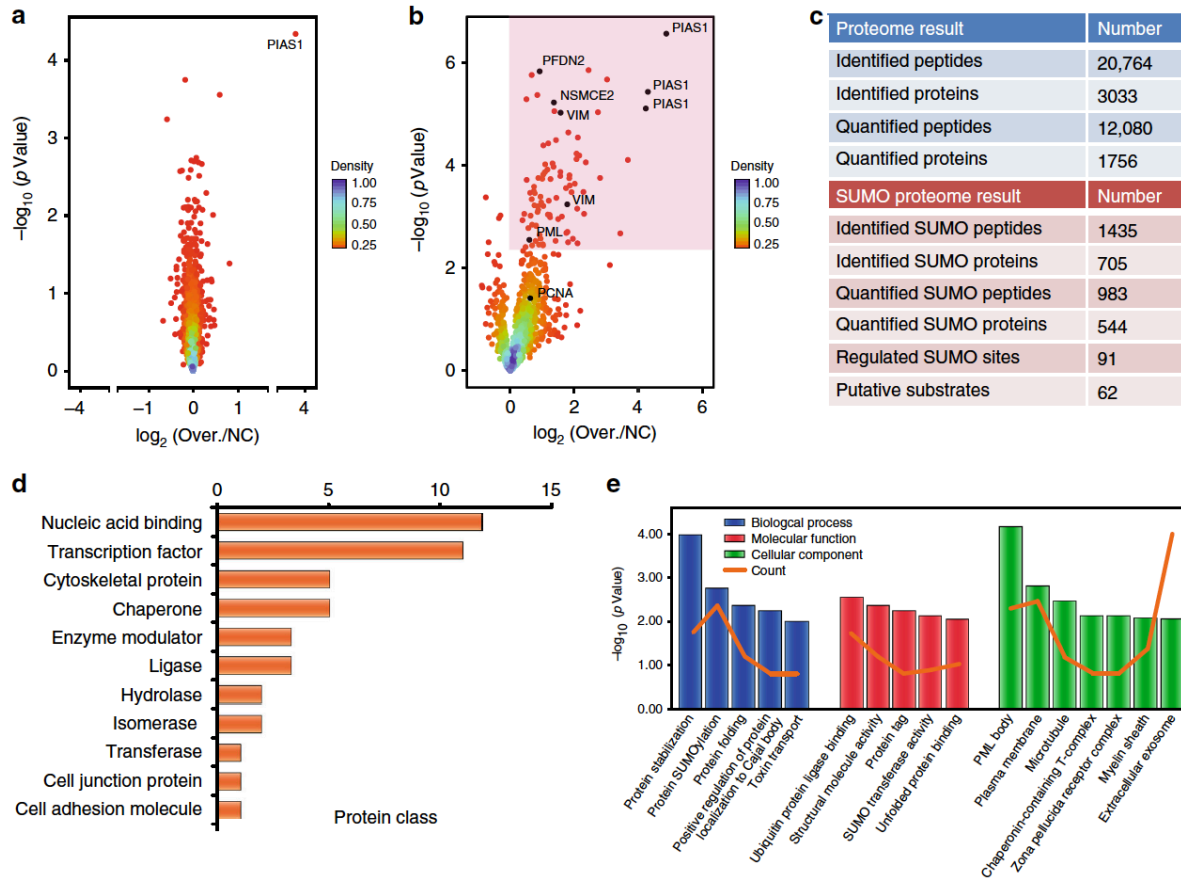
Figure 3-3 Mass spectrometry results and bioinformatic analyses of identified PIAS1 substrates.

(a) Volcano plots showing the global proteome changes in cells overexpressing PIAS1 (Over.) vs. control cells (NC). Individual proteins are represented by points. The area of the volcano plot where protein abundance changes are significantly regulated (Benjamini Hochberg corrected p-value of <0.05) are shaded in pink. (b) Volcano plots showing the global SUMOylation changes in cells overexpressing PIAS1 (Over.) vs. control cells (NC). Individual SUMOylation sites are represented by points. The area of the volcano plot where SUMO sites are significantly up-regulated (Benjamini Hochberg corrected p-value of <0.05) is shaded in pink. (c) Summary of identified and quantified peptides and proteins in both the proteome and SUMOylome experiments. (d) Functional classification of PIAS1 substrates using PANTHER (Protein Analysis Through Evolutionary Relationships) (http://www.pantherdb.org). (e) GO term enrichment distribution of the identified PIAS1 substrates using DAVID 6.8 (https://david.ncifcrf.gov/). Source data are provided as a Source Data file.

Previous SUMO proteome analyses indicated that under unstressed conditions, approximately half of acceptor lysine residues are found in the SUMO consensus and reverse consensus motifs [41]. Since SUMOylation is believed to occur at UBC9 consensus site without an E3 SUMO ligase, we surmised that PIAS1 mediated SUMOylation may occur at non-consensus

motifs. We therefore compared the amino acid residues surrounding the SUMOylation sites that are regulated by PIAS1 to those of the whole SUMO proteome (Supplementary Figure 3-4a). As anticipated, the sequences surrounding the PIAS1 mediated SUMOylation sites are depleted in glutamic acid at position +2 and depleted of large hydrophobic amino acids at position -1, consistent with the reduction of the consensus sequence. Indeed, E3 SUMO ligases appear to aid in the SUMOylation of lysine residues that reside in non-canonical regions. Furthermore, we investigated the local secondary structures and solvent accessibility of PIAS1 substrates surrounding SUMO sites using NetSurfP-1.1 software (Supplementary Figure 3-4b). We observed that one-third of PIAS1 regulated SUMOylation sites are located within α-helix and approximately one-tenth within β-strand. In contrast, the majority of SUMOylated lysine residues in the SUMO proteome are localized in coil regions. Taken together, these results support the notion that PIAS1 mediated SUMOylation preferentially occurs on structured regions of the protein, which may help substrate recognition by PIAS1. Additionally, we noted that PIAS1 mediated SUMOylation occurred primarily on solvent exposed lysine residues, which was also the case for the global SUMOylome. These results suggest that PIAS1 may not impart conformational changes to its substrate upon binding since it does not promote SUMOylation on lysine residues that would otherwise be buried within the core of the substrate. Overall, PIAS1 promotes the ability of UBC9 to SUMOylate lysine residues that are present in non-consensus sequences located on ordered structures of the proteins.

To better understand the cellular processes regulated by PIAS1, a STRING analysis was performed to analyze the interaction network of putative PIAS1 substrates. This network highlights the presence of highly connected interactors from PML nuclear body, transcriptional factors, cytoskeletal proteins and RNA binding proteins (Figure 3-4). PIAS1 was previously shown to colocalize to PML nuclear body and to regulate oncogenic signaling through SUMOylation of PML and its gene translocation product PML-RARα associated with acute promyelocytic leukemia (APL) [27].
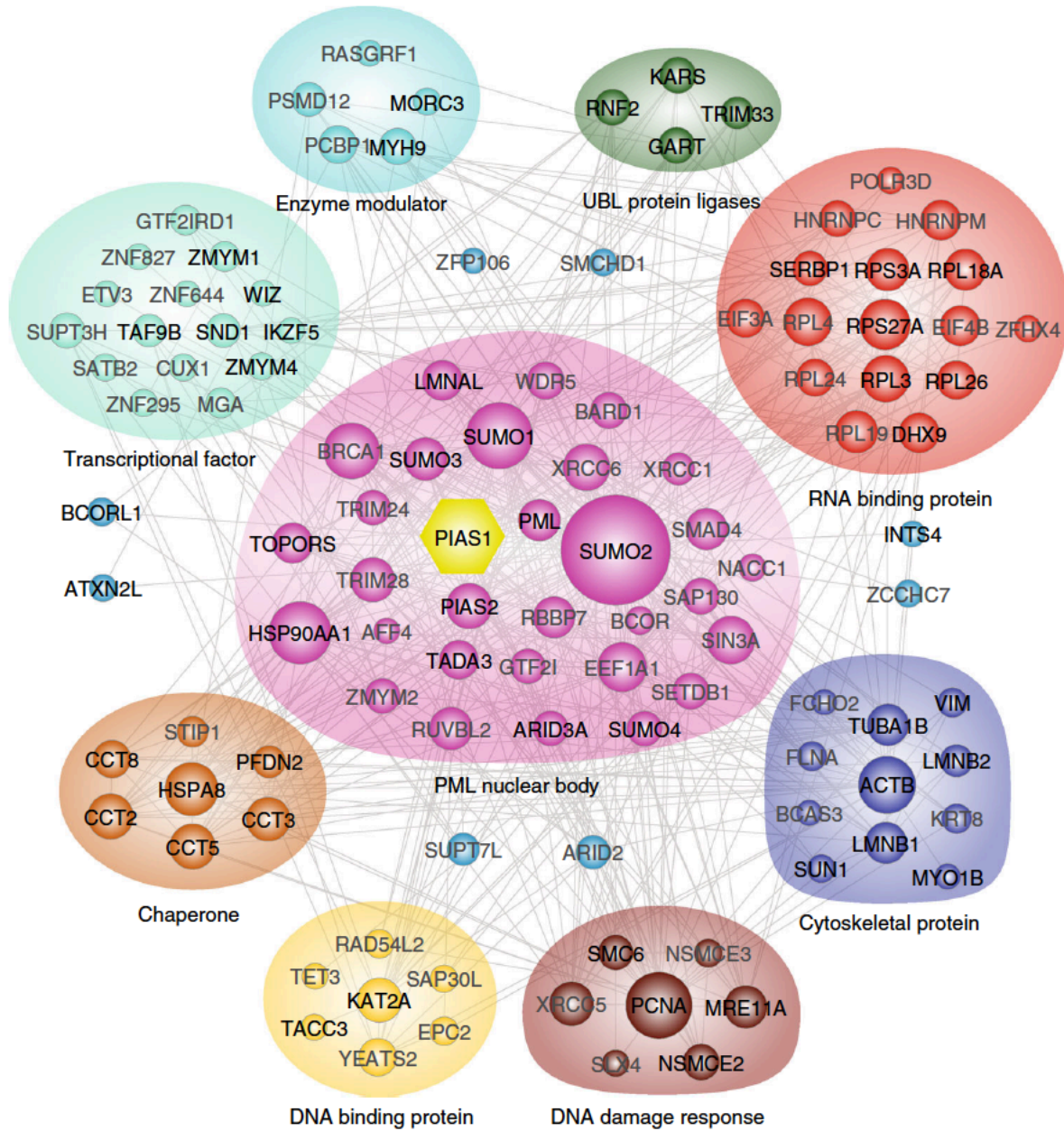
Figure 3-4 Protein-Protein Interaction Network of PIAS1 substrates.

STRING network of PIAS1 substrates and their interacting partners. Proteins are grouped according to their GO terms.

Interestingly, we also found that several putative PIAS1 substrates were associated with cytoskeletal organization, including β-actin (ACTB), α-tubulin (TUBA1B) and vimentin (VIM), in addition to several other intermediate filament proteins. The actin filaments, intermediate filaments, and microtubules that form the cytoskeleton of eukaryotic cells are responsible for cell

division and motility. They also help establish cell polarity, which is required for cellular homeostasis and survival [42]. Moreover, one SUMO site on β-actin (ACTB) and two of the five SUMO sites on α-tubulin (TUBA1B) that were found at non-consensus motif regions were regulated by PIAS1, suggesting an important role for PIAS1 in substrate protein dynamics during SUMOylation. Additionally, we evaluated the degree of evolutionary conservation of these modified lysine residues. Surprisingly, all SUMOylated lysine residues analyzed are highly conserved across different species (Supplementary Figures 3-5 and 3-6). In our data, we found intermediate filaments (IFs) to be major targets of PIAS1 among cytoskeletal proteins. Our results highlight that PIAS1 mediates the SUMOylation of the type III IF VIM protein at Lys-439 and Lys-445, both of which are located on the tail domain. Moreover, PIAS1 promotes the SUMOylation of several type V IF proteins (e.g. Prelamin A/C, Lamins B1 and B2) on their Rod domains (Supplementary Figure 3-7).

3.4.3 Validation of PIAS1 Substrates by in vitro SUMOylation Assays

Next, we selected E3 SUMO-protein ligase NSE2 (NSMCE2) and prefoldin subunit 2 (PFDN2), which were identified in SUMO proteomic experiments as putative PIAS1 substrates for further validation. We performed *in vitro* SUMOylation assays to confirm that these sites were regulated by PIAS1. For the reconstituted *in vitro* SUMO assay, we incubated individual SUMO substrates with SUMO-activating E1 enzyme (SAE1/SAE2), UBC9, SUMO-3 with or without PIAS1 in the presence of ATP. We also used PCNA, a known PIAS1 substrate, as a positive control. After 4h incubation at 37 °C, the western blots of each substrate showed either single or multiple bands of higher molecular weight confirming the SUMOylated products. Separate LC-MS/MS experiments performed on the tryptic digests of the *in vitro* reactions confirmed the SUMOylation of NSMCE2 at residues Lys-90, Lys-107, and Lys-125, and PFDN2 at residues Lys-94, Lys-111, Lys-132, and Lys-136. While UBC9 alone can SUMOylate these substrates, we noted an increasing abundance of SUMOylated proteins when PIAS1 was present, confirming that the E3 SUMO ligase enhanced the efficiency of the conjugation reaction (Supplementary Figure 3-8a). Interestingly, several SUMOylation sites that were regulated by PIAS1 on both NSMCE2 and

PFDN2 were not located within SUMO consensus motifs, further supporting the motif analysis of the large-scale proteomic data (Supplementary Figure 3-4a).

Furthermore, we examined whether PIAS1 contributes to substrate SUMOylation *in vitro* using a cell-based assay. HEK293-SUMO3m cells were co-transfected with Flag-NSMCE2 or PFDN2 and Myc-PIAS1. Co-transfected cells were subjected to immunoprecipitation with anti-Flag agarose gel, followed by western blot with an anti-His antibody. The SUMOylation of substrates was minimally detected when only transfecting Flag-substrates. In contrast, overexpression of PIAS1 under the same experimental conditions led to a marked increase in the SUMOylation of these substrates (Supplementary Figure 8b). These results further confirm our quantitative SUMO proteomics data and validate the proteins NSMCE2 and PFDN2 as *bona fide* PIAS1 substrates.

3.4.4 PIAS1 SUMOylation Promotes its Recruitment to PML Nuclear Body

Interestingly, our large-scale SUMO proteomic analysis identified five SUMOylation sites on PIAS1 (Figure 5a). Two of these sites (Lys-46, Lys-56) are located in the SAP domain, and may regulate the interaction of PIAS1 with DNA [16]. We also identified two SUMOylated residues (Lys-137 and Lys-238) located within the PINIT domain of PIAS1, potentially affecting its subcellular localization [15]. The last SUMOylated site (Lys-315) of PIAS1 is located next to an SP-RING domain, which may alter the ligation activity of PIAS1 [43]. Of note, PIAS1 contains a SIM, and previous reports indicated that this ligase can localize to PML nuclear bodies in a SIM-dependent manner with SUMOylated PML [44]. As PML also contains a SIM motif, we were interested in three out of the five SUMO sites on PIAS1: Lys-137, Lys-238 and Lys-315. Since PIAS1 is SUMOylated at several sites and PML contains a SIM, we surmised that reciprocal interactions could be mediated through SUMO-SIM binding. Accordingly, we constructed a PIAS1-GFP vector and used site-directed mutagenesis to specifically mutate the PIAS1 lysine residues that are SUMO-modified and are located within regions of PIAS1 that could interact with PML. As the SAP domain of PIAS1 is exclusively reserved for DNA binding, we excluded the SUMO-modified lysine residues in this domain when creating the mutant construct as it may affect its localization in a

PML independent fashion. We therefore created the variant constructs by mutating the codons for Lys-K137, Lys-238 and Lys-315 to arginine codons. Several mutant genes were created, including single mutants of each site and the triple mutant (PIAS1-GFP 3xKR). These mutant vectors were transfected into HEK293 SUMO3m cells and used to study the effects of SUMOylation at the various lysine residues on the PIAS1-PML colocalization. As evidenced by the immunofluorescence studies, approximately 49% of PIAS1-GFP-WT colocalized with PML (Figure 5b). Of all the single variants tested, significant changes in the colocalization of PIAS1 and PML were observed with the K238R and K315R alteration (Figure 5c). However, we noted a greater than 50% reduction in PIAS1-GFP–PML colocalization when all three sites were mutated, suggesting a possible cooperativity among these sites (Figures 5b and 5c). This functional redundancy may be required to ensure the proper localization of PIAS1 to PML nuclear bodies under different biological context. Also, the cooperative nature of multiple SUMOylation events to enhance affinity has been noted before, where the affinity of RNF4 for SUMO dimers is 10-fold higher than for the monomer [45].
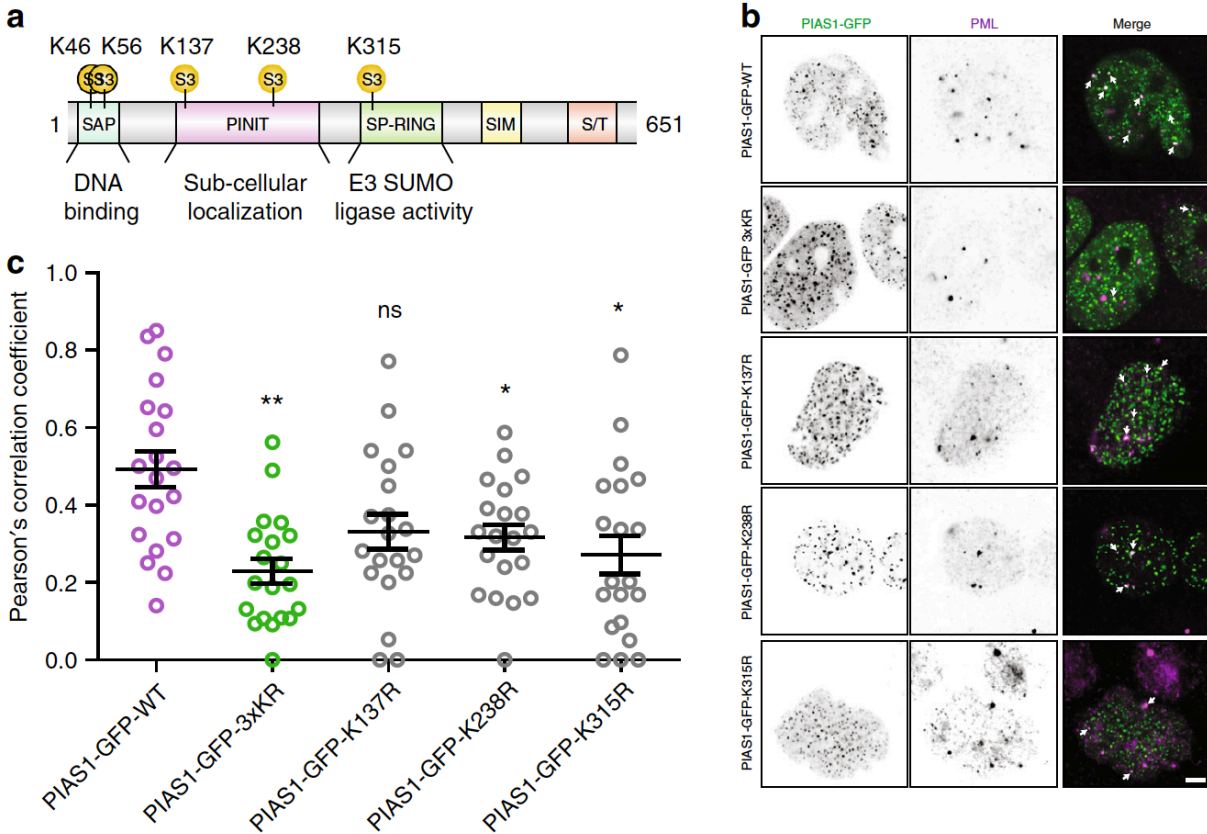
Figure 3-5 SUMOylation of PIAS1 promotes its PML localization.

(a) Distribution of SUMO sites identified on PIAS1. Three SUMO sites (K137, K237 and K315) were identified in the dataset. K137 and K238 are located in PINIT domain while K315 is located in SP-RING domain. (b) HEK293 SUMOm cells were co-transfected with PIAS1-GFP-WT, PIAS1-GFP-3xKR, PIAS1-GFP-K137R, PIAS1-GFP-K238R or PIAS1-GFP-K315R and immunofluorescence was performed with anti-PML. Scale bar: 2.5 μm. (c) Scatter graph showing the Pearson's correlation coefficient for the PIAS1-PML co-localization. Data represent the mean ± S.D., error bars represent S.D., ns, non-significant, *p<0.05, **p<0.01, Student's t-test, n=20 biologically independent cells/condition. Source data are provided as a Source Data file.

The fact that the co-localization of PML and PIAS1-GFP was not totally abrogated for the triple mutant might be explained by residual interactions between PML and the PIAS1-GFP [44]. Indeed, we have shown that several sites on PIAS1 are SUMOylated, which aids to localize PIAS1 to the PML bodies. However, PML itself is also heavily SUMOylated on several lysine residues. Therefore non-SUMOylated PIAS1 can still localize, albeit less efficiently, to PML nuclear bodies via the SIM that is located on PIAS1 and the SUMOylated moieties on PML. A similar phenomenon was reported by our group for the SUMO E2 protein UBC9 [46]. Taken together, these experiments confirmed that colocalization of PIAS1 at PML nuclear bodies is partly mediated by the SUMOylation of PIAS1 at Lys-137, Lys-238 and Lys-315 residues.

3.4.5 VIM SUMOylation Promotes Cell Migration and Motility

VIM is predominantly found in various mesenchymal origins and epithelial cell lines [47-49]. Increasing evidence shows that VIM plays key roles in cell proliferation [50], migration [51] and contractility [52]. Our data shows that two SUMO sites on VIM are regulated by PIAS1, both of which are located on the tail domain and are highly conserved across different species (Figure 6a). To confirm that VIM is selectively SUMOylated by PIAS1, we used CRISPR/Cas9 gene editing technology with sgRNA specific to each of the four PIAS E3 ligases (e.g. PIAS1, PIAS2, PIAS3 and PIAS4) and a scrambled sgRNA for the transfection in HEK293 SUMO3m cells. KO and control HEK293 SUMO3 cells were cultured in triplicate using light, medium, and heavy SILAC media (Supplementary Figure 3-9). Following NiNTA enrichment, SUMOylated proteins were digested on beads with trypsin and modified tryptic peptides were isolated by SUMO remnant

immunoaffinity purification prior to targeted LC-MS/MS analyses using an inclusion list to detect and identify individual isotopically labeled SUMOylated peptides of VIM (e.g. ETNLDSLPLVDTHSK*R and TLLIK*TVETR where * indicates SUMOylation site). MS/MS spectra of isotopically-labeled tryptic peptides were used to confirm identification of SUMOylated VIM peptides (Supplementary Figure 3-10). We also analyzed by LC-MS/MS in data-dependent acquisition the tryptic peptides from the flow through proteins to normalize protein abundance across the 6 different samples. These quantitative proteomics experiments revealed that PIAS1 selectively targeted the SUMOylation of VIM at K439 and K445.

To further investigate the function of PIAS1 mediated SUMOylation of VIM, we expressed a Flag-tagged VIM K439/445R double mutant (VIM^{mt}) that is virtually refractory to SUMOylation in HeLa cells, and compared the functional effects to cells expressing the wild-type Flag-tagged VIM (VIM^{wt}). We transfected Flag-VIM^{wt} and Flag-VIM^{mt} into HeLa cells and used the empty Flag vector as a negative control. At 48 h post-transfection, the cells were harvested, lysed in 8 M urea and protein pellets were separated on SDS-PAGE. The ensuing western blot results show that protein abundance between VIM^{wt} and VIM^{mt} are similar; yet, the SUMO level on VIM^{mt} is undetectable (Figure 3-6b). Moreover, transfecting PIAS1 considerably increased the level of SUMOylation of VIM^{wt}, supporting our proteomics experiments. Of note, transfecting PIAS1 along with VIM^{mt} promoted the SUMOylation of VIM^{mt}, albeit to a much lower degree than VIM^{wt}, improving the solubility VIM^{mt}.

We examined the effect of VIM^{wt} and VIM^{mt} expression on cell migration using the wound-healing assay. VIM^{wt} significantly promotes cell migration (Figure 3-6c), which is in line with the results obtained in HepG2 cells [53]. However, VIM^{mt} alone conferred no effect on cell migration, while transfecting PIAS1 with VIM^{mt} rescued the phenotypic effect (Figure. 3-6c-d). These results indicate that SUMOylation of VIM plays a role in cell growth and migration, presumably by regulating VIM IF function and/or formation.

To further investigate the function of PIAS1 mediated SUMOylation of VIM, we expressed a Flag-tagged VIM K439/445R double mutant (VIM^{mt}) that is virtually refractory to SUMOylation in HeLa cells, and compared the functional effects to cells expressing the wild-type Flag-tagged VIM

(VIM^{wt}). We transfected Flag-VIM^{wt} and Flag-VIM^{mt} into HeLa cells and used the empty Flag vector as a negative control. At 48 h post-transfection, the cells were harvested, lysed in 8 M urea and protein pellets were separated on SDS-PAGE. The ensuing western blot results show that protein abundance between VIM^{wt} and VIM^{mt} are similar; yet, the SUMO level on VIM^{mt} is undetectable (Figure 6b). Moreover, transfecting PIAS1 considerably increased the level of SUMOylation of VIM^{wt}, supporting our proteomics experiments. Of note, transfecting PIAS1 along with VIM^{mt} promoted the SUMOylation of VIM^{mt}, albeit to a much lower degree than VIM^{wt}, improving the solubility VIM^{mt}.

We examined the effect of VIM^{wt} and VIM^{mt} expression on cell migration using the wound-healing assay. VIM^{wt} significantly promotes cell migration (Figure 3-6c), which is in line with the results obtained in HepG2 cells [53]. However, VIM^{mt} alone conferred no effect on cell migration, while transfecting PIAS1 with VIM^{mt} rescued the phenotypic effect (Figure. 3-6c-d). These results indicate that SUMOylation of VIM plays a role in cell growth and migration, presumably by regulating VIM IF function and/or formation.
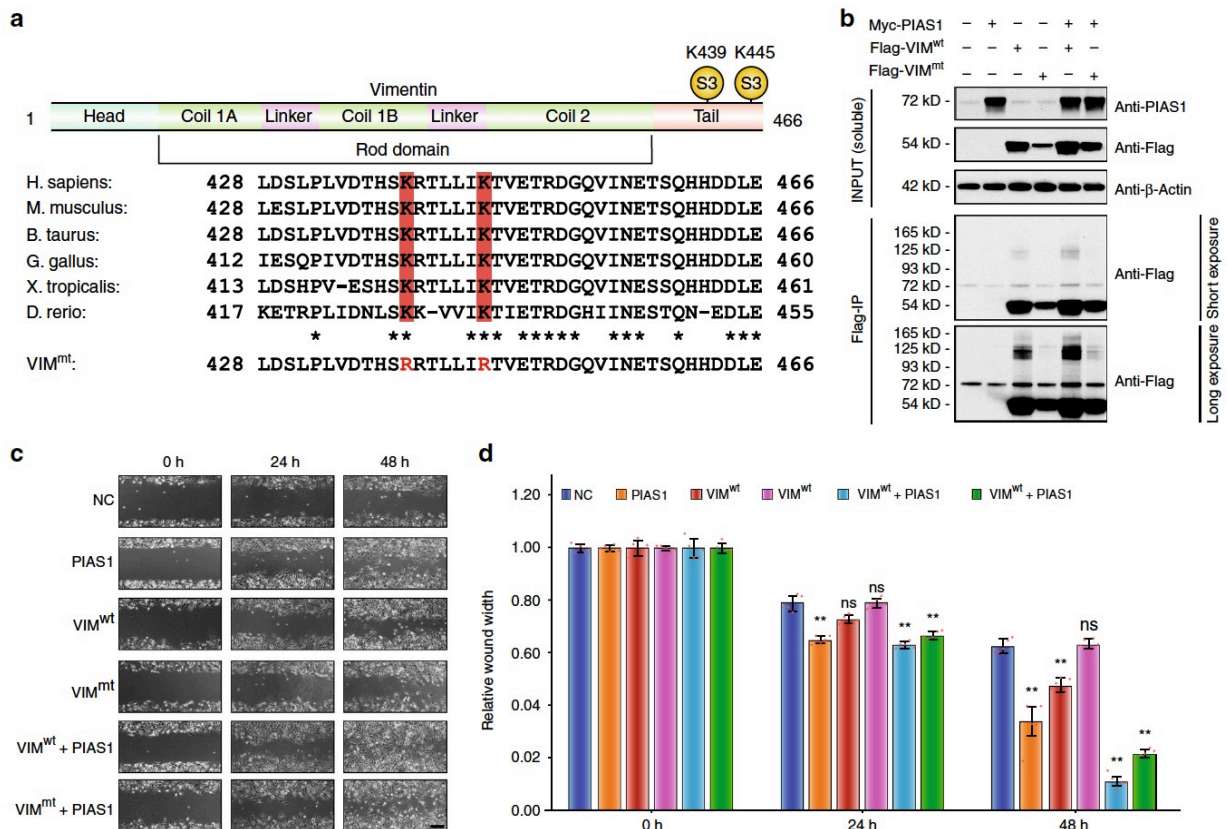
Figure 3-6 SUMOylation of Vimentin regulates its dynamic assembly.

(a) HeLa cells were transfected with an empty vector as a negative control, Flag-VIM^{wt} and Flag-VIM^{mt} and separated into RIPA soluble and insoluble fractions. VIM protein levels were examined by western blot. (b) MCF-7 cells were transfected with Emerald-wild-type vimentin (VIM^{wt}), and Emerald-Vimentin K439, 445R, double mutant (VIM^{mt}), and the proportion of unit-length filament (ULF), vimentin intermediate filament (VIF) and cytosolic vimentin between VIM^{wt} and VIM^{mt} was calculated under microscope. Data represent the mean ± S.D., error bars represent S.D., ns, non-significant, *p<0.05, **p<0.01, Student's t-test, n=4 biologically independent samples. (c) Fluorescence Recovery After Photobleaching (FRAP) assay of Emerald-wild-type vimentin (VIM^{wt}), and Emerald-Vimentin K439, 445R, double mutant (VIM^{mt}) in MCF-7 cells. Line plot shows average fluorescence at each time point ± S.D. Differences between values for VIM^{wt} and VIM^{mt} were not statistically significant at all time points by Student's t-test. n=7 biologically independent cells. (d) Model of the VIM dynamic assembly and disassembly. Vimentin is maintained in equilibrium between Unit-length Filament (ULF) and soluble tetramers. (Subunit exchange step). Vimentin filaments elongate by end-to-end annealing of ULF to form mature vimentin intermediate filament (VIF; annealing step). The phosphorylation-dependent shortening of VIF (severing step) involves phosphorylation on Ser 39 and Ser 56 of vimentin by several kinases, including Akt1. The short filaments can reanneal with another ULF to form new VIF (re-annealing step). However, the phosphorylated ULF are not amenable to the re-annealing process. These phosphorylated ULF products are subject to PIAS1-mediated SUMOylation, stimulating the dephosphorylation of the phosphorylated ULF, and subsequently reenter to either subunit exchange process or VIF maturation process. Source data are provided as a Source Data file.

Next, we investigated the function of VIM SUMOylation on the dynamics assembly of intermediate filaments (IFs). Both Keratin and lamin A IFs formation and solubility have been reported to be regulated by SUMOylation, while such properties have yet to be uncovered for VIM IFs [54]. For example, the SUMOylation of lamin A at Lys-201, which is found in the highly conserved rod domain of the protein, results in its proper nuclear localization [55]. Unlike lamin A SUMOylation, keratin SUMOylation is not detected under basal conditions. However, stress-induced keratin SUMOylation has been observed in mouse and human in chronic liver injuries. Additionally, keratin monoSUMOylation is believed to increase its solubility, while hyperSUMOylation promotes its precipitation [56].

We surmised that SUMOylation on VIM would alter its solubility, akin to the properties observed for keratin. Accordingly, we transfected an empty vector as a negative control, VIM^{wt}

and VIM^{mt} into HeLa cells and lysed the cells with a radioimmunoprecipitation assay (RIPA) buffer. Samples were fractionated into RIPA soluble and insoluble fractions. Western blot analysis of these samples shows that VIM^{wt} is preferentially located in the RIPA soluble fraction, while VIM^{mt} resides more in the insoluble fraction (Figure 3-7a). Clearly, SUMOylation of VIM drastically increases its solubility.

VIM filaments are also phosphorylated at multiple sites by several kinases [57], and this modification is associated with their dynamic assembly and disassembly [58-60]. In particular, phosphorylation sites located at the N-terminal head domain can impede the interaction between VIM dimer, thus preventing the formation of VIM tetramer necessary for the further assembly into filaments. Activation of Akt in soft-tissue sarcoma cells promotes the interaction of the head region of VIM and the tail region of Akt, resulting in the phosphorylation of VIM at Ser-39, further enhancing cell motility and invasion [55].

Next, we sought to examine if VIM SUMOylation alters its phosphorylation status, which could lead to changes in VIM IF dynamic assembly/disassembly. Accordingly, we separated protein extracts from VIM^{wt} and VIM^{mt} by SDS-PAGE, excised bands that corresponded to the soluble and insoluble VIM, and performed in-gel trypsin digestion followed by LC-MS/MS (Supplementary Figure 3-11). We identified several phosphorylated serine residues and one phosphorylated threonine residue on VIM (Supplementary Table 3-3), which have also been reported in the literature (S5, S7, T20, S22, S26, S29, S39, S42, S51, S56, S66, S72, S73, S83, S226, T258 and S459). All sites except those located on the C-terminal were found to be hyperphosphorylated in the insoluble VIM^{mt} compared to the wild-type counterpart. Interestingly, we observed an increase phosphorylation of S39 in the insoluble pellet of VIM^{mt}, a site known to be phosphorylated by Akt [55]. The observation that VIM is hyperphosphorylated at its N-terminus in VIM^{mt} suggests a possible cross-talk between SUMOylation and phosphorylation.

To further understand how SUMOylation affects VIM dynamics *in vivo*, we transfected Emerald-VIM^{wt} or Emerald-VIM^{mt} vectors in VIM null MCF-7 cells. VIM null cells were employed to eliminate the contribution of endogenous VIM on the VIM dynamics, which could mask the

phenotypic effects of our mutant. Using fluorescence microscopy we quantified the proportion of the various VIM structures for the two different Emerald tagged constructs. We found three major forms of VIM in the cells, which in accordance with the literature, were categorized as cytosolic, unit-length filament (ULF) and VIM intermediate filaments (VIFs) (Supplementary Figure 3-12). Statistical analysis of the proportion of the VIM structures revealed that the SUMO conjugation deficient VIM (VIM^{mt}) promoted the formation of ULF with a concomitant reduction in VIF formation compared to its wild-type counterpart (Figure 3-7b). Taken together these results indicate that SUMOylation of VIM promotes the formation of VIFs from the ULF building blocks.

Although VIM filament growth primarily relies on elongation by the longitudinal annealing of ULFs via end-to-end fusion, recent studies suggest that the subunit exchange of tetramers within these filaments does occur [61]. To determine if the SUMOylation of VIM affects the subunit exchange rate of these filaments we performed fluorescence recovery after photobleaching (FRAP) assays (Supplementary Figure 3-13). We monitored the recovery time of filament fluorescence up to 300 s after bleaching for both Emerald-VIM^{wt} and Emerald-VIM^{mt}, and noted no statistical difference in recovery between constructs (Figure 3-7c). This observation suggests that the SUMOylation of VIM is not involved in the subunit exchange of tetramers within filaments.
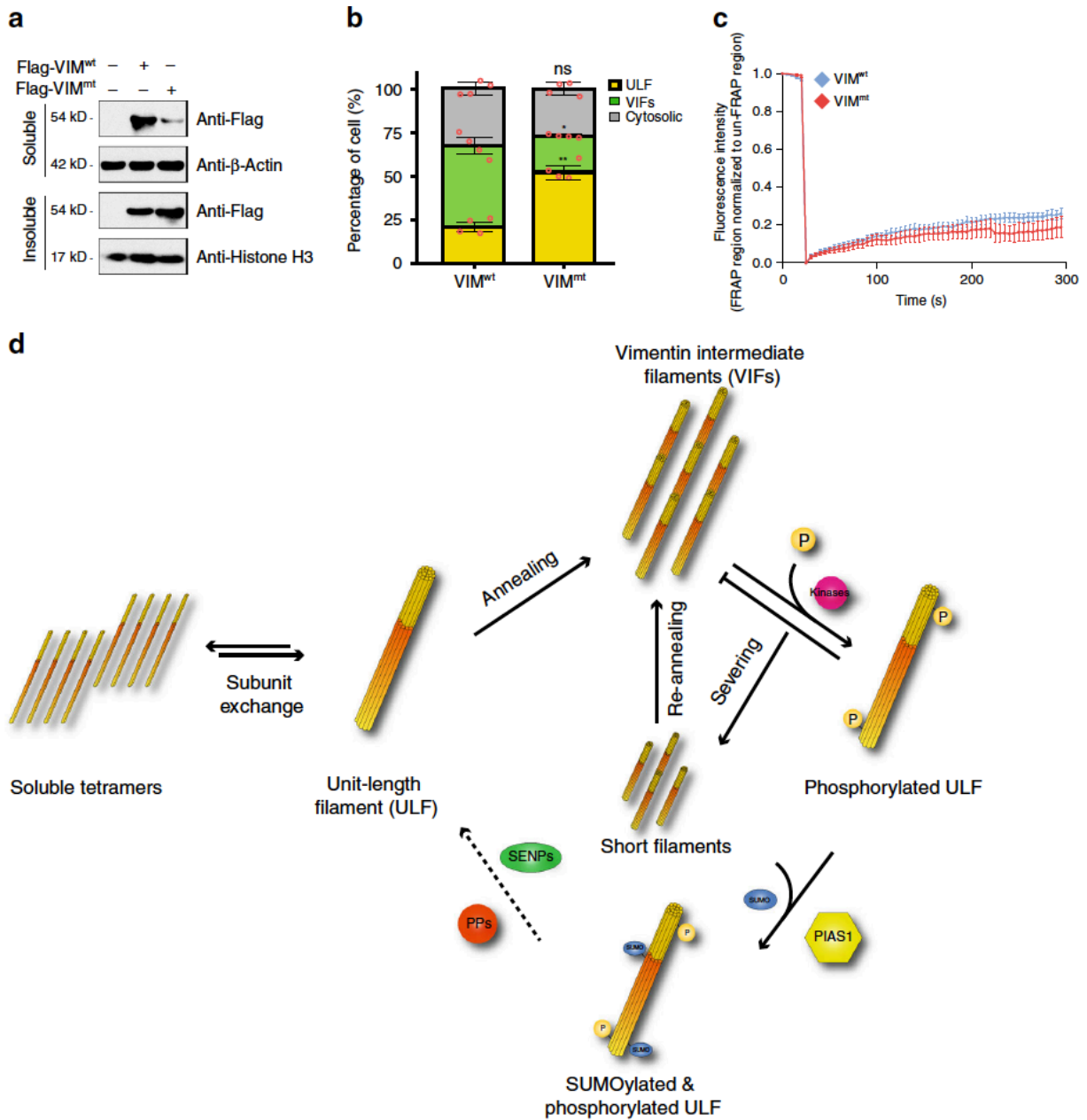
Figure 3-7 SUMOylation of Vimentin regulates its dynamic assembly.

(a) HeLa cells were transfected with an empty vector as a negative control, Flag-VIM^{wt} and Flag-VIM^{mt} and separated into RIPA soluble and insoluble fractions. VIM protein levels were examined by western blot. (b) MCF-7 cells were transfected with Emerald-wild-type vimentin (VIM^{wt}), and Emerald-Vimentin K439, 445R, double mutant (VIM^{mt}), and the proportion of unit-length filament (ULF), vimentin intermediate filament (VIF) and cytosolic vimentin between VIM^{wt} and VIM^{mt} was calculated under microscope. Data represent the mean ± S.D., error bars represent S.D., ns, non-significant, *$p<0.05$, **$p<0.01$, Student's t-test, n=4 biologically independent samples. (c) Fluorescence Recovery After Photobleaching (FRAP) assay of Emerald-wild-type

vimentin (VIM^{wt}), and Emerald-Vimentin K439, 445R, double mutant (VIM^{mt}) in MCF-7 cells. Line plot shows average fluorescence at each time point ± S.D. Differences between values for VIM^{wt} and VIM^{mt} were not statistically significant at all time points by Student's t-test. n=7 biologically independent cells. (d) Model of the VIM dynamic assembly and disassembly. Vimentin is maintained in equilibrium between Unit-length Filament (ULF) and soluble tetramers. (Subunit exchange step). Vimentin filaments elongate by end-to-end annealing of ULF to form mature vimentin intermediate filament (VIF; annealing step). The phosphorylation-dependent shortening of VIF (severing step) involves phosphorylation on Ser 39 and Ser 56 of vimentin by several kinases, including Akt1. The short filaments can reanneal with another ULF to form new VIF (re-annealing step). However, the phosphorylated ULF are not amenable to the re-annealing process. These phosphorylated ULF products are subject to PIAS1-mediated SUMOylation, stimulating the dephosphorylation of the phosphorylated ULF, and subsequently reenter to either subunit exchange process or VIF maturation process. Source data are provided as a Source Data file.

The model depicted in Figure 3-7d combines the results from the proteomic, immunofluorescence and FRAP assays and describes the molecular mechanism of PIAS1 mediated control of VIM dynamics. Under physiological conditions, vimentin is maintained in equilibrium between ULF and soluble tetramers. The formation of VIM ULF has been shown to occur spontaneously on the order of seconds in vitro showing that this arrangement is thermodynamically favorable and proceeds rapidly without the need for protein modifications [62]. VIM filaments elongate by end-to-end annealing of ULFs and eventually form mature VIM intermediate filaments (VIFs). This is followed by the breakdown of VIFs by severing, which involves phosphorylation on several residues found on the N-term of VIM [63, 64]. We show in this work that this hyper-phosphorylation occurs exclusively on the N-terminal of VIM in a SUMO-dependent mechanism. The truncated filaments can reanneal with another ULF to form larger VIFs. However, the phosphorylated ULF must be SUMOylated by PIAS1 to increase either the solubility or interaction with protein phosphatases, such as type-1 (PP1) and type-2A (PP2A) protein phosphatases, as shown by the large increase in VIM phosphorylation levels with the SUMO deficient vimentin construct [64]. These results suggest that the PIAS1-mediated SUMOylation of VIM stimulates the dephosphorylation of ULF and facilitate the reentry of the ULF into the VIF maturation process by annealing on growing VIFs. This dynamic assembly and

disassembly of VIFs thus involve the SUMOylation of VIM, a modification that also regulates the cell migration and motility (Figure 3-6).

3.5 Discussion

We report the functional effect of E3 SUMO ligase PIAS1 in HeLa cells, and determined that PIAS1 not only promotes cell proliferation, but also stimulates cell migration. PIAS1 has been extensively studied in other cancer lines, such as Human Prostate Cancer, where PIAS1 expression is increased and enhanced proliferation through inhibition of p21 [22]. In addition, other studies have also reported that PIAS1 may function as a tumor suppressor to regulate gastric cancer cell metastasis by targeting the MAPK signaling pathway [65]. Interleukin 11 was previously shown to reduce the invasiveness of HTR-8/SVneo cells via reduced ERK1/2 activation, PIAS1/3-mediated activated STAT3 (Tyr-705) sequestration, and a decrease in PIAS1 expression, leading to reduced expression of Fos and several major metalloproteinases (MMP2, MMP3, MMP9 and MMP23B) [66]. However, these studies were limited to individual PIAS1 targets to understand the regulatory mechanism. These targeted approaches sufficed to answer specific questions about PIAS1 mediated SUMOylation, but lack the depth to fully elucidate the function of PIAS1. A systematic approach to establish the global properties of PIAS1 as an E3 SUMO ligase and how these SUMOylation events alter substrate function are missing and needed. Indeed, such a method was never conceived due to the complex nature of quantitative SUMO proteomics. Global SUMO proteome analyses are challenging due to the low abundance of protein SUMOylation and the extremely large remnant that is retained on the modified lysine residues upon tryptic digestion. Moreover, proteomic workflows that are currently available to study SUMOylation require two levels of enrichment, which adversely affects the reproducibility of SUMO site quantitation.

We used a straightforward method for the identification of PIAS1 substrates by expanding on our previously described SUMO proteomics strategy [28]. This method combines SILAC labelling for reproducible quantitative proteomic analyses, E3 SUMO ligase protein overexpression, followed by SUMO remnant immunoaffinity enrichment. This workflow allows for the selective profiling of substrates and regulated SUMOylation sites of any E3 SUMO ligase. All the PIAS1 substrates identified in this work were analyzed using forward and reverse SILAC labelling under basal condition, which further increases the confidence of the identified

substrates. Notably, we observed that PIAS1 overexpression has a global effect on protein SUMOylation (Figure 3-3b). This is in part due to some PIAS1 substrates being directly involved in protein SUMOylation, such as PML, PIAS2, NSMCE2 and TOPORS. In addition, many of the identified substrates were found to participate in protein ubiquitination regulation, such as TRIM33 and RNF2, which may also affect the global protein SUMOylation through the interplay between SUMOylation and ubiquitination [28, 67]. As for other closely related PIAS family members, the regulatory function of PIAS1 extend beyond the SP-RING type SUMO ligase and can mediate protein interactions through non covalent SUMO binding. Indeed, the SUMOylation of direct substrates of PIAS1 can promote the modifications of other targets of the same protein complexes. Also, the complex formation maybe facilitated via the SAP and PINIT domains of PIAS1 through the interactions with the chromatin or the sub-cellular localization of binding partners [10, 48]. Thus, several factors affect protein SUMOylation, and some of the SUMOylation sites that increased upon PIAS1 overexpression may not be direct substrates of PIAS1, or may be SUMOylated to a different extent when PIAS1 is expressed at endogenous levels.

We identified five SUMOylation sites on PIAS1 itself (Lys-46, Lys-56, Lys-137, Lys-238 and Lys-315), suggesting a possible feedback mechanism that could keep SUMOylation levels in check. Our immunofluorescence studies show that SUMOylation of PIAS1 promotes its localization to PML nuclear bodies (Figures 3-5b-c). Interestingly, a recent paper that studied the substrates of RNF4 identified PIAS1 as a substrate of this SUMO-targeted ubiquitin ligase [68]. Moreover, RNF4 is localized to PML nuclear bodies, where it ubiquitylates SUMOylated proteins for their subsequent proteasomal degradation. Elevated levels of cellular SUMOylation may lead to an increase SUMOylation of PIAS1, prompting its localization to PML nuclear bodies and its degradation by RNF4 in a ubiquitin-dependent manner. This feedback mechanism used to regulate global SUMOylation may not be reserved solely for PIAS1. Indeed, other members of the PIAS family, as well as NSE2 and TOPORS, have been found to be SUMOylated at several lysine residues, while also being substrates of RNF4 [68].
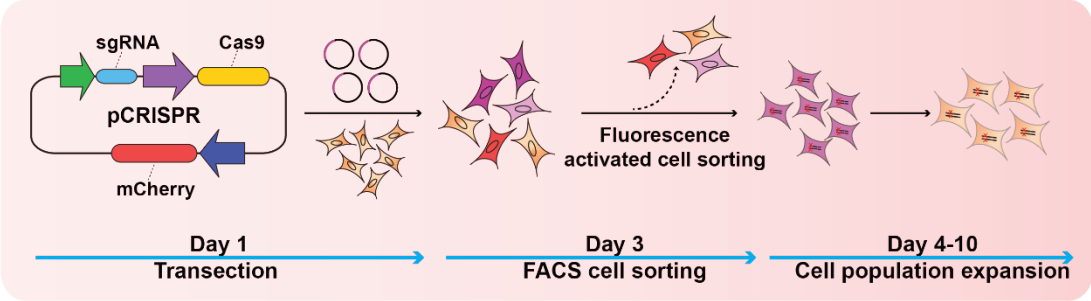
Notably, cytoskeletal proteins occupy a significant proportion of the identified PIAS1 substrates. Constituents of actin filaments, intermediate filaments and microtubules were all

183

found to be PIAS1 substrates. Interestingly, unlike UBC9 substrates that are typically SUMOylated on consensus motifs [7], the acceptor lysine residues found on these cytoskeletal proteins are highly conserved but are located in non-consensus sequence motif. These observations suggest that PIAS1 may act as an adaptor protein to change cytoskeletal protein turnover or dynamics by facilitating their SUMOylation. We uncovered that PIAS1 specifically SUMOylates K439 and K445 residues of VIM. This modification increased the solubility of VIM and is correlated with the uptake of ULF onto VIF in a phospho-dependent mechanism. VIM SUMOylation in turn favors cell proliferation and motility, which could lead to an increase in cancer cell aggressiveness. Although these findings could reveal the molecular mechanism of PIAS1 mediated VIM SUMOylation and its involvement in cancer cell aggressiveness, additional evidence is required to further understand the function of PIAS1-mediated SUMOylation on the other cytoskeletal proteins and how these cytoskeletal proteins collaborate during cell migration.

3.6 Acknoledgements

3.7 Supplementary Figures

Supplementary Figure 3-1 Overview of PIAS gene knockout by CRISPR/Cas9-based gene editing technology.

SUMO3 (human)   N-ter…RQIRFRFDGQPINETDTPAQLEMEDEDTIDVFQQQTGG^{92}

SUMO3 (mutant)   ..RNQTGG^{92}

HHHHHH

Supplementary Figure 3-2 Protein sequences of the endogenous SUMO3 and SUMO3m.
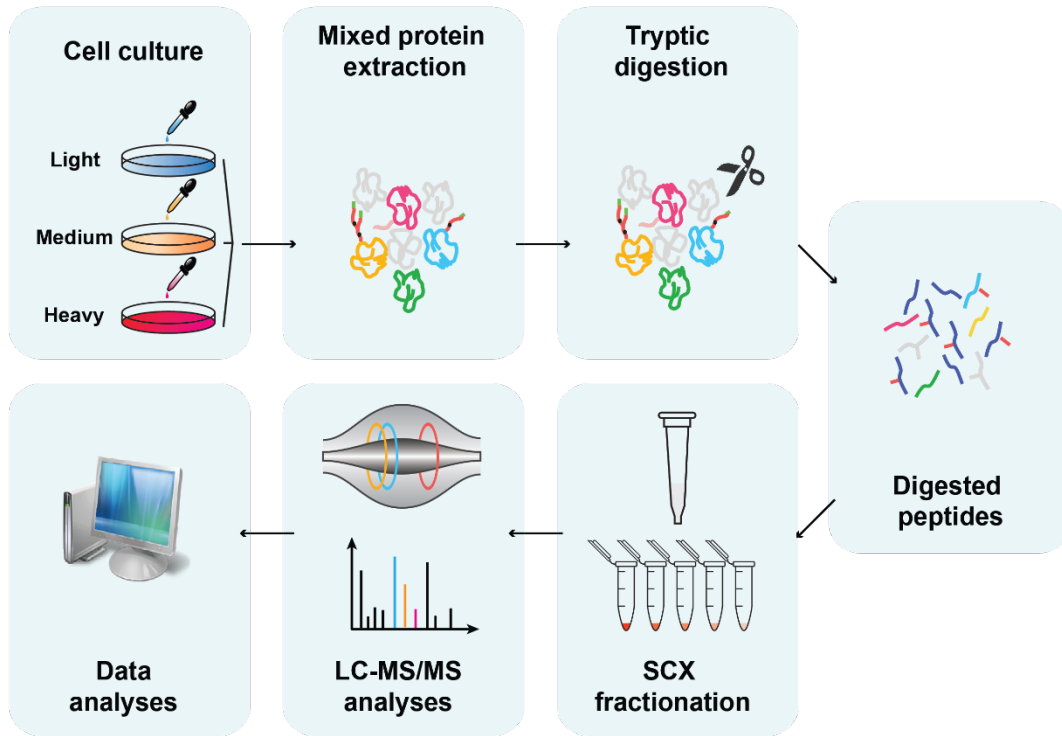
Supplementary Figure 3-3 Overview of proteome identification.
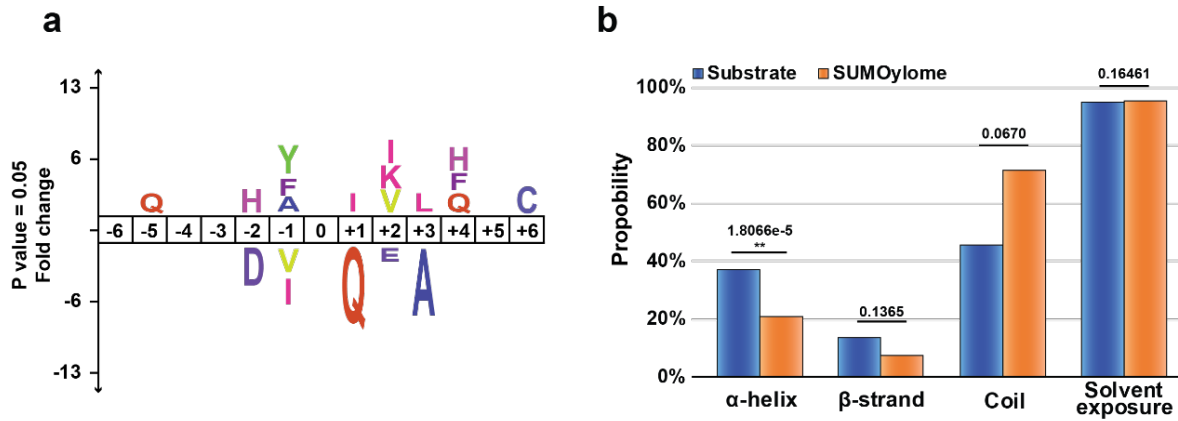
Supplementary Figure 3-4 Structural analysis of identified substrates.

(a) Icelogo of the amino acid sequence surrounding the PIAS regulated SUMO sites compared to the whole SUMO proteome. (b) Secondary structure predication of identified PIAS1 substrates vs identified SUMOylome, **p<0.01, Student's t-test.

Actin filaments　　　　　　**7-nm diameter**

Actin

```
H. sapiens:     100 PEEHPVLLTEAPLNPKANREKMTQIMFETFN 128
M. musculus:    100 PEEHPVLLTEAPLNPKANREKMTQIMFETFN 128
B. taurus:      100 PEEHPVLLTEAPLNPKANREKMTQIMFETFN 128
G. gallus:      100 PEEHPVLLTEAPLNPKANREKMTQIMFETFN 128
X. tropicalis:  100 PEEHPVLLTEAPLNPKANREKMTQIMFETFN 128
D. rerio:       100 PEEHPVLLTEAPLNPKANREKMTQIMFETFN 128
                    ******************************
```
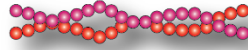
Supplementary Figure 3-5 Cartoon representation of the identified SUMOylation sites on Actin at Lys 115.

Protein sequence alignment of Actin across six different species showing that Lys 115 is highly conserved.

Microtubules   **25-nm diameter**

Tubulin

```
H. sapiens:    311 KYMACCLLYRGDVVPKDVNAAIATIKTKRSI 341
M. musculus:   311 KYMACCLLYRGDVVPKDVNAAIATIKTKRSI 341
B. taurus:     311 KYMACCLLYRGDVVPKDVNAAIATIKTKRSI 341
G. gallus:     311 KYMACCLLYRGDVVPKDVNAAIATIKTKRSI 341
X. tropicalis: 311 KYMACCLLYRGDVVPKDVNAAIATIKTKRTI 341
D. rerio:      311 KYMACCLLYRGDVVPKDVNAAIATIKTKRTI 341
                   ************************** *

H. sapiens:    355 INYQPPTVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYA 400
M. musculus:   355 INYQPPTVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYA 400
B. taurus:     355 INYQPPTVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYA 400
G. gallus:     355 INYQPPTVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYA 400
X. tropicalis: 355 INYQPPTVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYA 400
D. rerio:      355 INYQPPTVVPGGDLAKVQRAVCMLSNTTAIAEAWARLDHKFDLMYA 400
                   *********************************************
```
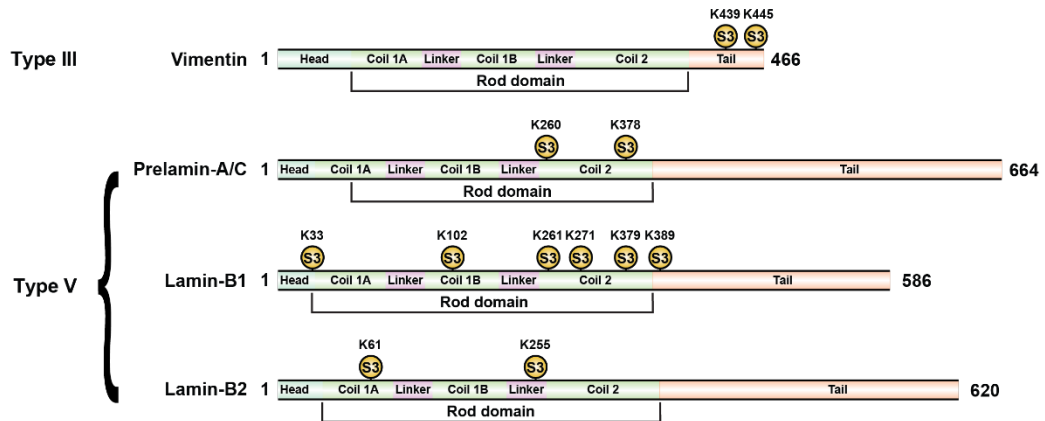
Supplementary Figure 3-6 Cartoon representation of the identified SUMOylation sites on Tubulin at Lys 326 and Lys 370.

Protein sequence alignment of Tubulin across six different species showing that all these lysines are highly conserved.

Supplementary Figure 3-7 Cartoon representation of the identified SUMOylation sites on different intermediate filament proteins.
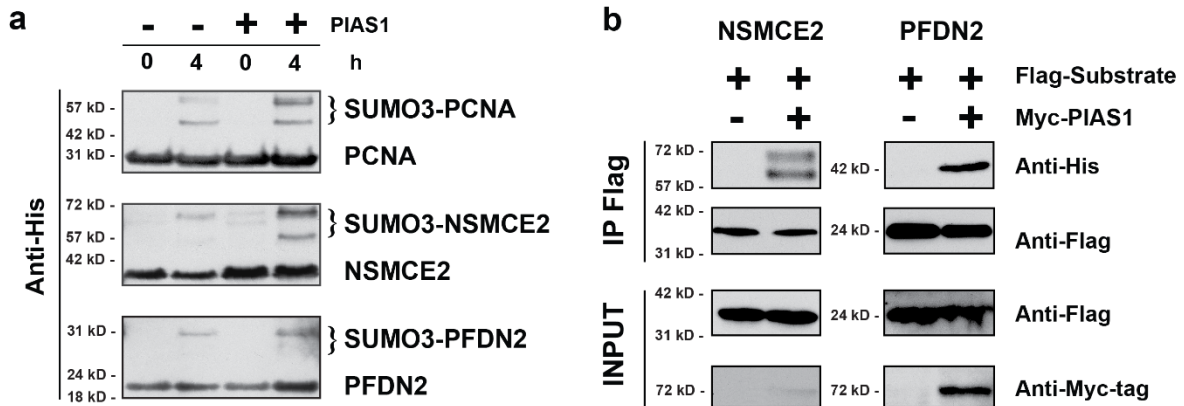
Supplementary Figure 3-8 Validation of SUMOylation on identified PIAS1 substrates.

(a) In vitro SUMOylation assay was performed with or without PIAS1 in a buffer containing SAE1/SAE2, UBC9, SUMO3, ATP and substrates. The samples were incubated at 37 °C for 4h and examined by western blot. In vitro SUMO assays show that PIAS1 enhances SUMOylation of PCNA, NSMCE2 and PFDN2. (b) HEK293 SUMO3m cells were co-transfected with the indicated vectors (top), immunoprecipitated with an anti-Flag antibody, and examined by western blot. SUMOylation of NSMCE2 and PFDN2 were also enhanced by PIAS1 in cells.
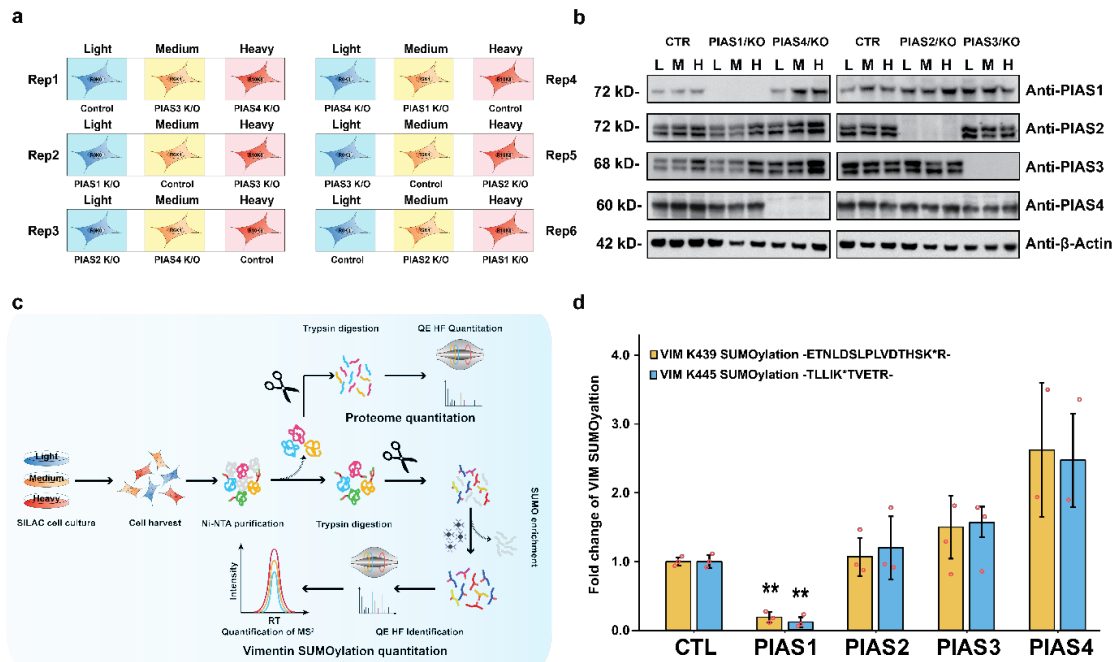
Supplementary Figure 3-9 Workflow for the quantification of vimentin SUMOylation following knockout of different PIAS E3 SUMO ligases.

(a) SILAC labeling strategy for each individual sample replicates. The strategy ensures that each cell line is in all three SILAC channels to eliminate bias caused by a given channel. (b) Endogenous PIAS protein expression in SILAC-labeled HEK293 PIAS-knockout cells were analyzed by western blot. The blot reveals that SILAC channels for each cell line express similar levels of PIAS and that the knockout for each cell line is specific to the respective isoforms. (c) Experimental workflow for the quantification of VIM SUMOylation by mass spectrometry. SUMOylated proteins were first enriched by Ni-NTA, digested on beads with trypsin and modified tryptic peptides were purified by SUMO remnant immunoaffinity purification prior to targeted LC-MS/MS analyses to quantify changes in vimentin SUMOylation. Targeted LC-MS/MS analyses were performed on a Q-Exactive HF mass spectrometer with an inclusion list to acquired MS/MS spectra of the $[M+3H]^{3+}$ precursor ions of the isotopically labeled SUMOylated peptides ETNLDSLPLVDTHSK*R and TLLIK*TVETR where * indicates SUMOylation site. We also analyzed by LC-MS/MS in data-dependent acquisition the tryptic peptides from the flow through proteins to normalize protein abundance across the 6 different samples. d) Fold change of VIM SUMOylation relative to control cells based on SILAC ratios and sample normalisation. VIM SUMOylation is largely abolished at K439 and K445 following CRISPR/Cas9 PIAS1 KO. n=3 biologically independent samples. Data represent the mean ± S.D., error bars represent S.D., **$p<0.01$, Student's t-test).
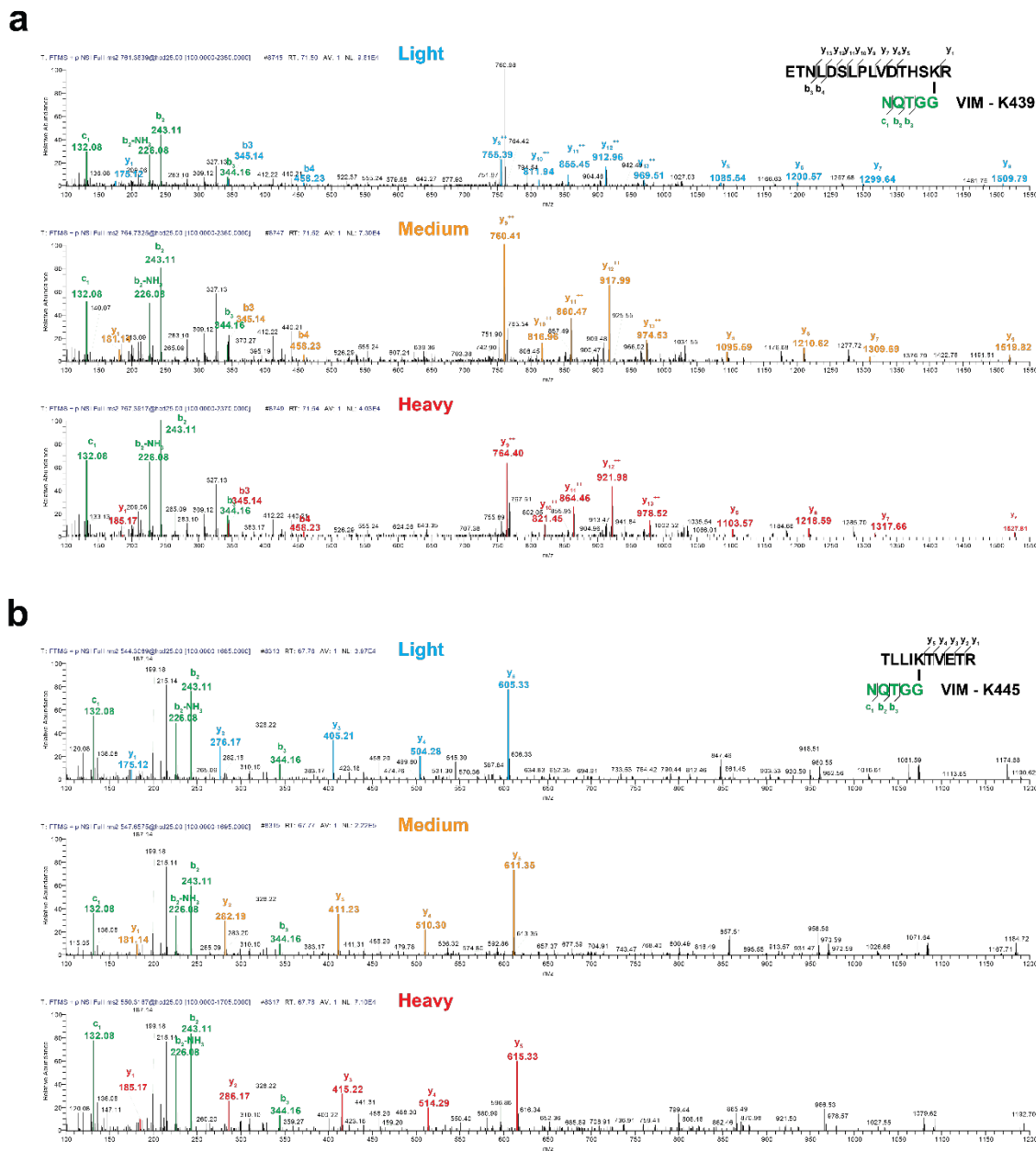
Supplementary Figure 3-10 Representative MS/MS spectra of SUMOylated vimentin peptides.

(a) ETNLDSLPLVDTHSK*R and (b) TLLIK*TVETR in each SILAC channel, where * designates SUMOylation site.

Spectra from each SILAC channel correspond well with the other channels.
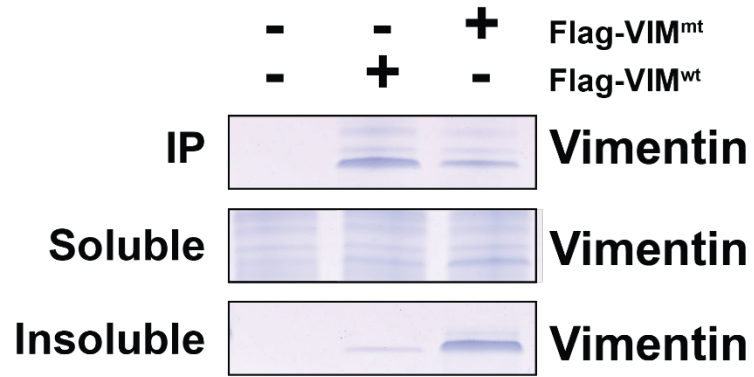
Supplementary Figure 3-11 SDS-PAGE gel fraction.

Flag-VIM^wt, Flag-VIM^mt and negative control from immunoprecipitation, soluble fraction and insoluble fraction used for LC-MS/MS analysis.
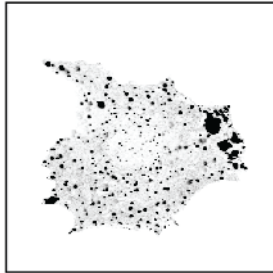
Unit-length filaments   **Intermediate filaments**   **Cytosolic**
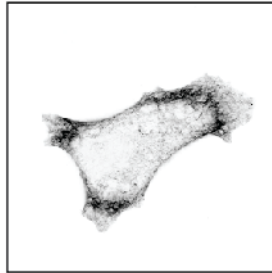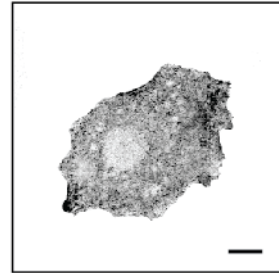
Supplementary Figure 3-12 Representative depiction of different forms of vimentin in MCF-7 cells.

Scale bar: 10 μm.
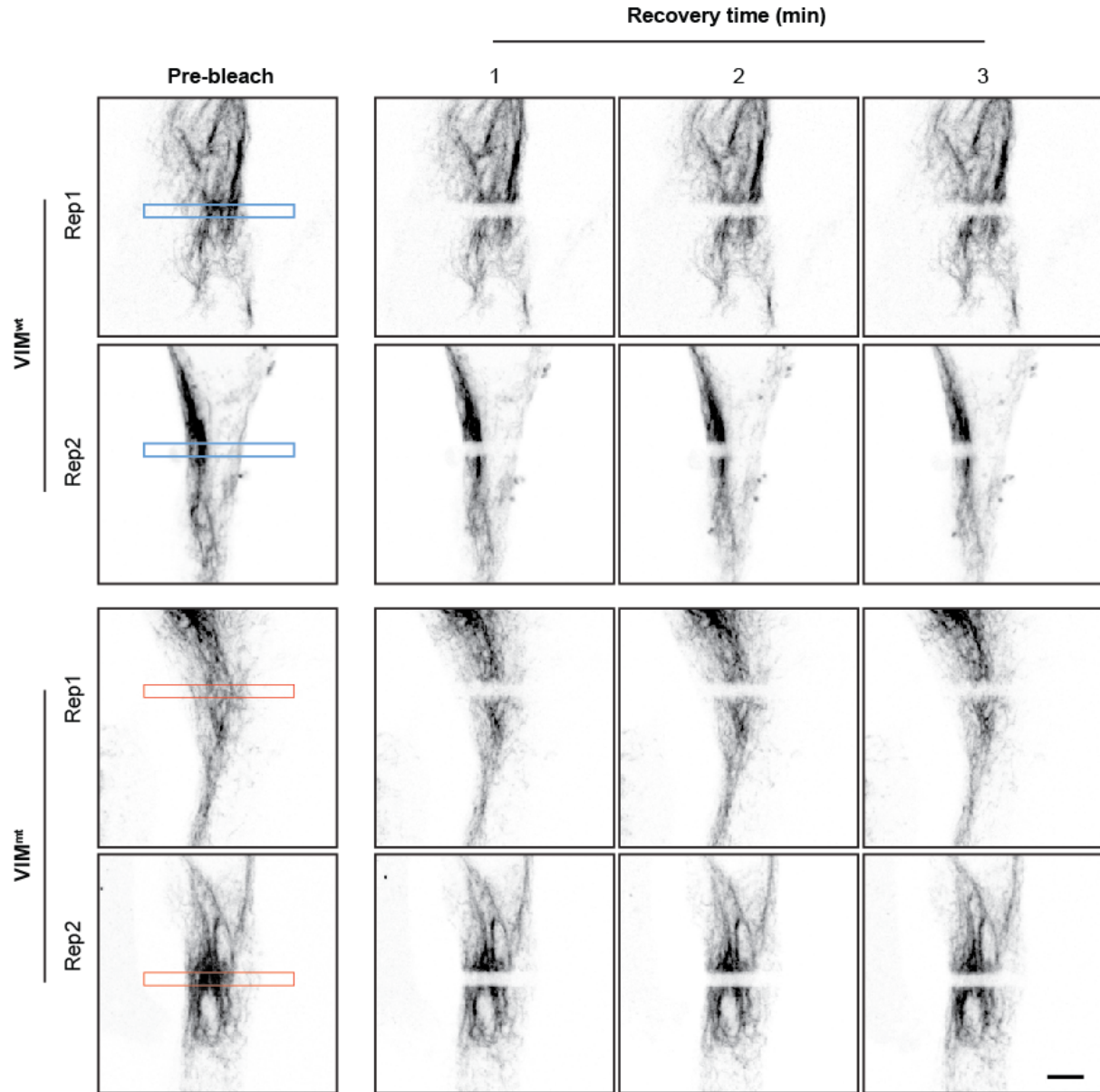
Supplementary Figure 3-13 FRAP assays of Emerald-VIM^{wt} and VIM^{mt} in MCF-7 cells.

Selected images of fluorescence recovery after bleaching are shown. Scale bar: 5 μm.

3.8 References

1.   Li, C., et al., *Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility.* Nat Commun, 2020. **11**(1): p. 834.

2.   Flotho, A. and F. Melchior, *Sumoylation: A Regulatory Protein Modification in Health and Disease*, in *Annual Review of Biochemistry, Vol 82*. 2013, Annual Reviews: Palo Alto. p. 357-385.

3.   Liang, Y.C., et al., *SUMO5, a Novel Poly-SUMO Isoform, Regulates PML Nuclear Bodies.* Scientific Reports, 2016. **6**: p. 15.

4.   Kunz, K., T. Piller, and S. Muller, *SUMO-specific proteases and isopeptidases of the SENP family at a glance.* J Cell Sci, 2018. **131**(6).

5.   Mukhopadhyay, D. and M. Dasso, *Modification in reverse: the SUMO proteases.* Trends in Biochemical Sciences, 2007. **32**(6): p. 286-295.

6.   Liu, B. and K. Shuai, *Regulation of the sumoylation system in gene expression.* Current Opinion in Cell Biology, 2008. **20**(3): p. 288-293.

7.   Sampson, D.A., M. Wang, and M.J. Matunis, *The small ubiquitin-like modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for SURIO-1 modification.* Journal of Biological Chemistry, 2001. **276**(24): p. 21664-21669.

8.   Johnson, E.S., *Protein modification by SUMO.* Annual Review of Biochemistry, 2004. **73**: p. 355-382.

9.   Shuai, K. and B. Liu, *Regulation of gene-activation pathways by pias proteins in the immune system.* Nature Reviews Immunology, 2005. **5**(8): p. 593-605.

10.   Rytinki, M.M., et al., *PIAS proteins: pleiotropic interactors associated with SUMO.* Cellular and Molecular Life Sciences, 2009. **66**(18): p. 3029-3041.

11.   Tan, J.A.T., et al., *Phosphorylation-Dependent Interaction of SATB1 and PIAS1 Directs SUMO-Regulated Caspase Cleavage of SATB1.* Molecular and Cellular Biology, 2010. **30**(11): p. 2823-2836.

12. Okubo, S., et al., *NMR structure of the N-terminal domain of SUMO ligase PIAS1 and its interaction with tumor suppressor p53 and A/T-rich DNA oligomers.* Journal of Biological Chemistry, 2004. **279**(30): p. 31455-31461.

13. Kipp, M., et al., *SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA.* Molecular and Cellular Biology, 2000. **20**(20): p. 7480-7489.

14. van den Akker, E., et al., *FLI-1 functionally interacts with PIASx alpha, a member of the PIAS E3 SUMO ligase family.* Journal of Biological Chemistry, 2005. **280**(45): p. 38035-38046.

15. Duval, D., et al., *The 'PINIT' motif, of a newly identified conserved domain of the PIAS protein family, is essential for nuclear retention of PIAS3L.* Febs Letters, 2003. **554**(1-2): p. 111-118.

16. Palvimo, J.J., *PIAS proteins as regulators of small ubiquitin-related modifier (SUMO) modifications and transcription.* Biochemical Society Transactions, 2007. **35**: p. 1405-1408.

17. Kahyo, T., T. Nishida, and H. Yasuda, *Involvement of PIAS1 in the sumoylation of tumor suppressor p53.* Molecular Cell, 2001. **8**(3): p. 713-718.

18. Gross, M., et al., *Distinct effects of PIAS proteins on androgen-mediated gene activation in prostate cancer cells.* Oncogene, 2001. **20**(29): p. 3880-3887.

19. Megidish, T., J.H. Xu, and C.W. Xu, *Activation of p53 by protein inhibitor of activated Stat1 (PIAS1).* Journal of Biological Chemistry, 2002. **277**(10): p. 8255-8259.

20. Liu, B., et al., *Inhibition of Stat1-mediated gene activation by PIAS1.* Proceedings of the National Academy of Sciences of the United States of America, 1998. **95**(18): p. 10626-10631.

21. Driscoll, J.J., et al., *The sumoylation pathway is dysregulated in multiple myeloma and is associated with adverse patient outcome.* Blood, 2010. **115**(14): p. 2827-2834.

22. Hoefer, J., et al., *PIAS1 Is Increased in Human Prostate Cancer and Enhances Proliferation through Inhibition of p21.* American Journal of Pathology, 2012. **180**(5): p. 2097-2107.

23. Puhr, M., et al., *PIAS1 is a determinant of poor survival and acts as a positive feedback regulator of AR signaling through enhanced AR stabilization in prostate cancer.* Oncogene, 2016. **35**(18): p. 2322-2332.

24. Rabellino, A., et al., *PIAS1 Promotes Lymphomagenesis through MYC Upregulation.* Cell Reports, 2016. **15**(10): p. 2266-2278.

25. Kadare, G., et al., *PIAS1-mediated sumoylation of focal adhesion kinase activates its autophosphorylation.* Journal of Biological Chemistry, 2003. **278**(48): p. 47434-47440.

26. Streich, F.C. and C.D. Lima, *Capturing a substrate in an activated RING E3/E2-SUMO complex.* Nature, 2016. **536**(7616): p. 304-+.

27. Rabellino, A., et al., *The SUMO E3-ligase PIAS1 Regulates the Tumor Suppressor PML and Its Oncogenic Counterpart PML-RARA.* Cancer Research, 2012. **72**(9): p. 2275-2284.

28. Lamoliatte, F., et al., *Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification.* Nature Communications, 2017. **8**: p. 11.

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

30. Cox, J., et al., *Andromeda: A Peptide Search Engine Integrated into the MaxQuant Environment.* Journal of Proteome Research, 2011. **10**(4): p. 1794-1805.

31. Feuermann, M., et al., *Large-scale inference of gene function through phylogenetic annotation of Gene Ontology terms: case study of the apoptosis and autophagy cellular processes.* Database-the Journal of Biological Databases and Curation, 2016: p. 11.

32. Mi, H.Y., et al., *PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements.* Nucleic Acids Research, 2017. **45**(D1): p. D183-D189.

33. Huang, D.W., B.T. Sherman, and R.A. Lempicki, *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic Acids Research, 2009. **37**(1): p. 1-13.

34. Colaert, N., et al., *Improved visualization of protein consensus sequences by iceLogo.* Nature Methods, 2009. **6**(11): p. 786-787.

35. Petersen, B., et al., *A generic method for assignment of reliability scores applied to solvent accessibility predictions.* Bmc Structural Biology, 2009. **9**: p. 10.

36. von Mering, C., et al., *STRING: a database of predicted functional associations between proteins.* Nucleic Acids Research, 2003. **31**(1): p. 258-261.

37. Szklarczyk, D., et al., *The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored.* Nucleic Acids Research, 2011. **39**: p. D561-D568.

38. Shannon, P., et al., *Cytoscape: A software environment for integrated models of biomolecular interaction networks.* Genome Research, 2003. **13**(11): p. 2498-2504.

39. Cline, M.S., et al., *Integration of biological networks and gene expression data using Cytoscape.* Nature Protocols, 2007. **2**(10): p. 2366-2382.

40. McManus, F.P., F. Lamoliatte, and P. Thibault, *Identification of cross talk between SUMOylation and ubiquitylation using a sequential peptide immunopurification approach.* Nature protocols, 2017. **12**(11): p. 2342-2358.

41. Hendriks, I.A., et al., *Uncovering global SUMOylation signaling networks in a site-specific manner.* Nature Structural & Molecular Biology, 2014. **21**(10): p. 927-936.

42. Goodman, S.R., *Medical cell biology*. 2007: Academic Press.

43. Hochstrasser, M., *SP-RING for SUMO: new functions bloom for a ubiquitin-like protein.* Cell, 2001. **107**(1): p. 5-8.

44. Brown, J.R., et al., *SUMO Ligase Protein Inhibitor of Activated STAT1 (PIAS1) Is a Constituent Promyelocytic Leukemia Nuclear Body Protein That Contributes to the Intrinsic Antiviral Immune Response to Herpes Simplex Virus 1.* Journal of Virology, 2016. **90**(13): p. 5939-5952.

45. Keusekotten, K., et al., *Multivalent interactions of the SUMO-interaction motifs in RING finger protein 4 determine the specificity for chains of the SUMO.* Biochem J, 2014. **457**(1): p. 207-14.

46.   McManus, F.P., et al., *Quantitative SUMO proteomics reveals the modulation of several PML nuclear body associated proteins and an anti-senescence function of UBC9.* Sci Rep, 2018. **8**(1): p. 7754.

47.   Lane, E.B., et al., *CO-EXPRESSION OF VIMENTIN AND CYTOKERATINS IN PARIETAL ENDODERM CELLS OF EARLY MOUSE EMBRYO.* Nature, 1983. **303**(5919): p. 701-704.

48.   Ramaekers, F.C.S., et al., *COEXPRESSION OF KERATIN-TYPE AND VIMENTIN-TYPE INTERMEDIATE FILAMENTS IN HUMAN METASTATIC CARCINOMA-CELLS.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1983. **80**(9): p. 2618-2622.

49.   Thomas, J.T., et al., *Human papillomavirus type 31 oncoproteins E6 and E7 are required for the maintenance of episomes during the viral life cycle in normal human keratinocytes.* Proceedings of the National Academy of Sciences of the United States of America, 1999. **96**(15): p. 8449-8454.

50.   Anastasi, E., et al., *Expression of Reg and cytokeratin 20 during ductal cell differentiation and proliferation in a mouse model of autoimmune diabetes.* European Journal of Endocrinology, 1999. **141**(6): p. 644-652.

51.   Chu, Y.W., et al., *EXPRESSION OF COMPLETE KERATIN FILAMENTS IN MOUSE L-CELLS AUGMENTS CELL-MIGRATION AND INVASION.* Proceedings of the National Academy of Sciences of the United States of America, 1993. **90**(9): p. 4261-4265.

52.   Bordeleau, F., et al., *Keratin contribution to cellular mechanical stress response at focal adhesions as assayed by laser tweezers.* Biochemistry and Cell Biology-Biochimie Et Biologie Cellulaire, 2008. **86**(4): p. 352-359.

53.   Hu, G., et al., *Impacts of CyhospitalclinE downstream vimentin on proliferation, invasion and apoptosis of hepatoma HepG2 cell.* International Journal of Clinical and Experimental Medicine, 2018. **11**(6): p. 5564-5571.

54.   Snider, N.T. and M.B. Omary, *Post-translational modifications of intermediate filament proteins: mechanisms and functions.* Nature Reviews Molecular Cell Biology, 2014. **15**(3): p. 163-177.

55.   Zhu, Q.S., et al., *Vimentin is a novel AKT1 target mediating motility and invasion.* Oncogene, 2011. **30**(4): p. 457-470.

56.   Snider, N.T., et al., *Keratin Hypersumoylation Alters Filament Dynamics and Is a Marker for Human Liver Disease and Keratin Mutation.* Journal of Biological Chemistry, 2011. **286**(3): p. 2273-2284.

57.   Goldman, R.D., et al., *Inroads into the structure and function of intermediate filament networks.* Journal of Structural Biology, 2012. **177**(1): p. 14-23.

58.   Helfand, B.T., et al., *Vimentin organization modulates the formation of lamellipodia.* Molecular Biology of the Cell, 2011. **22**(8): p. 1274-1289.

59.   Perez-Sala, D., et al., *Vimentin filament organization and stress sensing depend on its single cysteine residue and zinc binding.* Nature Communications, 2015. **6**.

60.   Satelli, A. and S.L. Li, *Vimentin in cancer and its potential as a molecular target for cancer therapy.* Cellular and Molecular Life Sciences, 2011. **68**(18): p. 3033-3046.

61.   Robert, A., et al., *Vimentin filament precursors exchange subunits in an ATP-dependent manner.* Proceedings of the National Academy of Sciences of the United States of America, 2015. **112**(27): p. E3505-E3514.

62.   Premchandar, A., et al., *Structural Dynamics of the Vimentin Coiled-coil Contact Regions Involved in Filament Assembly as Revealed by Hydrogen-Deuterium Exchange.* Journal of Biological Chemistry, 2016. **291**(48): p. 24931-24950.

63.   Chou, Y.H., et al., *Nestin promotes the phosphorylation-dependent disassembly of vimentin intermediate filaments during mitosis.* Molecular Biology of the Cell, 2003. **14**(4): p. 1468-1478.

64.   Eriksson, J.E., et al., *Specific in vivo phosphorylation sites determine the assembly dynamics of vimentin intermediate filaments.* Journal of Cell Science, 2004. **117**(6): p. 919-932.

65.   Chen, P., et al., *Protein inhibitor of activated STAT-1 is downregulated in gastric cancer tissue and involved in cell metastasis.* Oncology Reports, 2012. **28**(6): p. 2149-2155.

66.   Suman, P., et al., *AP-1 Transcription Factors, Mucin-Type Molecules and MMPs Regulate the IL-11 Mediated Invasiveness of JEG-3 and HTR-8/SVneo Trophoblastic Cells.* Plos One, 2012. **7**(1): p. 12.

67.   Cuijpers, S.A.G., E. Willemstein, and A.C.O. Vertegaal, *Converging Small Ubiquitin-like Modifier (SUMO) and Ubiquitin Signaling: Improved Methodology Identifies Co-modified Target Proteins.* Molecular & Cellular Proteomics, 2017. **16**(12): p. 2281-2295.

68.   Kumar, R., et al., *The STUbL RNF4 regulates protein group SUMOylation by targeting the SUMO conjugation machinery.* Nature Communications, 2017. **8**.

CHAPTER FOUR

4 SUMO proteomics and transcriptomics analyses identify PIAS-mediated regulatory networks involved in cell proliferation and migration

Chongyang Li[1,2], Cristina Mirela Pascariu[1], Trent Nelson[1,2], Zhaoguan Wu[1,3], Mathieu Courcelles[1], Simon Comtois-Marotte[1], Pierre Thibault[1,3,4*]

[1] Institute for Research in Immunology and Cancer, Université de Montréal, Montréal, Québec, Canada.
[2] Molecular Biology Program, Université de Montréal, Montréal, Canada.
[3] Department of Chemistry, Université de Montréal, Montréal, Québec, Canada.
[4] Department of Biochemistry and Molecular Medicine, Université de Montréal, Montréal, Québec, Canada.

Manuscript in preparation.

Authors Contribution:

CL and PT conceived of the project; CL performed phenotypic assay, proteomics and transcriptomics analysis. CMP, TN and CL generated the PIAS knockout cell lines and confirmed by western blot. MC, ZW and SCM assisted with bioinformatic analysis; CL and PT contributed to experimental design and analysis; CL wrote the manuscript. CL, TN and PT edited and reviewed the manuscript.

4.1 Abstract

Protein inhibitor of activated STAT (PIAS) are members of a group of proteins that act as transcription regulators and SUMO E3 ligases. In the triple-negative breast cancer cell line MDA-MB-231, PIAS proteins are overexpressed and PIAS knockout results in a reduction in both cell proliferation and migration. However, the molecular mechanisms underlying PIAS functions in cell proliferation and migration are largely unknown. Here, we leverage quantitative SUMO proteomics and transcriptomics to systematically explore the regulatory role of PIAS SUMO E3 ligases. We used gene editing to generate individual PIAS knock out in HEK293 cells. A total of 1422 SUMO peptides and over 32,000 transcripts were quantified at the SUMOylation and mRNA level, respectively. Among them, the differentially SUMOylated proteins and differentially expressed genes upon PIAS knockout were compared and analyzed using bioinformatics analyses. We identified several substrates involved in cell proliferation and migration that were regulated by the majority of PIAS members from both SUMO proteomic and transcriptomic data, suggesting a level of redundancy within the PIAS family members. Each PIAS also regulated cell proliferation and migration through a unique pool of candidates, such as CHD4 and CDC73 for PIAS1, PRDX1 for PIAS2, SSRP1 for PIAS3, and RBBP7 and CBX3 for PIAS4. The minimal overlap of regulated candidates identified from SUMO proteomic and transcriptomic data suggests different yet complementary mechanisms mediated by PIAS members. In addition, PIAS1-3 were also identified to be involved in DNA damage repair, chromosome structure and SUMO polychain formation through unique substrates. These results provide novel insights into both the redundant and specific regulatory mechanisms of cell proliferation and migration mediated by PIAS SUMO E3 ligases.

4.2 Introduction

Protein SUMOylation is the process by which a target protein is covalently modified by a small ubiquitin-like modifier (SUMO) protein [1]. Analogous to ubiquitin, SUMOylation occurs via an enzymatic cascade consisting of SUMO E1 activating enzymes SAE1/SAE2, E2 conjugating enzyme UBC9, and SUMO E3 ligases [2]. The principal family of SUMO E3 ligases is the protein inhibitor of activated STAT (PIAS) which play a key role in regulating various cellular processes, including most notably transcription, signal transduction, cell-cycle progression, protein SUMOylation and stabilization [3-7]. PIAS proteins are also implicated in immune regulation, DNA repair, cellular proliferation, cell survival, cell motility and migration [8, 9]. They were originally found to interact with and inhibit activated signal transducers and activators of transcription (STATs). More recently, these proteins have been shown to function as E3 ligases that promote the SUMO modification of a number of transcription regulators [10]. PIAS proteins were first discovered in *Saccharomyces cerevisiae* (Siz1/Siz2) and are evolutionarily conserved in eukaryotes [11, 12]. The family of mammalian PIAS consist of four genes giving rise to seven proteins, namely PIAS1, PIASxα, PIASxβ, PIAS3, PIAS3L, PIASy and PIASyE6⁻ (PIAS4) [12]. Sequence and structural analysis of PIAS proteins indicates a high degree of conservation within isoforms [12, 13]. Four conserved domains and two motifs, including the N-terminal scaffold attachment factor-A/B, acinus and PIAS (SAP) domain, the Pro-Ile-Asn-Ile-Thr (PINIT) motif, the central RING-finger-like zinc-binding domain (RLD), the highly acidic domain (AD), the SUMO-interacting motif (SIM), and the C-terminal serine/threonine-rich region (S/T) [14], have been identified in all PIAS proteins.

PIAS proteins have been most frequently implicated in tumorigenesis [15]. In human prostate cancer, PIAS1 functions as a target gene selective androgen receptor coregulator on chromatin and enhances proliferation through inhibition of p21 [16]. In the triple-negative breast cancer model cell line MDA-MB-231, PIAS1 was found to regulate tumour initiating stem cells by modulating the epigenetic status of several clinically relevant genes, including cyclin D2 (*CCND2*), estrogen receptor (*ESR1*), and breast tumor suppressor WNT5A (*WNT5A*) [17]. One multiple myeloma clinical study reported that expression levels of PIAS1 are negatively correlated with

the patient's 6-year survival rate, indicating that PIAS1 and the SUMO pathway may promote tumorigenesis and unfavorable prognostic [18]. Increased PIAS3 expression was observed in a variety of human cancers including lung, breast, prostate, colorectal and brain tumors [15]. In contrast, lower expression of PIAS3 protein was found to be correlated with poor survival of patients with gastric cancer [19]. Furthermore, microRNA-21 activates the STAT3-dependent signal pathway by inhibiting the function of PIAS3 and down-regulation of PIAS3 contributes to the oncogenic function of microRNA-21 in multiple myeloma [20]. Several studies have highlighted the relationship between hypoxic signaling and PIAS4 overexpression in cancers such as ovarian and pancreatic, implicating PIAS4 as part of the shift upon hypoxic signaling [21, 22]. Lastly, in acute promyelocytic leukemia, PIAS2-α mediated PML SUMOylation may have a role in promoting replicative senescence [13].

In addition to mediating cell proliferation and apoptosis in various cancers, PIAS proteins also regulate cell migration and invasion of cancer cells. For example, PIAS4 enhances epithelial cell migration through SUMOylation of the transcription factor C/EBPδ, leading to sequestration at the nuclear periphery and subsequent reduction of C/EBPδ transcriptional activity [23]. PIAS3-mediated GTP-Rac1 SUMOylation prolongs the Rac1-GTP activated state and further stimulates cell migration [24]. In MDA-MB-231 triple-negative breast cancer cells, PIAS1 and/or PIAS4 mediates PYK2 SUMOylation, subsequently triggering autophosphorylation, interaction with tyrosine kinase SRC, phosphorylation of paxillin, activation of ERK1/2, and promotion of cell migration [25]. Our recent large-scale quantitative SUMO proteomics analysis revealed that PIAS1 SUMOylates intermediate filament protein vimentin, which enhances vimentin solubility and accelerate intermediate filament disassembly and reassembly, further promoting HeLa cell migration and motility [14].

Although these studies have identified a role for protein SUMOylation in cancer development through cell proliferation and migration, the molecular mechanisms by which this is achieved remain unclear. There is thus a need to identify PIAS substrates that could be associated with tumor cell migration and invasiveness. To date, an increasing number of potential SUMO E3 ligase substrates has been reported, though most of these were analyzed individually

using biochemical and molecular biological assays [26-28]. More recently, a human proteome microarray-based activity screen has been used to identify putative substrates of specific SUMO E3 ligases and revealed that PIAS2 SUMOylation of PYK2, a member of focal adhesion kinase family, is essential for its activation and interaction with SRC [25, 29]. Although protein arrays provide a valuable approach to establish target specificity of SUMO E3 ligases they cannot provide information on the nature and sites of modification. To analyze the molecular components and regulatory mechanisms by which PIAS SUMO E3 ligases exert their action on cell proliferation and migration, we used SUMO proteomics and transcriptomics analyses to identify protein substrates and gene candidates upon knockout (KO) of individual PIAS genes. We identified a large subset of genes/proteins involved in cell proliferation and migration to be differentially expressed/modified, supporting the role of PIAS proteins on these cellular processes. To the best of our knowledge, this work represents the first combined transcriptomic and SUMO proteomic analysis of PIAS SUMO E3 ligase regulatory networks.

4.3 Methods

4.3.1 Cell Culture and Transfection

Human breast cancer cell line MDA-MB-231 was purchased from American Type Culture Collection (ATCC), and HEK293 stably expressing the 6xHis-SUMO3-Q87R/Q88N mutant (HEK293-SUMO3m) [30] were cultured in Dulbecco's modified Eagle's medium (HyClone) supplemented with 10% fetal bovine serum (Wisent), 1% L-glutamine (Thermo Fisher Scientific), 1% penicillin/streptomycin (Invitrogen) in 5% CO_2 at 37 °C.

For gene knockout, the CRISPR/Cas9-based gene knockout vectors pCRISPR were purchased from Genecopoeia, Inc. (Rockville, MD). We ordered three different sgRNAs for each PIAS gene. The sgRNA sequence for each gene is listed below. PIAS1-sgRNA1: 5'-TTCTGAACTCCAAGTACTGT-3', PIAS1-sgRNA2: 5'- GCCCTGCATTTGCTAAAGGC-3', PIAS1-sgRNA3: 5'-ACTTGAATGTACGTTGGGGA-3', PIAS2-sgRNA1: 5'-CAAGTATTACTAGGCTTTGC-3', PIAS2-sgRNA2: 5'-ATAAATGCAGCGCCCTCATC-3', PIAS2-sgRNA3: 5'-CTCATCAAGCCCACGAGTTT-3', PIAS3-sgRNA1: 5'-GCCCTTCTATGAAGTCTATG-3', PIAS3-sgRNA2: 5'- ATGGTGACATCAGGGTGCAC-3', PIAS3-sgRNA3: 5'-GCCCTTCTATGAAGTCTATG-3', PIAS4-sgRNA1: 5'-GGCTTCGCGCCGTAGTCTTAG-3', PIAS4-sgRNA2: 5'- GAAGCACGAGCTCGTCACCA-3', PIAS4-sgRNA3: 5'-GAGCTTCACC AGGCGGACTTC-3', and scrambled sgRNA: 5'-GGCTTCGCGCCGTAGTCTTA-3'. MDA-MB-231 cells or HEK293-SUMO3m cells were transfected with 1 μg pCRISPR vector or a scrambled sgRNA as negative control per million cells using JetPrime Reagent (Polyplus-transfection) according to the manufacturer's protocol. Single-cell sorting for clonal cell line development was achieved by Fluorescence-activated cell sorting (FACS) based on Red Fluorescent Proteins-mCherry signals 48h after transfection. mCherry positive cells were cultured and expanded for two weeks. PIAS knockout efficiency was confirmed by western blot.

4.3.2 Cell Proliferation Assay

The cell proliferation assay was conducted using cell proliferation reagent WST-1 solution (Roche). PIAS KO or CTL MDA-MB-231 cells were seeded into 96-well plates at a density of 1 × 10^3 cells/well. After 0, 2 and 4 days of cell culture, WST-1 reagents were added to each well, and

cells were further incubated at 37 °C with 5% CO_2 for 1h. Subsequently, the absorbance of the samples was measured using Nanoquant Infinite® 200 Pro spectrophotometer (TECAN) with an optical density (OD) at 450 nm. The background subtraction was performed using OD_{630}. Each experiment was performed with three instrumental replicates and three biological replicates. The relative numbers of viable cells for each condition were estimated based on the absorbance of (OD_{450}-$OD_{630)}$.

4.3.3 Migration Assays

Cell migration was assessed using a wound-healing assay. MDA-MB-231 PIAS KO and CTL cells were seeded as a single monolayer into 2-well culture-inserts (Ibidi) placed in a μ-slide 8 well plate (Ibidi). After removal of the inserts, cells were continuously cultured in a 37 °C thermostated chamber supplemented with 5% CO_2. The wound closure was captured every 5 min using a Zeiss LSM 700 confocal microscope over a period of 12h. The pictures were analyzed to calculate the relative area of the wound closure using "MRI Wound Healing Tool" plugin in ImageJ to further assess the would-healing rate.

4.3.4 SILAC Labeling and Protein Extraction

The parental HEK293-SUMO3m cells were cultured in DMEM (Thermo Fisher Scientific) supplemented with light (^{0}Lys, ^{0}Arg) amino acids. HEK293-SUMO3m-KO-CTL cells were grown in DMEM (Thermo Fisher Scientific) containing medium (^{4}Lys, ^{6}Arg), and different PIAS KO cells were separately grown in DMEM (Thermo Fisher Scientific) containing heavy (^{8}Lys, ^{10}Arg) isotopic forms of lysine and arginine (Silantes) for at least 6 passages to ensure full labelling. For each triple SILAC experiment, only the light channel was treated with 10mM MG132 for 16h to increase protein SUMOylation. The MG132-treated light channel serves as a signal booster for the other channels to trigger MS/MS of SUMOylated peptides. The number of cells combined from each channel was using the ratio of L: M: H = 1 : 4: 4, and the combined cells were washed twice with ice-cold PBS, lysed in NiNTA denaturing incubation buffer (6 M Guanidinium HCl, 100 mM NaH_2PO_4, 20 mM 2-Chloroacetamide, 5 mM 2-Mercaptoethanol, 10 mM Tris-HCl pH=8) and sonicated. Protein quantification was performed using micro Bradford assay (Bio-Rad).

4.3.5 Protein Purification, Digestion and Desalting

For each biological replicate, 9 mg of total cell extract (TCE) were incubated with 360 µL of NiNTA beads (Qiagen) at 4 °C for 16h. The beads were subsequently washed once with 12 mL of NiNTA denaturing incubation buffer, five times with 12 mL of NiNTA denaturing washing buffer (8 M urea, 100 mM NaH_2PO_4, 20 mM imidazole, 5 mM 2-Mercaptoethanol, 20 mM Chloroacetamide, 10 mM Tris-HCl pH=6.3) and twice with 10 mL of 50 mM ammonium bicarbonate. Protein quantification on beads was conducted using micro Bradford assay (Bio-Rad). Protein digestion on beads was carried out using a ratio 1:20 sequencing grade modified trypsin (Promega): protein extract in 100 mM ammonium bicarbonate at 37 °C overnight. To quench the reaction, 1% trifluoroacetic acid (TFA) was added. Sample desalting was performed on hydrophilic-lipophilic balance (HLB) cartridges (3cc, 60 mg) (Waters) and eluted in Low Protein Binding Collection Tubes (Thermo Fisher Scientific). Finally, samples were dried down by Speed Vac prior to LC-MS/MS analysis.

4.3.6 Immunoisolation of SUMO Peptide

MagReSyn® Protein A magnetic beads (RESYN BIOSCIENCES) were equilibrated with anti-K(NQTGG) antibody at a ratio of 1:2 (v/w) for 1h at 4 °C in PBS and then washed 3 times with 200 mM triethanolamine pH=8.3. For antibody/beads crosslinking, 10 µL of 5 mM DMP in 200 mM triethanolamine pH=8.3 was added per µl of bead slurry and incubated with inversion for 1h at room temperature. The reaction was quenched for 30min by adding 1 M Tris-HCl pH=8 to a final concentration 50 mM. Prior to immunoisolation, three washes with ice-cold PBS were applied on crosslinked beads. The desalted peptides were resuspended in 1 ml PBS containing cross-linked anti-K-NQTGG at a ratio of 1:2 (w/w). The peptides/beads were incubated for 1 h at 4 °C, followed with three washes of 1 mL of 1 × ice-cold PBS, once with 1 mL of 0.1 × ice-cold PBS, and last wash with cold MilliQ H_2O. The enriched SUMO peptides were eluted from the beads using 200 µL of 0.2% formic acid in water three times and dried down by Speed Vac.

4.3.7 SCX Fractionation

Strong cation exchange (SCX) stage tips (ThermoFisher Scientific) were conditioned with acetonitrile. SUMO peptides were redissolved in water containing 15% acetonitrile (ACN and 0.2% formic acid (FA) and loaded onto the conditioned SCX stage tip. Loaded peptides were washed once with an SCX wash buffer (10 mM ammonium acetate, 20% ACN, and 0.5% FA). Sequential peptide fractionation was performed using a series of SCX elution buffers with increasing concentration of ammonium acetate (125, 175, 250, 350, 1500 mM ammonium acetate in 20% ACN, 0.5% FA). All the eluted fractions were dried down by Speed Vac and subject to LC-MS/MS analysis.

4.3.8 Mass Spectrometry Analysis

Peptides were redissolved in 0.2% FA solution and analyzed by LC-MS/MS using a Q Exactive™ HF Hybrid Quadrupole-Orbitrap™ Mass Spectrometer (Thermo Fisher Scientific) coupled to a Proxeon Easy-nLC 1000. Samples were injected on a 300 μm ID × 5 mm trap and separated on a 150 μm × 20 cm nano-LC column (Jupiter C18, 3 μm, 300 A, Phenomenex). The separation on the LC system was performed using a linear gradient from 7 to 30% ACN, 0.2% FA over 105min at 600 nL/min. The acquisition parameters on the MS instrument were set as follow: For full MS scans, the scan window was from m/z 350 to m/z 1,500 with a resolution of 120,000 at m/z 200. A target AGC of 1E6 with a maximum injection time of 200ms was used. For MS/MS scans, HCD mode with a normalized collision energy of 25 was enabled and resolution of 30,000 with a Top 3 s method, a target AGC of 5E3, and a maximum injection time of 1,000ms was used. The MS/MS triggering threshold was set to 1E5, and the dynamic exclusion of previously acquired precursor was enabled for 20 sec within a mass range of ±0.8 Da.

4.3.9 Data Processing

MS raw data were analyzed using MaxQuant (version 1.6.2.10 and searched against UniProt/SwissProt human proteome database (http://www.uniprot.org/) released on June 2019). The maximum missed cleavage sites for trypsin was set to 3. The fixed modification was set using Carbamidomethylation (C) and acetylation (Protein N term), oxidation (M), deamination (NQ), and NQTGG (K) were set as variable modifications. The "match between runs" feature was enabled to correlate identification and quantitation results across different runs. The false

discovery rate was set to 1% for peptide, protein, and site identification. Only the SUMO sites with a localization probability of ≥ 0.75 were kept.

4.3.10 RNA extraction and sequencing

For HEK293-SUMO3m PIAS gene knockout cells, total RNA was isolated using TRIzol (Invitrogen) and purified using the RNeasy Mini kit (Qiagen) as recommended by the manufacturer. RNA integrity was determined by analysis on a 2100 Bioanalyzer (Agilent Genomics). RNA samples (500 ng of total RNA) were used to prepare cDNA libraries using the KAPA Stranded mRNA-Seq Kit (KAPA Biosystems). Libraries were then amplified using Truseq primers (Illumina). Amplified libraries (1 nmol/L) were used for single-end RNA-seq on the Illumina NextSeq 500 platform. Between 150-300 million reads were obtained after 75 cycles of single-end reads.

4.3.11 Quality control and transcriptome assembly

Raw Illumina SDS reads (fastq) were first processed using in-house R scripts. Sequences were trimmed for sequencing adapters and low quality 3' bases using Trimmomatic version 0.35 [31] and aligned to the reference human genome version GRCh38 using STAR version 2.7.1a [32]. Then gene expressions were obtained both as readcount directly from STAR as well as computed using RSEM [33] to obtain normalized gene and transcript level expression. Finally, DESeq2 version 1.22.2 [34] was used to normalize gene readcounts and filter out differentially expressed genes.

4.3.12 Bioinformatics Analysis

Classification of identified PIAS1 substrates was performed using PANTHER 16.0 (Protein Analysis Through Evolutionary Relationships) (http://www.pantherdb.org), which classifies genes and proteins by their functions [35, 36]. The PIAS-regulated SUMO substrates and genes were grouped into the biological process, molecular function and cellular component classes using DAVID Bioinformatics Resources 6.8 [37, 38]. The chromosome distribution information of identified transcripts was first extracted from annotated identification of transcriptomics analysis. Then the distribution was visualized using BioCircos package [39] in R software.

216

Regulatory networks of PIAS-regulated SUMO substrates and genes were built by searching against the STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) database version 11.0. Visualization of regulatory networks was then performed in Cytoscape v3.8.0 and sub-networks were grouped according to their GO terms. Evolutionary conservation analysis of SUMOylated lysine residues was assessed across twelve species, including *M. musculus*, *R. norvegicus, B. taurus, G. gallus, X. laevis, D. rerio, D. melanogaster, C. elegans, A. thaliana, O. sativa, S .cerevisiae and S. pombe*, using the ProteoConnections bioinformatics platform [40].

4.3.13 Western Blot

TCE was diluted in Laemmli buffer (10% (w/v) glycerol, 2% SDS, 10% (v/v) 2-mercaptoethanol and 62.5 mM Tris-HCl, pH=6.8) and then boiled for 5 min on a Dry Metal Heating Block Bath Incubator at 95°C. The protein was separated on a 4-12% precast SDS-PAGE (Bio-rad) followed by transfer onto nitrocellulose (NC) membranes. NC membranes were briefly stained with 0.1% Ponceau-S in 5% acetic acid to determine total protein content prior to blocking with 5% non-fat milk in Tris-buffered saline with Tween 20 (TBST) at RT for 1h. Membranes were subsequently incubated overnight with primary antibody at a 1:1000 prepared in 5% bovine serum albumin (BSA) in TBST at 4 °C. NC membranes were then incubated with peroxidase-conjugated anti-mouse or anti-rabbit IgG (Cell Signaling Technology) at a 1:5000 dilution prepared in 5% BSA in TBST for 1h at RT. Finally, the NC membranes were subjected to three washes with TBST for 10min each and revealed using ECL plus (GE Healthcare) as per the manufacturer's instructions. Chemiluminescence was captured on Blue Ray film (VWR). The following antibodies were used for western blot analyses: rabbit anti-PIAS1 Antibody (3550S, Cell Signaling Technology), rabbit anti-PIAS2 Antibody (ab126601, Abcam), rabbit anti-PIAS3 Antibody (9042S, Cell Signaling Technology), rabbit anti-PIAS4 Antibody (4392S, Cell Signaling Technology), rabbit anti-α-Tubulin Antibody (2144S, Cell Signaling Technology), rabbit anti-β-Actin Antibody (4970S, Cell Signaling Technology).

4.3.14 Statistical Analysis

To assess the difference among experimental groups, statistical analysis was conducted to analyze the significance using Student's t-tests. $p < 0.05$ was considered to be statistically significant. One asterisk and two asterisks indicate $p < 0.05$ and $p < 0.01$, respectively.

4.4 Results

4.4.1 CRISPR/Cas9-based gene editing ensures the PIAS KO specificity

To understand the physiological functions and differences between PIAS SUMO E3 ligases, cell lines of each individual PIAS gene knockout were generated in the breast cancer cell line MDA-MB-231 and used to perform phenotypic assays. All PIAS proteins contain four conserved functional domains that are ideal targets for sgRNA by CRISPR/Cas9 to obtain an efficient gene KO. Due to the high degree of protein sequence identity within these functional domains, it is crucial to choose regions unique to different PIAS genes in the open reading frame (ORF) to avoid targeting several PIAS genes simultaneously. Here we utilized a CRISPR/Cas9-based gene editing strategy followed by Fluorescence-activated Cell Sorting (FACS) to efficiently select monoclonal cell lines specific for each PIAS gene KO (Figure 4-1a). We designed three different sgRNAs for each PIAS gene that targeted different sequences corresponding to protein functional domains and inserted them into the pCRISPR vector, which constitutively expresses sgRNA, Cas9 nuclease, and mCherry fluorescence protein. Two days post-transfection, cells expressing the mCherry protein were sorted using FACS and single cells were seeded into 96-well plates for clonal cell line expansion. This workflow allows for a rapid KO screen without antibiotic selection and resulted in a >90% success rate among obtained clonal cell lines after two weeks. Next, we verified the individual PIAS knockout efficiency and specificity by western blot (Figure 4-1b). The western blot results confirmed that each PIAS gene was successfully knocked out in MDA-MB-231 cells without disturbing other PIAS genes.
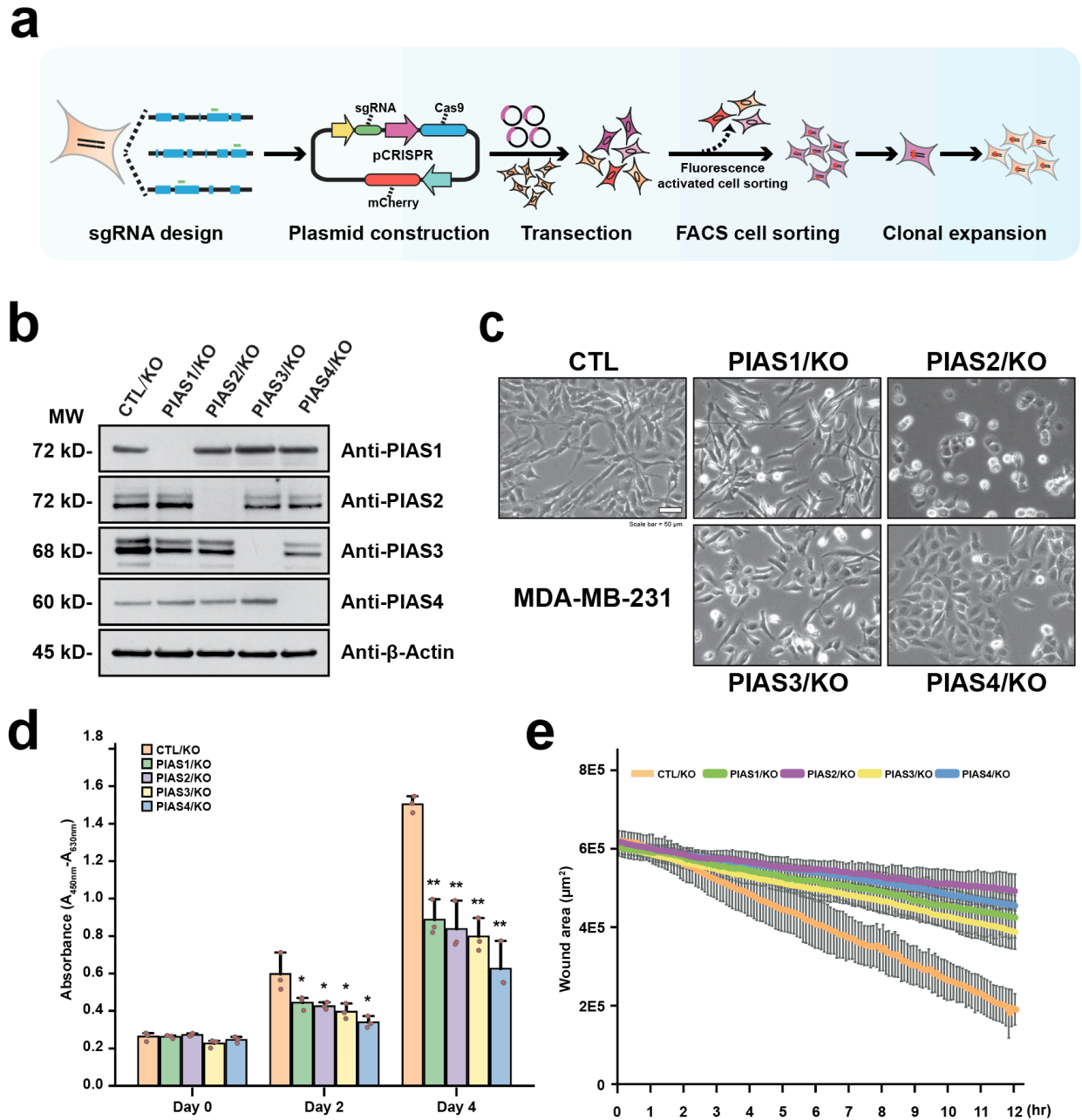
Figure 4-1 Functional effects of PIAS gene KO in breast cancer cell line MDA-MB-231.

(a) Workflow for individual PIAS gene KO in MDA-MB-231 cells using CRISPR/Cas9 gene-editing technology. (b) Western blot validation of individual PIAS gene KO efficiency and specificity. β-Actin was used as a loading control. (c) Cell morphology of MDA-MB-231 cells after individual PIAS gene KO under inverted microscope, scale bar: 50 μm. (d) WST-1 proliferation assay of MDA-MB-231 cells after individual PIAS gene KO. Bar plot illustrates the degree of proliferation for individual PIAS KO MDA-MB-231 cells up to 4 days. The negative control cells were generated using scrambled sgRNA, n=3 biologically independent samples. (e) Comparison of

cell migration ability of each PIAS KO MDA-MB-231 cell lines by wound-healing assay, n=3 biologically independent samples. Cell Data represent the mean ± S.D., error bars represent S.D., *p<0.05, **p<0.01, Student's t-test.

4.4.2 PIAS KO affect MDA-MB-231 cell proliferation and migration

We have previously reported that the expression of PIAS1 is positively correlated with increased levels of cell proliferation and migration in HeLa cells [14]. Here, we further examined the physiological effects of different PIAS gene KO in the breast cancer cell line MDA-MB-231. We first investigated cell morphological changes upon PIAS KO (Figure 4-1c). Interestingly, PIAS2 and 4 KO significantly altered cell morphology. PIAS2 KO cells lost their ability to form pseudopodia resulting in a more rounded shape, whereas PIAS4 KO cells presented a similar phenotype but to a lesser extent. Conversely, PIAS1 KO caused the cells to stretch into a longer and narrower shape, while no significant changes were observed for PIAS3 KO cells compared to control cells. Next, we evaluated the proliferation and migration ability of these KO cells. We seeded 1000 cells of each PIAS KO cell line with four technical and three biological replicates, and measured the proliferation rate up to four days using the WST-1 cell proliferation assay. The results indicated that different PIAS KO have a consistent influence on cell proliferation, each reducing the proliferative capacity of MDA-MB-231 cells. The reduced proliferation rate varied between the PIAS KO cells, with PIAS4 KO exhibiting the greatest impact, showing more than 60% reduction compared with control cells by day 4. PIAS1 KO had the least impact on the proliferation rate but still resulted in a 40% decrease relative to control cells by day 4 (Figure 4-1d). To determine the effect of PIAS KO on cell motility, we performed a wound-healing assay and measured the area of wound closure for a period of 12h. Surprisingly, cell motility was affected by all PIAS KO, despite there being no morphological changes observed for PIAS3 KO cells. We found PIAS2 KO cells migrated the slowest, as shown by approximately 10% wound recovery compared with control cells by 12h (Figure 4-1e). The reduction in cell migration upon PIAS1 KO in MDA-MB-231 cells is in line with previous findings in HeLa cells [14]. In addition, other PIAS KO show the same trend in MDA-MB-231 cells. Altogether, these results indicated that

each PIAS KO reduced cell proliferation and migration to varying degrees, thus suggesting that a role for SUMOylation in the observed phenotypes.

4.4.3 Integrative analyses of SUMO proteomics and transcriptomics upon PIAS KO in HEK293 SUMO3m cells

To better understand the mechanisms responsible for regulating cell proliferation and migration by different PIAS SUMO E3 ligases, we conducted a large-scale SUMO proteomics experiment to identify substrates of PIAS-mediated protein SUMOylation. Here, we used the CRISPR/Cas9 gene knockout approach described in Figure 4-1a to generate PIAS KO cells in HEK293 SUMO3m cell lines for SUMO proteomics. We obtained efficient and specific PIAS KO in HEK293 SUMO3m cells as confirmed by western blot (Figure 4-2a). Next, we performed a triple SILAC-based SUMO proteomics analysis modified from our previously published approach [14] to quantitatively profile the PIAS-mediated SUMO regulation networks in a site-specific manner. For each triple SILAC, HEK293 SUMO3m KO control cells were grown in media containing medium isotopic forms of lysine and arginine (^{4}Lys, ^{6}Arg), while PIAS KO cells were grown in heavy media (^{8}Lys, ^{10}Arg) under basal condition at 37 °C. In order to normalize quantitation results across all 12 biological replicates including 4 PIAS KO samples, we used parental HEK293 SUMO3m cells grown in light media (^{0}Lys, ^{0}Arg) for all light channels as an internal standard. In this case, the quantitation across samples can be normalized and technical variations introduced from sample preparation can be removed. Additionally, these cells were treated with MG132 for 16h prior to cell collection to boost the SUMO identification. The light channel allows the mass spectrometer to find high intensity precursor ions of SUMO peptides presented as the light version, and further triggers MS/MS fragmentation and identification for those ions with low intensities from medium and heavy channels. Of note, only ¼ amount of protein lysate from the light channel was combined with samples from medium and heavy channels, giving a ratio of 1:4:4 for each sample combination. This SILAC combination strategy prevents signal saturation from SUMO peptides in the light channel, which could cause an inaccurate quantitation for lower abundant SUMO peptides from medium and heavy channels.
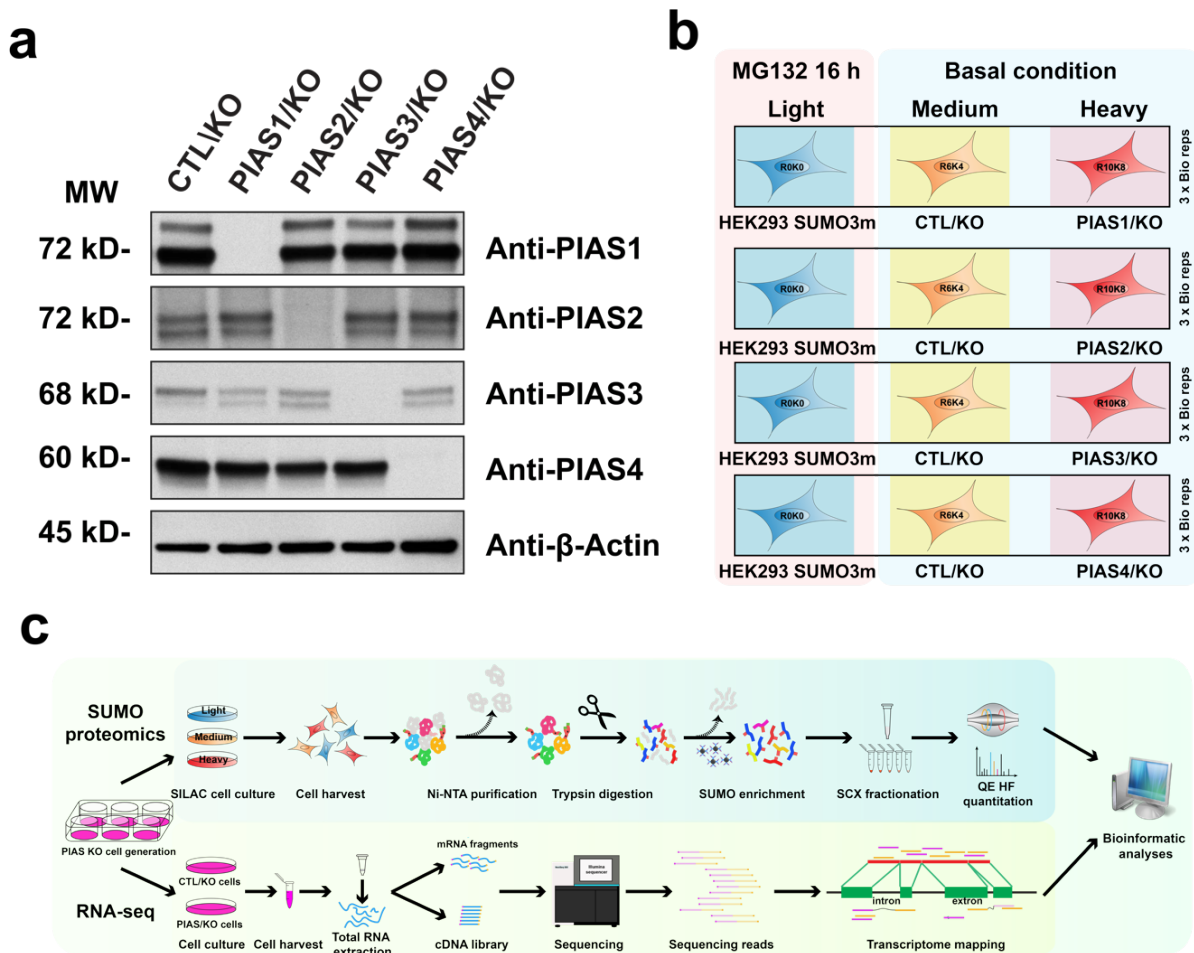
Figure 4-2 Overview for the integrative analyses of SUMO proteomics and transcriptomics on individual PIAS knockout HEK293 SUMO3m cells.

(a) Individual PIAS gene knockout in HEK293 SUMO3m cells was generated using the same workflow shown in Figure 4-1a. Western blot results confirm the individual PIAS gene KO efficiency and specificity in HEK293 SUMO3m cells. β-Actin was used as a loading control. KO control cells were generated using scrambled sgRNA. (b) A schematic diagram showing the triple SILAC labelling strategy. Parental HEK293 SUMO3m cells were cultured in the DMEM media containing light versions of lysine and arginine. The medium channel was used for labelling KO control cells, and heavy channels were used for individual PIAS KO cells. Only cells in the light channel were treated with MG132 for 16 h before collection. SILAC labelled cells were lysed and combined in a 1:4:4 ratio based on amount of protein, n=3 biologically independent samples. (c) Workflow of integrative analyses of SUMO proteomics and RNA-seq-based transcriptomics for determining PIAS regulation networks.

We took advantage of our optimized workflow to identify SUMO substrates regulated by each PIAS SUMO E3 ligase (Figure 4-2c). After pooling labeled cells, samples were subjected to

Ni-NTA pulldown for SUMO protein purification prior to tryptic digestion. Subsequently, immunoisolation was performed on the Ni-NTA purified peptide pool to enrich the SUMO peptides containing the NQTGG-remnant. Off-line SCX peptide fractionation was conducted to enhance peptide coverage and the depth of SUMO proteome analysis. Finally, all samples were injected on the Q Exactive HF LC-MS/MS System and raw data were searched using MaxQuant software [41].

To uncover the regulatory mechanisms on the transcription level by PIAS SUMO E3 ligases, we analyzed in parallel RNA-seq data of the same cells (Figure 4-2c). Total RNA was extracted from each PIAS KO HEK293 SUMO3m cells as well as control cells using Trizol. The mRNA fragments were then sequenced using the Illumina NextSeq 500 platform, and subsequently mapped to the reference human genome version GRCh38 using STAR version 2.7.1a [32]. Finally, the normalized gene expression, in Transcripts Per Million (TPM) values, was computed using RSEM [33], and differential gene expression analysis was performed with *DESeq2* package using R statistical software [42].

4.4.4 Overview of SUMO proteomic and transcriptomic results

To evaluate the reproducibility of our SUMO proteome and RNA-Seq datasets, we determine the Pearson correlation coefficients and obtained values over 0.9 across biological triplicates, indicating a high level of reproducibility of our analyses (Supplementary Figure 4-1 a and b). Surprisingly, the regulation similarity differs between PIASs on a SUMOylation level with a lower score, whereas gene expression upon different PIAS KO are quite similar with scores close to 1. We examined the similarity of individual SUMO proteomes and transcriptomes using principal component analysis (PCA) (Supplementary Figure 4-1 c and d). PCA generated four distinct groups from the SUMO proteome dataset and five distinct groups from the transcriptome dataset. Each group encompasses biological triplicates of significantly regulated substrates or genes upon each PIAS KO. These results highlight the diversity of regulation on both a SUMOylation and transcription level by different PIASs.

Next, we analyzed the substrates that are significantly regulated by SUMOylation upon PIAS KO. To ensure that changes in SUMOylation patterns were not attributed to variation of protein expression upon PIAS KO, we also analyzed the total proteome of each PIAS triple SILAC sample. In addition, the proteome dataset was used to define the fold change cut-off value for SUMOylated targets. We performed one sample t-test with a *P*-value threshold of 0.05 using Benjamini-Hochberg correction on the proteome dataset to determine the significant changes on the protein level upon each PIAS KO. We selected 2.5% on both up-regulated and down-regulated proteins and used the corresponding fold change (FC) as a cut-off value to define the significant changes on the SUMO proteome level. In total, we quantified 1422 SUMO peptides across all 12 samples (Supplementary Table 4-1). After one sample student's t-test and FC cut-off on the x-axis as shown in Figure 4-3a, PIAS2 KO led to more up-regulated than down-regulated SUMOylation, whereas the other three PIAS KO samples showed the opposite trend. In general, PIAS3 KO had the greatest impact on regulation with 174 downregulated and 82 upregulated SUMO sites. PIAS2 regulated 204 SUMOylation sites, of which 87 were downregulated and 117 were upregulated. PIAS1 and PIAS4 regulated similar numbers of SUMOylation sites, with PIAS1 downregulating 116 and upregulating 55 sites, and PIAS4 downregulating 129 and upregulating 44 sites. Interestingly, among all regulated SUMOylation sites, 30 sites were found to be common to all PIAS proteins and at least 67 sites were regulated by two different PIAS proteins (Figure 4-3d). These results indicate that PIAS SUMO E3 ligases share common and unique substrates thus highlighting the functional diversity within the PIAS-family members.
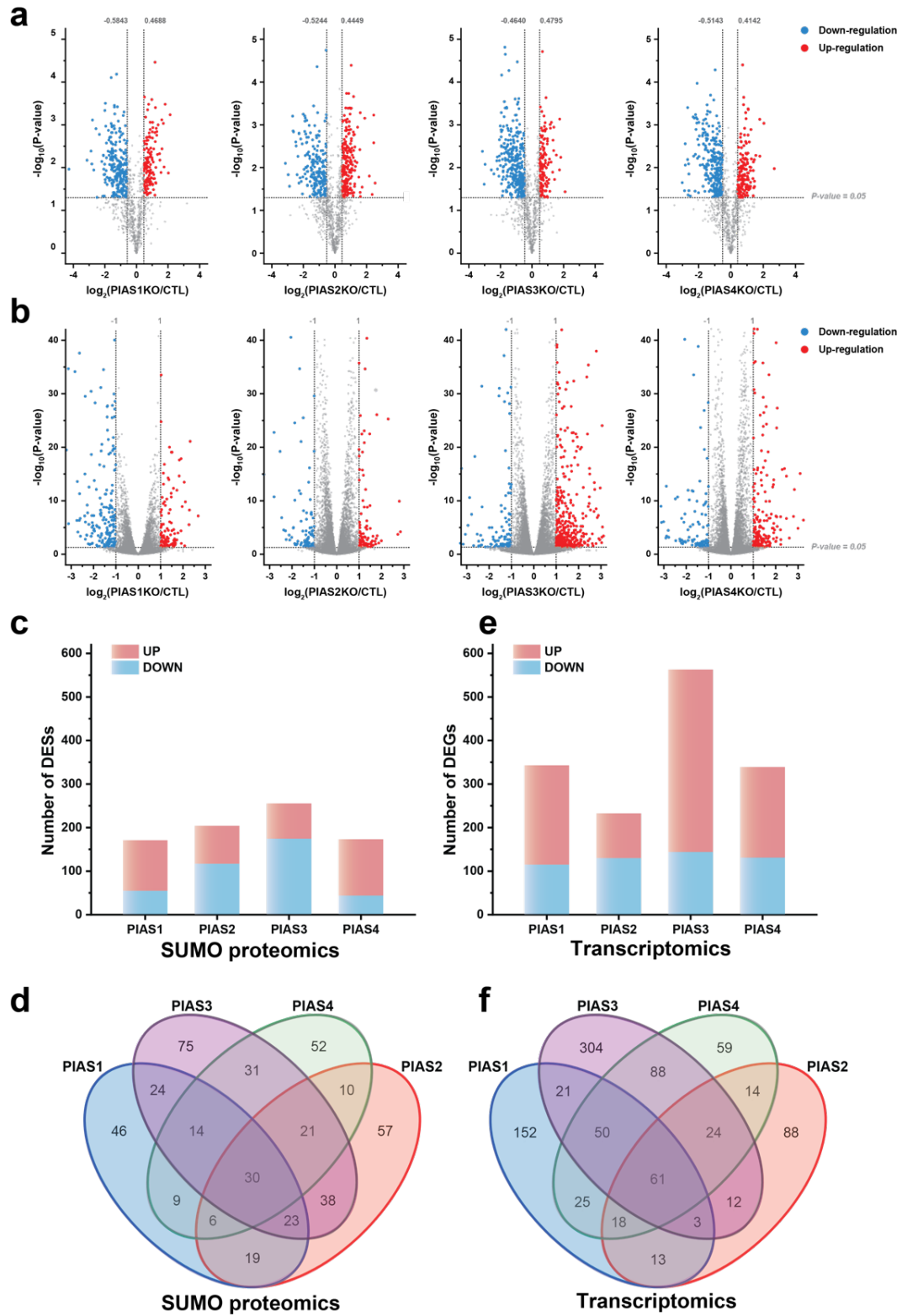
Figure 4-3 Results of SUMO proteomics and transcriptomics.

(a) Volcano plots showing the global SUMOylation changes in PIAS KO cells vs. control cells where dots represent individual SUMO peptides. The significantly regulated SUMOylation sites upon PIAS KO are highlighted in blue for down-regulation and in red for up-regulation. Fold change cut-off values are defined based on FDR 5% on proteome changes. Benjamini Hochberg corrected p-value of <0.05. (b) Volcano plots showing the global transcriptome changes in PIAS KO cells vs. control cells, where dots represent individual genes. The significantly regulated genes upon PIAS KO are highlighted in blue for down-regulation and in red for up-regulation. Log_2(Fold change) ≥ 1 or log_2(Fold change) ≤ -1 are applied on the x-axis. Benjamini Hochberg corrected p-value of <0.05. (c) Bar plots showing the numbers of differentially regulated SUMOylation sites upon individual PIAS KO. Down-regulated and up-regulated SUMOylation sites are highlighted in blue and in red, respectively. (d) Bar plots showing the numbers of differentially expressed genes upon individual PIAS KO. Down-regulated and up-regulated genes are highlighted in blue and in red, respectively. (e) Venn diagram showing common and unique SUMOylation sites affected by specific PIAS KO. (e) Venn diagram showing the common and unique genes regulated by specific PIAS KO.

For the transcriptomic data, we sequenced an average of 20M reads per sample, which gave approximately 70% overlapping genes across all sequenced biological triplicates including 4 PIAS KO samples and KO CTL samples. Further analysis showed that over 32,000 genes were covered across all PIAS KO samples. Significantly regulated genes were defined by a fold change ≥ 2 after normalization and a *P*-value threshold of 0.05 (Figure 4-3b, Supplementary Table 4-2—4-5). Interestingly, the number of down-regulated genes was comparable across PIAS proteins, ranging from 115 to 144; however, the number of upregulated genes differed greatly. Gene up-regulation upon PIAS1-4 KO ranged from 103 for PIAS2 KO to as many as 419 for PIAS3 KO, while PIAS1 and PIAS4 KO resulted in 228 and 208 gene up-regulations, respectively. Next, we compared the overlap of regulated genes by different PIASs (Figure 4-3f). Only six genes were found to be regulated by all PIAS proteins, with at least 117 genes regulated by two different PIASs. Additionally, PIAS3 regulated 304 unique genes while PIAS4 only regulated 59 unique genes.

To understand if PIAS regulated gene transcription and protein SUMOylation on a similar pool of candidates, we compared the differentially expressed SUMOylated proteins (DESPs) and differentially expressed genes (DEGs) upon PIAS KO. Intriguingly, the overlap between DESPs and DEGs was limited, only three common candidates, including SUMO3, MAFA and PROX1 (Supplementary Figure 4-2). This result suggested that important changes in regulation exist at

the post-transcriptional level. We also compared our data with those observed from protein microarray-based studies [25]. In total, 18 DESPs including SUMO1-4, UBR7, RBBP7, ATRX and SSRP1 were common to both studies. The low overlap observed here reflects the variation in protein expression and sensitivity of the different approaches used.

4.4.5 Gene Ontology term and pathway enrichment analysis of DESPs and DEGs

To gain further insight into the biological processes regulated by different PIAS SUMO E3 ligases on both SUMO proteomic and transcriptomic data, we performed a PANTHER classification analysis on DESPs and DEGs (Supplementary Figure 4-3). All identified DESPs and DEGs were grouped into categories according to their biologically relevant functions. Proteins involved in "gene-specific transcriptional regulator (PC00264)", "nucleic acid metabolism protein (PC00171)", "chromatin/chromatin-binding, or -regulatory protein (PC00077)" and "translational protein (PC00263)" regroup 87% of DESPs. However, with the exception of "gene-specific transcriptional regulator (PC00264)", 73% of DEGs are involved in entirely different categories, such as "metabolite interconversion enzyme (PC00262)", "protein modifying enzyme (PC00260)", "transporter (PC00227)", "intercellular signal molecule (PC00207)", "transmembrane signal receptor (PC00197)" and "protein-binding activity modulator (PC00095)".
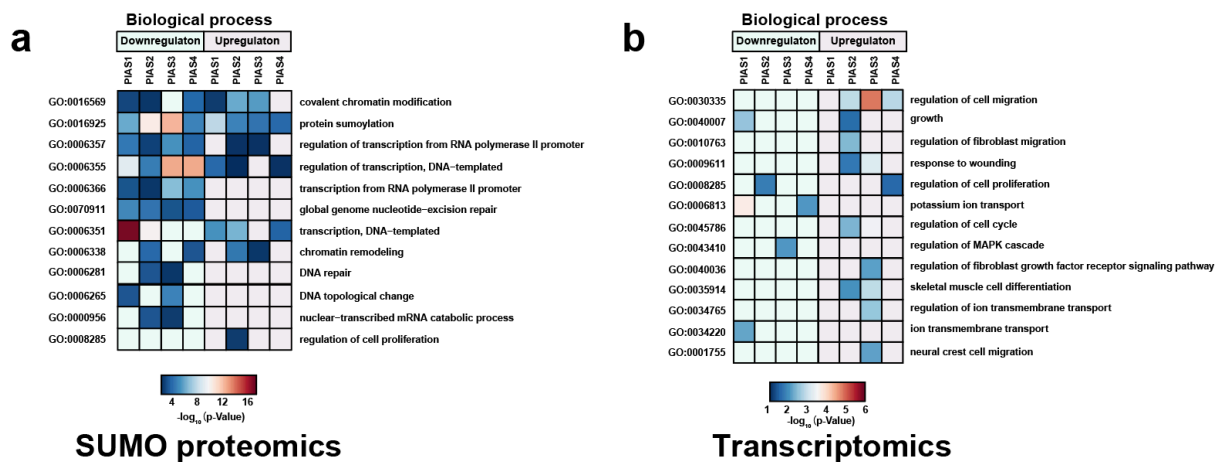

Figure 4-4 Gene Ontology (GO) term enrichment analysis of biological processes.

Heatmaps showing the enriched GO terms of biological processes for down-regulated and up-regulated candidates upon PIAS KO in (a) SUMO proteomics and (b) transcriptomics experiments. Downregulated candidates are shaded in light blue, while upregulated candidates are shaded in pink. The color intensity indicates the -log$_{10}$(p-Value) of enrichment.

To identify biologically relevant functions regulated by PIASs, we performed Gene Ontology (GO) term enrichment analysis of the downregulated and upregulated candidates from the integrative analyses (Figure 4-4 and Supplemental Figure 4-4). The heatmap in Figure 4-4 shows the enriched GO terms of biological processes of DESPs and DEGs from each PIAS KO. The GO terms "transcription regulation" and "protein SUMOylation" were significantly enriched in the SUMO proteome datasets across all PIAS members, suggesting that different PIASs may have similar regulatory mechanisms through gene transcription by regulating substrate activity or stability through SUMOylation. GO term of "regulation of cell proliferation" was enriched in DESPs for PIAS2, and DEGs for PIAS2 and PIAS4, respectively. In addition, GO terms related to cell migration were also enriched in DEGs of several PIASs. The GO terms "regulation of fibroblast migration" was enriched in PIAS2 DEGs, "response to wounding" was enriched in PIAS2 and PIAS3 DEGs, and lastly "regulation of cell migration" was enriched in DEGs of PIAS2, PIAS3, and PIAS4. The results of GO term enrichment for molecular functions and cellular compartment are shown in Supplementary Figure 4-4, and showed that distinct pathways were enriched for each PIAS at both mRNA and SUMO proteome levels.

The distinct GO term enrichment from the transcriptomic data raises the question of whether PIAS ablation has a specific or global effect on gene expression. To shed light on this question, we next compared the chromosomal distribution of the DEGs and their corresponding locations in the genome (Supplementary Figure 4-5). We found that all PIAS members regulate gene expression mainly through the protein coding regions, lncRNA, and retained intron. However, the relative portion of regulated regions differ significantly. For example, the protein coding and lncRNA contribute up to ~75% of the total regulation region (~40% and ~35% respectively). The retained intron (~18%) and nonsense mediated decay parts (~8%) make up the rest of the regulated regions. No significant differences in the proportion of regulated regions were observed between upregulated and downregulated genes though individual PIAS appeared

to affect the expression of specific genes (Supplementary Figure 4-5). Interestingly, we found that several transcription factors and repressors, including FOSL2, RLF, SNIP1 and ZBTB34, were SUMOylated by all PIAS proteins. For example, SUMOylation at K30 and K108 on the transcription repressor SNIP1 reduced its inhibitory effect on the expression of TGF-β target genes PAI-1 and MMP2, leading to an increase in TGF-β-regulated cell migration [43]. We thus suspect that PIAS mediated protein SUMOylation on transcription factors regulated specific gene expression on different chromosomes [44].

4.4.6 Regulation networks analysis

The phenotypic assays in Figure 4-1 c-e showed that PIAS KO reduced cell proliferation and migration of MDA-MB-231 cell to varying degree. To further understand the redundant and specific regulatory mechanism, we generated a regulation network incorporating the DESPs and DEGs using STRING and visualized by Cytoscape. In this network, each node represents a SUMO substrate, or a regulated gene which is connected to the corresponding PIAS. The node size represents the number of connections each substrate interacts with. As illustrated in Supplementary Figure 4-5, each PIAS has its own unique profile of protein SUMOylation and gene regulation, which forms a highly connected cluster within itself and between other PIAS clusters. Intriguingly, the connections between regulated proteins and genes within each cluster are weak, suggesting that distinct regulatory mechanisms are operative at transcriptional and post-translational levels.

Next, we investigated the commonly regulated networks in the center to better understand the functional redundancy among PIAS SUMO E3 ligases shown in Figure 4-5. Surprisingly, we found several highly connected interactors regulated by all PIAS members at either protein SUMOylation or transcriptional level. Among these commonly regulated networks, two groups of DESPs or DEGs were found to be involved in either cell proliferation, such as MKI67, TOP1 and TOP2A, or cell migration (e.g. KIF17). Interestingly, between these two groups, there are also several DESPs or DEGs involved in both biological processes (e.g. IGFBP3, DPYSL3, NRG3 and PDGFD). Also, we found that PIAS ligases regulated proteins involved in transcriptional

regulation, RNA processing, signal transduction, RNA processing and chromatin remodeling. In addition, several SUMO/Ubiquitin machinery enzymes were regulated by all PIAS proteins, which suggest a crosstalk between SUMO and ubiquitin involving PIAS SUMO E3 ligases.
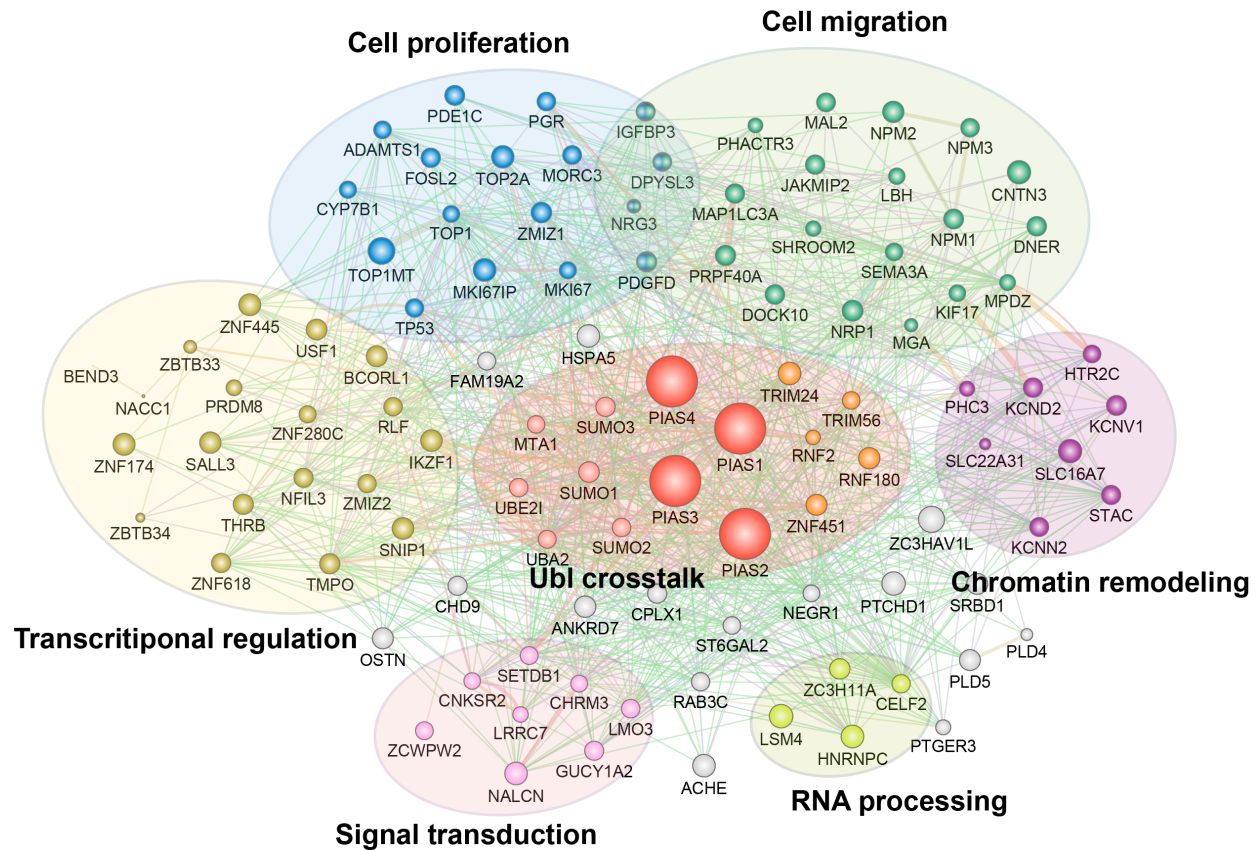

Figure 4-5 Common regulation network by all PIAS proteins.

STRING network showing the commonly regulated DESPs and DEGs by all PIAS proteins along with their interacting partners. Sub-networks are grouped according to their GO terms.

4.4.7 Regulation patterns by PIAS on SUMOylation level and transcription level

To establish trends between protein targets affected upon PIAS KO we generated a heatmap displaying the hierarchical clustering of differentially regulated SUMOylated proteins. In total, we identified 530 SUMOylated proteins, of which 30 were regulated by all four PIASs (Figure 4-6a). To extract groups of substrates specific to each PIAS, we also used fuzzy C-means (FCM) clustering and generate profiles of regulated clusters. We found four clusters of SUMO peptides that

displayed a specific downregulation upon individual PIAS KO. In addition, we also identified a cluster of substrates where PIAS1 and PIAS3 have similar down regulation pattern while PIAS2 and PIAS4 exhibit the opposite regulation trend. Interestingly, each cluster were found to contain unique regulated substrates involved in cell proliferation and migration (Figure 4-6b). For example, PIAS4 may regulate cell proliferation and migration through the SUMOylation of CBX3 and BRD4 [45-47]. These results indicate the regulation specificity of each PIAS SUMO E3 ligase on a sub-pool of substrates.
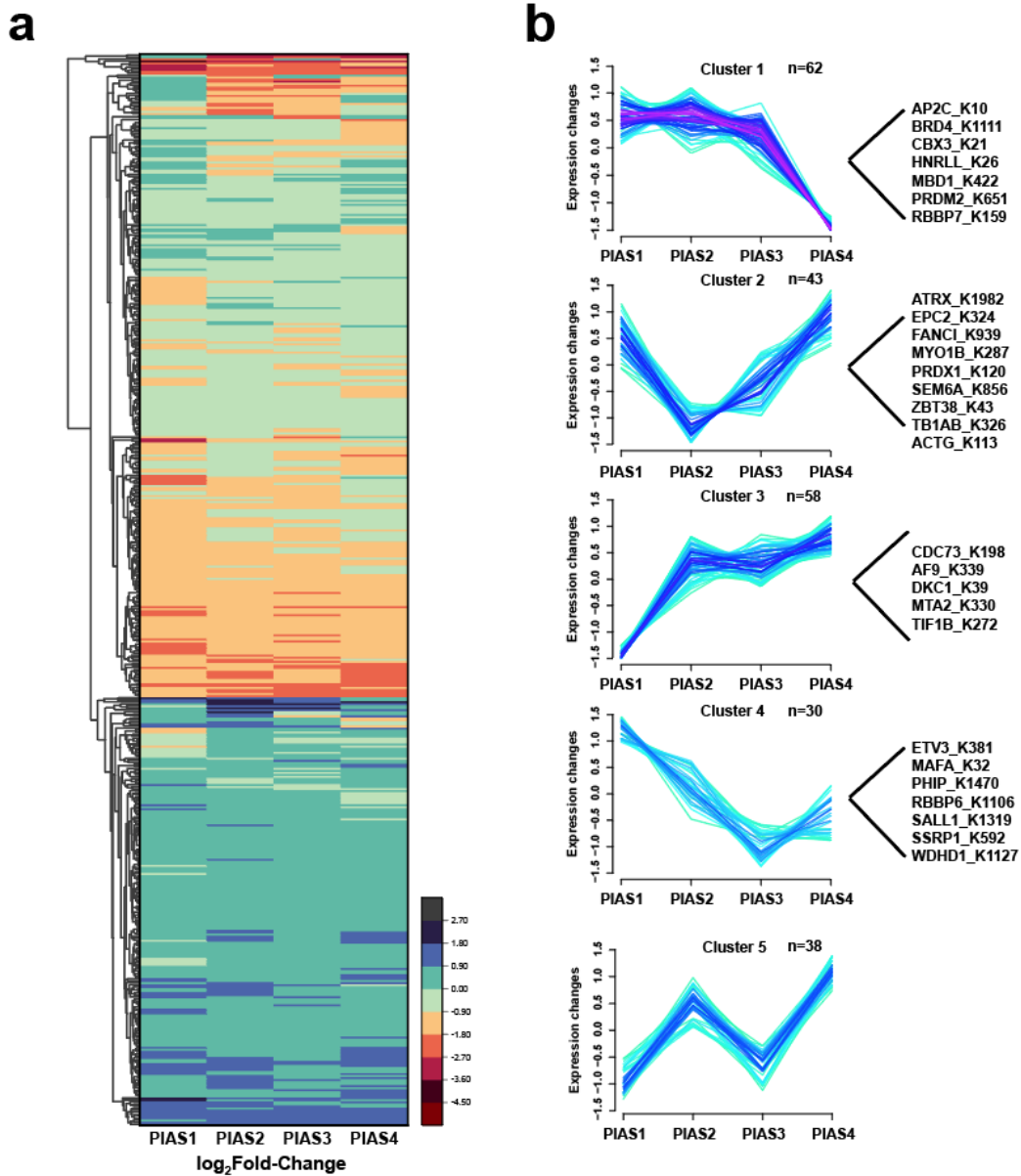
Figure 4-6 Regulation profiling of protein SUMOylation by different PIAS.

(a) Hierarchical clustering of normalized fold change for SUMOylation sites across different PIAS proteins. (b) SUMOylation sites were further clustered into six different groups using Fuzzy-C-means clustering based on their regulation profiles. The unique regulated candidates which are involved in cell proliferation and migration are listed beside each cluster.

Next, we evaluated the patterns of regulated transcripts upon individual PIAS KO (Supplementary Figure 4-7). The heatmap showed 932 regulated transcript profiles, 61 of which were common to all PIAS KO. Surprisingly, unlike the unique regulation patterns found in the SUMO proteomics dataset, the majority of clusters were found to be shared between two PIASs for down-regulation and two PIASs for up-regulation. For example, cluster 2 contains 146 transcripts down-regulated by PIAS1 and PIAS4, and up-regulated by PIAS2 and PIAS3. These results suggest that regulatory mechanisms on a SUMO level may be more specific to each individual PIAS; however, regulation on a transcriptional level appears to be more complementary between PIAS members.

4.4.8 Evolutionary conservation analysis of identified SUMO sites

All residues in a given protein are not equally important, and certain lysine residues may significantly affect protein function [48-50]. The rate of conservation is one of the most widely used approach to predict functionally important residues in protein sequences by comparing its occurrence in orthologues through evolution [51]. Although a large number of SUMOylated sites and their corresponding SUMO E3 ligases have been reported in various eukaryotes by different groups [52-54], the conservation of SUMOylated lysine sites in these organisms remains unclear. To determine the evolutionary conservation of the modified lysine residues, we used ProteoConnections [40] and compared all identified SUMOylated sites across twelve different species (Supplementary Table 4-6). We further classified the identified SUMO sites into three conservation levels based on the extent to which a site is conserved across species. Lysines conserved in more than 8 species were designated as having a "High" conservation rate. Lysines conserved in 5 to 8 or in 4 and fewer species were defined as "Medium" or "Low" conservation rate, respectively. As shown in Supplementary Figure 4-8a, the identified SUMOylome comprised

233

approximately 23% of highly conserved SUMO sites and 60% of medium conserved SUMO sites. In contrast, the regulated SUMO sites by PIAS1, PIAS2 and PIAS3 showed a greater proportion of highly conserved sites with a decreased percentage of sites defined as medium. The conservation rate of PIAS4 substrates was predominantly medium. Moreover, the percentage of PIAS4-regulated sites with a high conservation rate was the lowest compared to the other PIAS. In addition, the percentage of sites classified as having a low conservation rate was comparable across all PIAS-regulated sites. Of note, PIAS3 regulated almost 40% of SUMO sites which had a high conservation rate while only 45% of SUMO sites had a medium conservation rate, indicating a contrasting trend of conservation compared to the regulated sites of other PIAS SUMO E3 ligases. These observations suggest that although most SUMOylated lysines have at least medium conservation rate, they vary across different PIAS SUMO E3 ligases.

Next, we looked at substrates of each PIAS ligase that display a high conservation level to understand their regulation specificity. Of note, each PIAS ligase may also regulate cell proliferation and migration through very distinct SUMOylated substrates (Figure 4-7). We found that PIAS1 SUMOylate CDC73 at K198, a site located on the N-terminus of its β-Catenin binding domain. PIAS1 also SUMOylated MTA2 at K330 on the C-terminus of its SANT chromatin remodeling domain. CDC73 was reported to be involved in cell cycle and cell growth, and MTA2 was found to promote cell migration and invasion in gastric cancer cells [55]. It is worth noting that PIAS2 uniquely regulates SUMOylation on Actin, Tubulin and Myosin which may affect the cytoskeleton organization and further affect the cell proliferation, morphology and motility [56]. PIAS3 mediated SUMOylation of SALL1 at K1319 was reported to affect its nuclear localization and further regulate target gene expression [57]. We also identified CBX3 K21 as target of PIAS4. CBX3 recognizes methylated K9 histone H3 tails, leading to epigenetic repression, and its SUMOylation could affect its transcriptional silencing activity.
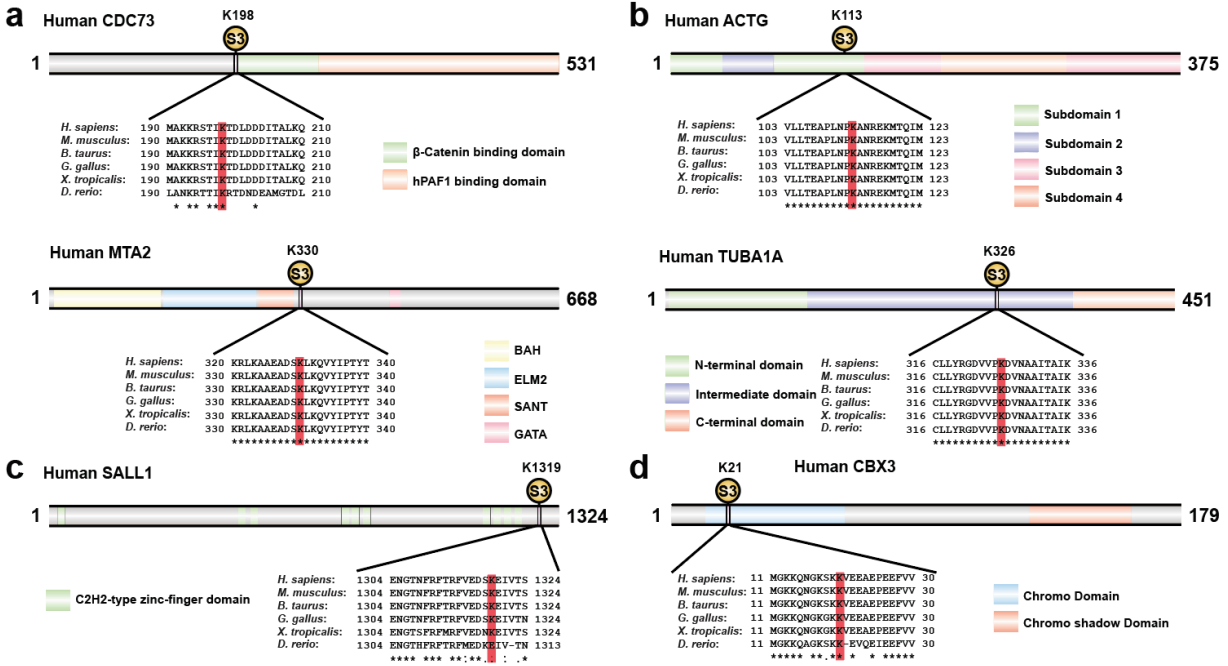
Figure 4-7 Evolutionary conservation analysis of unique substrates of each PIAS ligase.
(a) PIAS1 mediated SUMOylation of CDC73 at K198 and MTA2 at K330. (b) PIAS2 mediated SUMOylation of ACTG at K113 and TUB1A1 at K326. (c) PIAS3 mediated SUMOylation of SALL1 at K1319. (d) PIAS4 mediated SUMOylation of CBX3 at K21.

Furthermore, we assessed the regulation specificity of PIAS SUMO E3 ligases by selecting the substrates with highly conserved lysine residues that are uniquely regulated by each PIAS ligase (Supplementary Figure 4-8b). Surprisingly, although the regulated lysine residues differ across all PIAS substrates, there is a better overlap at the protein level which further confirms the regulation specificity of PIAS proteins throughout their SUMO E3 ligase activity. Interestingly, a closer examination of specific substrates revealed that each PIAS ligase also regulate unique biological processes through different substrates. For example, PIAS1 regulates DNA damage repair through multiple SUMOylation sites on PARP1 [58]. PIAS2 regulates chromosome remodeling through SUMOylating several sites on multiple histones [59]. PIAS3 regulates SUMO2, SUMO3 and SUMO4 on the lysine at position 33, required for polySUMO chain formation [60, 61].

4.5 Discussion

As SUMO E3 ligases, the Protein Inhibitor of Activated STAT (PIAS) family members have been reported to play multiple roles in regulating several cellular processes, such as protein stability, signal transduction, and DNA repair [1]. Several studies have also demonstrated that PIAS-family members participate in tumorigenesis and cancer-related processes, specifically in the processes of cellular proliferation and migration [17, 24, 27, 62]. However, most of these studies are based on individual investigation between a given PIAS protein and their potential targets, resulting in a lack of understanding of PIAS regulatory mechanisms in a system-wide manner. In addition, PIAS protein members share several conserved domains, making it more challenging to identify and discriminate redundancy and specificity across all PIAS members. The first global identification of PIAS SUMO substrates was performed using the HuProt™ microarray-based proteomic analysis in tandem with an *in vitro* SUMOylation assay [25, 29]. Although this approach has provided a general profile of PIAS SUMO E3 ligase substrate specificity, the in vitro SUMO condition increases the probability of false positive substrate identification. Previously, we have reported that overexpression of PIAS1 not only promotes HeLa cell proliferation but also cell migration. By applying a large-scale SUMO proteomics approach, we profiled the regulated SUMO substrates of PIAS1 [14]. In the present study, we extended our analysis to all human PIAS members and investigated their influence on cellular proliferation and migration in the triple-negative breast cancer cell line MDA-MB-231. By taking advantage of CRISPR/Cas-based gene knockout technology, we performed a parallel comparison of different PIAS-mediated regulatory mechanisms. Furthermore, the subsequent analysis using our modified SUMO proteomics approach enabled the discovery of low abundant SUMO peptides under basal conditions, and thus provided a comprehensive profiling of PIAS-regulated substrates in a site-specific manner.

In this study, we successfully quantified more than 1400 SUMO peptides under basal condition with the aid of a booster channel containing MG132-treated samples. Through the analysis of common and unique substrates, we highlighted the redundancy and specificity of PIAS proteins. The shared SUMO substrates identified in our data set revealed a certain degree of SUMO E3 ligase functional redundancy among PIAS members, particularly at the protein level.

Interestingly, the proliferation marker protein Ki-67 (MKI67) was found highly SUMOylated under basal condition and regulated by all PIAS members on 7 different sites (K1035, K1093, K2009, K2613, K2734, K2852, and K2967). This protein plays a key role in cell proliferation and is commonly used as a marker of cell proliferation [63, 64]. SUMOylation on Ki-67 may affect its subcellular localization during cell cycle progression [65]. Our observation may explain why PIAS KO negatively affects cell proliferation; however, the role of PIAS mediated SUMOylation on Ki-67 function is unclear. In addition, TOP1 and TOP2A were also SUMOylated by all PIASs, potentially impacting protein stabilization, leading to an inhibition of p53 and increased cell proliferation [66, 67]. Besides, other identified PIAS-mediated SUMO substrates include Nucleus accumbens-associated protein 1 (NACC1) [68, 69], Fos-related antigen 2 (FOSL2) [70, 71], Nucleophosmin (NPM1) [72], and Pre-mRNA-processing factor 40 homolog A (PRPF40A) [73], all of which play a role in either cell proliferation, cell migration/morphology, or cytoskeleton organization, however, more functional studies need to be conducted to assess the role of SUMOylation on those proteins. The importance of our site-specific approach is emphasized in our data of different PIAS SUMO E3 ligases that target unique sites on the same substrates. Intriguingly, we also identified unique mechanism of regulation for each PIAS protein involved in cell proliferation and cell migration, which may explain why different PIAS KO led to the reduction of MDA-MB-231 cell proliferation and migration to different degrees. Although there are several studies showing SUMO paralog preferences by PIAS members, for instance, PIAS4 preference for SUMO2 [74]; our approach is limited to substrates modified by the SUMO3 paralog due to our modified cell line.

PIAS family members are also known as transcriptional coregulators, thus it is necessary to analyze the PIAS-regulated transcriptome in an effort to gain a comprehensive understanding of PIAS-mediated regulation. Therefore, we also performed RNA-seq analysis on total mRNA of each PIAS knockout cell line to compare the regulatory mechanisms on a transcription level by different PIAS members and identify gene components involved in cell proliferation and migration. Transcriptomic data analysis revealed common DEGs upon different PIAS gene knockout involved in cell proliferation and migration, providing novel insights into the molecular mechanisms of these processes. Interestingly, PIAS members share a certain number of common

DEGs, suggesting a degree of regulatory redundancy in addition to what was observed in the SUMO-proteome experiments. We found that all PIAS regulated the gene expression of PDGFD [75], NRG3 [76], and DPYSL3 [77], ADAMTS1 [78], IGFBP3 [79] and LBH [80], all of which have been reported to be involved in cell proliferation. Additionally, IGFBP3 and LHB are also involved in the mitogen-activated protein kinase (MAPK) signaling pathway, which may suggest PIAS-mediated crosstalk between phosphorylation and SUMOylation [81, 82]. Among the commonly regulated genes, several were associated with cytoskeleton organization and cell migration, such as IGFBP3 [83], SEMA3A [84], DOCK10 [85], PDGFD [86], and SHROOM2 [87]. Interestingly, we also found that IGFBPS, PDGFD, NRG3, and DPYSL3 are involved in both cell proliferation and migration, all of which were shown to be regulated by all PIAS members. These and other DEGs demonstrate the role of PIAS SUMO E3 ligases in regulating cell proliferation and migration on transcription level, illustrating a novel aspect of PIAS-mediated regulation.

Although there are extensive changes in the transcriptome upon PIAS KO, the corresponding changes were not observed at the SUMO proteome level. This is in line with findings that describe the modest correlation between protein and mRNA levels [88, 89]. In addition, protein expression is also reported to be influenced by post-transcriptional regulatory mechanisms, making it necessary to analyze PIAS KO cells at both the transcriptomic and proteomic levels to gain a systems-level understanding. We integrated RNA-seq based transcriptomic analysis into a quantitative SUMO proteomics approach to investigate the regulation networks by PIAS members and elucidate specific regulatory mechanisms and the redundancy in cell proliferation and migration. In conclusion, this integrative approach of SUMO proteomic and transcriptomic analyses revealed multiple levels of regulation by PIAS members in cell proliferation and cell migration, providing a valuable data resource for further understanding the extent of regulation by PIAS SUMO E3 ligases.

4.6 Acknowledgements

4.7 Supplementary Figures
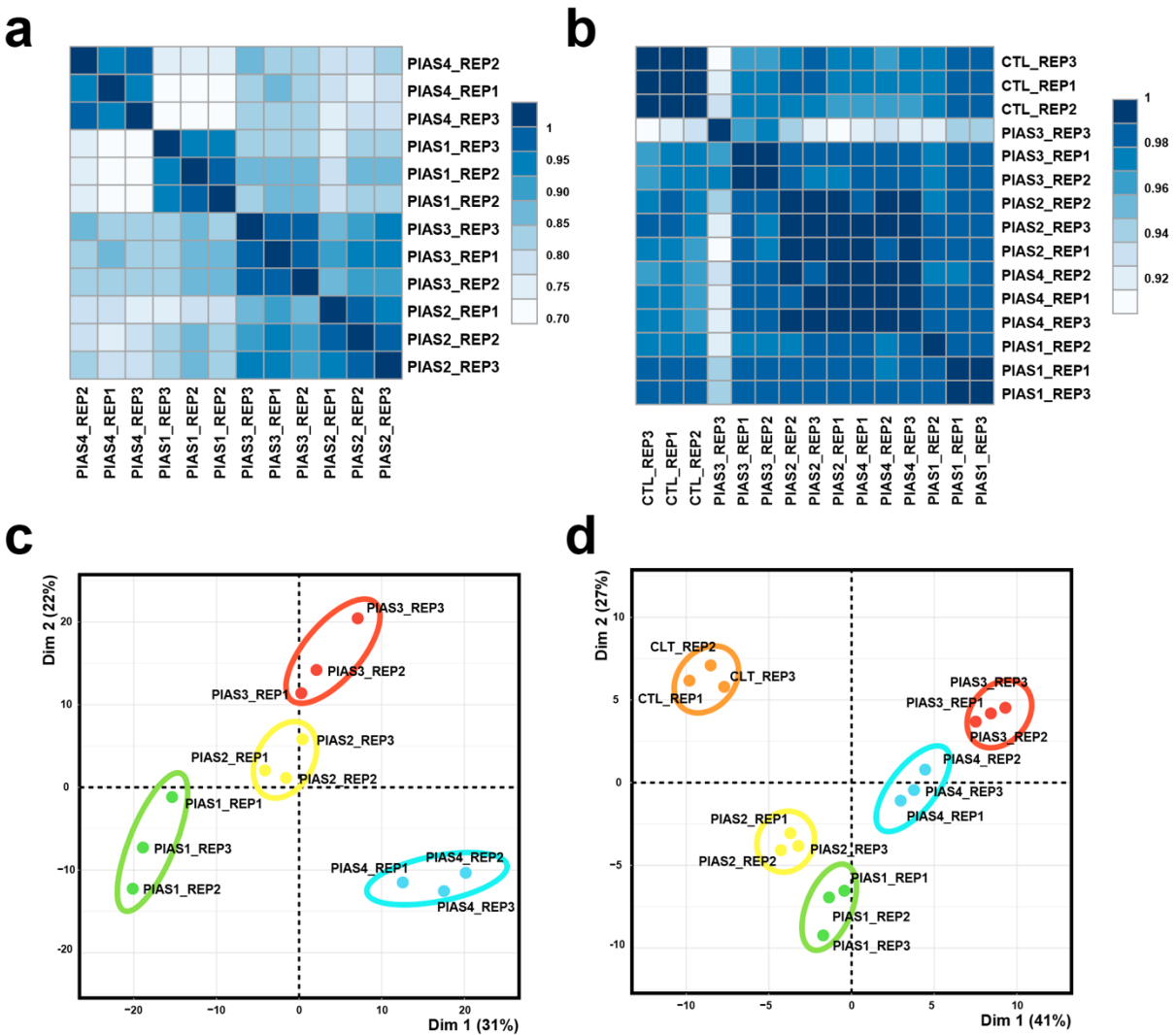
Supplementary Figure 4-1 Pearson's correlation coefficient and principal component analysis (PCA). Heatmaps showing the magnitude of the matrix among (a) DESPs regulated by each PIAS ligase and (b) DEGs upon each PIAS knockout along with the knockout control cells. PCA attributes the first two most significant components among (c) DESPs regulated by each PIAS ligase and (d) DEGs upon each PIAS knockout along with the knockout control cells.
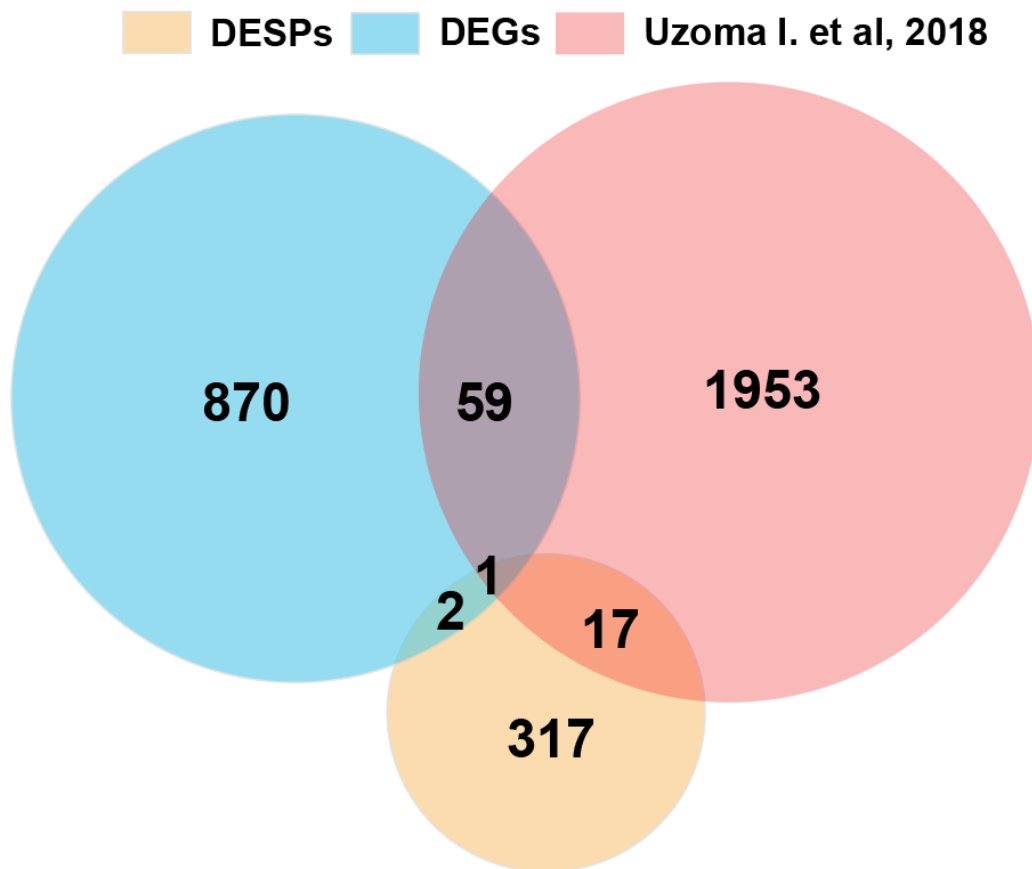
Supplementary Figure 4-2 Comparison of DESPs, DEGs and PIAS substrates identified from protein microarray. Venn diagram showing the comparison of total DESPs (highlighted in orange), total DEGs (highlighted in blue) identified in this study and PIAS substrates identified by protein microarray reported by Uzoma I. et al in 2018 (highlightedin pink).
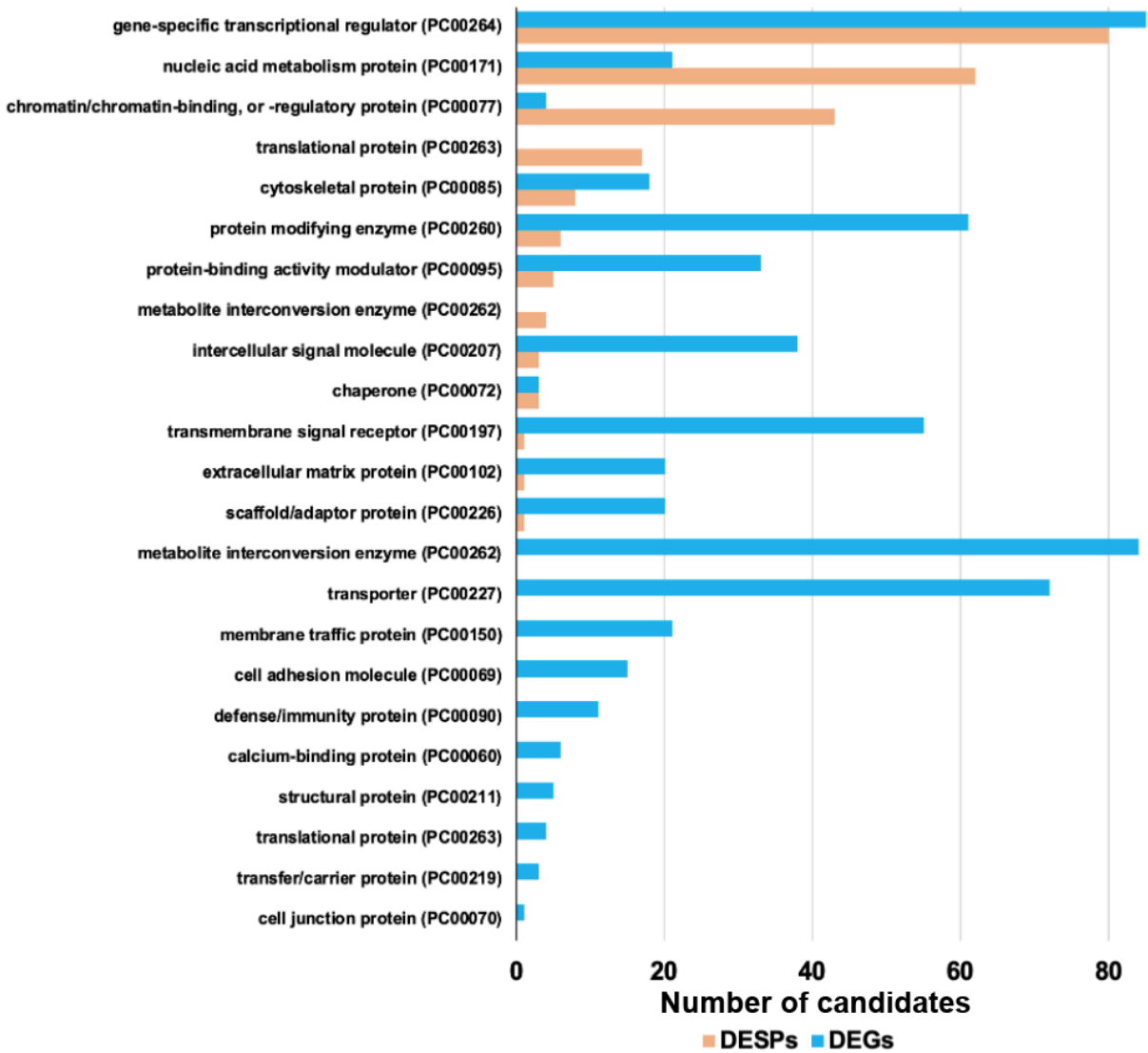
Supplementary Figure 4-3 PANTHER classification analysis.

Bar plot showing the protein categories that total DESPs (highlighted in orange) and total DEGs (highlighted in blue) are involved in.
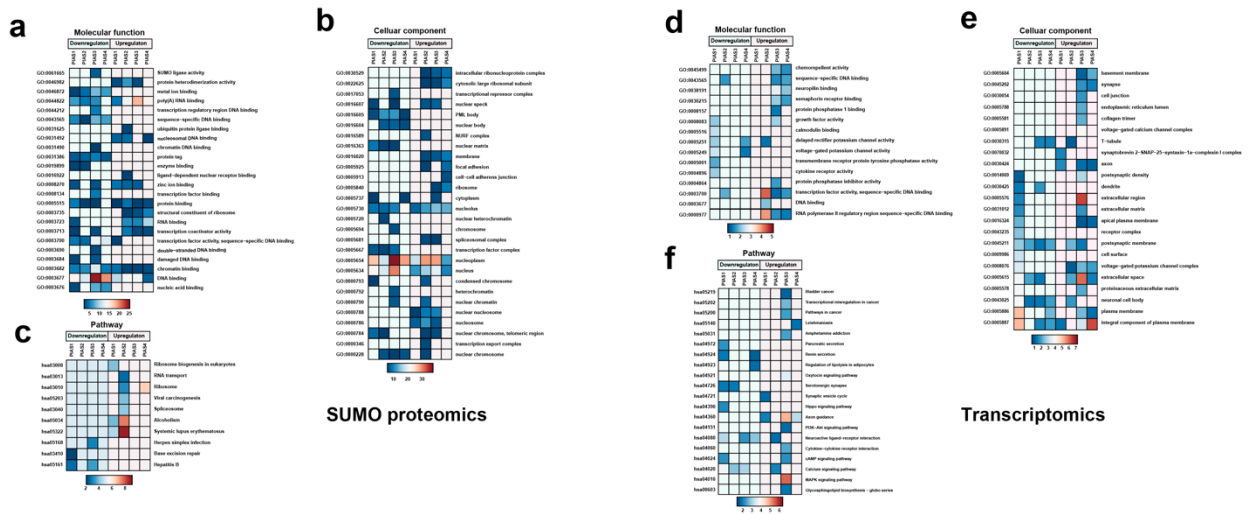
Supplementary Figure 4-4 Gene Ontology (GO) term enrichment analysis of DESPs and DEGs by each PIAS protein.

Heatmaps showing the enriched terms: (a) Molecular function, (b) Cellular component, and (c) Pathway are generated from SUMO proteomics analysis. (d) Molecular function, (e) Cellular component, and (f) Pathway are generated from transcriptomics analysis. Down-regulated candidates are shaded in light blue, while up-regulated candidates are shaded in pink. The colour intensities indicate the -log$_{10}$(p-Value) of enrichment.
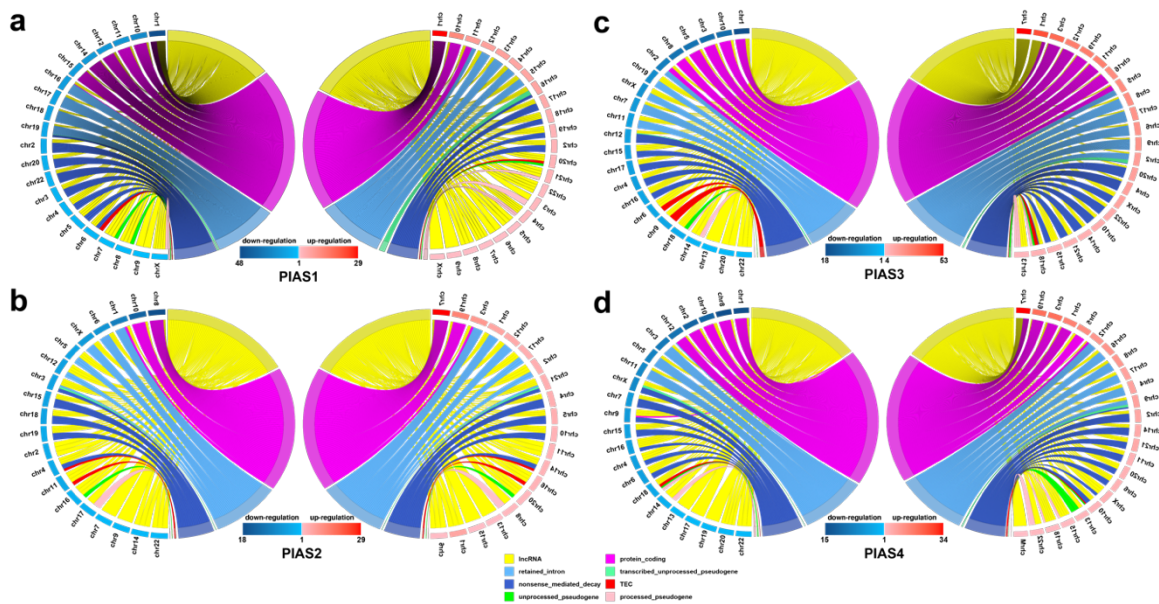
Supplementary Figure 4-5 PIAS-regulated gene distributions on the chromosome.

Circle plot showing regulated-gene distribution by (a) PIAS1, (b) PIAS2, (c) PIAS3, and (d) PIAS4. The chromosomes are labelled with different numbers highlighted in blue for the down-regulated gene distribution and red for up-regulated gene distribution for each panel. The colour intensities indicate the number of genes identified on a certain chromosome. Additionally, the different regulation regions on the genes are highlighted in yellow (lncRNA), purple (protein_coding), light blue (retained_intron), cyan (transcribed_unprocessed_pseudogene), dark blue (nonsense_mediated_decay), red (TEC), green (unprocessed_pseudogege), and pink (processed_pseudogene). The size of different color schemes represents the percentage of the corresponding regulation region on the genes.
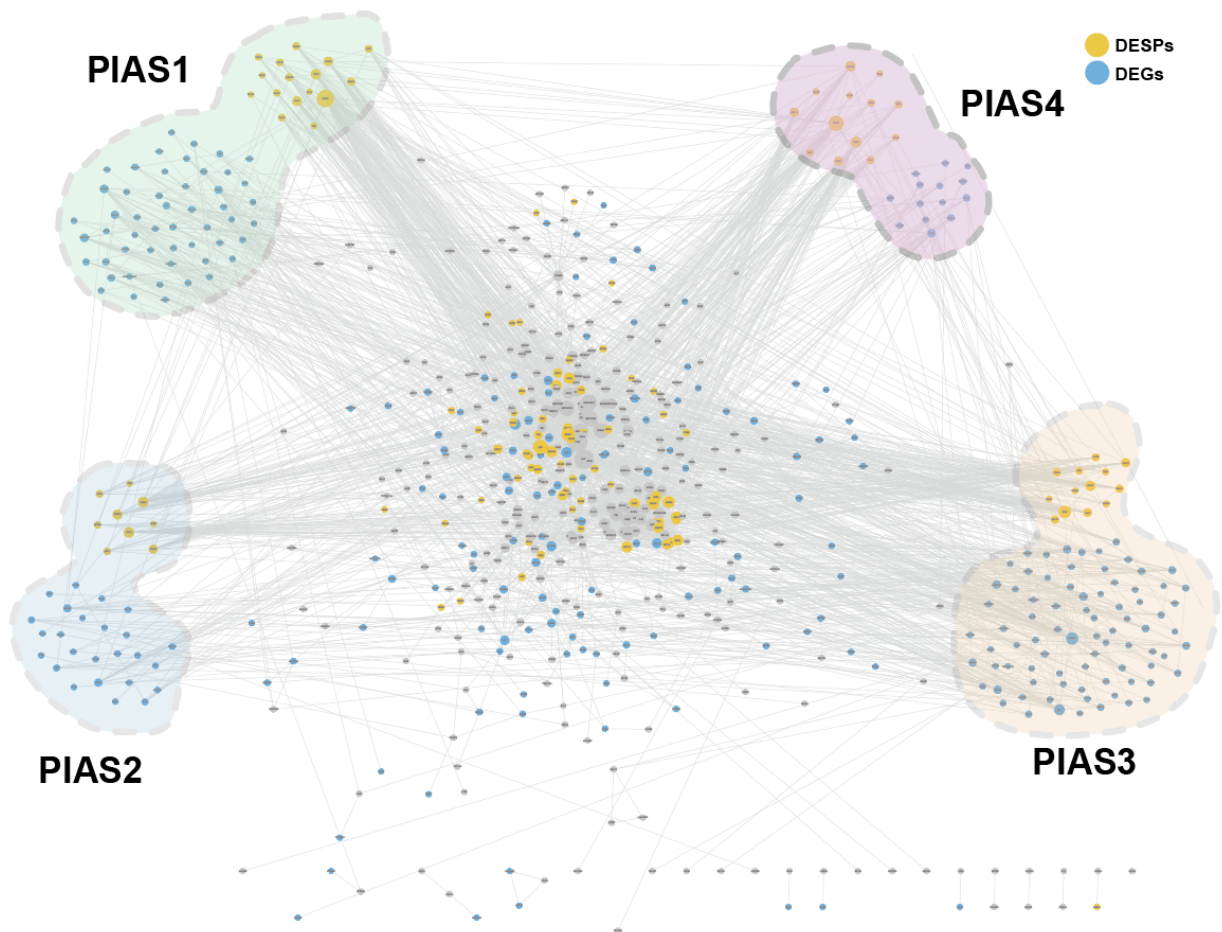
Supplementary Figure 4-6 PIAS-mediated regulation network.

STRING network showing the connections between each PIAS and the corresponding regulated candidates highlighted in orange on SUMOylation level and in red on transcription level. The colored edges depict the connection to an upstream PIAS protein. Many candidates are connected to more than one PIAS protein, thus revealing the overlap and redundancy between different PIAS proteins.
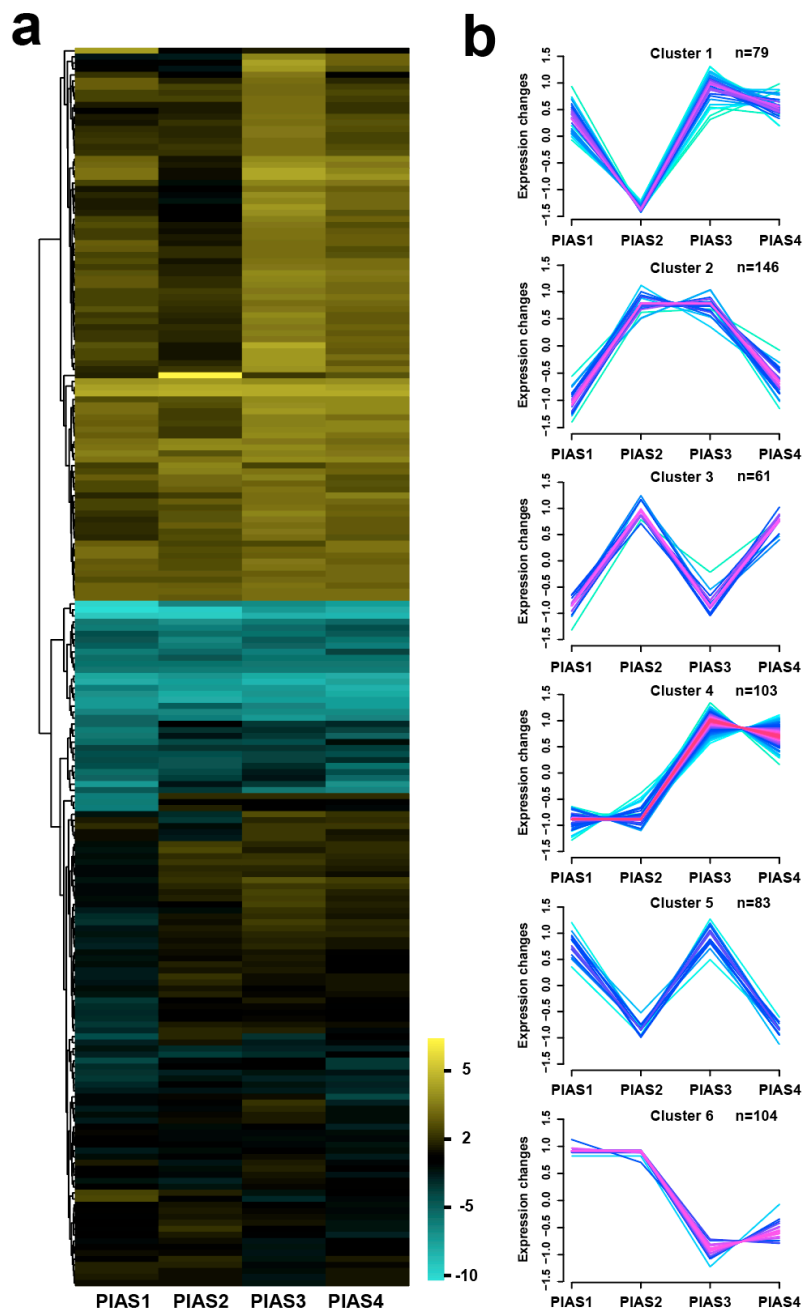
Supplementary Figure 4-7 Regulation profiling of the genes by different PIAS proteins.

(a) Clustered heatmap of normalized fold change for regulated genes across different PIAS proteins. (b) Genes were further clustered into six different groups using Fuzzy-C-means clustering based on their regulation profiles.
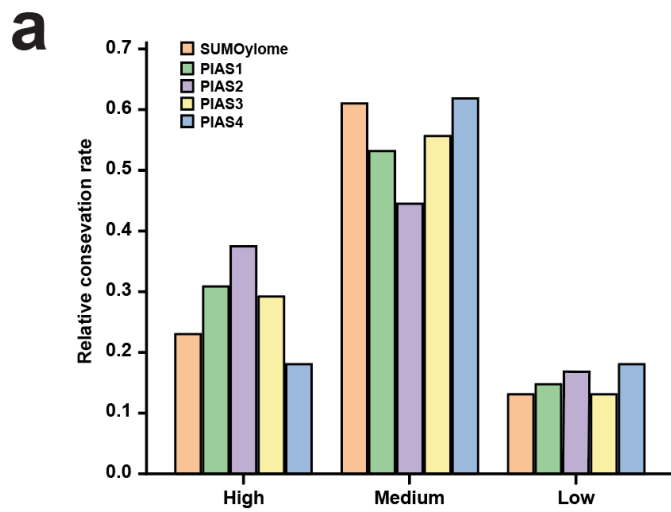
PIAS1	PIAS2	PIAS3	PIAS4
SMRCD_K724	H2B1D_K121	RS3A_K249	TOP2B_K605
CHD4_K1304	H4_K9	EZH2_K634	ZN724_K199
EAF6_K69	H32_K15/K24/K80	SF3B1_K413	IKZF5_K108
MTA2_K330	H2AY_K123	RL13_K174	SNUT1_K188
PARP1_K203/K249/K331	H2B3B_K6	TOP2B_K1307	XPC_K183
LARP1_K311	H2A1C_K96	SMCA5_K966	CHD4_K1606
UBP28_K99	H2B1K_K21	ZIC2_K253	ARI3A_K398
BRPF3_K457	SF3B1_K298	RL5_K220	RL1D1_K435
NOP58_K37	TBA1B_K326	SUMO3_K32	PSMD8_K298
SAP30_K214	H2B2F_K21	SUMO2_K33	BMS1_K415
ZBT11_K760	RBBP4_K160	EXOS9_K297	KCTD1_K252
HNRPM_K685	ACTG_K113	NOLC1_K647	PRDM2_K651
	RS17_K103	SUMO4_K33	AP2C_K10
	GDIA_K269	PIAS2_K443	

Supplementary Figure 4-8 Evolutionary conservation analysis of identified SUMO sites.

(a) Bar plot showing the percentage of DESPs regulated by each PIAS protein in each conservation level.

(b) The table listing the unique DESPs and their corresponding sites from high conservation level regulated by each PIAS protein.

4.8 References

1.   Celen, A.B. and U. Sahin, *Sumoylation on its 25th anniversary: mechanisms, pathology, and emerging concepts.* FEBS J, 2020. **287**(15): p. 3110-3140.

2.   Gareau, J.R. and C.D. Lima, *The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition.* Nat Rev Mol Cell Biol, 2010. **11**(12): p. 861-71.

3.   Chung, C.D., et al., *Specific inhibition of Stat3 signal transduction by PIAS3.* Science, 1997. **278**(5344): p. 1803-5.

4.   Liu, B., et al., *Inhibition of Stat1-mediated gene activation by PIAS1.* Proc Natl Acad Sci U S A, 1998. **95**(18): p. 10626-31.

5.   Sachdev, S., et al., *PIASy, a nuclear matrix-associated SUMO E3 ligase, represses LEF1 activity by sequestration into nuclear bodies.* Genes Dev, 2001. **15**(23): p. 3088-103.

6.   Liao, J., Y. Fu, and K. Shuai, *Distinct roles of the NH2- and COOH-terminal domains of the protein inhibitor of activated signal transducer and activator of transcription (STAT) 1 (PIAS1) in cytokine-induced PIAS1-Stat1 interaction.* Proc Natl Acad Sci U S A, 2000. **97**(10): p. 5267-72.

7.   Gross, M., et al., *Distinct effects of PIAS proteins on androgen-mediated gene activation in prostate cancer cells.* Oncogene, 2001. **20**(29): p. 3880-3887.

8.   Shuai, K., *Modulation of STAT signaling by STAT-interacting proteins.* Oncogene, 2000. **19**(21): p. 2638-44.

9.   Schmidt, D. and S. Muller, *PIAS/SUMO: new partners in transcriptional regulation.* Cell Mol Life Sci, 2003. **60**(12): p. 2561-74.

10.   Rogers, R.S., C.M. Horvath, and M.J. Matunis, *SUMO modification of STAT1 and its role in PIAS-mediated inhibition of gene activation.* J Biol Chem, 2003. **278**(32): p. 30091-7.

11.   Johnson, E.S. and A.A. Gupta, *An E3-like factor that promotes SUMO conjugation to the yeast septins.* Cell, 2001. **106**(6): p. 735-44.

12.   Rytinki, M.M., et al., *PIAS proteins: pleiotropic interactors associated with SUMO.* Cell Mol Life Sci, 2009. **66**(18): p. 3029-41.

13. Rabellino, A., C. Andreani, and P.P. Scaglioni, *The Role of PIAS SUMO E3-Ligases in Cancer.* Cancer Res, 2017. **77**(7): p. 1542-1547.

14. Li, C., et al., *Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility.* Nat Commun, 2020. **11**(1): p. 834.

15. Wang, L. and S. Banerjee, *Differential PIAS3 expression in human malignancy.* Oncol Rep, 2004. **11**(6): p. 1319-24.

16. Toropainen, S., et al., *SUMO ligase PIAS1 functions as a target gene selective androgen receptor coregulator on prostate cancer cell chromatin.* Nucleic Acids Res, 2015. **43**(2): p. 848-61.

17. Liu, B., et al., *PIAS1 regulates breast tumorigenesis through selective epigenetic gene silencing.* PLoS One, 2014. **9**(2): p. e89464.

18. Driscoll, J.J., et al., *The sumoylation pathway is dysregulated in multiple myeloma and is associated with adverse patient outcome.* Blood, 2010. **115**(14): p. 2827-34.

19. Li, J., et al., *PIAS3, an inhibitor of STAT3, has intensively negative association with the survival of gastric cancer.* Int J Clin Exp Med, 2015. **8**(1): p. 682-9.

20. Xiong, Q., et al., *Identification of novel miR-21 target proteins in multiple myeloma cells by quantitative proteomics.* J Proteome Res, 2012. **11**(4): p. 2078-90.

21. Sun, L., et al., *PIASy mediates hypoxia-induced SIRT1 transcriptional repression and epithelial-to-mesenchymal transition in ovarian cancer cells.* J Cell Sci, 2013. **126**(Pt 17): p. 3939-47.

22. Chien, W., et al., *PIAS4 is an activator of hypoxia signalling via VHL suppression during growth of pancreatic cancer cells.* Br J Cancer, 2013. **109**(7): p. 1795-804.

23. Zhou, S., et al., *PIASy represses CCAAT/enhancer-binding protein delta (C/EBPdelta) transcriptional activity by sequestering C/EBPdelta to the nuclear periphery.* J Biol Chem, 2008. **283**(29): p. 20137-48.

24. Castillo-Lluva, S., et al., *SUMOylation of the GTPase Rac1 is required for optimal cell migration.* Nat Cell Biol, 2010. **12**(11): p. 1078-85.

25.   Uzoma, I., et al., *Global Identification of Small Ubiquitin-related Modifier (SUMO) Substrates Reveals Crosstalk between SUMOylation and Phosphorylation Promotes Cell Migration.* Mol Cell Proteomics, 2018. **17**(5): p. 871-888.

26.   Gonzalez-Prieto, R., et al., *c-Myc is targeted to the proteasome for degradation in a SUMOylation-dependent manner, regulated by PIAS1, SENP7 and RNF4.* Cell Cycle, 2015. **14**(12): p. 1859-72.

27.   Wu, R., et al., *SUMOylation of the transcription factor ZFHX3 at Lys-2806 requires SAE1, UBC9, and PIAS2 and enhances its stability and function in cell proliferation.* J Biol Chem, 2020. **295**(19): p. 6741-6753.

28.   Yan, Y., et al., *SUMOylation of AMPKalpha1 by PIAS4 specifically regulates mTORC1 signalling.* Nat Commun, 2015. **6**: p. 8979.

29.   Cox, E., et al., *Identification of SUMO E3 ligase-specific substrates using the HuProt human proteome microarray.* Methods Mol Biol, 2015. **1295**: p. 455-63.

30.   Lamoliatte, F., et al., *Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification.* Nat Commun, 2017. **8**: p. 14109.

31.   Bolger, A.M., M. Lohse, and B. Usadel, *Trimmomatic: a flexible trimmer for Illumina sequence data.* Bioinformatics, 2014. **30**(15): p. 2114-20.

32.   Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

33.   Li, B. and C.N. Dewey, *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome.* BMC Bioinformatics, 2011. **12**: p. 323.

34.   Varet, H., et al., *SARTools: A DESeq2- and EdgeR-Based R Pipeline for Comprehensive Differential Analysis of RNA-Seq Data.* PLoS One, 2016. **11**(6): p. e0157022.

35.   Feuermann, M., et al., *Large-scale inference of gene function through phylogenetic annotation of Gene Ontology terms: case study of the apoptosis and autophagy cellular processes.* Database-the Journal of Biological Databases and Curation, 2016: p. 11.

36.   Mi, H.Y., et al., *PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements.* Nucleic Acids Research, 2017. **45**(D1): p. D183-D189.

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

38.   Huang, D.W., et al., *DAVID Bioinformatics Resources: expanded annotation database and novel algorithms to better extract biology from large gene lists.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W169-75.

39.   Cui, Y., et al., *BioCircos.js: an interactive Circos JavaScript library for biological data visualization on web applications.* Bioinformatics, 2016. **32**(11): p. 1740-2.

40.   Courcelles, M., et al., *ProteoConnections: a bioinformatics platform to facilitate proteome and phosphoproteome analyses.* Proteomics, 2011. **11**(13): p. 2654-71.

41.   Cox, J. and M. Mann, *MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification.* Nat Biotechnol, 2008. **26**(12): p. 1367-72.

42.   Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

43.   Liu, S., et al., *SUMO Modification Reverses Inhibitory Effects of Smad Nuclear Interacting Protein-1 in TGF-beta Responses.* J Biol Chem, 2016. **291**(47): p. 24418-24430.

44.   Rosonina, E., et al., *Regulation of transcription factors by sumoylation.* Transcription, 2017. **8**(4): p. 220-231.

45.   Wang, L., et al., *BRD4 inhibition suppresses cell growth, migration and invasion of salivary adenoid cystic carcinoma.* Biol Res, 2017. **50**(1): p. 19.

46.   Zhang, C., et al., *Cbx3 inhibits vascular smooth muscle cell proliferation, migration, and neointima formation.* Cardiovasc Res, 2018. **114**(3): p. 443-455.

47.   Tan, Y.F., et al., *Inhibition of BRD4 prevents proliferation and epithelial-mesenchymal transition in renal cell carcinoma via NLRP3 inflammasome-induced pyroptosis.* Cell Death Dis, 2020. **11**(4): p. 239.

48.   Creton, S. and S. Jentsch, *SnapShot: The SUMO system.* Cell, 2010. **143**(5): p. 848-848 e1.

49.   Escobar-Ramirez, A., et al., *Modification by SUMOylation Controls Both the Transcriptional Activity and the Stability of Delta-Lactoferrin.* PLoS One, 2015. **10**(6): p. e0129965.

50.   Santos, A., et al., *SUMOylation affects the interferon blocking activity of the influenza A nonstructural protein NS1 without affecting its stability or cellular localization.* J Virol, 2013. **87**(10): p. 5602-20.

51.   Urena, E., et al., *Evolution of SUMO Function and Chain Formation in Insects.* Mol Biol Evol, 2016. **33**(2): p. 568-84.

52.   Li, M., et al., *TRIM28 functions as the SUMO E3 ligase for PCNA in prevention of transcription induced DNA breaks.* Proceedings of the National Academy of Sciences of the United States of America, 2020. **117**(38): p. 23588-23596.

53.   Yang, W.S., et al., *In Vitro SUMOylation Assay to Study SUMO E3 Ligase Activity.* J Vis Exp, 2018(131).

54.   Yan, Y.L., et al., *DPPA2/4 and SUMO E3 ligase PIAS4 opposingly regulate zygotic transcriptional program.* PLoS Biol, 2019. **17**(6): p. e3000324.

55.   Zhou, C., et al., *MTA2 promotes gastric cancer cells invasion and is transcriptionally regulated by Sp1.* Mol Cancer, 2013. **12**(1): p. 102.

56.   Alonso, A., et al., *Emerging roles of sumoylation in the regulation of actin, microtubules, intermediate filaments, and septins.* Cytoskeleton (Hoboken), 2015. **72**(7): p. 305-39.

57.   Sanchez, J., et al., *Sumoylation modulates the activity of Spalt-like proteins during wing development in Drosophila.* J Biol Chem, 2010. **285**(33): p. 25841-9.

58.   Zilio, N., et al., *DNA-dependent SUMO modification of PARP-1.* DNA Repair (Amst), 2013. **12**(9): p. 761-73.

59.   Cubenas-Potts, C. and M.J. Matunis, *SUMO: a multifaceted modifier of chromatin structure and function.* Dev Cell, 2013. **24**(1): p. 1-12.

60.   Hendriks, I.A., et al., *Site-specific characterization of endogenous SUMOylation across species and organs.* Nat Commun, 2018. **9**(1): p. 2456.

61.   Theurillat, I., et al., *Extensive SUMO Modification of Repressive Chromatin Factors Distinguishes Pluripotent from Somatic Cells.* Cell Rep, 2020. **33**(1): p. 108251.

62.   Hoefer, J., et al., *PIAS1 is increased in human prostate cancer and enhances proliferation through inhibition of p21.* Am J Pathol, 2012. **180**(5): p. 2097-107.

63. Gerdes, J., et al., *Production of a mouse monoclonal antibody reactive with a human nuclear antigen associated with cell proliferation.* Int J Cancer, 1983. **31**(1): p. 13-20.

64. Dowsett, M., et al., *Assessment of Ki67 in breast cancer: recommendations from the International Ki67 in Breast Cancer working group.* J Natl Cancer Inst, 2011. **103**(22): p. 1656-64.

65. Sun, X. and P.D. Kaufman, *Ki-67: more than a proliferation marker.* Chromosoma, 2018. **127**(2): p. 175-186.

66. Li, F., et al., *Topoisomerase I (Top1): a major target of FL118 for its antitumor efficacy or mainly involved in its side effects of hematopoietic toxicity?* Am J Cancer Res, 2017. **7**(2): p. 370-382.

67. Liu, T., et al., *Mutual regulation of MDM4 and TOP2A in cancer cell proliferation.* Mol Oncol, 2019. **13**(5): p. 1047-1058.

68. Nakayama, K., et al., *A BTB/POZ protein, NAC-1, is related to tumor recurrence and is essential for tumor growth and survival.* Proc Natl Acad Sci U S A, 2006. **103**(49): p. 18739-44.

69. Nakayama, K., et al., *NAC-1 controls cell growth and survival by repressing transcription of Gadd45GIP1, a candidate tumor suppressor.* Cancer Res, 2007. **67**(17): p. 8058-64.

70. Xu, P., et al., *Hsa_circ_0001869 promotes NSCLC progression via sponging miR-638 and enhancing FOSL2 expression.* Aging (Albany NY), 2020. **12**.

71. Ji, C., et al., *Circ_0091581 Promotes the Progression of Hepatocellular Carcinoma Through Targeting miR-591/FOSL2 Axis.* Dig Dis Sci, 2020.

72. Liu, X., et al., *Nucleophosmin (NPM1/B23) interacts with activating transcription factor 5 (ATF5) protein and promotes proteasome- and caspase-dependent ATF5 degradation in hepatocellular carcinoma cells.* J Biol Chem, 2012. **287**(23): p. 19599-609.

73. Bai, S.W., et al., *Identification and characterization of a set of conserved and new regulators of cytoskeletal organization, cell morphology and migration.* BMC Biol, 2011. **9**: p. 54.

74. Azuma, Y., et al., *PIASy mediates SUMO-2 conjugation of Topoisomerase-II on mitotic chromosomes.* EMBO J, 2005. **24**(12): p. 2172-82.

75.   Ray, S., et al., *Platelet-derived growth factor D, tissue-specific expression in the eye, and a key role in control of lens epithelial cell proliferation.* J Biol Chem, 2005. **280**(9): p. 8494-502.

76.   Panchal, H., et al., *Neuregulin3 alters cell fate in the epidermis and mammary gland.* BMC Dev Biol, 2007. **7**: p. 105.

77.   Matsunuma, R., et al., *DPYSL3 modulates mitosis, migration, and epithelial-to-mesenchymal transition in claudin-low breast cancer.* Proc Natl Acad Sci U S A, 2018. **115**(51): p. E11978-E11987.

78.   Xu, Z., Y. Yu, and E.J. Duh, *Vascular endothelial growth factor upregulates expression of ADAMTS1 in endothelial cells through protein kinase C signaling.* Invest Ophthalmol Vis Sci, 2006. **47**(9): p. 4059-66.

79.   Ingermann, A.R., et al., *Identification of a novel cell death receptor mediating IGFBP-3-induced anti-tumor effects in breast and prostate cancer.* J Biol Chem, 2010. **285**(39): p. 30233-46.

80.   Liu, Q., et al., *Limb-bud and Heart Overexpression Inhibits the Proliferation and Migration of PC3M Cells.* J Cancer, 2018. **9**(2): p. 424-432.

81.   Mayo, J.C., et al., *IGFBP3 and MAPK/ERK signaling mediates melatonin-induced antitumor activity in prostate cancer.* J Pineal Res, 2017. **62**(1).

82.   Ai, J., et al., *A human homolog of mouse Lbh gene, hLBH, expresses in heart and activates SRE and AP-1 mediated MAPK signaling pathway.* Mol Biol Rep, 2008. **35**(2): p. 179-87.

83.   Yen, Y.C., et al., *Insulin-like growth factor-independent insulin-like growth factor binding protein 3 promotes cell migration and lymph node metastasis of oral squamous cell carcinoma cells by requirement of integrin beta1.* Oncotarget, 2015. **6**(39): p. 41837-55.

84.   Gehler, S., F.V. Compere, and A.M. Miller, *Semaphorin 3A Increases FAK Phosphorylation at Focal Adhesions to Modulate MDA-MB-231 Cell Migration and Spreading on Different Substratum Concentrations.* Int J Breast Cancer, 2017. **2017**: p. 9619734.

85.   Shin, S., et al., *ERK2 regulates epithelial-to-mesenchymal plasticity through DOCK10-dependent Rac1/FoxO1 activation.* Proc Natl Acad Sci U S A, 2019. **116**(8): p. 2967-2976.

86.   Wagsater, D., et al., *Effects of PDGF-C and PDGF-D on monocyte migration and MMP-2 and MMP-9 expression.* Atherosclerosis, 2009. **202**(2): p. 415-23.

87.   Yuan, J., et al., *SHROOM2 inhibits tumor metastasis through RhoA-ROCK pathway-dependent and -independent mechanisms in nasopharyngeal carcinoma.* Cell Death Dis, 2019. **10**(2): p. 58.

88.   Schwanhausser, B., et al., *Global quantification of mammalian gene expression control.* Nature, 2011. **473**(7347): p. 337-42.

89.   Wu, L., et al., *Variation and genetic control of protein abundance in humans.* Nature, 2013. **499**(7456): p. 79-82.

CHAPTER FIVE

5 Conclusions and Perspectives

5.1 Conclusions

Since their discovery in the mid-1990s, SUMO-family proteins have received a high degree of attention due to their various roles in cellular events [1]. As a PTM, SUMOylation is highly dynamic and reversibly modifies lysine residues on target proteins through a covalent bond [2]. The identification of SUMO substrates provides valuable insights to improve understanding of the biological functions in which SUMO proteins are involved [3]. The development of MS-based high throughput technology further extends the repertoire of SUMOylated substrates [4].

In this thesis, I reviewed different proteomic strategies for Ubl identifications developed in the past two decades which advanced the comprehensive discovery of Ubl substrates. As SUMO identification by mass spectrometry is more challenging due to its amino acid sequence at the C terminus, I also discussed the different strategies for SUMO identification in particular [5-10]. Among these strategies, the immunoisolation approach using the exogenous expression of SUMO mutants in the cells followed by mass spectrometry identification allows us to obtain precise SUMO site information [10, 11]. Our lab has developed a workflow using a two-step purification of SUMOylated substrates [12], which makes it possible to identify SUMOylation substrates by MS in a site-specific manner. By integrating metabolic labeling and gene expression changes using molecular biology tools into this workflow, I further extended its application into the identification of SUMO E3 ligase substrates.

In the first study, I observed that expression changes of SUMO E3 ligase PIAS1 had physiological impacts on HeLa cell proliferation and migration [13]. To further understand the regulatory mechanism of cell proliferation and migration by PIAS1, I combined triple SILAC labelling and PIAS1 overexpression with a previously developed large-scale SUMO proteomic workflow. To identify the *bona fide* PIAS1 substrates, I investigated the global SUMOylome changes quantitatively upon PIAS1 overexpression in basal condition. In 2016, I quantified 983 SUMO peptides on 544 SUMO proteins without the inhibition of proteosome degradation or any

other cell stimulation, which marks a milestone in SUMO site profiling. Intriguingly, overexpression of PIAS1 significantly upregulated global SUMOylome changes. A total of 91 SUMOylation sites on 62 proteins were identified as PIAS1 substrates including the known substrate, promyelocytic leukemia (PML) protein. A few substrates, including NSMCE2, PFDN2 and VIM were further validated *in vitro* or in cells using western blots. The bioinformatic analysis of PIAS1 substrates indicated that PIAS plays roles in a variety of biological processes, including protein stabilization, protein sumoylation and protein folding. The protein-protein interaction network of putative PIAS1 substrates highlighted the presence of highly connected interactors from PML nuclear bodies, transcription factors, cytoskeletal proteins and RNA binding proteins. Of note, PIAS1-mediated SUMOylation occurred primarily on solvent exposed lysine residues in non-consensus regions of the target proteins. In addition, PIAS1 was also identified to be SUMOylated at five different positions. Through site-directed mutagenesis of three modified lysines located within the PINIT domain or next to the SP-RING domain, immunofluorescence reveals that the interaction between PML-nuclear bodies and PIAS1 is SUMO-SIM dependent, and partly mediated by the SUMOylation of PIAS1 at Lys-137, Lys-238 and Lys-315 residues.

Interestingly, this study also indicated that PIAS1 regulated a pool of substrates which were associated with cytoskeletal organization including actin filaments, microtubules and several intermediate filament proteins. Among those identified intermediate filament proteins, vimentin was reported to play key roles in cell proliferation, migration and contractility. The two lysine residues located on the tail domain of vimentin were highly conserved across different species and were SUMOylated by PIAS1. Hence, I further investigated the function of PIAS1-mediated SUMOylation of vimentin. By conducting site-directed mutagenesis of those two lysine residues, I evaluated the migration ability of cells expressing mutant vimentin and studied the solubility and organization of vimentin intermediate filaments. All together, these results showed that PIAS1-mediated SUMOylation of vimentin increased its solubility and facilitated the maturation process of vimentin intermediate filaments. This dynamic assembly and disassembly of vimentin intermediate filaments thus regulated cell migration and motility.

The first study has proven the successful application of this modified workflow to the identification of SUMO E3 ligase substrates. However, as the human PIAS family has four members, PIAS1, PIAS2, PIAS3 and PIAS4, which share various conserved domains, it is more challenging to identify and discriminate regulation redundancy and specificity across all PIAS members. In addition to SUMO E3 ligase activity, PIAS-family members have also been reported to play multiple roles in gene transcription regulation. Thus, in the following studies, the functional evaluation of cell proliferation and migration were extended to all PIAS members. In breast cancer cell line MDA-MB-231, the individual PIAS gene knockout using CRIPSR/Cas9 gene-editing technology affected cell morphology and led to the reduction of cell proliferation and migration to a different degree. To generate a better understanding of the regulatory mechanisms across all PIAS members, I integrated an RNA-seq based transcriptomics analysis into the SUMO proteomics analysis upon individual PIAS KO by CRISPR. The integrative analyses allowed the comparison of different PIAS-mediated regulations at both the SUMOylation level as well as the gene transcription level. The quantitation of SUMO peptides was further improved in this study to a total number of 1422, among which, PIAS ligases shared a pool of commonly regulated substrates, while approximately 30% of substrates were unique to an individual PIAS ligase. At the transcription level, PIAS members regulated 61 gene transcripts in common and each PIAS also regulated a unique pool of gene transcripts. These results confirmed the regulation redundancy and were in line with the phenotypic assays. The genes uniquely regulated by each PIAS ligase predict the degree to which cell proliferation and migration decreased in MDA-MB-231 cells upon PIAS KO. Of note, the overlap between regulated SUMO candidates and gene candidates is very poor, which indicated the supplementary importance of this transcriptomic analysis. Among these identified candidates, Ki-67, FOSL2 and Nucleophosmin were involved in cell proliferation, cell migration, in the regulation of cell morphology and cytoskeletal organization, and were commonly regulated by all PIAS members at the SUMOylation level. We have successfully determined the regulation pattern of vimentin SUMOylation across all PIAS members by PRM-MS targeted analysis. Although the regulation patterns of these common substrates and unique substrates identified in this study were not yet validated, the established approach will soon be performed for this purpose. Apart from SUMOylation regulation,

ADAMTS1, Protein LBH, Semaphorin-3A, DPYSL3, SHROOM2, and DOCK10 were found to be involved in cell proliferation, migration, cytoskeleton orgnization and are regulated by all PIAS at the transcriptional level.

Throughout the above summarized studies, quantitative SUMO proteomics combined with gene overexpression or depletion have been successfully applied to the discovery of SUMO E3 ligase substrates. The integrative analyses of SUMO proteomics and transcriptomics further aid the understanding of PIAS-mediated regulatory mechanisms on cell proliferation and migration via a trans-omics view. These omics analyses enable the unraveling of key biological events in which SUMO E3 ligase is involved on a systems biology level.

5.2 Perspectives

Identification of SUMO E3 ligase substrates using a proteomics approach has been improved rapidly in the past decade [13, 14]. The development of gene editing technology further promotes the application of SUMO proteomics study in a wider range of research fields, specifically in the profiling of SUMO E3 ligase substrates [15-17]. However, due to the nature of the SUMOylation modification and the limitations of my workflow, there are many questions that cannot be easily answered. Thanks to the rapid development of cutting-edge MS technologies, significant improvement in proteome coverage is observed with each generation of instruments. Thus, future improvements on sample preparation workflow and data analysis can be further archived to gain more insights into the roles of SUMO E3 ligase.

Although SILAC-based quantitation provides accurate quantitation and minimizes technical variation throughout the sample preparation process, the number of available label channels limit the amount of different conditions that may be compared in one single MS injection. NeuCode is an alternative metabolic labelling quantitation method which enables higher SILAC multiplexing [18, 19]. This method has a similar principle to the conventional SILAC approach while using different isobaric amino acid isotopologs. Through different combinations of lysine with designated numbers of ^{13}C, ^{2}H, and ^{15}N, respectively, up to 8 conditions can be analyzed in a single injection. During cell culture, the proteins are metabolically labeled in different isotopic forms which can be further resolved using high resolution mass spectrometry.

Currently, the most popular method for SUMO proteomics study is still based on the combination of exogenous expression of SUMO mutants in a given cell line and imminoisolation of SUMO peptides. In this method, a gap still remains between the exogenous and endogenous SUMOylation due to a variety of variations that are introduced into cells. The cell-based nature of this model limits the research scope, thus tissue or animal samples cannot be assessed by this workflow. Although more than 50,000 SUMO sites have been identified from several groups and many antibodies for immunoisolation are commercially available, the cellular engineering of mutant SUMO expression limits it wider application [20]. Therefore, novel approaches need to

be developed for systematic identification of the endogenous SUMOylome. Recently, many different approaches, which are based on the wild-type alpha-lytic protease (WaLP) or the sequential digestion with different enzymes to generate relatively small remnants of endogenous SUMO peptides, enable the investigation of endogenous SUMOylome in any sample, including in primary tissues and clinical samples [5, 7]. More recently, the Rapid and deep-scale ubiquitylation profiling using TMT labeling (UbiFast) and Boosting to Amplify Signal with Isobaric Labeling (BASIL) strategy open a new avenue for improving SUMOylome analysis [21, 22]. UbiFast introduced the use of TMT, which not only increased the number of conditions to 12, but also improved the sensitivity and throughput during multiplexed analysis. BASIL strategy illustrated the introduction of samples with high intensities into one TMT channel as a booster, which amplified signal for the low abundance SUMO peptides existing in the samples. Thus, these additional labeling approaches have demonstrated that even for samples with very little material where traditional workflows do not provide sufficient sensitivity, the identification and quantification of SUMO peptides can still be tremendously improved.

In my SUMO proteomics dataset, I found that many other SUMO E3 ligases were also SUMOylated by PIAS ligases. Although the improved workflow discussed above will definitely reach another milestone of deeper SUMO discovery, it is still not possible to design a purely experimental strategy to discriminate the direct and indirect substrates of a given SUMO E3 ligase. Machine learning technology has been successfully applied in global phosphoproteomic profiling which enables discrimination of direct versus indirect kinase substrates [23]. Through training the algorithm to filter out a set of validated direct substrates from negative controls, it learns to computationally partition the identified phosphosites into direct and indirect classes. With more large-scale dataset available and more experimental evidences of features between SUMO E3 ligase and their direct substrates, it is very promising to discriminate the direct versus indirect substrates using a machine learning approach.

Additionally, The SUMO chain formation facilitated by SUMO E3 ligase needs to be further addressed [20]. The top-down proteomics approach is reported to be possible for identifying the different types of polychains, however, the throughput and sensitivity need to be improved

before the larger-scale application in SUMO chain discovery study. Most recently, the study of using Ub-clipping method provides a valuable strategy for studying the architecture of the polyubiquitin chains [24]. In this study, they used an engineered viral protease to cleave ubiquitin from substrates and polymers except for the diglycine motif that leaves a trackable remnant. Enrichment of particular types of ubiquitin chains is achieved using molecular engineering traps and facilitate the identification of the polymer branching points, acceptor lysines and chain composition. Thus, it is conceivable that a similar approach could be devised and applied to the study of SUMO chains.

5.3 References

1.   Matunis, M.J., E. Coutavas, and G. Blobel, *A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex.* J Cell Biol, 1996. **135**(6 Pt 1): p. 1457-70.

2.   Celen, A.B. and U. Sahin, *Sumoylation on its 25th anniversary: mechanisms, pathology, and emerging concepts.* FEBS J, 2020. **287**(15): p. 3110-3140.

3.   Verma, V., F. Croley, and A. Sadanandom, *Fifty shades of SUMO: its role in immunity and at the fulcrum of the growth-defence balance.* Mol Plant Pathol, 2018. **19**(6): p. 1537-1544.

4.   Hendriks, I.A. and A.C. Vertegaal, *A comprehensive compilation of SUMO proteomics.* Nat Rev Mol Cell Biol, 2016. **17**(9): p. 581-95.

5.   Hendriks, I.A., et al., *Site-specific characterization of endogenous SUMOylation across species and organs.* Nat Commun, 2018. **9**(1): p. 2456.

6.   Hendriks, I.A., et al., *Site-specific mapping of the human SUMO proteome reveals co-modification with phosphorylation.* Nat Struct Mol Biol, 2017. **24**(3): p. 325-336.

7.   Lumpkin, R.J., et al., *Site-specific identification and quantitation of endogenous SUMO modifications under native conditions.* Nat Commun, 2017. **8**(1): p. 1171.

8.   McManus, F.P., F. Lamoliatte, and P. Thibault, *Identification of cross talk between SUMOylation and ubiquitylation using a sequential peptide immunopurification approach.* Nat Protoc, 2017. **12**(11): p. 2342-2358.

9.   Lamoliatte, F., et al., *Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification.* Nat Commun, 2017. **8**: p. 14109.

10.   Lamoliatte, F., et al., *Large-scale analysis of lysine SUMOylation by SUMO remnant immunoaffinity profiling.* Nat Commun, 2014. **5**: p. 5409.

11.   Hendriks, I.A., et al., *System-wide identification of wild-type SUMO-2 conjugation sites.* Nat Commun, 2015. **6**: p. 7289.

12.   McManus, F.P., et al., *Quantitative SUMO proteomics reveals the modulation of several PML nuclear body associated proteins and an anti-senescence function of UBC9.* Sci Rep, 2018. **8**(1): p. 7754.

13.   Li, C., et al., *Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility.* Nat Commun, 2020. **11**(1): p. 834.

14.   Uzoma, I., et al., *Global Identification of Small Ubiquitin-related Modifier (SUMO) Substrates Reveals Crosstalk between SUMOylation and Phosphorylation Promotes Cell Migration.* Mol Cell Proteomics, 2018. **17**(5): p. 871-888.

15.   Mitra, S., et al., *Genetic screening identifies a SUMO protease dynamically maintaining centromeric chromatin.* Nat Commun, 2020. **11**(1): p. 501.

16.   Sun, Y., et al., *A conserved SUMO pathway repairs topoisomerase DNA-protein cross-links by engaging ubiquitin-mediated proteasomal degradation.* Sci Adv, 2020. **6**(46).

17.   Crowl, J.T. and D.B. Stetson, *SUMO2 and SUMO3 redundantly prevent a noncanonical type I interferon response.* Proc Natl Acad Sci U S A, 2018. **115**(26): p. 6798-6803.

18.   Potts, G.K., et al., *Neucode Labels for Multiplexed, Absolute Protein Quantification.* Anal Chem, 2016. **88**(6): p. 3295-303.

19.   Merrill, A.E., et al., *NeuCode labels for relative protein quantification.* Mol Cell Proteomics, 2014. **13**(9): p. 2503-12.

20.   Jansen, N.S. and A.C.O. Vertegaal, *A Chain of Events: Regulating Target Proteins by SUMO Polymers.* Trends Biochem Sci, 2020.

21.   Udeshi, N.D., et al., *Rapid and deep-scale ubiquitylation profiling for biology and translational research.* Nat Commun, 2020. **11**(1): p. 359.

22.   Yi, L., et al., *Boosting to Amplify Signal with Isobaric Labeling (BASIL) Strategy for Comprehensive Quantitative Phosphoproteomic Characterization of Small Populations of Cells.* Anal Chem, 2019. **91**(9): p. 5794-5801.

23.   Kanshin, E., et al., *Machine Learning of Global Phosphoproteomic Profiles Enables Discrimination of Direct versus Indirect Kinase Substrates.* Mol Cell Proteomics, 2017. **16**(5): p. 786-798.

24.   Swatek, K.N., et al., *Insights into ubiquitin chain architecture using Ub-clipping.* Nature, 2019. **572**(7770): p. 533-537.

Appendix - Scientific Contributions

Publications (selected highlights)

1. **Chongyang Li,** Cristina Mirela Pascariu, Trent Nelson, Zhaoguan Wu, Mathieu Courcelles, Simon Comtois-Marotte, Pierre Thibault, Integrative analyses of SUMO proteomics and transcriptomics identify PIAS-mediated regulation networks involved in cell proliferation and cell migration (Manuscript in preparation).

2. **Chongyang Li***, Trent Nelson*, Alfred C. O. Vertegaal, Pierre Thibault, Proteomic strategies for characterizing ubiquitin-like modifications, Nature Reviews Methods Primers, June 2021, * = equal contribution.

3. **Chongyang Li**, Francis P. McManus, Cédric Plutoni, Cristina Mirela Pascariu, Trent Nelson, Lara Elis Alberici Delsin, Gregory Emery, Pierre Thibault, Quantitative SUMO proteomics identifies PIAS1 substrates involved in cell migration and motility, Nature Communications, February 11, 2020.

4. **Chongyang Li**, Qian Xiong, Jia Zhang, Feng Ge, Lijun Bi, Quantitative proteomic strategies for the identification of microRNA targets. Expert Review Proteomics October 9, 2012.

5. Zhuo Chen*, Mingkun Yang*, **Chongyang Li**, Jia Zhang, Dianbing Wang, XianenZhang, Feng Ge, Phosphoproteomic analysis provides novel insights into stress responses in *Phaeodactylum tricornutum*, a model diatom, Journal of Proteome Research. April 18, 2014, * = equal contribution.

Conference presentations (selected highlights)

1. **Chongyang Li**, Francis P. McManus, Cristina Mirela Pascariu, Pierre Thibault, Quantitative SUMO proteome analyses identifies new target associated with cancer cell migration and motility. 11th Annual Symposium Canadian National Proteomics, Quebec City, May 2019.

2. **Chongyang Li**, Francis P. McManus, Cédric Plutoni, Cristina Mirela Pascariu, Trent Nelson, Lara Elis Alberici Delsin, Gregory Emery, Pierre Thibault, Quantitative proteomics identifies novel PIAS1 protein substrates involved in cell migration and motility. 17th Human Proteome Organization World Congress, Orlando, USA, September 2018.

3. **Chongyang Li**, Francis P. McManus, Cristina Mirela Pascariu, Pierre Thibault, Uncovering the Mechanisms of Action of SUMO E3 ligase PIAS1 by Quantitative Proteomics. 10th Annual Symposium Canadian National Proteomics, Vancouver, May 2018.

4. **Chongyang Li**, Francis P. McManus, Pierre Thibault, Identification of Novel E3 SUMO Ligase PIAS1 Substrates by Quantitative Proteomics. 65th ASMS Conference on Mass Spectrometry and Allied Topics, Indianapolis, USA, June 2017.

5. **Chongyang Li**, Francis P. McManus, Pierre Thibault, PIAS1-mediated SUMOylation of BAF57 is a critical regulator of cell growth and drug sensitivity in ovarian cancer cells. 15th Human Proteome Organization World Congress, Taipei, Taiwan, September 2016.

6. **Chongyang Li**, Francis P. McManus, Pierre Thibault, Identification of Novel E3 SUMO Ligase PIAS1 Substrates by Quantitative Proteomics, 7th Annual Symposium Canadian National Proteomics, Montreal, April 2015.