CHARACTERIZING THE SUMO PATHWAY IN HUMAN HEALTH AND DISEASE

by

Danielle Marie Bouchard

A dissertation submitted to Johns Hopkins University in conformity with the requirements for the degree of Doctor of Philosophy

Baltimore, Maryland

December 2020

Abstract

The small ubiquitin-related modifier (SUMO) regulates nearly every aspect of cellular function, from gene expression in the nucleus to ion transport at the plasma membrane. As such, misregulation of SUMO pathway enzymes are implicated in human cancers, neurodegeneration and inflammatory diseases, among others. Despite this knowledge, many questions remain unanswered, including, how can one ~15kDa protein contribute to so many various and fundamentally different cellular functions and diseases? And, can we efficaciously target the SUMO pathway for treatment of associated diseases? If so, how might we identify the best patients to treat with such therapies? The work presented in this thesis uses a multi-faceted approach to address these questions. First, we dive into an exploration of the five SUMO paralogs (SUMO1-5) to understand how they can collectively regulate such diverse biological functions. One potential explanation is that the paralogs have unique and non-redundant cellular functions, though comprehensive evidence to support this was lacking. We therefore performed a systematic analysis of the literature, SUMO paralog expression in various human tissues, and CRISPR paralog knock-out cell lines, which each provided evidence for tissue and paralog-specific functions. Analysis of the knockout cell lines revealed non-redundant roles for the SUMO1 and SUMO2 paralogs in regulating responses to various cellular stressors, nuclear body integrity, gene expression, and cellular morphology. Collectively, this work defines unique roles for the paralogs in diverse cellular processes, thus shedding light on how one pathway can be implicated in various cellular functions and diseases, and simultaneously providing a foundation for the development of precise SUMO paralog-targeting therapies. To aid in the selection of patients for treatment with such therapies, we developed a bioinformatics workflow to analyze the expression levels of SUMO

pathway enzymes in cancerous versus normal tissues. We present a published example of this workflow where we revealed that expression levels of the SENP1 SUMO protease are unchanged in pancreatic cancer, thus indicating that SENP1 is not a predictive biomarker for this particular disease. Together, with our work on paralog-specific functions, we have provided insights essential for realizing the full therapeutic potential of the SUMO pathway.

Advisor: Dr. Michael J. Matunis

Primary Readers: Dr. Valeria Culotta, Dr. Alan Meeker and Dr. Hongkai Ji

Secondary Readers: Dr. Philip Jordan and Dr. Mark Kohr

Acknowledgements

I could write an entirely separate thesis full of Acknowledgements for my colleagues, friends and family. Each and every one of them have played a role in helping me get to this point – which is having achieved my dream of getting a Doctorate in Biochemistry and Molecular Biology, and from the top School of Public Health in the world, at that! However, I will limit myself to these next few pages and hope that I am able to do justice to all of the wonderful people that I have been fortunate to surround myself by throughout my life.

I'd like to begin by thanking my most recent mentor, Dr. Mike Matunis. I knew from my first interactions with Mike during the interview/recruitment weekend that I wanted to join his lab. Not only was his research fascinating, but I knew that he would be a great mentor. One of my goals in coming to graduate school was to learn more about proteins and protein interactions, and I certainly met that goal by joining a lab that studies protein post-translational modifications. Additionally, having the opportunity to train in the lab of the person who *discovered* the SUMO post-translational modification is something that I will be telling everyone for the rest of my life (and so will my family, sorry Mike). I really appreciate Mike giving me the freedom to explore my own projects, first to develop an IHC method for my pancreatic cancer project, and then later to learn how to code for the SUMO paralog project. I couldn't have asked for a better thesis research experience or mentor, and for that I am eternally grateful!

As for my past and current colleagues in the Matunis Lab, it was a pleasure to work with each and every one of them. I learned so much during my rotation working with Hana, and will always think of her whenever I have to do complicated calculations while completely exhausted, write an abstract or give a scientific presentation. Working with Hana was like a mini-PhD experience wrapped up into a few quick months, and I strive to be even a fraction of the mentor that Hana was to me during that time. I also want to thank Christine for being such a great and supportive friend and mentor as I started in the lab, and as I became a more independent researcher. I learned so much from her, and always looked forward to it being her time to bring snacks to lab meeting. We were lucky to inherit Justin relatively early on in my training, and I always enjoyed talking about games, Student Assembly, keratins, SUMO, etc. between our lab benches. More recently, it has been a pleasure to work with Wei and Shuying, and to see them each grow into their individual projects. I'd like to thank Caitlin for being my first PhD rotation student, and for taking over the beer cart along with Fergan. I'd also like to thank Nyla for bringing so much enthusiasm and positive energy to the lab while working on her high school research project. It was a lot of fun working at the bench with her, and I am constantly impressed at her grasp of sumoylation and her desire to do research, and I cannot wait to see what the future holds for her! Similarly, I am thankful that I had the opportunity to work with Allison and Anthony, though primarily over Zoom, and I also cannot wait to see where their training takes them.

Another great feature of the Matunis Lab is that is shares a physical lab space with the Culotta Lab. I am so thankful for the friendships and scientific mentoring I have had with Chynna, Sabrina, Griffin, Courtney, Natalie, Ed and of course my twin and coffee partner, Angelique. They made every day in the lab even more fun by being such an

energetic and great group of people. Beyond Matunalotta lab brunches, birthday celebrations and game nights, I am also so appreciative for their help with IHC and qRT-PCR, and am so happy that our labs were adjoined! Of course, I would also like to thank Val for allowing me to come to lab holiday dinners and to enjoy the lab latte machine, beyond her role as a member of my thesis committee. On that note, I would also like to thank all of my thesis committee members throughout the years for their support and guidance.

Outside of shared lab spaces, I would like to thank the BMB administrators, staff, faculty and my BMB friends. Shannon, Mystee and Erika all played a large role in helping me get acclimated to Baltimore and BMB and I will always appreciate how much they helped me and my conversations with them. I'd like to thank Ashi for being such a fantastic guiding force throughout everything crazy that 2020 has brought, including letting me stay in her Baltimore condo at the beginning of the pandemic. Jennifer, for bravely taking me on as her first ever rotation student. It was such a fun experience to work directly with a PI at the bench, and I learned so much about protein purification, general lab organization and am thankful for the friendship that we've built over the years. As for my fellow BMBers, I'd first like to thank Grace for being such a great friend and always being down for a foodie adventure. Steve, for hosting so many fun game nights and allowing me to come to Game of Thrones nights, even though I had no idea what was going on. Ed, for being my walking partner for over four years! Also, for hosting so many fun game nights and getting such a great group together for Ladies and Gents. I'd like to thank Haobo and Zach for being such great friends during the pandemic, and for constantly checking on me while I was hiding out at Ashi's. Lyle, for pseudo-taking me on as a rotation student in Anthony's lab, and showing me that

it is okay to take a break from studying and to have a healthy work-life balance. Running the historical Insoluble Fraction beer cart became a big part of my graduate school experience, and through that I met some of my best friends in Baltimore, Noa and Ashka. I am forever grateful to have met these amazing and brilliant women who made Baltimore truly feel like home. In addition to my Baltimore friends, I of course want to acknowledge my partner Mat, who I am so fortunate to have met at James Joyce during the Hopkins Residents welcome party that neither of us almost went to. I'd like to thank him for all of his support and encouragement these past three years, even as he has been busy finishing up his own residency and fellowship training, taking board exams and applying for jobs all across the country. He has made the transition to Raleigh relatively painless and I'm excited for our next adventure here. I'd also like to thank Mat's family, Sam, Shirly, Miles and Elena for their encouragement and support these past few years, in addition to the APIs who made Baltimore such a fun place to live!

Prior to moving to Baltimore, I was fortunate to work in the Biotech industry in the San Francisco Bay Area. There, I would like to thank Dr. Frank Andel for taking a chance and hiring an HPLC column-selling Account Manager into a lab position at Novartis Vaccines and Diagnostics. I later had the opportunity to work with Dr. Jerry Tso and Dr. Katy Barglow, who were each incredible mentors and really encouraged me to pursue my dream of a graduate education. Outside of work, I used to rock climb almost every night at Planet Granite with some of my very best friends: Liz, Jon, Adam, David, Joanna, Cameron, Anand, James, Claire, Wallace, Amanda, Mandy, Sam, Kristin, Nick, Megan, and Josh. I'd really like to thank them all not only for making my time in San Francisco so much fun, but for encouraging me to pursue this dream and for

having faith that I could do it. They all made leaving San Francisco one of the hardest decisions of my life and I continuously miss them and our old climbing routine.

Lastly, and most importantly, I would like to thank my family for all of their continued support and encouragement. First, I'd like to thank my sister, Cathryn, and my brother-in-law, Cody, for being such fun people and for always encouraging me. It's hard living across the country from each other, but I love knowing that they are just a phone or video call away. I'd next like to thank my dad, Heidi, Haley and Maddy for their collective support these past few years. My dad has an amazing sense of design that he uses to build beautiful homes in Montana. I like to think I inherited some of those design skills, though I use mine to make graphs and colorful data visualizations.

I am next so excited to thank my mom, who has played such a huge role in helping me achieve all of my dreams throughout my entire life. She helped me get through the endless problem sets, science fair projects and essays required of elementary, middle and high school students, and we even managed to have fun doing them together. Even the nights leading up to my defense seminar she was helping me with my talk, which was such a special experience that I wouldn't change for anything. It was a particularly rewarding experience now that she is back home safe and sound after traveling through the pandemic to come out and support me and to make sure that we could celebrate together in person. I'm so happy that she was able to be here to watch me defend and was even in the room when my committee told me that I passed! My mom is one of the most hard-working people I know, and this degree is really as much for her as it is for me. I of course also need to thank Terry and Patty for their fun visits to Baltimore and for taking me out to sushi.

Last but not least, I am so excited to thank and acknowledge my grandparents, Grammy and Papa, who have been encouraging and supporting me since the very beginning. They were always the first people I would call after receiving my report card to tell them my grades, and it is now with so much pride that I can tell them that I have finally finished my PhD! Although I wish we could have all celebrated this particular milestone together, I am so appreciative that they were able to attend my White Coat Ceremony in Baltimore, tour the lab, and meet Mike a couple of years ago.

In conclusion, I feel like the luckiest person in the world to be surrounded by so many incredible mentors, friends and my family. I wouldn't be here without the collective support and encouragement from each and every one of them and so from the very bottom of my heart – thank you, I won't let you down!

Table of Contents

Contents

List of Tables

List of Figures

Chapter 1: Introduction

Introduction to Ubiquitin and Ubiquitin-like Proteins

A curious protein with two N-termini but only a single C-terminus was discovered in the mid-1970's (1). This unusual Y-shaped protein was not an antibody, but a histone H2A protein that was covalently attached to a 76-amino acid polypeptide called ubiquitin. Ubiquitin has since become the classic example of a protein that through covalent attachment to other proteins, functions as a post-translational modification (PTM). PTMs greatly expand the diversity and functional capabilities of target proteins, and modification by ubiquitin is no exception. Although ubiquitin was first described as a signal for targeting proteins for proteasomal degradation, our understanding of its signaling capacity has grown to include regulation of the cell cycle, cellular stress response, DNA repair, protein synthesis and transcriptional regulation, among others (1, 2). The discovery of ubiquitin as a PTM has led to the subsequent identification of approximately a dozen novel ubiquitin-like proteins (UBLs) (3). There are two broadly defining characteristics of UBLs: they are reversibly conjugated onto lysine residues of target proteins through an enzymatic cascade, and they share a similar 3-D architecture. Crystal structures of ubiquitin and UBLs have revealed that they have a distinctive 3-D structure defined as a beta-grasp fold. This name comes from 5 anti-parallel beta-strands that appear to "grasp" a single helical structure (3, 4). One such UBL, the small ubiquitin-related modifier (SUMO), is the focus of the work presented in this thesis.

Introduction to the SUMO Pathway

SUMOs are conserved throughout all eukaryotes, from single-celled yeast to humans and plants. They function as post-translational protein modifications in a reversible process called sumoylation. Like ubiquitin, they modify thousands of proteins and thereby regulate nearly every aspect of cellular function, from control of gene expression and genome integrity in the nucleus, to mitochondrial fission and ion channel activity in the cytoplasm (3, 5). Like most UBLs, SUMOs are covalently conjugated onto the lysine residues of target proteins through an E1, E2, E3 enzymatic cascade, and can be removed by proteases (Figure 1A). Unlike ubiquitin, of which there is a single modifier, the human genome encodes for five distinct SUMO paralogs (SUMO1-5) that have varying sequence homology (Figure 1B). These divergent sequences impart the paralogs with unique molecular features that raise the intriguing question about whether the SUMO paralogs have unique and non-redundant functions. If the paralogs do have unique functions, what are they and what is the overall significance? These questions are partly the focus of my thesis work and will be explored in Chapters 2 and 3, following a general introduction to sumoylation and the SUMO pathway.

SUMO Conjugation

When SUMO was first discovered 25 years ago, it was described based on its striking similarities to ubiquitin (6, 7). Like ubiquitin, SUMO is first produced as an immature precursor protein with a C-terminal peptide extension is that cleaved to expose a conserved di-glycine motif, which is required for activation. The cleaved (mature) protein is activated by an ATP-requiring heterodimeric E1 enzyme, SAE1-UBA2. The

E1 adenylates the exposed C-terminal glycine residue to form a high energy thioester intermediate between SUMO and the catalytic cysteine of the E1 (8). SUMO is then transferred to the active site cysteine of the sole SUMO E2 conjugating enzyme, Ubc9 (3, 9). Ubc9 can then either directly conjugate SUMO onto target proteins, or transfer SUMO to an E3 ligating enzyme, which facilitates ligation of SUMOs onto target proteins. The most well-characterized E3 ligases include the Protein Inhibitor of STAT (PIAS) family of RING-related proteins, and RANBP2 (10).

While the conjugation pathway is mostly analogous to ubiquitylation, the number of involved enzymes varies greatly. For instance, the human ubiquitin conjugation pathway has two E1s, 35 E2s and more than 300 E3 ligases, with some estimates ranging up to 700 E3s (11, 12). Of note, the hundreds of ubiquitin E3 ligases are required to determine target protein selection and specificity. In contrast, the SUMO pathway consists of a single E1 activating enzyme, a single E2 conjugating enzyme and approximately a dozen E3 ligating enzymes that can enhance the sumoylation reaction but have broad substrate specificity (Figure 2) (13-15). This relative simplicity raises the question of whether sumoylation may be a less complex and diverse PTM as compared to ubiquitylation.

Ubiquitin and SUMO are similarly conjugated onto lysine residues of target proteins. Interestingly, although the ubiquitin pathway has been studied quite extensively, there is no clear consensus motif for ubiquitin conjugation, though it is often mapped to intrinsically disordered regions of target proteins (16). SUMOs, however, have a well-defined consensus motif, ψKxE, where ψ is a bulky hydrophobic residue (such as isoleucine, leucine or valine), K is the target lysine residue, x is any amino acid, and E is the negatively charged glutamic acid residue (17). This consensus motif facilitates

interactions between the target protein and Ubc9 (17), and proteomics studies suggest that approximately 70% of SUMO-modified lysines are within this motif (18). Intriguingly, sumoylation appears to become more promiscuous under conditions of stress, such as proteasomal inhibition and heat shock, as exemplified by the finding that under these conditions only ~23% of SUMO modified lysines are within this motif (19).

SUMO Deconjugation

Modification by both ubiquitin and SUMO can similarly be reversed through the catalytic activity of isopeptidases. Humans express 6 SUMO-specific proteases, SENP1-3 and SENP5-7 (20). Of note, SENP stands for Sentrin-specific protease, because SUMO was also named Sentrin after the discovery that it functioned similarly to a Sentry. In this case, the Sentry was guarding the cell death pathway based on the finding that SUMO binds to the death domain of Fas (21). Similarly, the proteases are guards of the pool of activated and free SUMOs. This is because SENPs cleave the C-terminal peptide extension of precursor SUMOs to reveal the di-glycine motif needed for activation, and also cleave the covalent isopeptide bond formed between SUMOs and substrate proteins (22).

SENPs have conserved C-terminal active site cysteines, but divergent N-terminal domains that our lab and others have shown determine unique subcellular localizations and interacting partners (Figure 3) (20, 23-27). Interestingly, evaluation of evolutionary sequence relationships between the SENPs reveals pairwise similarities between SENP1-SENP2, SENP3-SENP5, and SENP6 -SENP7, suggesting both overlapping and unique functions of the SENPs (20). Consistent with

this, SENP1 and SENP2 both localize to the nucleus and to nuclear pores (28), however they regulate sumoylation uniquely during mitosis (24). Furthermore, work from our lab has found that SENP2 uniquely localizes to cytoplasmic membranes, suggesting a unique role for SENP2 in regulating sumoylation at such membranes (29). Moreover, SENP3 and SENP5 localize to the nucleolus, where SENP3 regulates ribosome biogenesis and SENP5 is involved in mitochondrial fragmentation during mitosis (22, 30, 31). Lastly, SENP6 and SENP7 predominately localize to the nucleoplasm, where SENP6 is required for inner kinetochore assembly during mitosis (20, 32).

In addition to their involvement in regulating various essential cellular processes, SENPs also display SUMO paralog preferences. For instance, SENP1 and SENP2 are the only SENPs that remove the SUMO1 C-terminal peptide extension, although they can also process SUMO2 and SUMO3 (20). However, SENP2 is more efficient than SENP1 at processing SUMO2 and SUMO3 (33). Moreover, the SENPs display paralog-specific preferences for cleavage of isopeptide bonds between the SUMO paralogs and target proteins, as discussed more thoroughly in Chapter 2.

Taken together, SENPs are critical regulators of the SUMO pathway. They regulate the pool of active SUMO, and spatiotemporally remove SUMO from target proteins (20). To that end, misregulation of SENPs are implicated in numerous cancers, and have thus become therapeutic targets as well as inspiration for prognostic biomarkers, as discussed in more detail at the end of this chapter (15, 34). In Chapter 4, I also present my published work assessing to utility of SENP1 as a biomarker for pancreatic cancer (35).

SUMO Paralogs

Perhaps one of the most significant differences between the ubiquitin and SUMO pathways is in the complexity of the modifiers themselves. In contrast to the single ubiquitin protein, humans express multiple SUMO proteins, SUMO1-5. The five SUMO paralogs contain a unique unstructured N-terminal tail that is not present in ubiquitin (36). As such, although “small” is in their name, SUMOs are actually bigger than ubiquitin and average approximately 101 amino acids in length, as compared to the 76 amino acids of ubiquitin. Similar to ubiquitin, SUMO paralogs are produced as immature precursor proteins. After processing by SENPs to remove the C-terminal peptide extensions, they have varying homologies. For instance, SUMO2 and SUMO3 share ~97% peptide sequence identity and are often referred to as SUMO2/3. Interestingly, although SUMO2 and SUMO3 only differ by 3 amino acids in the N-terminus, these paralogs have unique biochemical features, which will be discussed in the next chapter (37). In contrast, SUMO1 only shares ~45% sequence identity with SUMO2/3. Moreover, SUMO4 shares 85% sequence identity with SUMO2/3 and SUMO5 is the most similar paralog to SUMO1, with 88% shared homology (Figure 1B).

These sequence differences suggest functional diversification of the paralogs. Indeed, these intrinsic differences in amino acid sequences result in varying signaling, interaction and modification motifs within each paralog. For instance, SUMO2/3 can be efficiently sumoylated at an internal lysine (K11), enabling the formation of polymeric SUMO2/3 chains on substrates (38). These SUMO2/3 chains have unique functional consequences, such as being recognized by the RNF4 SUMO targeted

ubiquitin ligase (STUbL) (39, 40). Through this PTM crosstalk, SUMO2/3 have the unique potential to target proteins for degradation through the ubiquitin-proteasome system. In fact, it has been hypothesized based on these STUbLs, that SUMO2/3 conjugation and the ubiquitin-proteasome system are tightly integrated and interact in a cooperative manner (18). From an evolutionary standpoint, it is possible that the two pathways and interactions between them evolved as eukaryotes became more complex. This is in contrast to SUMO1, which is thought to be conjugated as a mono-SUMO1 motif onto substrate lysine residues. However, SUMO1 has also recently been shown to act as a "cap" that terminates poly-SUMO2/3 chains (39, 41). Interestingly, these capped chains are not efficiently targeted for degradation by RNF4, but appear to be uniquely recognized and targeted for degradation by a STUbL called Arkadia/RNF111 (39). Thus, SUMO1, SUMO2/3 and mixed SUMO1-SUMO2/3 chains are distinct signals with unique cellular consequences.

In contrast to other SUMOs, SUMO4 possesses a proline residue in the C-terminus that may prevent processing by the SENPs, thus rending SUMO4 incapable of being activated and subsequently conjugated onto target proteins (42). It is therefore proposed that SUMO4 acts through non-covalent interactions with target proteins. However there is some confusion about this, given a subsequent study that found SUMO4 is conjugated to substrate proteins under conditions of stress, such as serum starvation (43). Consistent with this finding, approximately 90 SUMO4 substrates were identified, and many of them were chaperones and transcription factors involved in the regulation of cellular stress. Thus, two studies suggest that SUMO4 may function in cellular stress response pathways (44). Additionally, polymorphisms of the SUMO4 gene have been linked to Type 2 diabetes in multiple ethnic groups, which is mechanistically attributed to the SUMO4 modification of IKBα ultimately turning on NF-

κB regulated genes (45). Collectively, more information is needed regarding the functional attributes and mechanisms of action associated with SUMO4. This includes a need for more data about when, or if, SUMO4 is conjugated onto target proteins.

Lastly, there is only one biological function attributed to SUMO5, which involves regulating promyelocytic leukemia (PML) nuclear bodies (46). SUMO5 expression is highly tissue specific and limited to the testes and peripheral blood leukocytes (46). This suggests potential tissue-specific functions of SUMO5, though these have yet to be explored. Of note, SUMO5 is a predicted pseudogene of SUMO1, and as such, they differ by only 12 amino acid residues, many of which are located towards the N-terminus. Collectively, more work is needed to assess the functional contributions of SUMO4 and SUMO5 to the SUMO pathway. As such, the remainder of this chapter will highlight the molecular attributes of the SUMO1-3 paralogs. Chapters 2 and 3 will then focus on a review of evidence for SUMO paralog-specific functions, and on our systematic discovery of non-redundant paralog-specific functions for SUMO1 and SUMO2.

SUMO non-covalent interactions and PTM crosstalk

Yet another similarity with ubiquitin is the ability of the SUMO paralogs to interact non-covalently with target proteins that contain SUMO interacting motifs (SIMs) (47). SIMs in target proteins contain a hydrophobic core (for example, V/I-X-V/I-V/I) and are often flanked by acidic or serine residues (48, 49). As revealed by structural studies, the hydrophobic core of the target SIM interacts with a groove of SUMO formed by a β-sheet and part of the α-helix (49-51). Interestingly, sequence differences between SUMO1 and SUMO2/3 are found in the second β-sheet and the α-helix, which

corresponds to the SIM interacting domains (52, 53). Functionally, charged lysine residues in this region are critical for the intrinsic ability of SUMO2 to function as a transcriptional repressor, as discussed in more detail in Chapter 2 (54). Moreover, SIMs have been identified in specialized SUMO E3 ligases, termed E4 elongases, which elongate SUMO2/3 chains (10, 55). One such E4, ZNF451, catalyzes chain formation through a tandem SIM region and has been associated with stress-induced SUMO2/3 conjugation (10).

Lastly, the SUMO modifiers themselves can be regulated by other PTMs, such as phosphorylation (36), acetylation and ubiquitination (19), further adding to the complexity of the SUMO pathway and further highlighting the importance of its proper regulation. As explored in Chapters 2 and 3, understanding how the unique molecular attributes of the paralogs contribute to their specific functions will help us better understand why we have evolved to have multiple SUMO proteins.

SUMO mechanisms of action

The unique abilities of the individual SUMO paralogs to form chains or interact noncovalently with target proteins provides the SUMO pathway with a diversity of signals that regulate many essential processes. But what do these signals do to regulate such diverse and important cellular functions? Consistent with many PTMs, covalent or non-covalent modification by SUMO alters the overall surface and therefore the interaction domains of target proteins. This can have a wide range of effects, such as creating or blocking protein-protein interaction domains or creating competition with other PTMs for target lysine residues. This, in turn, can affect the localization, stability, activity and interacting partners of the target protein (Figure 4).

This is exemplified in nearly every SUMO interaction, with a few key examples highlighted here.

First, SUMO-SIM interactions facilitate the formation of protein complexes, such as those used to mediate DNA double-strand break repair, regulate transcription and form nuclear bodies. For instance, once sumoylated, SIMs in the promyelocytic leukemia (PML) protein interact non-covalently with SUMO on itself and other SUMO-modified proteins, to form PML-nuclear bodies (PML-NBs) (9, 56-59). The SUMO-dependent formation of PML-NBs in turn regulates a variety of cellular functions, such as genome maintenance, telomere lengthening, the stress response, DNA repair, transcription, epigenetic modifications and the immune response (58-67).

Secondly, sumoylation can block protein-protein interaction motifs. This is exemplified by SUMO modification of the Forkhead Box protein M1 preventing its homo-dimerization and thus auto-repression, thereby increasing transcription (68). Cross-talk between the ubiquitin and SUMO pathways through the recognition of poly-SUMO2/3 chains by STUbLs, such as RNF4, can also target SUMO substrates for degradation. This is observed in DNA double-strand break repair, for instance, where SUMO and RNF4 are required for proper turnover of DNA repair factors (69). Conversely, sumoylation can also enhance the stability of target proteins through modification at shared ubiquitylation lysine motifs. This is exemplified by the SUMO-mediated protection of Mdm2 and IκBα, which are inhibitors of p53 and NF-κB, respectively. In the absence of SUMO modification, each of these proteins are instead ubiquitylated and consequently degraded by the proteosome (70, 71).

Lastly, sumoylation of proteins can directly alter their conformation, as is the case with thymine DNA glycosylase (TDG), an enzyme involved in nucleotide excision repair and DNA demethylation. Covalent and non-covalent modification of TDG by SUMO facilitates a conformational change that has been proposed to mediate release of TDG from DNA, thus allowing for subsequent factors to repair the DNA (72, 73). Of note, although structural studies support this model, an *in vivo* assay developed by our lab and collaborators found that sumoylation is not required for its enzymatic turnover, as SUMO conjugation and SUMO binding TDG mutants were equally able to mediate base-excision repair (74).

In summary, there are a number of well-defined effects attributed to sumoylation, with specific consequences determined in large measure through context dependent effects on protein-protein interactions.

The SUMO Pathway in Human Health and Disease

Many SUMO-regulated processes, such as DNA repair, gene regulation and proteotoxic stress are misregulated in cancer, stroke, neurodegenerative diseases and heart disease. Thus, sumoylation is implicated in the development and progression of these non-communicable diseases that collectively kill tens of millions of people each year (64, 75-84). How can sumoylation be involved in so many various diseases? To begin, SUMO pathway components, such as the conjugating and deconjugating enzymes, are often misregulated. Since these enzymes are essential in maintaining the proper balance of SUMO modified versus un-modified target proteins, this misregulation can have dire consequences, as discussed below. Moreover, the proteins that are modified by SUMO are themselves critical in many disease-related

processes. Thus, similar to the multi-faceted nature of many diseases, the role sumoylation plays in them is also varied. In this section, I focus primarily on the role of sumoylation in human cancers, since that is most relevant to my work presented in Chapters 3 and 4. However, I also provide a broad overview of the role of sumoylation in neurodegenerative and inflammatory diseases. I conclude this section with a summary of current therapies targeting the SUMO pathway, followed by a brief perspective on the future of SUMO-targeting therapies.

Sumoylation in Cancer

The first direct link between sumoylation and cancer emerged with the discovery that sumoylation is essential for the successful treatment of patients who have Acute Promyelocytic Leukemia (APL) (85, 86). APL develops from a chromosomal translocation that occurs between the promyelocytic leukemia (PML) and retinoic acid receptor alpha (RARα) loci. The resulting oncogenic PML/RARα fusion protein is a potent repressor of nuclear hormone receptor signaling and also a disrupter of PML nuclear bodies (PML-NBs) (86). The functional consequence of this is repression of myeloid differentiation, which is a hallmark of leukemias (87, 88). APL is successfully treated with arsenic trioxide (As_2O_3), which mechanistically enhances the sumoylation of PML/RARα, specifically in the form of poly-SUMO2/3 chains. These chains are recognized by RNF4, which targets PML/RARα for proteasomal degradation (15, 85). Thus, treatment with As_2O_3 releases the differentiation block in APL by allowing PML and RARα to function normally, and leads to remission rates ranging from 72% up to 92%, depending on the study (89). Therefore, despite being administered as a single agent, As_2O_3 is a potent therapy for treatment of APL, and its effectiveness is attributed in large part to SUMO.

Many cancers display changes in the sumoylation status of SUMO substrates that are involved in DNA repair, cell division and cellular signaling, which are also processes that are often misregulated in cancer. Changes in the sumoylation status are often attributed to altered levels of SUMO pathway enzymes, which are essential in regulating the proper balance of SUMO-modified versus unmodified protein targets (15). Changes in the sumoylation status of these target proteins directly impacts cancer-related processes, such as pro-survival signaling, inflammation and metastasis (76, 77). For instance, elevated levels of Ubc9, the sole SUMO E2 conjugating enzyme, results in enhanced tumor formation in breast, colon, lung, liver, prostate, and head and neck cancers (76, 90, 91). Ubc9 overexpression contributes to cancer progression by promoting migration and invasion of breast cancer cells, likely through enhanced sumoylation of proteins involved in these process, or through SUMO-regulated changes in transcription (92). Consistently, elevated levels of Ubc9 are associated with more aggressive breast cancers and may be a predictor of chemoresistance (93). It is therefore unsurprising that elevated levels of Ubc9 are also associated with poor prognoses for many cancers in which it is overexpressed (76, 93).

Altered expression of the PIAS E3 ligases have also been reported in breast, gastric, ovarian, pancreatic, prostate and non-small cell lung cancers (14, 76). For instance, expression analysis of multiple PIAS genes has revealed downregulated expression of PIAS2 and PIAS3 in breast cancer tissues as compared to adjacent noncancerous tissues, but no change in PIAS1, and only a slight increase in PIAS4 (94). To dig deeper, the researchers stratified tumor types by hormone receptor expression, and found that PIAS1-3 had significantly decreased levels in estrogen receptor positive

(ER+) samples as compared to ER- negative samples, thus concluding that PIAS gene expression could determine responses to breast cancer antihormone therapies (94). Although a direct mechanism of action has not been proposed, through their activity as SUMO E3 ligases, PIAS proteins regulate the transcription of genes involved in cellular proliferation, differentiation and survival, thus their misregulation could be associated with pathogenesis (95). In line with this, the PIAS1 protein is overexpressed in human prostate cancer tissues and cell lines (96-98). PIAS1 facilitates sumoylation of the androgen receptor (AR), which is a critical regulator of prostate cancer pathogenesis (96, 99). Downregulation of PIAS1 leads to reduced cellular proliferation and colony formation, attributed to a G0/G1 cell cycle arrest (98, 99). Mechanistically, sumoylation of AR by SUMO1-3 has a negative effect on AR transactivation in prostate cancer cell lines (100). Thus, it is unclear how elevated levels of PIAS1, which presumably enhance the sumoylation of AR, are advantageous to prostate cancer cells. More work is needed to facilitate our understanding of AR regulation, but the many roles of PIAS proteins, including their regulation of JAK/STAT and other signaling pathways involved in prostate cancer, could be a driving force behind changes in PIAS1 levels in cancer cells (101, 102).

Similar to the conjugating enzymes, misregulation of SENPs has also been implicated in the development and progression of multiple cancers (103-111). One well-studied SENP, SENP1, is overexpressed in multiple myeloma, neuroblastoma, prostate, thyroid and bladder cancers (76). In prostate cancer, SENP1 has been correlated with cancer aggressiveness and metastatic potential, and as such, can be used as a prognostic biomarker for this disease (34, 104, 109). Mechanistically, these findings are attributed partially to the induction of hypoxia-inducible factor 1 alpha (HIF1α) dependent signaling pathways (105). HIF1α is de-sumoylated by SENP1, which

increases its stability and transcriptional activity and thus promotes cancer cell survival (105). Of note, SENP1 regulation of HIF1α is described in more detail later in this chapter. Moreover, knockdown of endogenous SENP1 in prostate cancer cells inhibits cellular proliferation through desumoylation of the SMAD4 tumor repressor, which in turn promotes an epithelial to mesenchymal transition (EMT) driven by E-cadherin (104, 109). Thus, SENP1 regulates multiple critical cancer signaling pathways and its misregulation has important consequences. A study has also reported a correlation between increased levels of SENP1 and the development and progression of pancreatic ductal adenocarcinoma (112). This study formed the basis of my first thesis project and publication, which is presented in Chapter 4 (35).

Other SENPs are also implicated in cancer. For instance, SENP3 is overexpressed in oral squamous cell carcinoma, colon and gastric cancers (76). Like SENP1, SENP3 also regulates EMT. Mechanistically, SENP3 desumoylates the Forkhead box protein C2 transcription factor, which induces expression of EMT- promoting genes in gastric cancer cells (113). On the other hand, SENP2 is downregulated in cancers such as bladder, liver, osteosarcoma and chronic lymphocytic leukemia (CLL) (106, 114, 115). The low expression levels observed across various cancer types suggests that SENP2 is a tumor suppressor. Consistent with this suggestion, SENP2 negatively regulates cellular proliferation and migration in both CLL and osteosarcoma through regulation of β-catenin and SOX9 transcription factor stability, respectively (114, 115). Intriguingly, our lab has identified interactions between SENP2 and cytoplasmic membranes (29). Whether control of sumoylation at membranes is associated with pathogenesis is unknown but could be interesting to explore.

Collectively, although there is a trend towards SUMO pathway enzyme up-regulation in numerous cancers, decreased expression of PIAS in breast cancer tissues and SENP2 in various cancer cell lines demonstrate that this is not always the case. This suggests that in certain cellular contexts, SUMO regulators may either contribute to or inhibit pathogenesis. What is consistent, however, is the loss of properly balanced sumoylation in each of these contexts. Balanced sumoylation is essential for maintaining cellular homeostasis, as revealed by the numerous cancers associated with misregulated sumoylation, and the requirement of properly regulated sumoylation in many organisms, as discussed in Chapter 2 (111, 116). Moreover, since many SUMO targets are critical regulators of cancer-related processes, as discussed in the following section, misregulation of the SUMO pathway components has serious implications for cancer development and progression (76, 111).

SUMO regulation of tumor suppressors and oncogenes

The protein product of the breast and ovarian cancer susceptibility gene, BRCA1, is a RING finger E3 ubiquitin ligase, and also a SUMO substrate and SUMO binding protein (117, 118). BRCA1 functions as a tumor suppressor through mediating accurate DNA repair, which is critical in preventing the initiation of many cancers (119). As such, mutations in the BRCA1 RING domain are associated with increased risk of breast and ovarian cancer. Our lab has shown that RAP80, a protein that is required for stabilizing BRCA1 at sites of DNA double-strand breaks, recruits BRCA1 to these sites through non-covalent binding of RNF4 catalyzed hybrid SUMO-ubiquitin chains (118). Moreover, sumoylation also enhances the interaction of BRCA1 with a ubiquitin E2 conjugating enzyme at sites of DNA repair, subsequently enhancing its activity as an E3 ligase (76, 117). Mutations in the BRCA1 RING domain lead to the loss of

ubiquitin E3 ligase activity and SUMO-mediated localization of BRCA1 to sites of DNA damage. Thus, although the role of SUMO in regulating mutated and pathogenic BRCA1 is not fully understood, the fact that SUMO-regulated BRCA1 activity and localization are lost in the mutant protein suggests that sumoylation may be relevant in the development, progression and perhaps treatment of cancers with BRCA1 mutations (117).

The Ras and Myc oncogenes are frequently misregulated in pancreatic, lung, colon and breast cancers, and are also involved in driving their progression (120). Intriguingly, genome-wide shRNA screens have identified the SUMO activating and conjugating enzymes as synthetic lethal partners for both of these oncogenes, as discussed below, further highlighting the importance of the SUMO pathway in human cancers (15).

The Ras family of small GTPases are signal transducing molecules that act downstream of growth factor receptors and regulate essential processes such as cellular proliferation and motility. Activating mutations in Ras family proteins, namely KRAS, are found in many aggressive cancers and unfortunately for patients, have evaded successful targeting with chemotherapeutic agents (121). In an effort to identify other targets that effect KRAS-driven cancers, and might ideally be easier to target therapeutically, a synthetic lethal shRNA screen was performed in two independent human colorectal cancer cell lines. From this screen, the SUMO activating and conjugating enzymes, SAE1, UBA2 and Ubc9, were all identified as synthetic lethal partners with KRAS (122). It was subsequently found that the SUMO pathway, notably Ubc9, is required for KRAS driven transformative growth of colon cancer in mice and *in vitro* assays (123).

Of relevance to Chapters 2 and 3 of this thesis, the authors of the KRAS study directly explored the contributions of the individual SUMO1, SUMO2 and SUMO3 paralogs on the colony formation ability of KRAS mutant cancer cells and found intriguing differences. For instance, depletion of SUMO1 only modestly decreased the viability of KRAS mutant cells and had little effect on colony formation size. In contrast, co-depletion of SUMO2/3 (targeted by individual SUMO2 and SUMO3 shRNAs) significantly reduced the colony formation ability of KRAS mutant cells as compared to WT, though they also only slightly effected viability. Moreover, the authors also looked at anchorage independent growth, which is a method used to assess the metastatic potential of cancer cells, and again found unique and non-redundant paralog-specific functions (124). Specifically, they demonstrated that loss of SUMO1, SUMO2, SUMO2+SUMO3 and SUMO1+SUMO2+SUMO3 each significantly reduced anchorage independent colony formation, but loss of SUMO3 alone did not have a significant effect. This suggests non-redundant roles for SUMO1 and SUMO2, but not SUMO3, in this process. Interestingly, when rescue experiments were performed, re-expression of SUMO1 rescued SUMO1 colony number and size phenotypes but was unable to rescue SUMO2 phenotypes. Consistently, only the SUMO2 phenotypes were rescued upon SUMO2 re-expression. Taken together, these findings reveal that the SUMO1 and SUMO2 paralogs play functionally distinct roles in the colony formation ability of KRAS mutant cells, with SUMO2 appearing to have a broader role in clonogenic growth (123).

A similar screen was used to identify UBA2, a subunit of the SUMO E1 activating enzyme, as a synthetic lethal partner in Myc-driven tumors in human mammary epithelial cells (125). Mechanistically, they found that loss of UBA2 leads to mitotic

defects, likely through the mis-regulated sumoylation of proteins involved in mitosis, as further discussed in the next chapter. Lastly, they also found that low levels of the E1 in breast cancer tumors correlated with longer metastasis-free patient survival, again highlighting the importance of sumoylation in cancer progression and as a potential therapeutic target, as discussed at the end of this chapter (15, 125).

Sumoylation in Neurodegenerative Disease

A common feature of neurodegenerative diseases is the accumulation of misfolded and aggregated cellular proteins. Individual neurodegenerative diseases are identified and classified by the proteins found in these aggregates, and the localization of the aggregates. For instance, the huntingtin protein is observed in nuclear aggregates in Huntingtin's disease, and amyloid-β is found in deposits in the brain of patients with Alzheimer's disease (81, 126). The SUMO paralogs localize to these deposits and aggregates, and many neurogenerative disease-associated proteins are sumoylated. Thus sumoylation is clearly linked to neurodegenerative diseases (78, 81, 127). For instance, the mutated huntingtin protein (Htt) can be modified by both SUMO1 and SUMO2 in vitro (84, 128). Moreover, conjugation of SUMO1 to the same lysine targeted by ubiquitylation increases the abundance, stability and toxicity of mutant Htt in an *in vivo* Drosophila model and in cell culture (126). Consequently, decreased levels of SUMO1 modification are predicted to have a protective role in Huntington's disease (126).

Moreover, SUMO2/3 regulates the degradation of misfolded ataxin-7 (polyQ-ATXN7), which is the hallmark protein of a neurodegenerative disorder called spinocerebellar ataxia type 7 (SCA7) (127, 129). PolyQ-ATXN7 aggregates form intranuclear neuronal

inclusions, which contain transcriptional regulators, proteasome subunits, PML, ubiquitin and SUMO (126, 129). Mechanistically, *in vivo* and *in vitro* experiments suggest that misfolded polyQ-ATXN7 is specifically recognized and modified by poly-SUMO2/3 chains. These chains are recognized by the STUbL, RNF4, which targets the mutant protein for degradation. It has been suggested that as neurons age and the proteasome becomes less active, degradation is compromised, leading to the formation of nuclear aggregates in SCA7 patients (129). Possible age-related declines in sumoylation may also contribute to disease progression. In summary, there is more work to be done to study the roles of sumoylation in neurodegenerative diseases, but the evidence thus far suggests it would be a worthwhile effort.

Sumoylation in Inflammatory Disease

In addition to cancer and neurodegenerative diseases, sumoylation is also involved in inflammatory and autoimmune diseases. This is due to SUMO regulation of Type I interferons (IFNs), which are produced upon detection of viral nucleic acids (130). Despite the essential role of IFNs in defending the host from viral infection, IFNs are also misregulated in many inflammatory and autoimmune disorders, such as Systemic Lupus Erythematosus (131). SUMO-regulated IFN responses have been observed at the levels of global sumoylation and at the paralog-specific level, as described below.

First, using differentiated bone marrow cells derived from $Ubc9^{-/-}$ conditional knockout mice, an increase in pro-inflammatory mediators were observed as compared to cells from $Ubc9^{+/+}$ mice. Moreover, the $Ubc9^{-/-}$ cells stimulated with lipopolysaccharide, tumor-necrosis factor and other pattern-recognition receptor agonists had exacerbated inflammatory responses. To demonstrate that this was not cell-type specific, the

authors also confirmed these findings in Ubc9$^{-/-}$ cells derived from bacteria infected mouse embryonic fibroblasts, and in a human monocyte cell line. Taken together, these findings reveal that sumoylation has as critical role in negatively regulating inflammation *in vivo*. Mechanistically, it was found that sumoylation silenced the normally constitutive expression of the gene encoding IFN-β, and restrained the activation of *Ifnb* by Toll-like receptor ligands (130). A second group looked more closely at the individual roles of the SUMO paralogs in regulating the inflammatory response. Using human monocytes, they found that loss of SUMO2 and SUMO3 together resulted in a significant increase in IFNs, and thereby a subsequent increase in IFN-stimulated genes. Thus, this work specifically demonstrated that SUMO2 and SUMO3 are essential negative regulators of inflammation. Of note, using targeted lysine mutations (K11R and 5KR), it was found that the poly-SUMO2/3 chain forming abilities of SUMO2 and SUMO3 were not required for this response. Moreover, the authors sought to identify the pathways responsible for triggering this IFN response in SUMO2/3-deficient cells through looking at canonical IFN-inducing pathways and transcription factors, such as nucleic acid sensing and metabolism pathways (STING and MAVS), the TBK1-related kinases, and IRF3 and IRF7 transcription factors. Interestingly, the IFN response caused by loss of SUMO2/3 is independent of all canonical IFN-inducing pathways, suggesting a distinct mechanism of IFN regulation. One possibility the authors suggest is that alternative transcription factors besides IRF3 and IRF7 are regulated by SUMO2/3, and in SUMO deficient cells, they trigger an IFN response (132). Consistent with these findings, we also identified many up-regulated IFN-related genes upon knockout of SUMO2 in U2OS cells, as discussed in Chapter 3.

Lastly, and further consistent with the role of sumoylation in the Type I IFN response, sumoylation also regulates host viral responses (133). For instance, upon infection with influenza, host proteins involved in transcription, mRNA processing, RNA quality control and DNA damage repair become sumoylated (134). Thus, sumoylation is an important regulator of viral responses and the immune system, and its misregulation could have important consequences.

Sumoylation in Chronic Mountain Sickness

Interestingly, there are multiple single nucleotide polymorphisms (SNPs) in the *SENP1* gene that are associated with susceptibility to chronic mountain sickness (CMS) in Andean highlanders (15, 135-138). The first SNP reported, rs7963934, is located in intron 6 and involves a cytosine to a guanine (C/G) change (137) (Figure 5). This SNP has been identified as being protective against CMS patients from the Peruvian Andes (135).

A subsequent study identified 66 differential SNPs between CMS and non-CMS individuals, and found that cells from CMS individuals had increased expression of SENP1 at both the gene and protein levels when grown under hypoxic conditions (137, 138). Of note, when grown under hypoxic conditions, the CMS patient-derived cells had an approximate 60% increase in CD235a, an erythroid marker used as a proxy for red blood cell levels, as compared to cell lines derived from non-CMS and sea-level residing individuals. Moreover, when SENP1 expression was depleted by shRNAs in these same cell lines, the proportion of CD235a fell to <1% of the level of CMS cells. Consistently, SENP1 overexpression in the non-CMS cell lines increased the proportion of CD235a to 40% more than non-CMS derived cells. Taken together,

these findings suggest that SENP1 may play a functional role in the observed increase in red blood cells that occurs in CMS patients (138).

A possible, and likely, molecular explanation for the association of *SENP1* SNPs and CMS is based on the regulation of HIF1α by SENP1. Under normoxic conditions, HIF1α is localized in the cytoplasm, where it is hydroxylated and subsequently degraded by ubiquitin. Under hypoxic conditions, HIF1α is no longer hydroxylated and localizes to the nucleus, where it is sumoylated. SUMOylated HIF1α is subsequently targeted for degradation by ubiquitin, unless SENP1 stabilizes the transcription factor by removing SUMO. Desumoylated HIF1α dimerizes and turns on hypoxic genes such as erythropoietin (Epo), glucose transporter (Glut-1) and vascular endothelial growth factor (VEGF) (15, 105). Taken together, SNPs in SENP1 are associated with CMS, and likely contribute to an up-regulation of SENP1 in CMS patients. Given that this association has been shown by multiple independent groups using *in vivo* (cell culture) assays and rigorous bioinformatic analyses, it could be worth exploring as a therapeutic target.

Sumoylation as a Therapeutic Target

Collectively, the involvement of sumoylation in diseases ranging from cancer to viral infection have led to the emergence of the SUMO pathway as an attractive therapeutic target (76, 139). As such, there is currently a SUMO E1 inhibitor, TAK-981, that is in four concurrent Phase I clinical trials, all for the treatment of various human cancers (140-143). Timely, the most recently initiated trial is recruiting cancer patients who have also tested positive for COVID-19, a viral infection caused by the novel SARS-CoV-2 coronavirus that has caused a global pandemic, which we are currently still

living through (144). While we hope TAK-981 will be successful in treating cancer patients and possibly COVID-19, it is likely that there will be cytotoxic effects due to inhibiting the conjugation of all SUMO paralogs. Therefore, it might be prudent to develop a more selective SUMO paralog inhibitor, for instance, one that only targets SUMO2/3 if the goal is to increase the IFN response. At a molecular level, this could be achieved through designing a drug that targets the SIM interaction domain of SUMO2/3, since it is distinct from that of SUMO1 (49). Moreover, SIM interacting domain synthetic peptides (Affimers) and “monobodies” have already been developed and are used in the lab, thus providing viable proof of concept designs (49, 145).

Beyond targeting the SUMO paralogs, it could also prove advantageous to selectively target the SUMO E3 ligases and deconjugating enzymes that display paralog specificity (20, 76). For instance, it has been suggested that E3 ligases may be effective targets for neurodegenerative diseases, however there are currently no PIAS-inducing therapies in the clinic (126). As such, designing viral vectors that selectively target the brain where they can express PIAS proteins could be a potentially efficacious strategy, especially to increase the SUMO2/3 modification of mutated ATXN7. Interestingly, since IFN-β upregulates both PML and SUMO2/3 conjugated proteins, it could also be a viable therapeutic agent for proteasomal degradation of mutant ATXN7 (129). Moreover, as SENP1 is upregulated in many human cancers, researchers have started developing targeted inhibitors. Although they have shown success killing prostate cancer cells in the lab, optimization is needed before going into patients (146). In summary, through studying the basic biology of sumoylation and enhancing our understanding about its contributions to misregulated processes driving numerous diseases, we are meeting the vision of the Johns Hopkins Bloomberg School of Public Health: Protecting Health, Saving Lives - Millions at a time.

Overview of Thesis Work

The work described in this thesis centers on the SUMO pathway, starting with a focus on the unique roles of the paralogs, and ending with the use of a SUMO specific isopeptidase as a biomarker for cancer therapy. Chapter 2 highlights evidence from the literature for SUMO paralog specific functions. These paralog specific functions are further explored and described in Chapter 3, where we use CRISPR-Cas9 to knock out SUMO1 and SUMO2 from human cancer cells in order to systematically evaluate the cells for non-redundant paralog functions. Our findings provide evidence for unique and non-redundant paralog-specific roles. More specifically, we identify unique roles for the paralogs in regulating cellular morphology, nuclear body formation, gene expression and in the cellular stress response. Collectively, our data reveal that loss of SUMO2 has a more robust effect on biological function than loss of SUMO1, though loss of SUMO1 is not without consequence. This suggests that therapeutically targeting the paralogs more precisely could be an advantageous strategy as compared to non-specific targeting of all five SUMO paralogs.

We then develop a workflow for identifying the most optimal cancer patients for treatment with a SUMO inhibitor, as discussed in Chapter 4. Since misregulated SENP levels in human cancers are often reported, we used SENP1 expression as a proxy for total SUMO levels. Consistent with the literature, we found that elevated levels of SENP1 indeed lead to a corresponding decrease in levels of SUMO conjugated proteins (76), and thus hypothesize that patients with elevated levels of SENP1 might be more sensitive to treatment with a SUMO inhibitor. Although we are still in the process of testing this hypothesis, we were able to successfully develop an easy to

use workflow for accessing the utility of an annotated protein as a biomarker for common human cancers. We used this workflow to evaluate SENP1 gene and protein expression in pancreatic cancer cell lines and human tissues. This disease was selected based on a previous report that SENP1 was overexpressed in pancreatic cancer. However, we used validated cell lines and robust bioinformatics resources and concluded that SENP1 is not overexpressed and thus not likely to be an effective biomarker for pancreatic cancer. Our aim is that, in future studies, this workflow can be used to identify human cancers where the SUMO pathway is mis-regulated, thus providing predictive biomarkers for treatment of cancers with a SUMO inhibitor. Taken together, this work highlights the importance of the SUMO pathway, from the paralogs, to the SUMO pathway enzymes, in human health and disease, and opens many doors for future explorations.

Figures

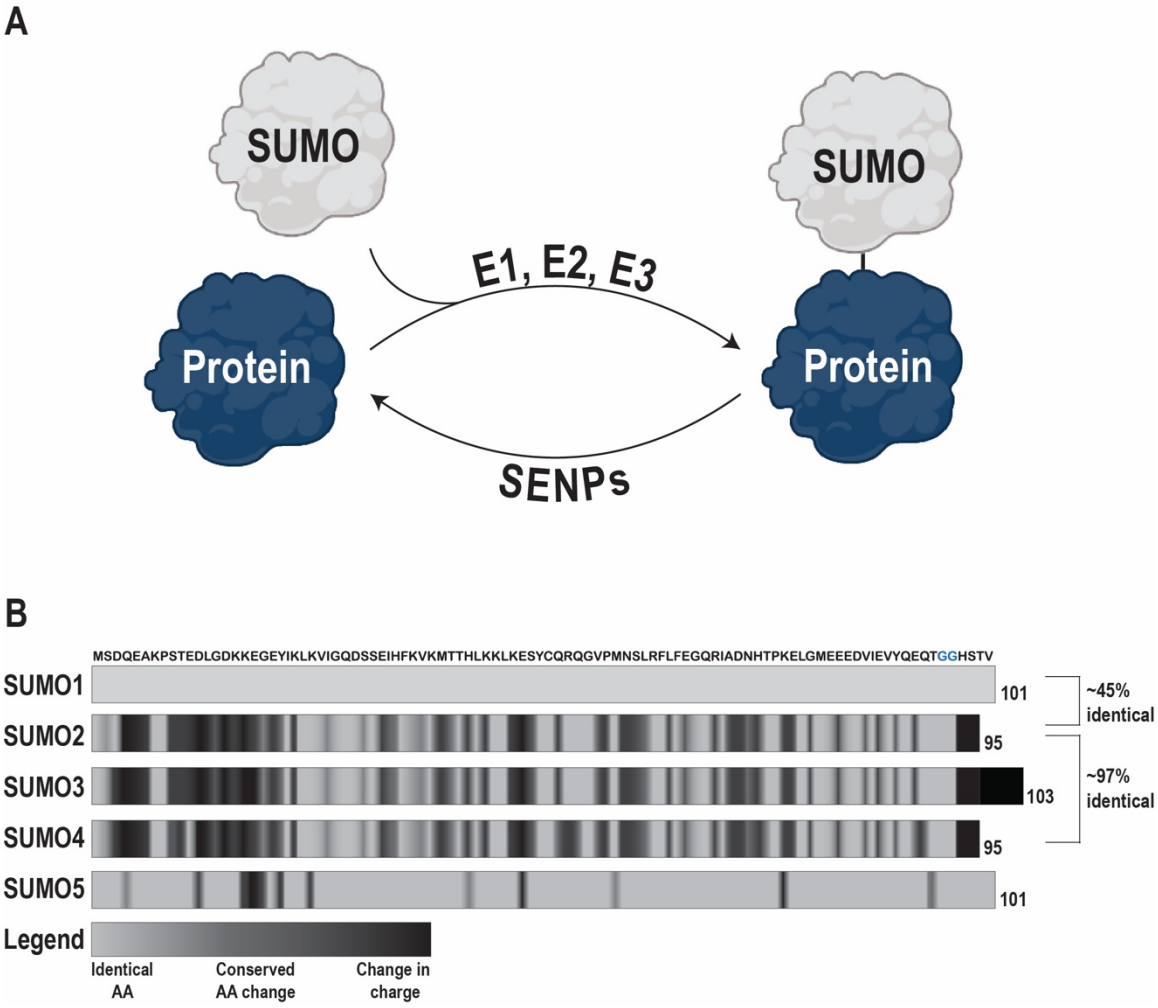


Figure 1. Overview of the SUMO pathway and paralog homologies. A) Small ubiquitin related modifiers (SUMOs) are covalently attached to substrate proteins through an E1, E2, E3 enzymatic cascade. This modification is reversible through the catalytic activity of SUMO proteases, called SENPs. B) The five human SUMO paralogs, SUMO1-5, have varying amino acid homologies. In this figure, the SUMO2-5 amino acid sequences are compared to SUMO1, where the dark lines represent a non-conserved amino acid residue. As shown, SUMO2/3/4 are approximately 97% identical, whereas SUMO1 only shares ~45% homology with SUMO2/3/4. SUMO5 is a predicted pseudogene of SUMO1 and shares ~90% homology with SUMO1.

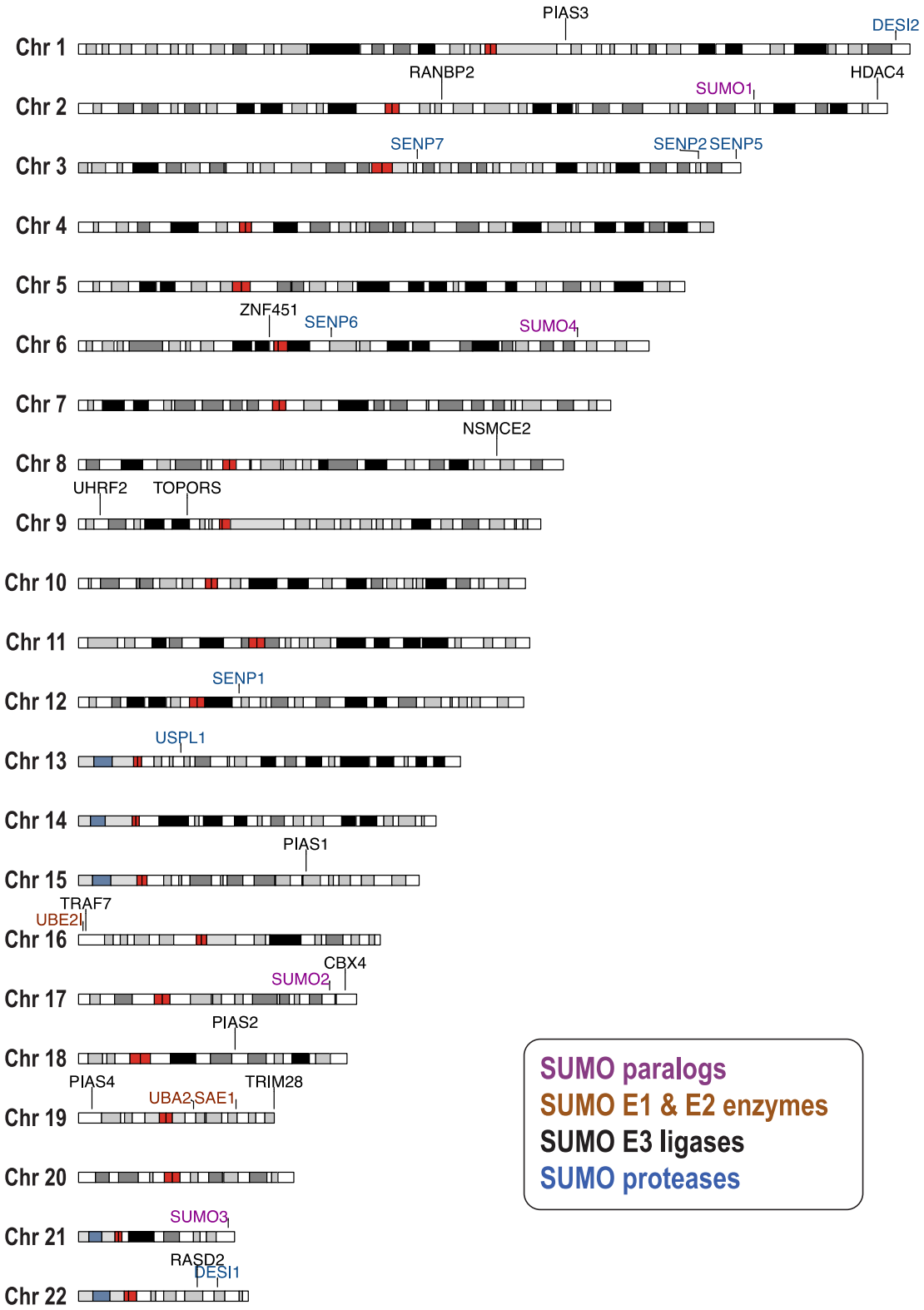


Figure 2. Genomic Location of SUMO Pathway Components. The SUMO paralogs (SUMO1-4), E1, E2 and E3 enzymes, and SUMO proteases mapped to the human genome.

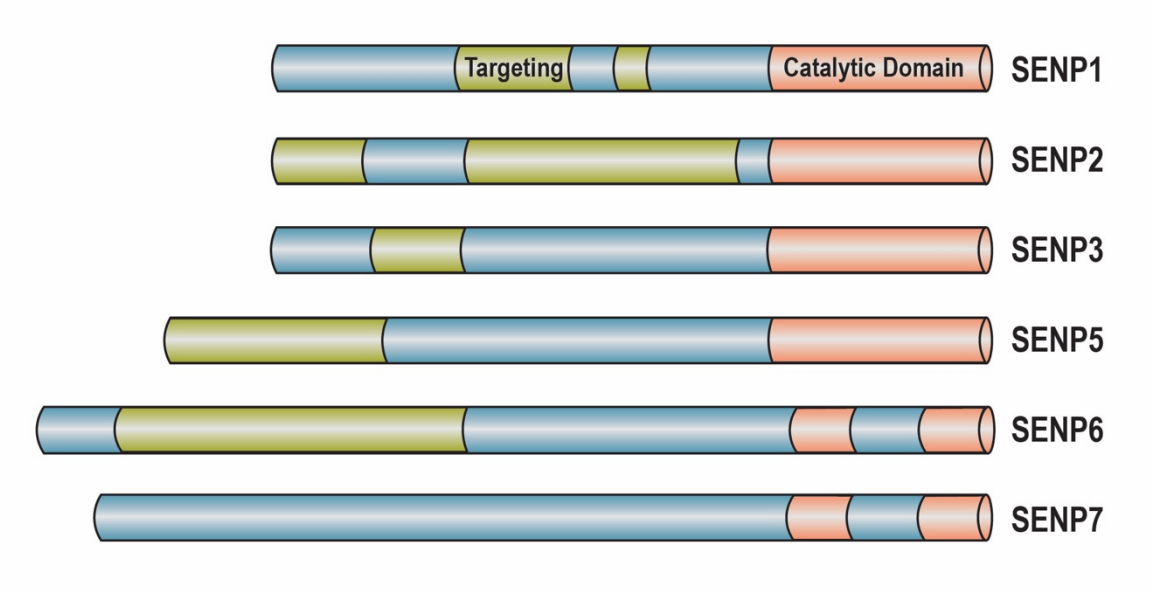


Figure 3. SUMO Protease Schematics. Humans have six SUMO proteases (SENPs). SENP1-3 and SENP5 have conserved catalytic domains (orange). The SENPs have divergent N-termini that contain unique targeting signals (green), such as nuclear localization signals, nuclear pore complex interaction domains, SUMO interacting motifs and phosphorylation sites.

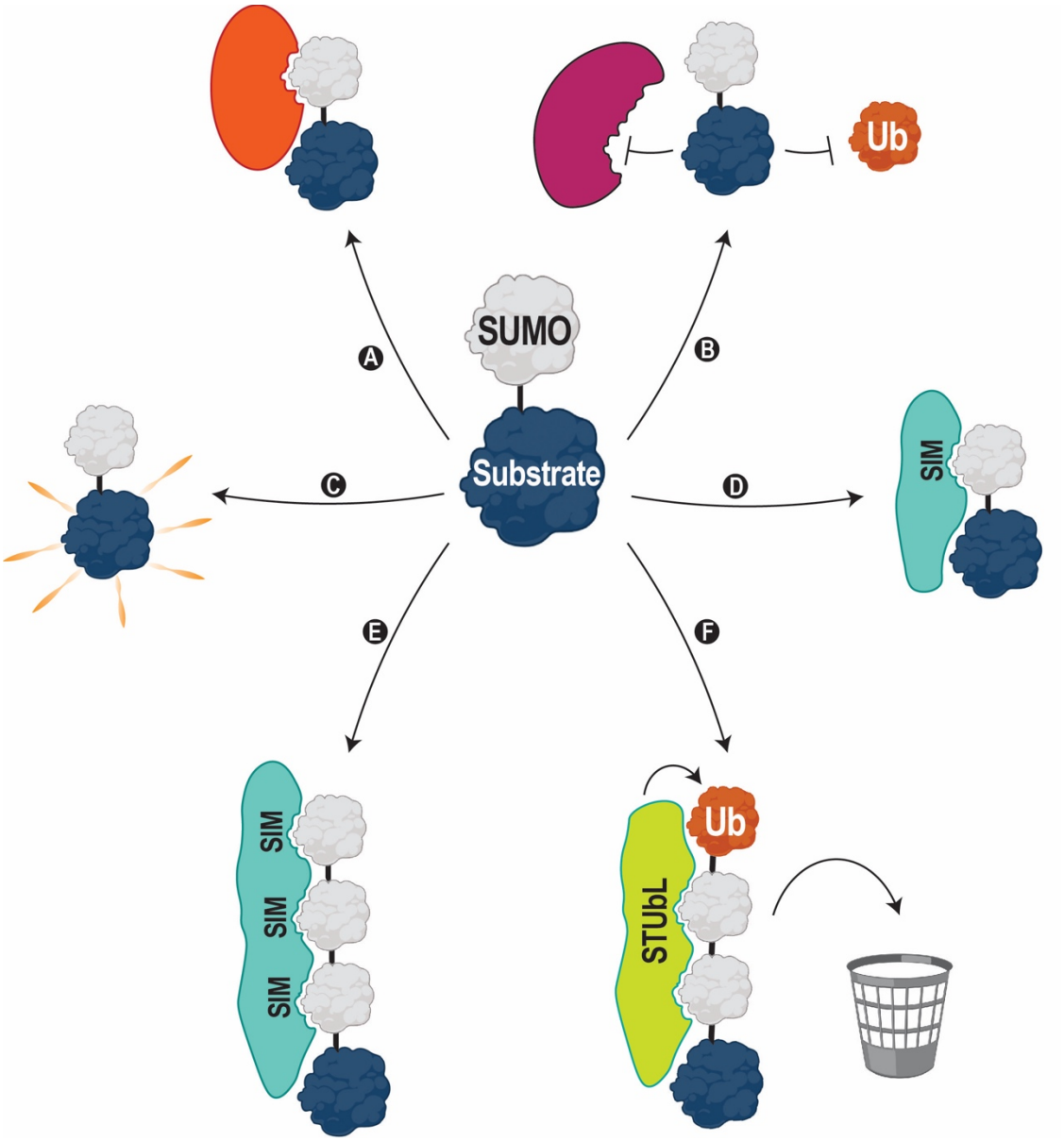


Figure 4. Substrate modification by SUMO has diverse effects. SUMO conjugation onto a target substrate protein, as shown in the center, can have a variety of biological effects. A) SUMO modification can form new protein interaction domains. B) Conversely, SUMO modification can block, or inhibit protein interaction domains. This can be sterically, or through blocking other PTMs that modify lysine residues, such as ubiquitin (Ub). C) SUMO modification can activate a target protein. D) SUMO modification can facilitate non-covalent interactions with proteins that contain a SUMO Interacting Motif (SIM), or E) many SIMs. This can have stabilizing, or degradation effects. F) Modification by poly-SUMOs allow for cross-talk between the SUMO and ubiquitin pathways, mediated by SUMO targeted ubiquitin ligases (STUbLs), which can add ubiquitin to a SUMO chain and subsequently target the protein for degradation.

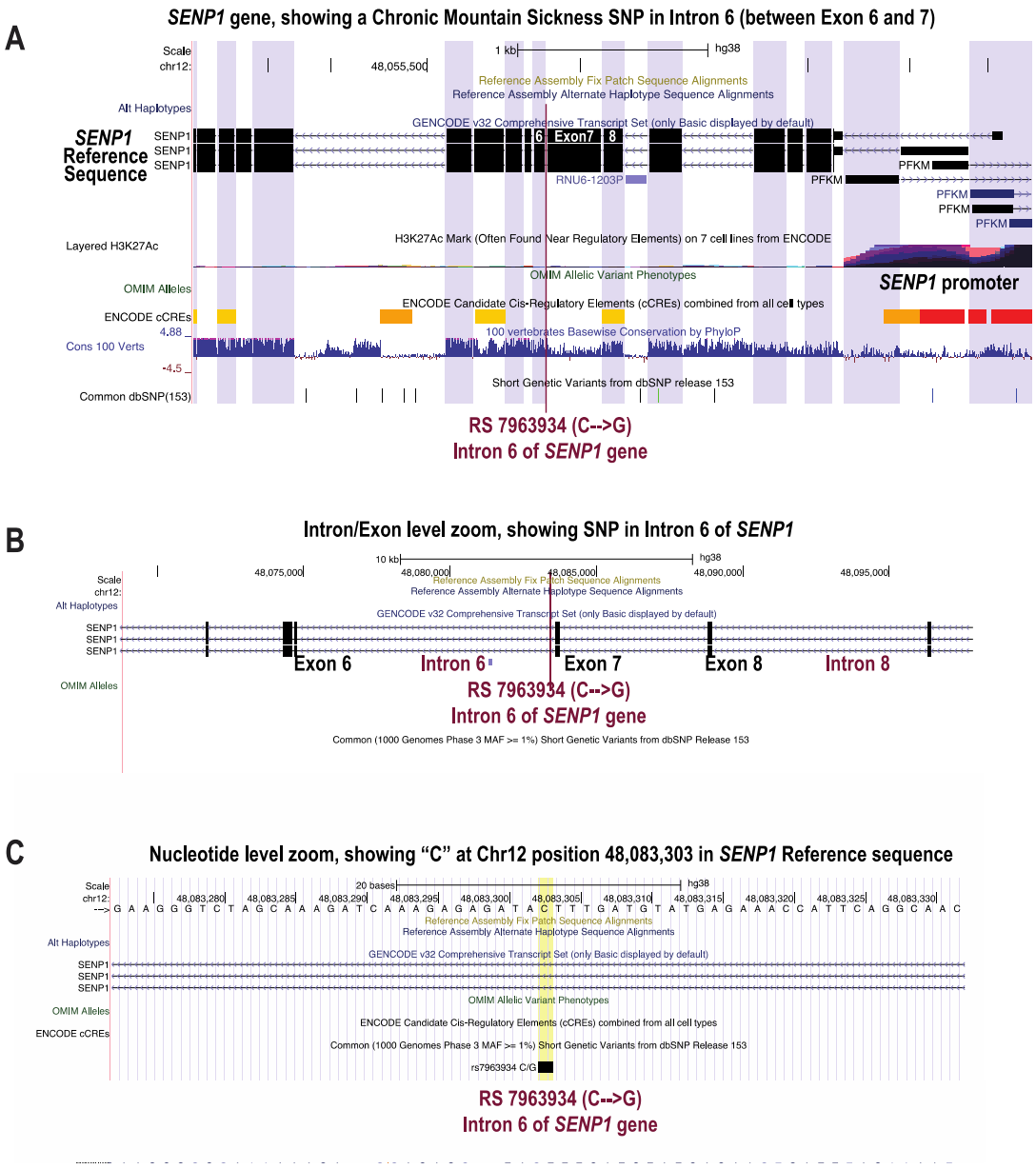


Figure 5. A SENP1 SNP in Chronic Mountain Sickness. Views of a SNP in the *SENP1* gene that are associated with chronic mountain sickness. A) A view of the *SENP1* gene, showing the location of SNP RS7963934 in Intron 6, as marked by a magenta line. B) A zoomed-in view of the SNP, showing its location in Intron 6, but near Exon 7, again marked with a magenta line. C) A nucleotide view of the SNP, showing a cytosine in the reference sequence (highlighted in yellow).

Chapter 2: The Unknown Knowns of Sumoylation – Evidence for SUMO paralog-specific functions

Abstract

Previous explorations of the SUMO pathway in yeast, mice and human cells have revealed conserved functions for sumoylation in chromosome segregation, nuclear architecture, development, the cellular stress response and transcription. These findings have revealed varying requirements for the SUMO paralogs in diverse biological processes, indicating that the paralogs have unique and non-redundant cellular functions. Although a handful of studies have directly examined paralog-specific functions, many fundamental questions still remain about why multiple SUMO paralogs have evolved and what their unique functions are. To better position ourselves to answer these questions, we turned to the literature to uncover the unknown knowns of the SUMO paralogs. More specifically, we looked for evidence of SUMO paralog specific functions in diverse model systems and in various biological processes. This chapter starts with an overview of what we know about SUMO modifiers in commonly studied model organisms, revealing a requirement for sumoylation in these systems. We then highlight studies that have explored the SUMO paralogs using cell culture conditions and *in vitro* assays. From studies in these systems, we report evidence for paralog-specific roles in the stress response, DNA repair, nuclear body integrity, mitosis and transcription. These roles are mediated through paralog-specific modification of unique substrate proteins. The functional outcomes of modification by different SUMO paralogs are complex but center around varying effects on protein-protein interactions and protein stability. Our experimental approach to enhancing our understanding of paralog-specific biological functions is presented in Chapter 3.

Evidence from Non-Human Model Organisms

Although the work in this thesis focuses on human SUMOs, much of what we know about these modifiers has come from studies in non-human organisms. Thus, it is important to highlight some of the molecular functions and features of the SUMO modifiers from commonly studied eukaryotic species, as summarized in Figure 6.

Yeast have a single SUMO, called Smt3. Human SUMO1-3 paralogs each share approximately 50% sequence identity with Smt3 (147). In *Saccharomyces cerevisiae*, mutations in the *Smt3* gene, or in the E1 or E2 activating and conjugating enzyme genes, leads to non-viable yeast. These mutants arrest at the G2/M phase of the cell cycle, highlighting the importance of sumoylation for viability and in cell cycle regulation (148). Moreover, a systematic analysis of Smt3 mutants performed by our lab noted that some mutations in the Smt3 SIM-binding surface were lethal, thus revealing the critical importance of non-covalent SUMO-SIM interactions in organism survival (147). Lastly, similar to human SUMO2/3, Smt3 can also form chains. While Smt3 chain formation is not required for viability, it does enhance the response to numerous stress conditions (147). Of note, invertebrates such as *Drosophila* and *Caenorhabditis elegans* each have one gene coding for a single SUMO paralog (Figure 6) (149). In *C. elegans*, SUMO is called *smo-1* and its deletion is lethal (150).

Interestingly, the *Arabidopsis thaliana* plant has eight SUMO paralogs (AtSUMO1-8), though only four (AtSUMO1,2,3 and 5) appear to be functional PTMs (148, 151). In contrast to the human paralogs, AtSUMO1 and AtSUMO2 are the most closely related,

sharing 89% homology, whereas AtSUMO2 and AtSUMO3 only share 48% homology (152). Moreover, AtSUMO1 and AtSUMO2 have internal consensus site lysine residues that allow for efficient chain formation, analogous to human SUMO2/3. In contrast, AtSUMO3 lacks this internal modification site, which is similar to human SUMO1 (153). Interestingly, although AtSUMO1 and AtSUMO2 are the most closely related, co-deletion of both paralogs results in plant death. However, single deletions of AtSUMO1 or AtSUMO2 has no effect on plant development, thus revealing redundant functions of AtSUMO1 and AtSUMO2 in this critical process (151, 153). Lastly, it is proposed that the AtSUMOs are more functionally divergent than their human orthologues (152). This is based on findings that non-conserved residues on the surface of AtSUMO3 and AtSUMO5 are less efficiently recognized by the *A. thaliana* E1 activating enzyme (AtE1) and are also deficient at forming non-covalent interactions with the AtE2 conjugating enzyme. This results in AtSUMO3 and AtSUMO5 being less efficiently conjugated to other proteins compared to AtSUMO1 and AtSUMO2. The overall consequence of this is that AtSUMO1 and AtSUMO2 are evolving to be the preferred SUMO paralogs (152).

Consistent with the requirement of sumoylation for viability in yeast, worms and plants, sumoylation is also essential for mouse embryogenesis (63). This essential function was first demonstrated by the inability to generate mice upon deletion of the single E2 conjugating enzyme, Ubc9. A closer examination of the developmental timeline of $Ubc9^{-/-}$ mice found that the embryos successfully develop to the blastocyst stage and undergo uterine implantation, but then die before embryonic day 7.5, presumably as cellular pools of maternal Ubc9 are depleted (63). Thus, sumoylation is essential in the post-implantation phase of development. What about paralog-specific roles in this

process? Similar to humans, mice have four SUMO paralogs (SUMO1-4), of which, the contributions of SUMO1-3 have been studied in mouse development (46, 154).

One study in mice aimed to better understand the distinct functional roles of the highly homologous SUMO2 and SUMO3 paralogs, since they only differ by three amino acids. Analysis of SUMO2 and SUMO3 null mice revealed that SUMO2 is essential for embryonic development whereas SUMO3 is not (155). Closer examination of the SUMO2 null mice found that they died around embryonic day 10.5. In contrast, SUMO3 null mice were recovered in expected Mendelian ratios, were fertile and lacked any overt phenotypic abnormalities. Analysis of total *SUMO1-3* mRNA levels from embryonic day 7.5 and 8.5 mice revealed that *SUMO2* accounted for >70% of total SUMO mRNA, in contrast to *SUMO3*, which only accounted for ~3%. The remaining approximately 20% was from *SUMO1*. It was thus proposed and later revealed that the overall loss of total SUMO in the SUMO2 KO mice resulted in embryonic death, which was due specifically to the loss of the predominant SUMO2 isoform. Unfortunately, the authors did not try to rescue the SUMO2 embryonic lethal phenotype with overexpression of SUMO3, thus it remains unclear whether elevated levels of SUMO3 would functionally compensate for SUMO2. Interestingly, the authors also looked at *SUMO1* and *SUMO3* mRNA levels across adult wildtype mice and found that *SUMO3* levels increase to almost 20% of total SUMO mRNA across multiple tissues, suggesting that SUMO3 may have unique and relevant biological functions post-development (155). Taken together, loss of SUMO2 is embryonic lethal and is not compensated for by SUMO1 or SUMO3. Whether the essential requirement for SUMO2 is due to functionally distinct properties, or due to its elevated gene expression levels, remains to be investigated (155).

While loss of SUMO2 is detrimental in mice, and loss of SUMO3 does not appear to have any developmental or phenotypic defects, the story is significantly less clear with regard to loss of SUMO1. The confusion began in 2006 when a 5 year old girl with cleft-lip and palate was found to have a mutation in the *SUMO1* gene that led to decreased SUMO1 mRNA and protein levels in patient samples. Given that sumoylation regulates many developmentally important proteins, an association was made between the loss of SUMO1 and the orofacial defect (156). To further study this finding *in vivo*, SUMO1 mutant mice were bred using embryonic stem cells, and a small fraction of heterozygous mutant mice were reported to have orofacial defects (157). However, a subsequent study using unspecified mice strains, but bred using 2 different embryonic stem cells (one of which was from the original paper), were unable to replicate these findings. In fact, they found that both strains of the SUMO1 null mice were produced at expected Mendelian ratios and were phenotypically normal. Thus, they came to the conclusion that SUMO1 is not required for development, nor for proper orofacial development, owing to functional compensation by the other SUMO paralogs. Of note, they suggested that previously reported phenotypic observations may have resulted from an off-target effect (158).

The finding that SUMO1 is dispensable for mouse embryonic development was supported by a second independent research group that generated SUMO1 knockout (KO) mice using a well-documented and conventional KO of the *SUMO1* locus following homologous recombination. They found that these mice developed normally and did not have any overt phenotypes, and again concluded that SUMO1 is dispensable for development (159). However, a subsequent study from this same group found that adult SUMO1 KO mice fed a high fat diet gained less weight and had smaller and fewer adipocytes than WT littermates. Mechanistically, they found that

SUMO1 regulates adipogenesis through affecting the transcriptional activity of a nuclear receptor that mediates insulin sensitivity (160). Further adding to the complexity surrounding the role of SUMO1 function, a final independent research group studied two previously tested strains of SUMO KO mice, and although they did not observe orofacial defects, they did find congenital heart defects in both strains (79). Taken together, the findings reveal that although loss of SUMO1 is not embryonic lethal, it is not without consequence. While the precise consequences require further exploration, it is likely that the context, such as mice strains and knockout method, have important confounding effects. A final note about cleft palate, SUMO1 does not appear to be implicated in this orofacial defect, as an independent research group genotyped and performed haplotype association studies on over 400 Central-European controls and patients with cleft-palate, and found that none of the analyzed *SUMO1* single nucleotide polymorphisms had significant associations with the defect (161).

In summary, studies in non-human model organisms have revealed regulatory roles for sumoylation in development, cell cycle, the stress response and transcription. In organisms with a single SUMO paralog, loss of this modifier is lethal. In organisms with multiple paralogs, there appears to be a minimal level of total SUMO expression that is required for viability, and possible critical functions for individual paralogs. Taken together, this indicates that sumoylation is an essential PTM for many organisms. So then, what do we know about the biological roles of the SUMO paralogs in humans?

Evidence from Human Cell Culture and *In Vitro* Experiments

Consistent with findings from other eukaryotes, studying the regulatory roles of sumoylation in cultured human cell lines has also revealed important functions. For instance, sumoylation is involved in chromosome dynamics, gene regulation, DNA repair, recombination, nuclear import, nuclear body integrity, the cellular stress response and the immune response (66, 76, 78, 134, 162-165). Despite this knowledge, what we know about the contributions of the individual paralogs to these important biological processes is less clear. This section therefore highlights studies that have taken into consideration the unique attributes of the SUMO paralogs in human cell lines and using *in vitro* assays. Specifically, this section includes studies that evaluated paralog specific preferences of various SUMO pathway enzymes and target proteins, and that explored the functions of the SUMO paralogs in various biological processes. This section has a particular emphasis on paralog-specific functions in transcription, because of the high percentage of SUMO substrates that are transcription factors and chromatin remodeling proteins, and also because of my contributions to the work presented in Chapter 3, which focuses heavily on the unique roles of SUMO1 and SUMO2 in gene expression. This section concludes with a discussion highlighting some caveats of past methodologies, and how these methods have potentially confounded our understanding of the unique roles of the SUMO paralogs. Collectively, this section highlights the background, significance and impact of our work systematically identifying non-redundant paralog specific functions presented in Chapter 3.

Paralog-Specific Preferences of the SUMO Pathway Enzymes

The SUMO conjugation and deconjugation enzymes display paralog-specific preferences, thus providing a mechanism for paralog-specific regulation of target proteins. For instance, the PIAS1 E3 ligase has a preference for enhancing the conjugation of SUMO2 onto target proteins, whereas the RanBP2 E3 ligase has a preference for SUMO1 (10). RanBP2 in particular enhances the transfer of SUMO1 from Ubc9 to target proteins such as SP100 and HDAC4 (166, 167). Furthermore, RanGAP1, a nuclear pore complex protein, is selectively modified by SUMO1 *in vivo* (6). This paralog specificity is achieved through a protective high affinity interaction of SUMO1 with RanBP2, which selectively protects SUMO1 from being cleaved by isopeptidases (168). Lastly, an E4 elongase, ZNF451, has a preference for SUMO2/3. This preference is attributed to dual SIMs in ZNF451 which interact with the N-terminus and SIM interaction surface of SUMO2/3 (10). In each of these examples, paralog-specific SUMO-SIM interactions appear to underlie specificity.

SUMO proteases also demonstrate paralog-specific preferences. For instance, although SENP1 and SENP2 can each de-conjugate SUMO1-3 from substrate proteins *in vitro*, SENP1 is more efficient at proteolytically processing SUMO1 into its active form. This is dictated by differences in electrostatic surface interactions of the different paralogs with SENP1 (169). Moreover, despite general conservation among the C-terminal catalytic domains of individual SENPs (Figure 3), SENP6 and SENP7 have the most divergent catalytic domains, which partially dictate their preferences for deconjugating SUMO2/3 from target substrates (20). This selectivity is guided by the interaction of a C-terminal loop in SENP6/7 with two polar residues that are found uniquely on the surface of SUMO2/3 (N68 and D71). When the analogous residues

are swapped between SUMO2/3 and SUMO1, SENP6 and SENP7 then display a preference for SUMO1 over SUMO2/3 (170).

Taken together, differences in the SUMO surfaces directly contribute to paralog specificity regulated by SUMO conjugating and deconjugating enzymes. Moreover, the finding that SUMO pathway enzymes have preferences for the paralogs provides support for the hypothesis that the paralogs act as distinct cellular signals that can be individually regulated.

Paralog-Specific Modification of SUMO Substrates

Proteomics and cell-based assays have demonstrated that SUMO1 and SUMO2/3 can be conjugated to unique proteins (19, 171-174). For instance, an early mass spectrometry (MS) study using tagged SUMO1 and SUMO3 identified 122 novel SUMO substrates. Consistent with known SUMO-regulated functions, many of these proteins were transcription factors, nucleic acid binding proteins and cellular structural components. Relevant to SUMO paralogs having distinct functions, only 27 of these proteins were modified by both SUMO1 and SUMO3, and nearly twice the number of SUMO1 substrates were identified as SUMO3 (171). Two other independent MS studies also found that 40-50% of identified proteins were uniquely modified by SUMO1 compared to approximately 10% that were uniquely modified by SUMO2 (172, 173). Recent proteomics studies have focused on the identification of SUMO2/3 substrates under conditions of stress, as opposed to identifying unique SUMO1 and SUMO2/3 modified substrates (175-177). Moreover, a 2016 comprehensive analysis of all SUMO proteomics studies combined SUMO1 and SUMO2/3 datasets together, thereby identifying thousands of sumoylated proteins and sumoylation sites, but

unfortunately not revealing unique insights about paralog-specific target proteins (19). Thus, although new proteomics technologies are being developed to study sumoylation, there is an unmet need for a modern and systematic analysis of paralog-specific SUMO substrates. Identifying substrates that are preferentially modified by SUMO1 or SUMO2/3 could improve our overall understanding of the important biological processes that they regulate. However, it is important to note that modification by SUMO1 or SUMO2/3 on the same protein can also have diverse functional effects, thus identifying proteins that can be modified by multiple paralogs is also of interest. As discussed in the following sections, SUMO modification of overlapping or unique paralog specific substrates regulates many important cellular processes, including the stress response, DNA repair, nuclear body integrity, mitosis and transcription.

SUMO Paralogs and the Stress Response

Many SUMO2/3 modified proteins are involved in the cellular stress response. This was first observed when vertebrate cells heat shocked at 43˚C displayed a large increase in SUMO2/3 modified proteins and a corresponding decrease in free SUMO2/3, revealing that previously unmodified proteins were being conjugated by SUMO2/3 upon heat shock. Moreover, upon recovery at 37˚C, conjugated and free SUMO2/3 levels went back to pre-stress levels, revealing that SUMO2/3 modification of proteins in response to heat shock is reversible. In contrast, levels of free and conjugated SUMO1 did not change upon heat shock treatment. However, most SUMO1 is already conjugated under normal growth conditions, thus the role of SUMO1 in response to stress is less clear and therefore requires further exploration (178).

To see whether the increased SUMO2/3 modification of proteins was specific to heat shock, or a conserved reaction among other stressors, cells were also treated with hydrogen peroxide as an oxidative stressor, or ethanol and sodium chloride as osmotic stressors. Consistent with general stress response functions, an increase in SUMO2/3 conjugated proteins, similar to those obtained under heat shock conditions, was observed (178). Proteins modified in response to various stressors have been identified as factors involved in apoptosis, folding and degradation of proteins and DNA repair, among others (179, 180). These factors are often preferentially modified by poly-SUMO2/3 chains, where they serve as signals for ubiquitylation and degradation, or conversely, as signals for stabilizing protein complexes (10, 180). For instance, poly-SUMO2/3 chains stabilize protein complexes involved in gene expression and post-transcriptional modification of mRNAs coding for survival factors in response to heat shock (180).

Accumulation of misfolded or mis-targeted proteins in the endoplasmic reticulum (ER) is a consequence of proteotoxic stress. An important protein quality control mechanism to prevent this is ER-associated degradation (ERAD), where misfolded proteins in the ER are delivered to the cytoplasm for proteosome degradation (181). The ubiquitin specific protease 25 (USP25) is a component of the ERAD pathway and can be sumoylated with both SUMO1 and SUMO2/3. However, USP25 is more efficiently modified by SUMO2/3 due to a higher affinity interaction of the USP25 SIM with SUMO2/3, and there is evidence that USP25 is modified by poly-SUMO2/3 chains (47, 182). The functional consequence of poly-SUMO2/3 modification is impaired binding and hydrolysis of ubiquitin chains, though the consequence of this specifically in ERAD has not been established (47). Lastly, an important protein known to be

affected by ERAD is the cystic fibrosis related protein, CFTR (183). Interestingly, CFTR can also be modified by both SUMO1 and SUMO2/3, however the functional consequences of the different modifications appear to be unique. For instance, it is suggested that SUMO1 modification stabilizes CFTR, whereas modification by SUMO2/3 targets the misfolded protein for RNF4-mediated proteasomal degradation (184). Taken together, this reveals that poly-SUMO2/3 chains act as recruitment signals with diverse biological outcomes, such as stabilization or degradation of proteins and protein complexes. The spatiotemporal requirement for polymeric-SUMO2/3 chains in various biological capacities is not fully understood and therefore requires further exploration.

SUMO Paralogs and DNA Repair

Many cellular stressors cause damage to DNA. Consistent with the role of poly-SUMO2/3 chains acting as a stress response signal, the chains also serve to recruit repair factors to sites of DNA lesions (10). For instance, DNA damage caused by UV irradiation promotes the formation of poly-SUMO2/3 chains on the ATRIP repair factor (185). Sumoylated ATRIP enhances its localization to sites of DNA damage. Moreover, poly-SUMO2/3 chains facilitate ATRIP interactions with other repair factors, such as ATR, replication protein A 70 (RPA70) and the MRN complex, thereby facilitating the DNA repair process (185, 186). Since poly-SUMO2/3 chains interact with SIMs in RNF4, sumoylation also enhances the degradation of DNA repair proteins, such as MDC1 and RPA (69). This is a critical step in DNA repair because RPA needs to be replaced by homologous recombination factors, which is facilitated by the SUMO-dependent proteasomal degradation of RPA (69).

The Bloom syndrome protein, a RecQ related DNA helicase (BLM) is also involved in homologous recombination and DNA repair, where it helps repair damaged replication forks (187). Our lab has shown that BLM is preferentially modified by SUMO2/3, as compared to SUMO1, *in vitro* and *in vivo* (188, 189). This paralog-selective sumoylation is determined by two BLM SIMs that preferentially bind non-covalently to SUMO2 relative to SUMO1. The non-covalent interaction between SUMO2 and BLM likely enhances the interaction of SUMO2-charged Ubc9 with BLM, thus leading to its preferential modification by SUMO2 (188). Sumoylation regulates BLM function in repairing stalled replication forks through mediating protein-protein interactions with Rad51, another DNA repair protein. Interestingly, Rad51 was found to bind equally and non-covalently to both SUMO1 and SUMO2, and SUMO2 was shown to have a strong positive effect on BLM binding to Rad51 (data on SUMO1 was not reported) (187). In summary, through modification of a diverse set of proteins, or through diverse functional outcomes through modification of the same protein, the SUMO paralogs are important regulators of DNA repair.

SUMO Paralogs and Nuclear Bodies

Nuclear bodies are distinct, membrane-less nuclear microenvironments enriched with functionally related proteins that carry out specific processes (190). The promyelocytic leukemia nuclear body (PML-NB) is found in most cell lines and many tissues. Although the precise function of PML-NBs is complex, they contribute to the regulation of transcription factor activity, maintenance of genome stability and antiviral responses (58, 59, 62, 191). While the PML protein is the defining component of PML-NBs, SUMO1 and SUMO2/3 are also important resident proteins (190). Of functional importance, the PML protein has three sumoylation sites that allow for modification by

SUMO1, SUMO2 and SUMO3, and a SIM that facilitates non-covalent interactions with other SUMO-modified resident PML-NB proteins (190, 192). Sumoylation regulates the formation of PML-NBs through a “seeding” mechanism, whereby SUMO and SUMO-modified PML serve as nucleators that initiate formation (192). Beyond its role in PML-NB formation, sumoylation also regulates PML-NB integrity. For instance, loss of sumoylation leads to a decrease in PML-NBs in embryonic cells derived from *Ubc9* knockdown mice and from SUMO1 KO mice (63, 66).

As introduced in Chapter 1, PML fused to RARα forms an oncogenic protein that initiates acute promyelocytic leukemia, which is successfully treated with arsenic trioxide (86, 89, 193). Intriguingly, SUMO1 and SUMO2/3 have unique contributions in the arsenic-triggered degradation of PML-NBs. For instance, loss of SUMO1 delays arsenic-induced PML degradation and stabilizes PML conjugates, revealing that SUMO2/3 cannot compensate for the loss of SUMO1 in this process. Conversely, loss of SUMO2/3 leads to increased arsenic-induced PML degradation and a loss of PML conjugates. When both SUMO1 and SUMO2/3 were knocked-down simultaneously, neither PML conjugates nor arsenic-induced degradation was observed (85). Taken together, these findings reveal that SUMO1 and SUMO2 non-redundantly regulate arsenic-induced degradation of PML. Consistently, we also reveal non-redundant functions for SUMO1 and SUMO2 in PML-NB integrity, as explored in the following chapter.

The death-associated protein 6 (DAXX) is a transcriptional repressor, and also another resident protein of PML nuclear bodies. DAXX can be modified by both SUMO1 and SUMO2, though it preferentially binds SUMO1, as demonstrated through *in vitro* binding assays and NMR structural studies (194, 195). The preference for non-

covalent binding of SUMO1 is enhanced by the phosphorylation of serine residues that flank a SIM in DAXX, and this SUMO1-SIM interaction helps promote the preferential conjugation of SUMO1 to DAXX (196). Functionally, sumoylation affects the localization of DAXX to condensed chromatin and to transcription factors such as Smad4, the androgen receptor and the glucocorticoid receptor (GR). Intriguingly, and as discussed in more detail below, SUMO modification represses the activity of these transcription factors, suggesting that SUMO modified DAXX may have a role in transcriptional repression. Indeed, both reporter and endogenous gene expression assays using a DAXX SIM mutant found that DAXX inhibits GR transcriptional activity through binding to SUMO-modified GR (195). In summary, it is clear that SUMO1 and SUMO2/3 are each important in regulating nuclear body components. However, whether the paralogs have unique and non-redundant roles in nuclear body integrity, and what these unique roles are, remains to be answered. In the following chapter, we examine nuclear body integrity in SUMO1 and SUMO2 knock out cells and reveal evidence for non-redundant paralog-specific roles.

SUMO Paralogs and Mitosis

Sumoylation is essential for proper chromosome segregation in eukaryotes ranging from yeast to humans. This has been demonstrated in human cells using a global SUMO inhibitor, which results in the cells failing to complete cell division due to mitotic defects (197). This finding is also consistent with studies using shRNA or RNAi knockdown of SUMO conjugating enzymes (64, 198). With regard to paralog-specific functions in this process, our lab has revealed that SUMO1 and SUMO2/3 have unique localizations to chromatin during mitosis. For instance, SUMO1 localizes to the mitotic spindle and spindle midzone, whereas SUMO2/3 localizes to centromeres and

condensed chromosomes (198). These differences in localization suggest that different sets of proteins are being modified by SUMO1 and SUMO2/3 during mitosis. In line with this, SUMO1 is required for RanGAP1 localization to the mitotic spindle during metaphase (199). Furthermore, our lab has found that the centromere-associated protein E (CENP-E), BubR1 and Nuf2 kinetochore-associated proteins are modified selectively by SUMO2/3, but not SUMO1. In the case of CENP-E, this modification is dependent upon intact CENP-E SIMs that bind non-covalently to polymeric-SUMO2/3 chains, which are essential for proper CENP-E localization and functions at kinetochores (198). A subsequent group later found that sumoylation regulates the stability of the kinetochore through reversible SUMO2/3 modification of the CENP-H/I/K complex. Regulation of this complex is mediated by competing activities of the SUMO protease SENP6, and the SUMO targeted ubiquitin ligase, RNF4. In the absence of SENP6, RNF4 recognizes the poly-SUMO2/3 chains and subsequently ubiquitylates and targets CENP-I for proteasomal degradation. Conversely, through deconjugating poly-SUMO2/3 chains from CENP-I, SENP6 prevents RNF4 mediated degradation of the CENP-H/I/K complex, thereby stabilizing the kinetochores (32). Having intact kinetochores is required for the proper alignment of chromosomes to the metaphase plate, thus, SUMO2/3 is an important regulator of this process (32, 198).

Of relevance, our lab also recently found that a subunit of another critical mitotic regulator, the anaphase promoting complex 4, is sumoylated by both SUMO1 and SUMO2/3. However, our findings that sumoylation is required for a timely transition from metaphase to anaphase did not systematically compare the effects of the paralogs during this process (200). Given the evidence for diverse paralog-specific signals in regulating large complexes of proteins, it would be interesting to further

investigate the roles of the individual paralogs in regulating the anaphase promoting complex. In summary, the effects of sumoylation are complex. Beyond acting as signals for recruitment or localization, they can also act as stabilization or degradation signals, though an enhanced understanding of this requires further study.

SUMO Paralogs and Chromatin

Sumoylation has important roles in regulating chromatin structure, gene expression and genome integrity (201-204). Although, the individual contributions of the SUMO paralogs in these processes are not fully defined, evidence suggests the existence of paralog-specific functions. For instance, SUMO1 and SUMO2 are both associated with chromatin, as observed by chromatin immunoprecipitation coupled to DNA sequencing (ChIP-seq) in WI38 human fibroblasts. While both paralogs were heavily enriched at transcription start sites (TSS) in these cells, only 2/3 of all identified sites were enriched with both SUMO1 and SUMO2. Thus, 1/3 of identified SUMO1 binding sites in these cells were unique to SUMO1, and similarly, ~1/3 of identified SUMO2 sites were unique to SUMO2. Consistent with SUMO paralogs having unique functions, these findings reveal that a large portion of genomic sites display unique associations with SUMO1 or SUMO2 (202). Of note, many of the loci bound by SUMO1 and SUMO2 in this study coded for histone and protein biogenesis genes, where sumoylation was found to be repressive based on gene expression profiling. Consistently, another ChIP-Seq study also found that SUMO1 is enriched at TSS of genes coding for ribosomal protein subunits and translation factor genes (205). However, this study found that SUMO1 is associated with transcriptional activation of these genes. Mechanistically, it was identified that SUMO1 modification of the scaffold attachment factor B recruits polymerase II to these genes, thus enhancing their

transcription (205, 206). In Chapter 3, we analyze the transcriptome of SUMO1 and SUMO2 KO cells and find that the most differentially expressed genes in the SUMO1 KO cells are repressed, consistent with an activating role for SUMO1 in gene expression.

SUMO Paralogs and Transcription

Sumoylation is an important regulator of gene expression (201, 207-209), as demonstrated by the hundreds of transcription factors (TF) and chromatin remodeling proteins that are SUMO substrates (175, 210). A number of early studies evaluated the effects of either SUMO1, SUMO2 or SUMO3 modification on transcription factor function, but they often did not directly compare the effects of individual paralogs. However, the handful of studies that did investigate signaling capabilities of SUMO paralogs provided functional evidence for paralog-specific effects. Taken together, these findings defined a few common themes surrounding the effects of SUMO paralogs in regulating gene expression, as detailed below.

Early explorations of the functional consequences of SUMO-modified transcription factors often revealed a repressive phenotype (208, 211, 212). For instance, the Sp3, p300 and c-Jun transcription factors, and androgen and glucocorticoid receptors were all repressed when modified by SUMO1 (211, 213-215). As noted in Table 1, many of these studies either did not evaluate the effects of SUMO2 or SUMO3, or did not follow up on the functional implications of the other SUMO paralogs modifying these TFs. Thus, the findings from these studies did not necessarily indicate that SUMO1 uniquely functions as a repressive modifier, though it was thought to be for many years. In contrast, subsequent studies found that the p53 and CMV IE2-p86 transcription factors

are activated upon SUMO1 modification. The opposing effects of SUMO modification on TFs suggested that the effects may be context dependent (216, 217).

For instance, the corepressor for the DNA-binding protein REST (CoREST1) binds directly and noncovalently to SUMO2, but not SUMO1. As CoREST1 bridges binding between a lysine-specific histone demethylase and deacetylase, SUMO2 modification of unknown factors facilitates formation of a repressive complex that alters chromatin structure and gene expression (218). Moreover, although early studies of the androgen receptor (AR) in prostate cancer cell lines found that it was SUMO1 modified, it was subsequently found to be modified by SUMO1, 2 and 3 using overexpression conditions. Interestingly, the paralogs had different effects on transcription depending on the cell line. For instance, while SUMO1 and SUMO3 had a negative effect on AR transactivation in a prostate cancer cell line derived from bone metastases that lack an endogenous androgen receptor (PC-3), SUMO2 did not have an effect. In contrast, in prostate cancer cell lines that express endogenous AR (LnCap), SUMO2 and SUMO3 enhanced AR transactivation whereas SUMO1 had no effect (100).

Lastly, studies using gene fusion-reporter assays in cell culture found that SUMO2 has a stronger intrinsic repression activity than SUMO1 both *in cis* and *in trans* (54, 178, 215). Intriguingly, these functional differences are attributed to conserved basic residues found on the exposed Ub-fold of both SUMO1 and SUMO2. Mutations in these residues leads to a similar increase in reporter activity for both SUMO paralogs, suggesting that they utilize a structurally similar surface to mediate their effects on transcription (54). However, differences in surrounding residues must alter the molecular and structural environment enough to impart paralog-specific functional

differences, since SUMO1 does not inhibit transcription to the same extent as SUMO2. Consistent with this, it was found that a “hydrophobic hole”, formed between the B-sheet and a-helix, is substantially deeper and shifted in SUMO1 as compared to SUMO2. Thus differences in the shape and location of this cavity may contribute to the lower intrinsic repression potential of SUMO1 as compared to SUMO2, though this remains to be formally tested (54). Also of note, the repressive effect of SUMO2/3 is independent of its ability to form chains, as mutations of the internal consensus motif do not affect transcriptional repression activity (215).

How are the SUMO paralogs able to regulate the expression of such diverse TFs? A few mechanisms of action have been described (201), and they are consistent with the ability of SUMOs to mediate both covalent and non-covalent protein-protein interactions (Figure 4). For instance, sumoylation consensus sites for conjugation have been identified in the inhibitory domains of many TFs. The consequence of SUMO modification at these sites is the formation of a new interaction surface which recruits transcriptional regulators, often found to be repressors, to the TF (211). Moreover, sumoylation also recruits transcriptional regulators to regulatory elements within genes. For instance, eukaryotic genes are surrounded by activating and repressing regulatory elements, which are also regulated by activating and repressing multi-protein complexes. These regulatory elements function in concert, resulting in robust gene expression changes, and are thus termed synergy control (SC) motifs (214). Consistent with sumoylation bringing together protein complexes, such as the ones involved in transcriptional repression, these SC motifs contain sumoylation consensus sites (215). Modification by SUMO paralogs within these SC motifs selectively inhibits synergistic activation, as demonstrated by enhanced activity upon disruption of acceptor lysines in hormone receptors and TFs (214, 215, 219). Lastly,

chromatin remodeling factors are recruited to promoters in a SUMO-dependent manner (212). This includes important regulators such as histone deacetylases (HDAC1, 2 and 4), histone demethylases (LSD1), histone methyltransferase (SETDB1), lysine specific demethylases (KDM5B and KDM5C) (18), nucleosome remodeling ATPases (Mi-2) and chromatin-associated proteins (HP1 and L2MBTL1) (212). This recruitment is driven through non-covalent interactions of SUMOs with SIMs in the chromatin modifying enzymes. Taken together, these findings suggest that sumoylation plays a central role in coordinated histone modifications and chromatin structure important for regulation of gene expression (207).

Although early studies suggested a repressive phenotype associated with SUMO modification of TFs, subsequent studies revealed more complex regulation, suggesting that the effects of the paralogs on sumoylation are context dependent. More recently, it has also been suggested that sumoylation of TFs may function upstream of bringing together repressor and corepressor complexes. Specifically, it has been suggested that sumoylation facilitates interactions between TFs and specific chromatin loci (217). This idea is based on the finding that in the absence of sumoylation, eukaryotic TFs bind to numerous non-specific sites, suggesting that sumoylation aids in binding site selection (217, 220). Taken one step further, it has been suggested that prior to TF sumoylation, TFs bind promiscuously and with reduced specificity throughout the genome. This ensures that all functional sites become bound, but then is reliant, in part, upon sumoylation to either increase the specificity of binding, or to promote the release of the TF through mediation of protein-protein interactions (217). Sumoylation could affect TF conformation and DNA binding affinity, as proposed for the turnover of TDG from DNA abasic sites (73), or to promote interactions with other factors that influence DNA binding. In either case, sumoylation

acts as a specificity factor for TF binding. More work is required to understand the roles of the individual paralogs in this process. To that end, we performed a systematic analysis of the global effects of gene expression in the absence of SUMO1 or SUMO2 and identified unique and non-redundant roles for the paralogs in transcription regulation, as described in Chapter 3.

Caveats to Human Cell Culture and *In Vitro* Experiments

As mentioned, early studies of sumoylation focused on the role of a single SUMO paralog, such as SUMO1 or SUMO2/3, on a particular protein of interest. Furthermore, these studies followed a somewhat generic, yet consistent, workflow in which proteins of interest were shown to be modified by SUMO1 or SUMO2/3 using immunoprecipitation assays and western blots. Next, many researchers then identified the modified residue using targeted mutational studies, which involved changing the presumed lysine to an arginine or other amino acid. A caveat to this approach is that many PTMs modify lysines, including acetylation, ubiquitylation, and sumoylation (by the other SUMO paralogs) (208). Thus, observed downstream affects could potentially be due to the loss of other affected PTMs.

Moreover, to explore the function of SUMO on the modified protein of interest, mutant proteins were often tagged and then overexpressed in cells. For instance, exogenously tagged SUMO paralogs, mutant forms of SUMO paralogs and tagged or mutant SUMO substrates were often overexpressed in many of the highlighted studies. There are numerous caveats to these experimental designs. First, an obvious issue is the use of mutant or tagged proteins, which may not accurately recapitulate what the non-mutant, wildtype (WT) proteins are doing in human tissues or cells. For

instance, mass spectrometry analysis of SUMO paralogs and SUMO targeted proteins is often performed using an expressed His-tagged form of SUMO. Although this provides high yields of SUMO modified proteins for MS identification, elevated levels of SUMO may behave differently than endogenous levels (36). Additionally, these tags or mutations might create or abolish binding and interaction domains, subcellular localizations, or normal protein activity. Moreover, there are known artifacts associated with overexpression studies, such as unequal expression of the exogenous protein across cells, which could potentially mask or exaggerate phenotypes. Consistent with this, transient knockdowns, which were also used in many of the highlighted studies, often do not completely rid the biological system of the protein of interest, meaning incomplete or masked phenotypes may be present under these conditions. Lastly, although many studies used *in vitro* assays to demonstrate target protein sumoylation, this finding may not recapitulate *in vivo* requirements for a specific paralog. This is exemplified by RanGAP1, which as previously mentioned, is preferentially modified by SUMO1 *in vivo* (6), but can be also modified by SUMO2 *in vitro* (47, 50).

To circumvent many of these concerns, the work presented in the next two chapters explores endogenous expression levels of genes and proteins, as much as possible. This is paired with large scale sequencing studies from human patients, and knockout studies using CRISPR-Cas9 technology. Importantly, and differing from many of the early studies, we used a systematic approach in our discovery of unique and non-redundant SUMO paralog specific functions. However, we also acknowledge the inherent issues with CRISPR-Cas9 and other caveats to our study design, as discussed in the end of Chapter 3.

Conclusions

In summary, many essential cellular processes, from ER-associated degradation in the cytoplasm to transcription in the nucleus, are regulated by proteins that are selectively modified by SUMO1 or SUMO2/3. These proteins are recognized as substrates for paralog-specific modification at the level of the E3 ligases and SUMO proteases, as well as through non-covalent interactions of substrate proteins with SIMs. The functional outcomes of these paralog specific modifications are complex and have important and vastly different biological consequences (Figure 4). Enhancing our understanding of the spatiotemporal regulation and consequences of SUMO paralog modification is important to understanding both the diverse cellular effects of sumoylation and its consequences for human health and disease.

Figures

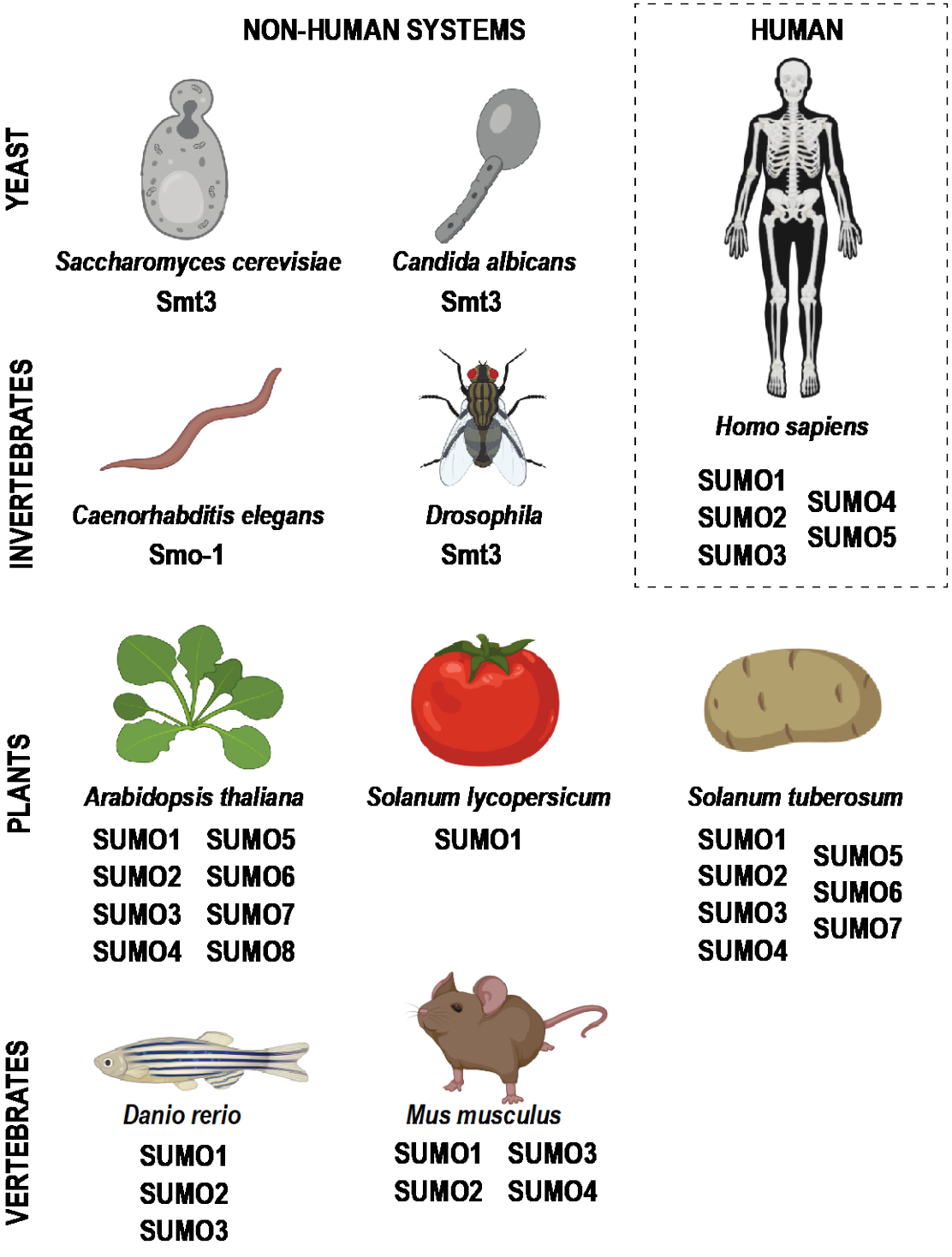


Figure 6. SUMO proteins are conserved across eukaryotic model organisms. The names of the SUMO modifiers are listed to show the number of paralogs in each organism.

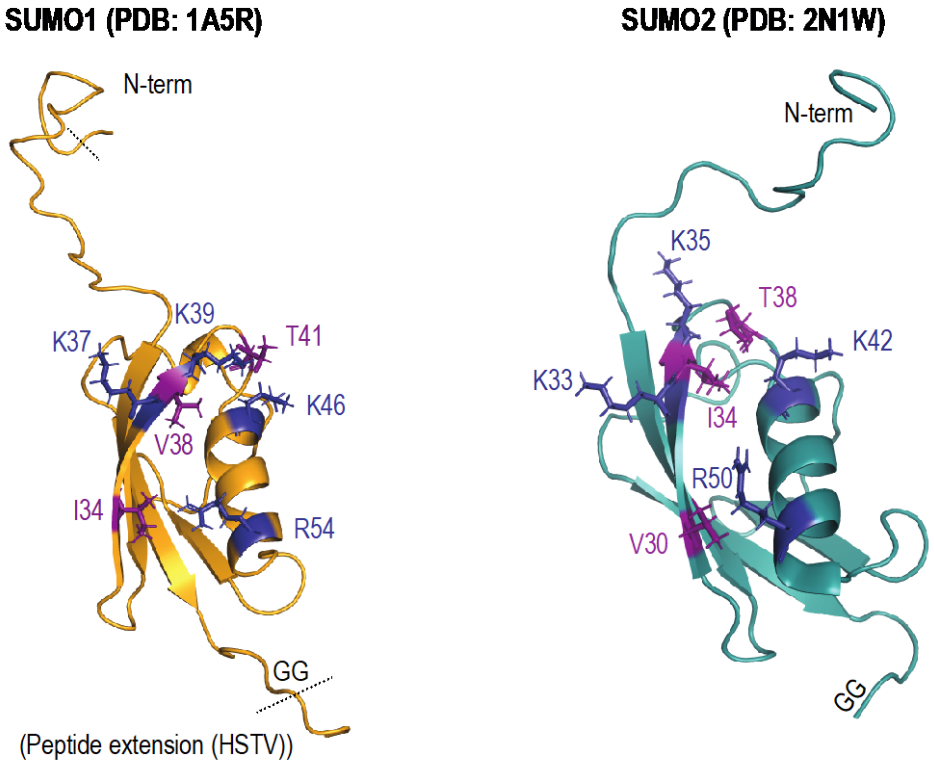


Figure 7. Key residues of SIM interaction domains involved in transcriptional regulation. 3-D structures of SUMO1 (left) and SUMO2 (right). The highlighted residues have been implicated in transcriptional repression. Specifically, the four basic residues in SUMO2 (K33, K35, K42 and R50, blue) were identified as critical for transcriptional inhibition using a linear non-cleavable fusion of SUMO2 to a Gal4 promoter (54). Analogous residues are found in SUMO1, as labeled. Of note, the SUMO1 structure includes two additional amino acids (GS) at the N-terminus, and also contains the precursor residues at the C-terminus (GSTV), as marked with a dashed line. In contrast, the SUMO2 structure is the complete processed sequence with no additions.

Table 1. Examples of sumoylated transcription factors.

TF/chromatin remodeler	SUMO paralog	Effect	Mechanism of action	Other SUMO paralogs studied?	References
Sp3	SUMO1	Repression	Alters localization	No	(211)
c-Jun	SUMO1	Repression	N/A	No	(70)
p300	SUMO1	Repression	HDAC6 recruitment	SUMO2,3*	(215, 221)
Pax3	SUMO1	Activation	Alters DAXX repressor localization	No	(222)
p53	SUMO1	Activation	N/A	No	(223, 224)
IκBα	SUMO1	Repression	PTM competition	No	(225)
Androgen Receptor	SUMO1	Repressive/ Context dependent	SC motif**	No / SUMO2,3	(213) / (100, 220)
Glucocorticoid Receptor	SUMO1	Repressive	SC motif**	SUMO2	(226, 227)
cEBP/α	SUMO1	Repressive	SC motif**	SUMO3*	(203, 228)
HSP2	SUMO1	N/A	Alters DNA binding	SUMO2*	(229)
CMV-IE2-p86	SUMO1	Activation	Alters protein interactions	SUMO2/3*	(209)
Elk-1	SUMO1	Repressive	Alters Localization	SUMO1,2,3*	(230) / (231)
CoREST1	SUMO2	Repression	Localizes LSD1 & HDAC at promoters	SUMO1	(218)
FOXC1	SUMO2/3	Repression	SC motif**	SUMO1*	(232)
FoxM1	SUMO2	Activation	Alters self-interaction domains	No	(68)
Ikaros	SUMO1	Activation	Alters binding with repressors	No	(233)

* Authors demonstrated that TF could be modified by other paralogs, and did not explain why they chose to focus on the selected paralog for downstream studies.

** Sumoylation site mapped to a synergy control (SC) motif. See text for details.

Chapter 3: Characterization of Human SUMO Knockout Cell Lines Reveals Paralog-Specific Roles in Regulating Cellular Morphology, Cellular Stress Responses and Gene Expression

This chapter is in preparation for submission as a publication. Danielle Bouchard is the first author and contributed 6 of 9 figures (Corresponding to thesis figures 8-9, 12, 14-16)

Danielle Bouchard, Wei Wang, Wei-Chih Yang, Shuying He,

Anthony Garcia and Michael J. Matunis

Department of Biochemistry and Molecular Biology

Johns Hopkins University Bloomberg School of Public Health

Baltimore, MD 21205

Abstract

The small ubiquitin-related modifiers (SUMOs) regulate nearly every aspect of cellular function, from gene expression in the nucleus to ion transport at the plasma membrane. In humans, the SUMO pathway has five SUMO paralogs with sequence homologies that range from 45% to 97%. SUMO1 and SUMO2 are the best studied paralogs, and also the most distantly related. To what extent SUMO1, SUMO2 and the other paralogs impart unique and non-redundant effects on cellular functions, however, has not been systematically examined and is therefore not fully understood. Knockout studies in mice have revealed conflicting requirements for the SUMO paralogs during development and studies in cell culture have relied largely on transient paralog overexpression or knockdown. To address the existing gap in understanding, we first analyzed SUMO paralog gene expression levels in normal human tissues and found unique patterns of SUMO1-3 expression across 30 tissue types, suggesting paralog-specific functions in adult human tissues. To systematically identify and characterize unique and non-redundant functions of the SUMO paralogs in human cells, we next used CRISPR-Cas9 to knock out SUMO1 and SUMO2 expression in osteosarcoma (U2OS) cells. Analysis of these knockout cell lines revealed specific functions for SUMO1 and SUMO2 in regulating cellular morphology, PML nuclear body structure, responses to proteotoxic and genotoxic stress, and control of gene expression. Collectively, our findings reveal non-redundant regulatory roles for SUMO1 and SUMO2 in controlling essential cellular processes and provide a basis for more precise SUMO-targeting therapies.

Introduction

Small ubiquitin-related modifiers (SUMOs) function as post-translational protein modifications that regulate a broad range of cellular functions including chromosome segregation, DNA damage repair, gene expression, cellular stress responses, mitochondrial fission and ion channel activity (234). At the molecular level, many similarities exist between the SUMO and ubiquitin protein modification pathways. Like ubiquitin, SUMOs are conjugated to other proteins through an enzymatic cascade involving an E1 activating enzyme, an E2 conjugating enzyme, and E3 ligases. Sumoylation of most proteins is also highly dynamic and reversible through the action of SUMO-specific isopeptidases. Similar to ubiquitin, SUMO is also recognized as a cellular signal and promotes protein-protein interactions between modified substrates and downstream effector proteins. However, the SUMO and ubiquitin pathways diverge at the level of the modifying proteins themselves. In contrast to a single ubiquitin protein, most multicellular organisms, including plants, vertebrates and basal insects, express multiple SUMO paralogs (151, 155, 235). Despite their expansion and conservation across species, the functional significance of SUMO paralogs and why they evolved remains an important question for the field.

Humans express five SUMO paralogs, SUMO1-5, that share 45-97% sequence identity. Of the paralogs, SUMO1-3 are the most widely studied. Following processing, SUMO2 and SUMO3 share ~97% peptide sequence identity and are thus often referred to as SUMO2/3. In contrast, SUMO1 shares only ~45% sequence identity with SUMO2/3, suggesting that these paralogs may have unique properties and can be recognized as distinct signals. Consistent with this, a number of studies have identified proteins that interact preferentially with SUMO1 or SUMO2/3 through variant SUMO

interacting motifs (SIMs) (188, 198, 218). In addition, SUMO2/3 contains an internal consensus site lysine at position 11 that allows for efficient assembly of SUMO2/3 polymeric chains (38). Among other possible functions, SUMO2/3 polymeric chains are recognized by SUMO targeted ubiquitin ligases, and can thereby target proteins for degradation through the ubiquitin-proteasome system (39, 40). SUMO1 lacks a consensus site lysine and thus has reduced potential to form polymeric chains. The ability to associate differentially with SIM-containing proteins and to form polymeric chains may be defining features that distinguish SUMO1 from SUMO2/3 function, although this remains to be formally tested. SUMO4 and SUMO5 are the least well understood, and limited studies suggest that both paralogs have restricted expression to specific tissues (43, 46). As such, our work primarily focuses on the SUMO1-3 paralogs.

At the organismal level, genetic knockout studies in vertebrates have provided conflicting results on the essential functions of individual SUMO paralogs. Whereas SUMO1 expression is uniquely required for development in *Xenopus laevis*, SUMO1, SUMO2 and SUMO3 are each dispensable for development in zebrafish (236). In mice, SUMO2 is essential for embryonic development, but SUMO3 is not (155). Moreover, functions for SUMO1 in mice are less clear, as it has been reported to be both critical and dispensable for embryonic development (157-159). Studies focused on the roles of the paralogs in development are further complicated by the fact that they do not reveal possible essential functions post-development. In this regard, otherwise normal SUMO1 knockout mice have dramatically different responses to a high fat diet (160). Consistently, unique roles for the SUMO paralogs post-embryonic development are supported by studies at the tissue level, including in the placenta,

intestine, eye and brain. In each of these tissues, SUMO1 and SUMO2/3 exhibit remarkably different expression and localization patterns (237, 238).

At the cellular level, numerous studies have also documented SUMO1 and SUMO2/3 specific effects on gene regulation. For instance, SUMO2/3, but not SUMO1, regulates a repressive complex that mediates chromatin structure and transcriptional changes that are important for cell-type specific gene expression (218). Moreover, it was reported that fusing SUMO1 or SUMO2 to the glucocorticoid receptor differentially affects transcription activation in transfected cells (215). Other lines of evidence supporting non-redundant roles for SUMO1 and SUMO2/3 in regulating cellular functions include their unique subcellular localizations and dynamics in cultured mammalian cells (198-200), apparent differential activation by cellular stress (178), evolution of paralog-specific E3 ligases and isopeptidases (10, 13, 166, 169), and identification of distinct target substrates through proteomic studies (19). It should be noted that one limitation of many of these studies has been a reliance on protein overexpression. Collectively, the available data justify a more detailed characterization of SUMO paralogs and their functions.

In this study, we analyzed SUMO paralog expression levels in human tissues and cell lines using publicly available gene expression data and found supporting evidence for paralog-specific functions across a wide range of normal human tissues. Using the CRISPR-Cas9 system, we knocked out SUMO1 and SUMO2 paralog expression individually in human U2OS cells. Systematic analysis of these knockout cell lines revealed unique and non-redundant functions for SUMO1 and SUMO2 in control of cell morphology, stress responses, PML nuclear body assembly and gene expression. Together, our findings provide insights into unique and non-redundant functions of

SUMO1 and SUMO2 in human cells and provide a foundation for further exploration of these functions.

Results

Evaluating SUMO Paralog Expression in Human Tissues and Cell Lines

Functional contributions of the SUMO paralogs may be reflected in their relative expression levels across various cell lines and tissues. We therefore turned to publicly available data repositories to explore *SUMO1-4* gene expression levels and patterns in approximately 500 cell lines and 30 human tissues. We first turned to the Broad Institute's Cancer Cell Line Encyclopedia (CCLE) (239) to explore *SUMO1-4* expression in cancer cell lines derived from bone, breast, liver, lung, ovary, pancreas and thyroid tissues. We found that *SUMO2* expression was consistently ~37% higher than *SUMO1* and *SUMO3*, which had similar levels (Figure 8A). In addition, the relative expression values of the paralogs across cell lines were similar, despite varying tissue origins. Of note, *SUMO4* was consistently expressed at near-zero levels, and thus is not shown. Since our studies involve U2OS osteosarcoma cells, as described below, we specifically looked at *SUMO1-4* levels in bone cancer cell lines and U2OS cells. Here, we found expression patterns and values consistent with those from the other analyzed cancer cell lines (Figure 8B, U2OS in insert).

We next compared *SUMO1-4* gene expression levels in normal human tissues using data from the Genotype-Tissue Expression (GTEx) project. In contrast to observing consistent paralog expression patterns across cancer cell lines, human tissue data revealed varying levels of paralog expression across tissue and organ types (Figure 8C). Notably, although *SUMO4* has been reported to have tissue-specific expression

in the kidney (42) and in lymph nodes, we found near-zero levels in all analyzed tissues, including the kidneys; data from lymph nodes was not available (Supplemental Figure 1). *SUMO1-3*, however, has varying expression levels. For instance, we found that *SUMO1* expression was higher than *SUMO2* in 10 of the 30 analyzed tissues, with the largest difference in expression occurring in the liver, followed by the adrenal gland and muscle (Figure 8D). *SUMO2* expression was higher than *SUMO1* in 16 of the 30 tissues, with noticeably elevated levels in reproductive organs, such as the ovaries, uterus, cervix uteri, vagina, fallopian tube, and the testis. Of note, many of the reproductive tissues had the lowest numbers of available samples, as labeled in Figure 8D. Lastly, *SUMO3* expression levels were higher than both *SUMO1* and *SUMO2* in 23 of the 30 analyzed tissues. These varying patterns of *SUMO1-3* expression suggest non-redundant, paralog-specific functions in adult human tissues.

Generation of SUMO1 and SUMO2 KO Cells using CRISPR-Cas9

To allow for a more systematic identification and characterization of unique and non-redundant functions of SUMO paralogs in human cells, we used the CRISPR-Cas9 system to individually knock out SUMO1 and SUMO2 expression in U2OS human osteosarcoma cells. Of the highly similar SUMO2 and SUMO3 paralogs, we chose to focus on SUMO2 given its apparent higher level of expression compared to SUMO3 in human cell lines and subsequent supporting evidence that SUMO3 protein levels are low in U2OS cells (Figure 8A, Figure 9C). We confirmed heterozygous biallelic SUMO gene knockouts using Sanger DNA sequencing (Figure 9A, Supplemental Figure 2, Supplemental Figure 3). More precisely, we confirmed a deletion of the first exon that interfered with transcription initiation in both alleles of the *SUMO1* gene, and

the creation of a premature stop codon in the second exon of both *SUMO2* alleles in the respective KO cells.

RNA-sequencing analysis revealed that paralog mRNA expression patterns in WT cells were similar to those observed in other cancer cell lines using data from CCLE. Specifically, we found that SUMO2 was the most highly expressed paralog, followed by SUMO1 and SUMO3, and lastly SUMO4, which had negligible expression (Figure 9B). In the SUMO1 KO cells, *SUMO1* mRNA decreased by >99% as compared to WT values, with a very small increase in *SUMO2* (< 3%) and a two-fold decrease in *SUMO3*. In the SUMO2 KO cells, we observed a 65% decrease in *SUMO2* signal as compared to values in WT cells. Based on observed protein levels (below), we hypothesize that the higher than expected levels of *SUMO2* mRNA may be due to detection of mutant transcripts (Figure 9A and B). Also of note, we observed a 17% increase in *SUMO1* and a 63% increase in *SUMO3* mRNA expression in the SUMO2 KO cells.

To evaluate SUMO1 and SUMO2/3 protein expression in the KO cells, we used immunoblot and immunofluorescence microscopy assays with SUMO1 and SUMO2/3 specific antibodies. Both assays revealed undetectable levels of SUMO1-modified proteins and severely diminished levels of SUMO2/3-modified proteins in the respective KO cell lines (Figure 9C and D). Since the SUMO2/3 antibody recognizes an epitope common to both SUMO2 and SUMO3 (198), residual signal in the SUMO2 KO cells is a likely indicator of the relatively low level of SUMO3 protein expression. Taken together, we generated viable cell lines with undetectable levels of SUMO1 protein expression and severely reduced levels of SUMO2/3 expression.

Characterization of Morphological Changes of SUMO KO Cells

Immunofluorescence microscopy revealed unique changes in the morphology of SUMO2 KO cells as compared to WT and SUMO1 KO cells (Figure 9D). Specifically, ~50% of SUMO2 KO cells exhibited a fibroblast-like morphology with an elongated and bipolar shape that contrasted with the primarily polygonal and epithelial-like WT and SUMO1 KO cells. Notably, fewer than 5% of WT and SUMO1 KO cells exhibited a fibroblast-like morphology (Figure 10A). To investigate whether the change in morphology was due specifically to the loss of SUMO2, we constructed SUMO2 KO rescue cell lines with stable, constitutive SUMO2 re-expression or SUMO1 overexpression by plasmid transfection and single-cell cloning. SUMO2 and SUMO1 protein levels in the rescue cell lines (S2KO+S2 and S2KO+S1, respectively) were assessed by immunoblotting and immunofluorescence microscopy (Figure 10B and C). Semi-quantitative measurements of relative SUMO protein levels revealed a near complete restoration of SUMO2/3 expression in the S2KO+S2 cells and a near two-fold increase in SUMO1 expression in the S2KO+S1 cells as compared to endogenous WT levels. A visual inspection of the rescue cell lines by immunofluorescence microscopy revealed that re-introduction of SUMO2 appeared to restore the epithelial-like morphology of WT cells, whereas SUMO1 overexpression had no effect (Figure 10A and D).

To assess the morphological changes of the knockout and rescue cell lines more quantitatively, we used FIJI image processing software (240) to analyze the average aspect ratio, area and circularity of the cells. Consistent with our visual inspection, quantitative measurements revealed a near two-fold increase in the approximate length to width ratio (aspect ratio, 3.30 vs 1.77 vs 1.63) of SUMO2 KO cells as

compared to WT and SUMO1 KO cells, respectively. Additionally, we observed a decrease in the average cell area (1092μm^2 vs 1705μm^2 vs 1710μm^2) and circularity (0.46 vs 0.63 vs 0.70) of the SUMO2 KO cells compared to WT and SUMO1 KO cells (Figure 11A). No significant differences were observed in the average aspect ratio or cell area between SUMO1 KO and WT cells, although SUMO1 KO cells were slightly more circular in comparison to WT cells (Figure 11A). Further assessment of the rescue cell lines confirmed that re-introducing SUMO2 rescued the morphology changes (Figure 11). In contrast, changes in morphology were not rescued in the S2KO+S1 cell line, indicating that SUMO1 overexpression cannot functionally compensate for the loss of SUMO2 (Figure 11). Collectively, these results reveal a unique and paralog-specific function for SUMO2 in regulating cell morphology.

Cell Cycle Analysis of SUMO KO Cells

Previous studies have shown that SUMO1 and SUMO2/3 have unique associations with mitotic chromosomes and that sumoylation of key mitotic regulators is required for timely cell cycle progression (23, 24, 32, 198-200, 241). We therefore used flow cytometry to gain insights into possible differences in the cell cycle dynamics of WT and SUMO KO cells. Using this approach, we found a nearly identical distribution of cells in the G0/G1, S and G2/M phases of the cell cycle in WT, SUMO1 and SUMO2 KO cell lines (Figure 12A). Of interest, we detected a population of cells that had a greater than 2N DNA content specifically in the SUMO2 KO cells. Quantitative analysis revealed that this population of >2N cells was significantly higher in the SUMO2 KO cells as compared to WT, whereas no other significant differences between cell lines were identified (Figure 12B).

Characterization of PML-NBs in SUMO KO Cells

Sumoylation has important roles in regulating the assembly and function of promyelocytic leukemia nuclear bodies (PML-NBs) (9, 162, 242). In particular, sumoylation is thought to affect the phase separation of proteins that underlies the formation of these membraneless organelles (243) although the individual functions of SUMO1 and SUMO2 in this process are less clear. We therefore analyzed the number and size of PML-NBs in WT and SUMO KO cells using antibodies specific for PML and another resident PML-NB protein, DAXX, coupled with immunofluorescence microscopy. Consistent with non-redundant roles for the paralogs in PML-NB assembly and function, we observed a significant decrease in the number of nuclear bodies detected with both PML and DAXX in the SUMO1 and SUMO2 KO cells (Figure 12C and D). In agreement with the literature (58) we observed a mean of approximately 13-14 PML-positive foci per nucleus in WT cells, with a reduction to approximately 6-7 in SUMO1 and SUMO2 KO cells per nucleus (Figure 12D). Moreover, although DAXX was only detected in a subset of PML-NBs, a similar decrease in DAXX-positive foci was observed in the KO cells as compared to WT. Specifically, an average of 4-5 DAXX-positive foci were detected per nucleus in WT cells, whereas only 1-2 foci were detected in SUMO1 KO cells and 0-1 per nucleus in SUMO2 KO cells. Surprisingly, the decrease in PML-positive foci in SUMO2 KO cells was not rescued by re-introducing SUMO2 expression, whereas the number of DAXX positive foci was partially restored (Figure 12D). These findings reveal non-redundant roles for SUMO1 and SUMO2 in affecting PML-NB assembly or integrity and suggest that loss of SUMO function may have irreversible effects on factors underlying their number and size. Further studies will be needed to determine whether the observed

decrease in NBs are due to lack of formation of PML-NBs, or if the stability is compromised in the absence of SUMO1 and SUMO2.

Characterization of Cellular Stress Responses in SUMO KO Cells

MTT Assay and Mitochondrial Function

Sumoylation has important functions in regulating cellular stress responses, including DNA damage and replication stress, oxidative stress and protein misfolding caused by heat shock or other insults (244). To investigate the individual requirements for SUMO1 and SUMO2 in response to cellular stressors, we challenged WT, SUMO1 KO and SUMO2 KO cells with a variety of stress conditions and measured cell survival using an MTT assay. The MTT assay provides a quantitative readout of cell viability and mitochondrial function, as the signal is dependent on mitochondrial respiration (245). We first validated the MTT assay for linearity and found that the readout is a linear function of cell number for WT and SUMO KO cells (Figure 13A). We also noted, however, that the MTT readout for the SUMO2 KO cells was consistently lower compared to WT and SUMO1 KO cells. Taking advantage of the SUMO2 rescue cell lines described above, we found that re-introduction of SUMO2 partially rescued the reduced MTT signal of SUMO2 KO cells, whereas overexpression of SUMO1 increased the signal above WT values (Figure 13A). These findings suggest that SUMO1 and SUMO2 may have unique and non-redundant roles in regulating the number or function of mitochondria.

Proteotoxic Stress Response

To investigate the functions of SUMO1 and SUMO2 in response to proteotoxic stress, we treated cells with varying doses of two different drugs, Azetidine-2-carboxylic acid (AZC) and Eeyarestatin I. AZC is a proline amino acid analog that causes protein misfolding when incorporated into newly synthesized polypeptides (246). WT and SUMO KO cells were treated with varying doses of AZC for 72 hours and cell viability was measured using the MTT assay (Figure 13B). We observed a dose-dependent decline in WT cell viability, demonstrating drug toxicity, and similar dose-dependent declines were also observed in SUMO1 and SUMO2 KO cells. These findings suggest that the SUMO paralogs do not have an obvious effect on the response to protein misfolding caused by AZC, or that SUMO1 and SUMO2 are functionally redundant in this response.

Eeyarestatin I (EerI) is an inhibitor of protein translocation into the endoplasmic reticulum (ER), in part through inhibition of Sec61 and the p97 AAA+ ATPase. It also inhibits ER-associated degradation (ERAD), which targets misfolded proteins in the ER for degradation through the ubiquitin proteasome pathway (182, 247). Treatment of cells with EerI leads to the accumulation of ubiquitylated proteins in the cytoplasm, but whether sumoylation also plays a role in the response to EerI has not been previously tested. WT and SUMO KO cells were therefore treated with varying doses of EerI for 48 hours and viability was measured using the MTT assay (Figure 13 C). SUMO2 KO cells were uniquely sensitive to EerI and sensitivity was most pronounced at 2 μM, where SUMO2 KO cell viability was reduced to 30%, compared to ~60% in WT and SUMO1 KO cells. To investigate whether sensitivity to EerI was due specifically to the loss of SUMO2, we performed dose-response assays using SUMO2

KO rescue cell lines. Surprisingly, re-introduction of SUMO2 and overexpression of SUMO1 both rescued the enhanced sensitivity of SUMO2 KO cells to EerI. Thus, differences in the relative expression levels of SUMO1 and SUMO2 may influence their roles in the cellular response to EerI.

Genotoxic Stress Responses

To study the functions of SUMO1 and SUMO2 in response to genotoxic stress, we investigated the sensitivity of WT and SUMO KO cells to treatment with hydroxyurea (HU), a drug that inhibits ribonucleotide reductase and causes DNA replication arrest and double-strand breaks (248). Cells were treated with varying doses of HU for 72 hours and viability was measured using the MTT assay. This analysis revealed that SUMO1 and SUMO2 KO cells were similarly less sensitive to HU as compared to WT (Figure 13D). SUMO1 and SUMO2 KO cells showed equal resistance at doses of HU below 400μM, whereas SUMO2 KO cells exhibited slightly greater resistance at doses above 400μM. The reduced sensitivity of SUMO2 KO cells to HU was rescued by reintroducing SUMO2 but not by overexpressing SUMO1 (Figure 13D). These findings indicate that SUMO1 and SUMO2 have non-redundant functions in promoting cell survival in the presence of HU-induced DNA replication stress.

Lastly, the reduced toxicity of HU may be due to mechanisms that limit its effect on nucleotide biosynthesis and DNA replication arrest, or on mechanisms that operate downstream of replication arrest. To distinguish between these possibilities, we assessed the effect of HU treatment on cell cycle progression by measuring cell growth over 4 days in the presence or absence of 700 μM HU (Figure 13E). Compared to untreated cells, which all exhibited exponential growth, all treated cells showed a near

complete inhibition of proliferation. Thus, WT and SUMO KO cells exhibited similar cell cycle arrests in response to HU treatment, consistent with expected inhibition of DNA replication. This indicates that the reduced sensitivity of SUMO KO cells to HU is due to effects downstream of replication arrest.

Transcriptomics Profiling of SUMO KO Cells

Hundreds of studies have examined the effects of SUMO-modified transcription factors and chromatin remodeling proteins on the expression of target or reporter genes, however there is limited data on the roles of the SUMO paralogs in regulating global gene expression (201, 216, 249). To address this gap in knowledge, we analyzed the effects of SUMO1 and SUMO2 knockout on the transcriptome of U2OS cells.

Summary of Findings and Validation of Results

Using RNA sequencing (RNA-seq), we identified a combined total of 10,336 genes that were differentially expressed in SUMO1 and SUMO2 KO cells as compared to WT (Figure 14 A). These differentially expressed genes (DEGs), account for a remarkable 70% of all identified genes. Of the identified DEGs, 42% (4,343 genes) were uniquely affected in SUMO2 KO cells and 20% (2,068 genes) were unique to the SUMO1 KO cells (Figure 14 A). Despite the differences in number of affected genes, nearly equal numbers of DEGs were up and down regulated in each cell line. Collectively, these findings are consistent with sumoylation playing a profound role in affecting gene expression and provide evidence that SUMO1 and SUMO2 perform unique and non-redundant functions that affect both activation and repression of different subsets of genes.

To help focus our analysis, we tightened the significance threshold to only include genes with a $\log_2$ fold-change ≥2 (equivalent to a 4-fold change). This more stringent parameter resulted in a combined total of 861 DEGs, and highlighted a more prominent role for SUMO2 in affecting gene expression, as 95% of these DEGs were unique to the SUMO2 KO cells (Figure 14 A). Consistently, many DEGs with the greatest fold changes in SUMO1 KO cells overlapped with SUMO2 KO cells (Figure 14B).

To validate our RNA-seq findings, we selected a subset of up and down-regulated DEGs and tested gene expression by qRT-PCR. We found a strong correlation (R^2 > 0.9) between the assays for both SUMO KO cell lines (Supplemental Figure 5E). To further validate that the robust gene changes observed in SUMO2 KO cells were due specifically to the loss of SUMO2, we also analyzed gene expression levels in both of the SUMO2 KO rescue lines (S2KO+S2 and S2KO+S1) by qRT-PCR. We first used qRT-PCR to quantify SUMO1 and SUMO2 mRNA levels in all tested cell lines to confirm SUMO1 and SUMO2 re-expression in the rescue lines (Figure 14C). We then analyzed 16 DEGs and found that expressing SUMO2 in the SUMO2 KO (S2KO+S2) cells resulted in near-WT levels of gene expression. Interestingly, expression values in the S2KO+S2 cells often went beyond WT values and in the opposite direction of the SUMO2 KO cells (Figure 14D and E), suggesting that SUMO2 has a strong effect on the expression of these genes. Further in support of the observed gene expression changes resulting from a direct loss of SUMO2, we also found that gene expression values were not rescued when SUMO1 was overexpressed in SUMO2 KO cells (S2KO+S1) (Figure 14C-E). Moreover, gene expression levels from these cells were nearly indistinguishable from SUMO2 KO cells, demonstrating that SUMO1 is unable to compensate for the loss of SUMO2 in regulating gene expression.

Karyoplot Analysis

To identify possible patterns or clusters of genes affected by the loss of SUMO1 and SUMO2, we next mapped the KO cell DEGs from our RNA-seq analysis that had a >4-fold change in expression to the human genome (Figure 15A and B). SUMO1 KO DEGs were randomly scattered throughout the genome, with the exception of genes clustered near the end of chromosome 2 and a cluster of histone genes on chromosome 6. Notably, these and other more significantly affected DEGs in the SUMO1 KO cells were down-regulated, as represented by the larger blue dots on the karyoplot. In contrast, SUMO2 KO DEGs were more equally up and down-regulated. In addition, “hotspots” of up and down-regulated SUMO2 KO DEGs were observed throughout the genome, including the same histone gene cluster on chromosome 6 that was observed in SUMO1 KO cells. Closer examination of these histone genes revealed opposing effects of SUMO1 and SUMO2, as they were down-regulated in the SUMO1 KO cells but up-regulated in the SUMO2 KO cells (Figure 15C). Lastly, we found that the SUMO2 KO DEGs often occurred at regions of high gene density, as represented by the gray density plot under each chromosome, whereas there was no such clear association with SUMO1 KO DEGs. Of note, DEGs identified in SUMO1 and SUMO2 KO cells were equally distributed between positive and negative sense strands of the genome and among genes of varied lengths.

Gene Set Enrichment Analysis

To explore the cellular functions associated with genes affected by the loss of SUMO1 or SUMO2, we turned to the Broad Institute’s Gene Set Enrichment Analysis (GSEA) (250). Significantly enriched gene sets were characterized into five broad categories:

nucleus-related, transcription and signaling, cellular stress response, immune response and cell morphology (Figure 16A). Intriguingly, although a majority of these gene sets were enriched in both the SUMO1 and SUMO2 KO cells, the same gene sets often contained genes with opposing expression levels. For instance, histone modification gene sets were heavily enriched with down-regulated genes in the SUMO1 KO cells (blue dots), and up-regulated genes in the SUMO2 KO cells (red dots). This is in-line with our previous observation that histone gene expression decreases in SUMO1 KO cells yet increases in SUMO2 KO cells, and further reveals that the paralogs can have opposing effects on gene expression.

A similar trend of shared gene sets with opposing expression levels was also observed for the transcription and signaling, cellular stress response and immune response categories. The immune response categories were of particular interest in light of recent discoveries highlighting the importance of sumoylation in the immune response (164, 251, 252). Consistent with the literature, we found an enrichment in innate immune response gene sets, such as interferon (IFN) α and γ responses, interleukin-signaling and viral genome integrity. A closer look at the data revealed that a majority of IFN-α response genes are up-regulated in the SUMO2 KO cells (Figure 16B). These same genes were also mostly up-regulated in the SUMO1 KO cells, but to a lesser extent (Figure 16B). Conversely, IFN-γ response genes are almost all down-regulated in the SUMO1 KO cells, but have mixed expression in the SUMO2 KO cells (Figure 16B). Lastly, cell morphology-related gene sets were of interest because of the previously described cell morphology phenotype observed uniquely in the SUMO2 KO cells. Notably, individual gene sets in this category, including the extracellular matrix and epithelial to mesenchymal gene sets, were uniformly down-regulated in the SUMO2 KO cells (Figure 16). Taken together, these findings reveal that SUMO1 and

SUMO2 have unique and non-redundant roles in regulating a broad range of genes. Moreover, they reveal a dominant role for SUMO2 in regulating gene expression, and opposing effects of SUMO1 and SUMO2 on expression of identical genes.

Discussion

Vertebrates express five SUMO paralogs whose individual functions remain to be fully understood. In this study, we revealed evidence that the paralogs have unique and tissue specific functions through analysis of SUMO1 and SUMO2 KO cells, and analysis of gene expression data from human tissues. Systematic analysis of the KO cells revealed paralog-specific phenotypes, such as varying responses to cellular stress, unique gene expression patterns and non-redundant roles in nuclear body integrity, as summarized in Table 2. Moreover, we observed morphological changes that were unique to the SUMO2 KO cells. Re-expression of SUMO2 in the SUMO2 KO cells rescued the cellular morphology, gene expression, and response to genotoxic and proteotoxic stress phenotypes. In contrast, overexpression of SUMO1 in the SUMO2 KO cells did not rescue these phenotypes. This indicates that SUMO1 is unable to functionally compensate for SUMO2 in these processes, thereby revealing that the paralogs have unique and non-redundant cellular functions. This section provides some perspective regarding the molecular mechanisms behind the phenotypes described in this paper.

First, it is possible that the observed changes in cellular morphology, gene expression, nuclear body integrity and responses to cellular stress are due to an upstream reaction to loss of the paralogs, and therefore may not be the direct result of loss of SUMO1 or SUMO2. For instance, it is possible that the gene expression changes in the SUMO2

KO cells are due to loss of SUMO2-modified transcription or chromatin remodeling factors, or they could also be the result of a global stress response due to the loss of SUMO2. Analysis of differentially expressed genes by RNA-sequencing identified an enrichment in some specific stress response pathways, notably related to DNA repair. However, the lack of differentially expressed heat shock proteins, ubiquitin-proteasome pathway components, and other canonical stress-response genes suggests that loss of the individual paralogs does not elicit a generalized stress response.

Moreover, the observed changes in cellular morphology, gene expression, nuclear body integrity and responses to cellular stress may also be interconnected. For instance, we observed defects in PML-NBs in both KO cell lines, which could affect downstream PML-NB regulated processes, such as genome maintenance, the stress response, DNA repair and transcription (58-67). Additionally, changes in transcription could offer a potential explanation for the observed SUMO2 KO cell morphology changes. In support of this, we observed changes in the transcription of a broad spectrum of genes in the SUMO2 KO cells that could contribute to changes in cell morphology. These include genes involved in the epithelial to mesenchymal transition (EMT), the extracellular matrix, integrin-cell surface interactions and genes coding for Wnt-family proteins. Notably, canonical EMT-related genes were predominantly down-regulated in the SUMO2 KO cells, suggesting that SUMO2 may regulate a non-canonical EMT-like process. Lastly, multiple studies have revealed that the SUMO paralogs modify intermediate filaments, microtubule associated proteins and actin regulatory proteins (175, 253). Collectively, since sumoylation regulates cytoskeletal proteins, signaling pathways and gene expression, the morphology changes observed

in the SUMO2 KO cells are likely the sum result of the effects of sumoylation at multiple levels, from genes to proteins.

With regard to the observed changes in the global transcriptome of the KO cells, previous studies have enhanced our understanding of how sumoylation regulates gene expression. For instance, many studies have identified that SUMO1 or SUMO2/3 modification of a specific transcription or chromatin remodeling factor alters its activity, localization, stability or binding partners (201, 207, 208, 211). Many of these studies also noted that SUMO modification of these factors was associated with a decrease in gene expression, which may be driven by sumoylation-mediated recruitment of co-repressors or co-repressor complexes (218, 249). Conversely, other studies have found that sumoylation of a specific transcription factor or chromatin remodeling protein enhances gene expression, thereby indicating that the effects of sumoylation are likely context dependent and promoter specific (207, 208, 218). In line with this reasoning is a relatively new hypothesis which suggests that sumoylation aids in transcription factor binding-site selection (217). However, the effects of the SUMO1 and SUMO2 paralogs on regulating global gene expression has yet to be reported, thus unique paralog-specific roles in transcription remain poorly understood. To that end, we uncovered three important findings about the paralog-specific roles of SUMO1 and SUMO2 in transcription using RNA-seq.

First, we identified uniquely altered genes in both of the KO cells, revealing that the SUMO1 and SUMO2 paralogs each have non-redundant roles in transcription. Consistent with a previous ChIP-Seq study, which reported nearly double the number of SUMO2 binding sites as compared to SUMO1 in proliferating human fibroblasts (202), we found almost twice as many uniquely altered genes in the SUMO2 KO cells

than in the SUMO1 KO cells. Furthermore, we identified nearly 15-times the number of unique genes with a four-fold change in expression in the SUMO2 KO cells as compared to the SUMO1 KO cells. The finding that genes with larger changes in expression levels occurred in the SUMO2 KO cells is consistent with the finding that SUMO2 can more potently inhibit transcription of the glucocorticoid receptor as compared to SUMO1 (215). Moreover, re-expressing SUMO2 in the SUMO2 KO cells rescued gene expression changes to WT-like levels. Conversely, overexpressing SUMO1 in the SUMO2 KO cells did not affect gene expression values, revealing that SUMO1 is unable to functionally compensate for SUMO2. Taken together, these findings indicate that although both SUMO paralogs have important and non-redundant roles in regulating gene expression, SUMO2 has a broader role in this essential cellular process.

Secondly, although sumoylation was once considered a repressive modification, we identified nearly equal numbers of up and down-regulated genes in the SUMO2 KO cells. In contrast, we identified nearly twice as many down-regulated genes in the SUMO 1 KO cells, but only for genes with a four-fold change in expression. Taken together, this data reveals that SUMO2 can function as either an activator or repressor, whereas SUMO1 may function more as an activator for the limited number of genes that it regulates.

Finally, gene set enrichment analysis (GSEA) revealed that although a majority of gene sets were enriched in both the SUMO1 and SUMO2 KO cells, the same gene sets often contained genes with opposing expression levels. Specifically, among a majority of GSEA categories, we observed a clear trend of down-regulated genes in the SUMO1 KO cells and up-regulated genes in the SUMO2 KO cells. We were

interested in exploring this further, specifically for the histone and immune-response gene categories, as discussed below.

As mentioned, we observed an increase in histone gene expression levels in the SUMO2 KO cells, and a corresponding decrease in the SUMO1 KO cells. Consistent with this finding, previous ChIP-seq analysis revealed that both SUMO1 and SUMO2 bind to histone gene promoters (202). Moreover, histone gene expression levels significantly increase upon knockdown of the Ubc9 SUMO E2 conjugating enzyme or the SUMO E3 ligase, PIASY (202). Our findings reveal that this previously observed increase in histone gene levels may be due specifically to the loss of SUMO2 modified regulators. Histone gene expression is tightly controlled by the cell cycle, with an approximate 35-fold increase in expression occurring specifically during S-phase (254). The lack of a significant difference in S-phase distribution between the SUMO KO cells, as quantified by flow cytometry (Figure 12A and B), indicates that the observed changes are being driven by something other than the cell cycle. Another potential mechanism could involve SUMO1 and SUMO2 mediated recruitment of transcriptional activators or repressors to specific genomic loci (201, 218). Interestingly, histone genes are regulated by a complex of proteins that form at histone loci, called histone locus bodies (255). Given the observed defects in PML-NB integrity in both the SUMO1 and SUMO2 KO cells, it is possible that the paralogs may have unique yet conserved functions in regulating membrane-less organelles, such as PML-NBs and histone locus bodies, however, further work is needed.

Genes involved in immune response processes were also uniquely up-regulated in the SUMO2 KO cells and down-regulated in the SUMO1 KO cells. These findings are consistent with previous studies that have reported an increase in immune response

genes upon abrogation of global sumoylation (132, 251, 252). There is also evidence that sumoylation can regulate innate immunity by altering the production of type I IFNs upon viral activation (164, 256). With regard to the paralogs, it has been reported that the combined loss of SUMO2 and SUMO3 drives a potent type I interferon (IFN) response, indicating that SUMO2 and SUMO3, but not SUMO1, are redundant and potent negative regulators of the type I IFN response (132). Furthermore, genes involved in systemic lupus erythematosus, an autoimmune disease with a potential association of innate immune response processes in disease pathogenesis (131), were significantly upregulated and enriched in the SUMO2 KO cells. These findings have important public health implications, especially in light of a global sumoylation inhibitor, TAK-981, that is currently in four concurrent clinical trials. This SUMO E1 inhibitor is being evaluated in the clinic for metastatic solid tumors, head and neck cancers and relapsed CD20-positive Non-Hodgkin Lymphoma, and for cancer patients who test positive for COVID-19 (140-143). The predicted success of this inhibitor is based, in part, on the finding that it increases the type I IFN response and thus helps the body attack the cancer through potential immune-activating and anti-tumor activities. The collective findings from our study and others suggest that selectively targeting SUMO2 and SUMO3 may be sufficient to illicit a type I IFN response. In summary, the SUMO paralogs have unique and non-redundant functions in regulating thousands of unique genes. More work is needed to understand the mechanism driving this regulation, and to better understand the specific contexts for SUMO activating and repressing functions.

What paralog-specific attributes could help explain their unique and non-redundant functions in regulating gene expression, nuclear body integrity, cellular morphology and responses to cellular stress? Defining features are likely to include differences in

the SUMO1 and SUMO2 SIM interaction domains that confer paralog-specific interactions (52, 53, 145, 257). For instance, lysine residues unique to the SUMO2 SIM interaction domain are critical for the intrinsic ability of SUMO2 to function as a strong transcriptional repressor, relative to SUMO1 (54, 178, 215). This repressive ability is attributed, in part, to SUMO2/3 specific binding and localization of transcriptional repressors and repressive complexes to promoters, which is mediated through SIM interactions (207, 218). Non-covalent paralog-specific interactions also mediate, in part, the cellular stress response and nuclear body integrity, thus differences in SUMO-SIM interactions could help explain other observed phenotypes of SUMO1 and SUMO2 KO cells as well (188, 258).

Moreover, the ability of SUMO2/3 to efficiently form polymeric-SUMO2/3 chains represents a major difference from SUMO1, which is preferentially conjugated to substrates as a monomer. Polymeric-SUMO2/3 chains can act as either a degradation signal, or function to stabilize protein complexes (10, 180). Under conditions of stress, polymeric-SUMO2/3 chains are quickly and reversibly formed on target proteins involved in stress response pathways, such as heat shock proteins, apoptotic factors and DNA repair proteins (40, 69, 178). These target proteins are likely recruited to sites of stress-induced damage, such as DNA double strand breaks or mis-targeted proteins, in a SUMO-dependent manner. Intriguingly, we report that in the absence of SUMO1 and SUMO2, cells are more resistant to genotoxic stress than WT cells. This suggests non-redundant roles for the paralogs in the repair of DNA double strand breaks, potentially through paralog-specific spatiotemporal regulation of repair factors. One hypothesis is that monomeric SUMO1 modification may occur upstream of SUMO2/3 modification, and function to recruit repair factors to sites of damage, as observed by the SUMO1 modification of DNA repair protein FEN1 (259). This could

subsequently be followed by polymeric-SUMO2/3 modification, which through the activity of SUMO targeted ubiquitin ligases, such as RNF4, function to turn over repair factors, as observed for PCNA (10, 69, 260). Given that thousands of proteins are affected by the loss of SUMO1 and SUMO2, the mechanisms behind this observation are likely to be complex and require further exploration (19). Of note, such future studies would be enhanced by SUMO1 rescue cells, however our efforts to generate them thus far have not been successful, possibly due to toxic effects of SUMO1 misregulation (116).

Whether the observed phenotypes are a direct or indirect response to loss of the SUMO paralogs, or due to cross-talk between biological processes, our data indicate that the paralogs have unique and non-redundant roles in many cellular functions. An important next step will be to explore molecular mechanisms underlying paralog-specific changes by going back into WT cells. More specifically, what are the effects of SUMO2 on the cytoskeleton and regulatory pathways that control cell morphology? How do SUMO1 and SUMO2 individually sensitize cells to treatment with hydroxyurea? What are the targets and consequences of SUMO1 and SUMO2 modification on proteins localized to PML-NBs? Do SUMO1 and SUMO2 mediate association of transcription factors with unique regulatory proteins at gene promoters?

Ultimately, it will also be important to explore the unique functions of SUMO paralogs in various tissues. Our gene expression analysis revealed that relative SUMO paralog expression levels are similar across hundreds of cell lines, regardless of tissue source, which is in contrast to varying paralog levels across normal human tissues. For instance, average SUMO2 expression levels are consistently higher than SUMO1 levels across all analyzed cell lines from multiple tissue sources, including the liver.

However, in normal human tissues, SUMO1 expression is higher than SUMO2 in the liver and nine other organs. This suggests that the cellular environment strongly influences SUMO paralog expression, with cells in culture potentially adapting SUMO paralog expression levels optimized for growth conditions in 2-D culture. This highlights the importance of using the appropriate system to study functions of sumoylation. While basic information about the paralogs can be inferred using cell culture, as scientists have been doing for years and as we demonstrated in this study, the roles of the paralogs in specific tissues will likely need to be studied within an organism. For instance, diverse expression patterns of SUMO pathway genes in the mouse retina and in human placental tissues have been identified and used to define biologically significant roles for the SUMO paralogs in these tissues (154, 237).

Lastly, *SUMO3* expression levels were higher in a majority of normal human tissues than *SUMO1* and *SUMO2*. This was an unexpected finding, based on the absence of any obvious phenotypes in SUMO3 KO mice, and the apparent low levels of SUMO3-modified proteins in cell culture (Figure 9) (155). However, an observed increase in SUMO3 expression levels in adult mice as compared to young mice suggests that SUMO3 is important in post-developmental processes (155). Given the high degree of sequence and structural similarities between SUMO2 and SUMO3, this observation is certainly interesting and reveals that the functional contributions of SUMO3 are worth further exploration. In summary, we have provided evidence that SUMO1 and SUMO2 have unique roles in regulating cellular morphology, nuclear body integrity, response to various cellular stressors and global gene expression. Our work provides a foundation for future medical and scientific SUMO-focused endeavors aimed at providing new avenues for targeted therapies.

Acknowledgements

We would like to thank all current and previous members of the Matunis lab for helpful comments and insight shared throughout the course of this study. This work was supported by a fellowship from the Taiwan National Science Council (awarded to WCY.), the cancer training grant (awarded to DMB) and a grant from the National Institutes of Health (GM060980, awarded to MJM).

Materials and Methods

Human Cell Line and Tissue Expression Analysis

Normalized gene RPKM values from 528 cancer cell lines were downloaded from The Broad Institute Cancer Cell Line Encyclopedia (CCLE) using the 02-JAN-2019 release (239). Student's t-tests were used for pairwise comparisons and p-values are listed in the legend of Figure 8. Normal human tissue data were downloaded as normalized gene transcript per million (TPM) values from the Genotype-Tissue Expression (GTEx) Project Version 8, which is supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH, and NINDS. The number of samples available for each tissue are labeled in the figure. Heatmaps were made in Rstudio, using ggplot2 and the gganatogram package (261).

CRISPR-Cas9 Genetic Knockout and Sequencing Validation

Gene specific knockout of SUMO1 and SUMO2 in U2OS cells using CRISPR-Cas9 was performed according to a previously published protocol (262). In brief, single guide

RNA (sgRNA) was designed by the CRISPR design tool (http://crispr.mit.edu) as following: SUMO1 5'-TCCCTCCTCCCTGCGCGAAG-3';SUMO2 5'-CCTCACCTGTCGTTCACAAT-3'. sgRNA was cloned into pSpCas9(BB)-2A-GFP vector using BpiI enzyme sites (Thermo Scientific), and the vector was transiently transfected into U2OS cells using X-tremeGENE HP reagent (Roche) according to manufacturer's protocol. Transfected cells expressing GFP were sorted as single cells into 96-well plates by FACS at The Bloomberg Flow Cytometry and Immunology core. CRISPR-Cas9 introduced mutations were identified using the Clonetech Guide-it Indel Identification Kit (Clonetech Catalog Number: 631444), following the user manual. Genomic primer sequences for SUMO1 and SUMO2 are listed in Table 4**,** with Clonetech regions of homology indicated. Ten individual SUMO1 and SUMO2 colonies were sent for Sanger DNA sequencing at the Johns Hopkins University Genetic Resources Core Facility. Aligned sequence reads surrounding the mutation sites are in Supplemental Figure 2 and 3. U2OS WT, SUMO1 KO and SUMO2 KO cells were confirmed free of mycoplasma contamination using the Promokline PCR Mycoplasma Test Kit I/C (Part Number: PK-CA91-1024) following the vendor protocol.

Cell Lines and Cell Culture Conditions

U2OS WT, SUMO1 KO, SUMO2 KO, SUMO2KO+SUMO2 and SUMO2KO+SUMO1 cells were grown at 37^0C, 5% CO_2 in Gibco DMEM (PN: 11965-092) supplemented with 10% Fetal bovine serum (FBS, Atlanta Biologics PN: S11550).

Generation of Stable Rescue Cell Lines

pTWIST CMV Puro plasmids encoding precursor SUMO1 or SUMO2 expression were purchased from TWIST Biosciences (Supplemental Figure 4). $2x10^5$ SUMO2 KO cells grown overnight were transfected with 1ug of pTWIST plasmids using Lipofectamine 2000, following their standard protocol (Invitrogen Lot no. 1881535). Fresh Gibco DMEM medium with 10% FBS was supplied after 6 hours of incubation. Puromycin (Sigma Cat No: P8833) selection was performed 48 hours post-transfection at a final concentration of 2ng/mL for 4-5 days. Stable rescue cell lines were obtained by single-cell cloning and maintained in 1ng/mL puromycin-containing DMEM medium for a month, and then grown in standard conditions, per above. SUMO paralog expression levels in the rescue cell lines were validated by immunofluorescence microscopy and immunoblotting with corresponding antibodies, per Table 5.

Immunoblotting and Semi-Quantification of SUMO Levels

Cells were lysed with 2x Laemmli Buffer (4% SDS, 20% glycerol, 125mM Tris-Cl pH6.8, 10% 2-mercaptoethanol, 0.02% bromophenol blue) and denatured at 95°C for 5 minutes. 10μL of whole cell lysate was loaded onto 10% SDS-polyacrylamide gels, run at 120V for 2 hours in a Biorad Mini-Protean vertical electrophoresis chamber, and transferred onto PVDF membrane in a Biorad Mini Trans-Blot cell at 100V, 4°C for 2 hours. Membrane was briefly washed with 1xTBS and blocked in 5% milk (in 1xTBS) for 1 hour at room temperature with gentle shaking. Membranes were incubated with anti-SUMO primary antibodies (SUMO1: [1:1000]; SUMO2: [1:800]; Tubulin: [1:10,000]) overnight at 4°C, washed 3x10 minutes with 1xTBS-T, incubated with HRP conjugated secondary antibodies ([1:10,000]) in 5% milk for 1 hour at room temperature, followed by washing for 3x10 minutes with 1xTBS-T. Membranes were

activated using Amersham ECL prime western blotting detection reagent (Cat No. 45-002-401) and exposed using a medical film processor (Konica Minolta Medical & Graphic Inc. Model: SRX-101A).

Developed films were scanned and imported into FIJI image processing software for semi-quantitative measurements of relative SUMO paralog expression (240). Images were converted into grayscale and SUMO signal intensities were measured using methods described by https://lukemiller.org/index.php/2010/11/analyzing-gels-and-western-blots-with-image-j/. SUMO signal densities were normalized to corresponding tubulin loading controls and calculated by dividing knockout and rescues cell values by WT values. Graphs were created using Prism software.

Immunofluorescence Microscopy and Quantitative Cellular Morphology Analysis

For SUMO immunofluorescence staining and SUMO+tubulin co-staining, cells were seeded at $2.0x10^5$ cells/coverslip in a 6-well dish and grown overnight. Harvest and staining of the cells were performed at room temperature as follows: cells were washed with 1xPBS and fixed in 4% paraformaldehyde for 20 minutes, followed by permeabilization in 0.05% Triton-X-100 (in 1xPBS) for 20 minutes. Cells were then incubated with anti-SUMO and anti-tubulin primary antibodies (Table 5) for 1 hour in a humidity chamber, washed in 1x PBS-T, and incubated with Alexa fluorescent secondary antibodies (Table 5) for 40 minutes. Coverslips were then mounted using Fluoroshield Mounting Medium with DAPI (Abcam, ab104139). Microscopy images were taken using an upright Zeiss Observer Z1 fluorescence microscope with an Apotome VH optical section grid. Non-saturated representative images showing

SUMO protein levels and morphology of each cell line were taken using a 40x objective and exported as tif files from the AxioVision Release 4.8 software.

Quantitative analysis of cellular morphology was performed as above, except cells were seeded at $1.8x10^5$ cells/coverslip in a 6-well dish. Quantitative measurements were done in FIJI software (240) by manually outlining individual cells. Cellular aspect ratio, area and circularity were measured using built-in measurement functions in the FIJI analyze menu, and images were calibrated into microns before measuring. For each cell line, at least 800 cells from 3 independent experiments were measured and data were imported into Prism software for statistical analysis. The Kruskal-Wallis non-parametric test was used for statistical comparison and calculation of p-values. Corresponding p-values are displayed on each graph for clarity.

Flow Cytometry

$4.5x10^6$ cells were fixed in ice cold 70% EtOH and stained with Propidium Iodide in triplicate, following a standard protocol (263). Cells were analyzed using a BD LSRII flow cytometer at the JHU Flow Core. BD FACSDiva acquisition software was used to acquire $5.0x10^4$ single events per sample, and G0 cells were centered on 100. FloJo version 10.6.1 was used for analysis. Statistics were calculated using an ANOVA and p-values are labeled on the final plots, made using ggplot in RStudio (261).

Quantitative Nuclear Body Imaging and Analysis

Cells were seeded at $2.5x10^5$ cells/coverslip in a 6-well dish and grown overnight. Cells were rinsed in room temperature (RT) 1x DPBS (Gibco PN: 14190-144), fixed in 3.5%

paraformaldehyde for 7 minutes at RT, permed in 0.05% Triton-X-100, for 20 minutes at RT and incubated for 1 hour in a RT humidity chamber with primary antibodies Table 5. Cells were washed in 1x PBS-T, and incubated for 35 minutes in a RT humidity chamber with secondary antibodies (Table 5). Microscopy images were taken using an upright Zeiss Observer Z1 fluorescence microscope with an Apotome VH optical section grid. Non-saturated 16-bit gray images were exported from the AxioVision Release 4.8 software and opened in FIJI (240). Nuclei (DAPI) and foci (dsRED and/or GFP) signal thresholds were set using the RenyiEntropy algorithm and the Speckle Inspector function of the Biovoxxel plug-in was used to quantitate (264, 265): the number of foci per nuclei, foci signal intensity and foci perimeters. Non-parametric Wilcoxin test was used to calculate p-values in R and graphs were generated using ggplot2.

Cellular Viability Analysis

2 x 10^3 cells per well were plated into 96-well plates in 100µL of media and grown overnight. Each cell line was seeded in triplicate for each dose of the drug treatment. After treatment, each well was washed once with 1x DPBS, and then 100 µL of DMEM (without phenol red, Gibco 21063-029) containing 10% FBS and 10 µL of 12mM MTT (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide) were added to each well, including a negative control of 10 µL of the MTT solution added to 100 µL of medium alone. After 4 hours of incubation at 37°C with 5% CO_2, 80 µL of the media was removed from each well, and 50 µL of DMSO was added to solubilize the metabolized insoluble formazan product. After incubation at 37°C for 10 minutes, 96-well plates were analyzed using a plate reader (BioTek Synergy HT) and Gen5 microplate software (2.00.17) at an absorbance of 540 nm. For the analysis, the

negative control signal was deducted from all wells, before relative cell viability was calculated for each treated group as a percentage of the untreated group for each cell line. Reagent information is in Table 3.

Transcriptome Analysis and Data Visualization

Three biological replicates of each authenticated and validated U2OS parental, SUMO1 KO and SUMO2 KO cell lines were sent to Novogene Co., Ltd for total RNA isolation, QC and library preparation. Transcriptomic data for the three cell lines with three biological replicates were generated using an Illumina HiSeq 4000 sequencer. Paired-end reads were obtained with a read depth of over 60 million reads per sample. The reads were cleaned and mapped to the reference genome using STAR, HTseq, Cufflink and Tophat programs. This resulted in a total of 48,162 Ensembl reads. Of those, 23,758 were mapped to Entrez gene IDs and used in downstream analysis. Genes were filtered to keep those that had approximately 10 or more read counts in at least all three replicates of one cell type, which resulted in a total of 14,999 genes for downstream analysis (266). Reads were evenly distributed throughout the genome and both a multidimensional scaling (MDS) plot and calculated Pearson correlations among replicates demonstrated highly consistent read counts with minimal variance between biological replicates (Supplemental Figure 5).

The Rstudio Karyoplotter package was used to visualize mapped DEGs along human chromosomes (267). Pathway enrichment analysis using GSEA (version 4.0.3) from the Broad Institute and visualization of the data using Cytoscape and EnrichmentMap were performed following published protocols (250, 268). The GSEA algorithm calculates GO enrichments from a list of global gene expression values, not just those

that meet a specific threshold criteria. Specifically, GSEA used the expression values of the same 14,999 genes from the SUMO1 and SUMO2 KO cells to generate two ranked lists of genes (one for each KO cell as compared to WT), which were then tested for gene set enrichments. For the analysis, .gct data of FPKM values for all 3 replicates for all 3 samples, and a phenotypes .cls file were loaded into GSEA. 1000 permutations were used, collapsed to match the Human NCBI Entrez GENE ID MSigDBv.7.1.chip platform, otherwise default settings were used. Results were compiled and analyzed in RStudio using GeneEnricher. Cytoscape (version 3.8.0 using Java 11.0.6) was used to visualize Interferon and collagen gene sets.

qRT-PCR for DEG Validation and Rescue Experiments

Cells were seeded at 5.0×10^5 cells/well in 3.5mm dishes and grown overnight. Total RNA was extracted using the Sigma GenElute Mammalian MiniPrep kit (Sigma PN: RTN70) following the vendor's protocol. Extracted RNA was analyzed by nanodrop for concentration and purity. cDNA was generated using the New England BioLab ProtoScript First Strand cDNA Synthesis Kit (NEB PN: E6300S) with 250ng of RNA and following the vendors protocol. Poly-$d(T)_{23}$ VN primers were used to generate cDNA for all genes except for histone genes, for which we used random-hexamer primers which were also included in the NEB kit. We used random hexamer primers since histone genes do not contain polyA tails, and we wanted to properly verify our RNA-seq findings. The qPCR reaction was performed using Applied Biosystems PowerUp SYBR green master mix (PN: A25742) following the vendors recommended protocol. qPCR runs were performed using Applied Biosystems Quant Studio with Quant Studio v1.3.1 software. Gene expression was calculated from three biological

replicates, each run in triplicate, using the Ct method and GAPDH as a validated housekeeping gene. A list of primer sequences can be found in Table 4.

Figures

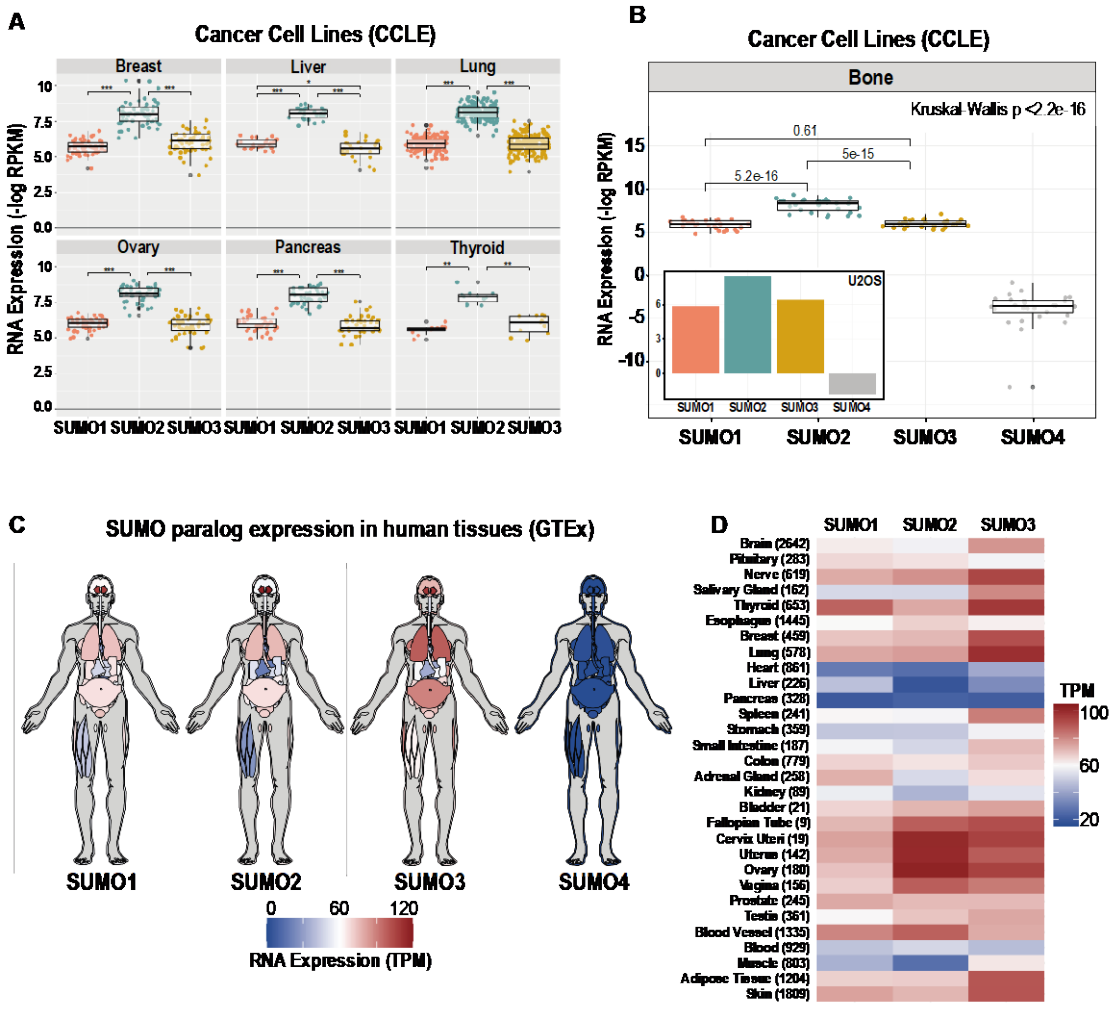


Figure 8. SUMO paralog expression levels vary among human cell lines and tissues. A) SUMO paralog expression from 467 human cancer cell lines derived from 6 different tissues. Significant P-values (<0.05) were determined using a Students t-test, where: * = 0.027; ** = 2.8e-6; *** <1.6e-14. B) SUMO paralog expression from 61 human bone cancer cell lines. Individual P-values were calculated using a Students t-test and are labeled. The inlay data is specifically from U2OS cells, no statistics are reported since expression values come from one cell line. Data in A and B come from CCLE and are reported as -$\log_{10}$ transformed RPKM values. C) Anatomical heat maps of *SUMO1-4* mRNA expression (Transcripts per Million (TPM)) in normal human tissue data from GTEx. D) A detailed heatmap of *SUMO1-3* expression in normal human tissues, ordered generally from the head down. The number of samples in A,B and D are labeled in parentheses.

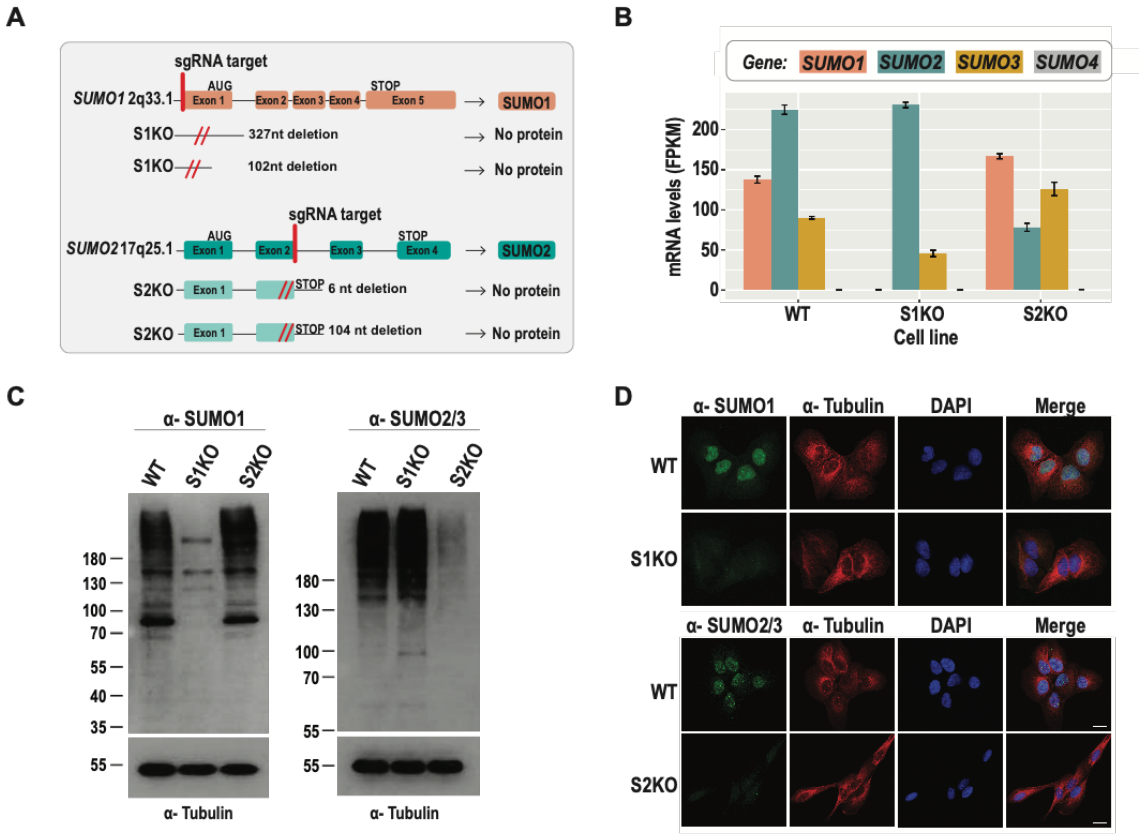


Figure 9. Generation of SUMO1 and SUMO2 KO cell lines. A) Schematic of CRISPR-Cas9 strategy and sanger sequencing results. B) *SUMO1-4* mRNA expression values in WT and KO cells, measured by RNA-sequencing. C) Representative western blots of WT and KO whole cell lysates show significant reduction in SUMO1 and SUMO2/3 conjugated proteins in respective KO cells. Tubulin is used as a loading control. D) Immunofluorescence microscopy images show loss of SUMO1 or SUMO2/3 specific signal in WT and respective KO cells. Cells are co-stained with tubulin to show cellular morphology.

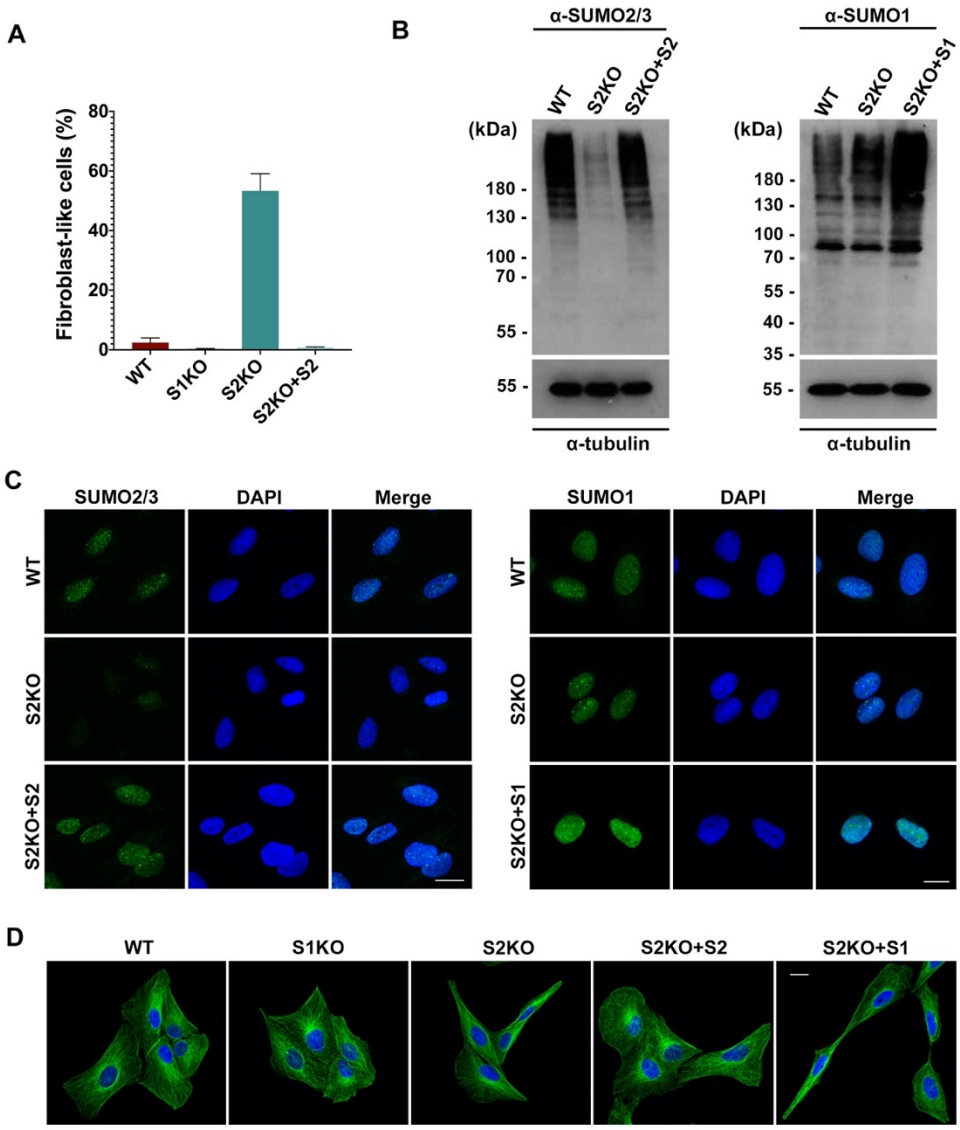


Figure 10. SUMO2 has a unique role in regulating cellular morphology. A) Percentage of fibroblast-like cells counted from >2400 cells from three independent experiments as analyzed by immunofluorescence microscopy. B) SUMO2 re-introduction in S2KO+S2 cells, and SUMO1 overexpression in S2KO+S1 cells were validated by immunoblotting using antibodies specific for SUMO2/3 or SUMO1. Tubulin was used as a loading control. C) Expression levels of SUMO2/3 and SUMO1 (green) were assessed by immunofluorescence microscopy use specific antibodies and DAPI (blue) counterstaining. D) Morphology was analyzed by anti-tubulin (green) immunofluorescence microscopy with DAPI (blue) counterstaining. Scale bars: 20μm.

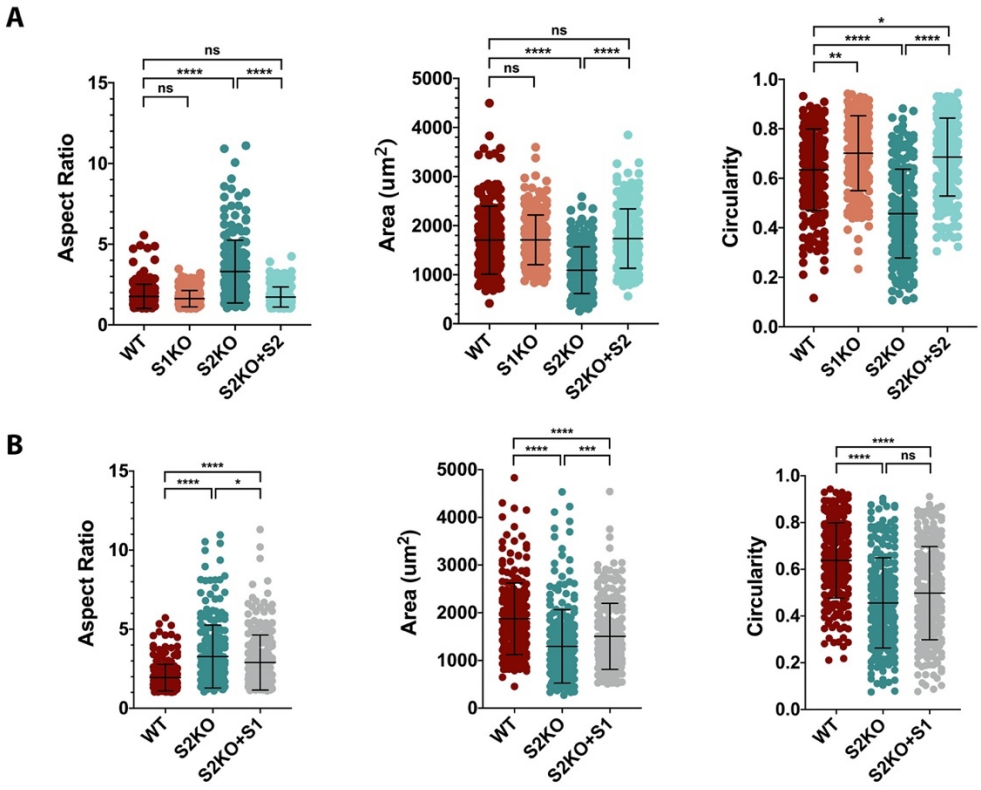


Figure 11. Rescue of S2KO Cells Reveals SUMO-specific role in cellular morphology. A) Quantitative shape analysis revealed changes in the average cellular aspect ratio, area and circularity in indicated cell lines. Over 200 cells from three independent experiments were analyzed for each cell line. Error bars represent standard deviation. P values were calculated by a Kruskal-Wallis test using where: **** p= 0.0001, *** p = 0.001, * p = 0.05, ns =p>0.05. B). Aspect ratio, area and circularity of WT, S2KO and S2KO+S 1 cell lines were analyzed as in A).

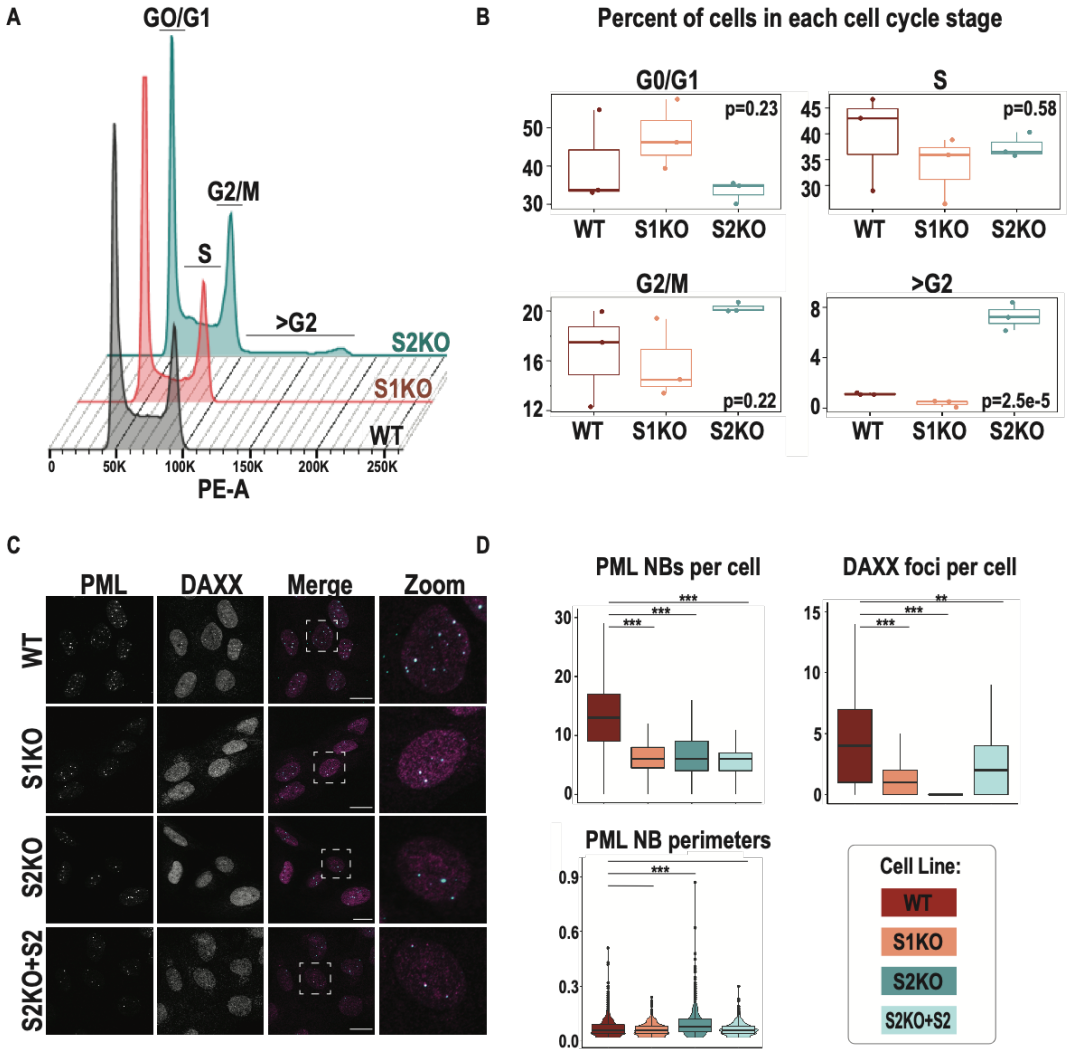


Figure 12. Cell cycle and nuclear body regulation in SUMO KO cells. A) Representative overlay of flow cytometry histograms from WT, S1KO and S2KO cells. B) Quantitation of the percentage of cells in G0/G1, S, G2/M and >G2 cell cycle stages. P-values, labeled, were calculated using an ANOVA. C) Representative immunofluorescence microscopy images using antibodies specific for PML and DAXX. A merged image with PML in teal and DAXX in magenta is shown in the third panel. A zoomed-in view of a merged cell (outlined) is shown in the final panel. D) Quantitation of PML nuclear bodies, DAXX foci and PML NB perimeter size estimates are shown for each cell type. The number of cells analyzed per cell line was between 174-195. P-values for each cell line as compared to WT were calculated using an unpaired Wilcoxon test, where: ** = p=1x10-8; and *** = p <2.2x10-16.

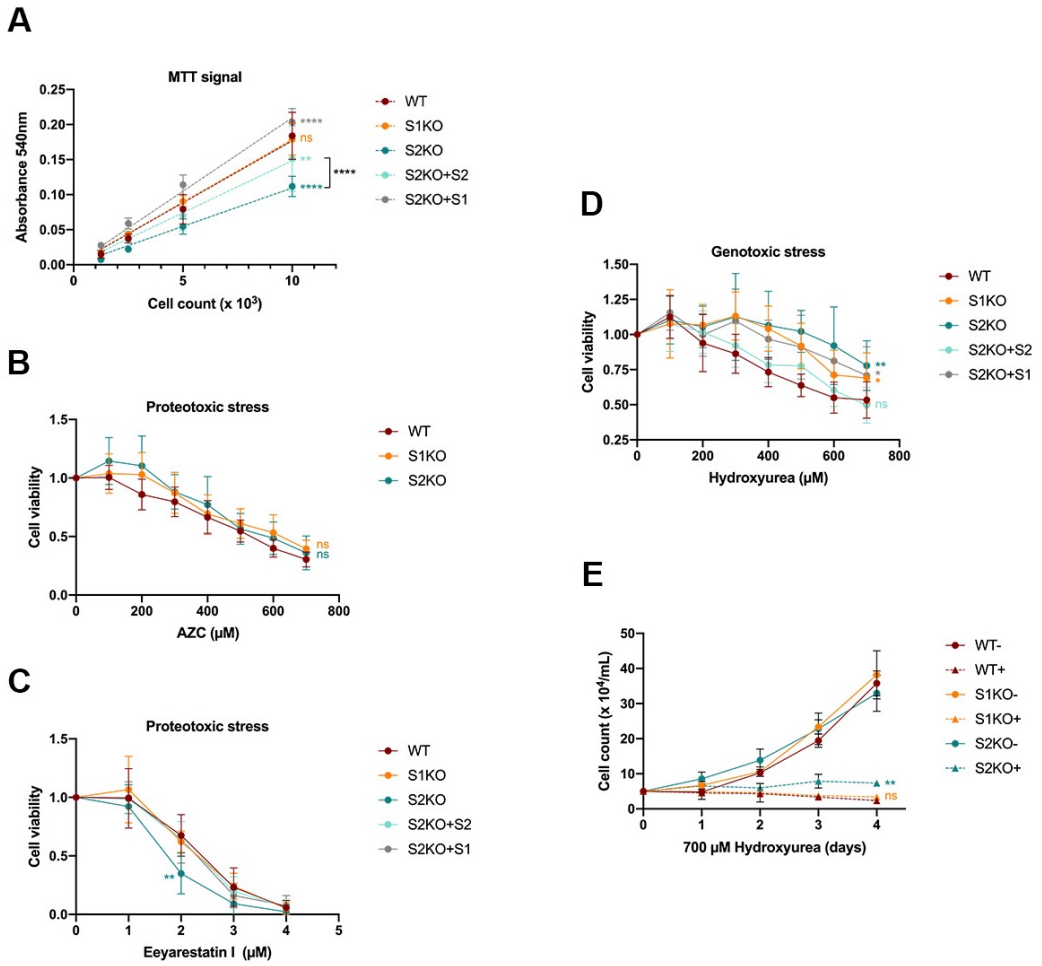


Figure 13. SUMO paralogs have non-redundant functions in response to cellular stress. A) Baseline MTT assay of cells at various confluency. Simple linear regressions were calculated for each cell line and slopes for each regression were compared, with the R^2 reported as 0.91 (WT), 0.96 (S1 KO), 0.94 (S2KO), 0.81 (S2KO+S2), and 0.96 (S2KO+S1). B,C,D) Cell viability measurements. Cells were treated with indicated doses of B) AZC for 72 hours, C) Eeyarestatin I for 48 hours, and D) HU for 72 hours followed by MTT assays. Relative cell viability is calculated as the fraction of MTT signal at each dose of the drug compared to untreated control cells. E) Cell growth analysis. Cells were treated with 700μM HU (dash line) or without HU (solid line) for up to 4 days. Viable cells were counted at each time point and plotted. Error bars: standard deviation, n=3; ns: not significant; * p<0.05, ** p<0.01, **** p<0.0001.

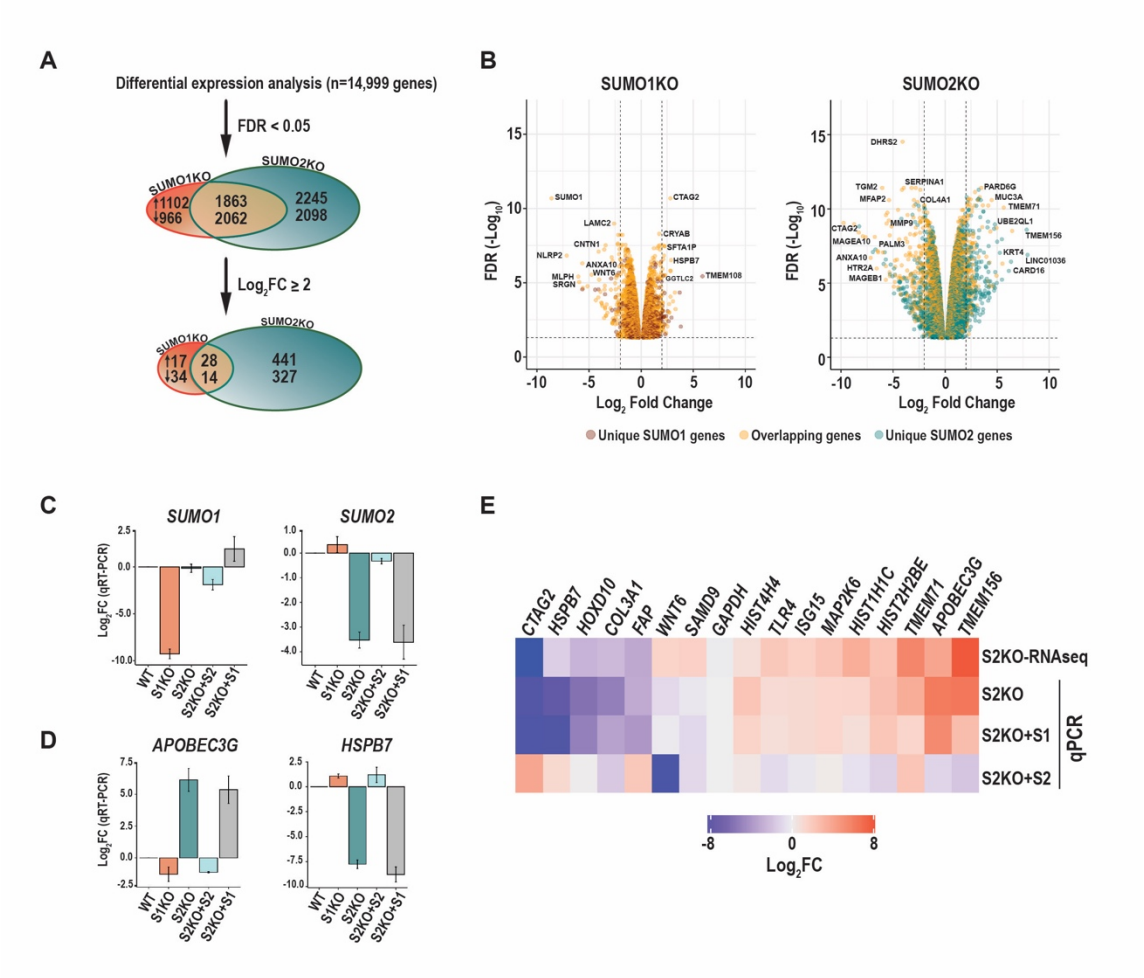


Figure 14. Overview of RNA-sequencing results of SUMO KO cells. A) Venn Diagram showing the numbers of unique and overlapping up and down-regulated differentially expressed genes (DEGs) in the SUMO1 KO and SUMO2 KO cells, at two significance thresholds: False Discovery Rate (FDR) <0.05 and FDR < 0.05 + Log_2 fold change (FC). B) Volcano plots of unique and overlapping SUMO1 and SUMO2 KO cell DEGs. The horizontal dashed line represents FDR <0.05 and the vertical dashed lines represent Log_2 fold change values of -2 and +2. C) Validation of SUMO1 and SUMO2 gene expression values by qRT-PCR in WT, KO and rescue cell lines. D) Representative bar plots of Log_2FC expression values of up and down-regulated genes, tested by qRT-PCR. E) Heatmap summarizing SUMO2 KO and rescue cell line Log_2FC values for genes tested by RNA-seq and qRT-PCR.

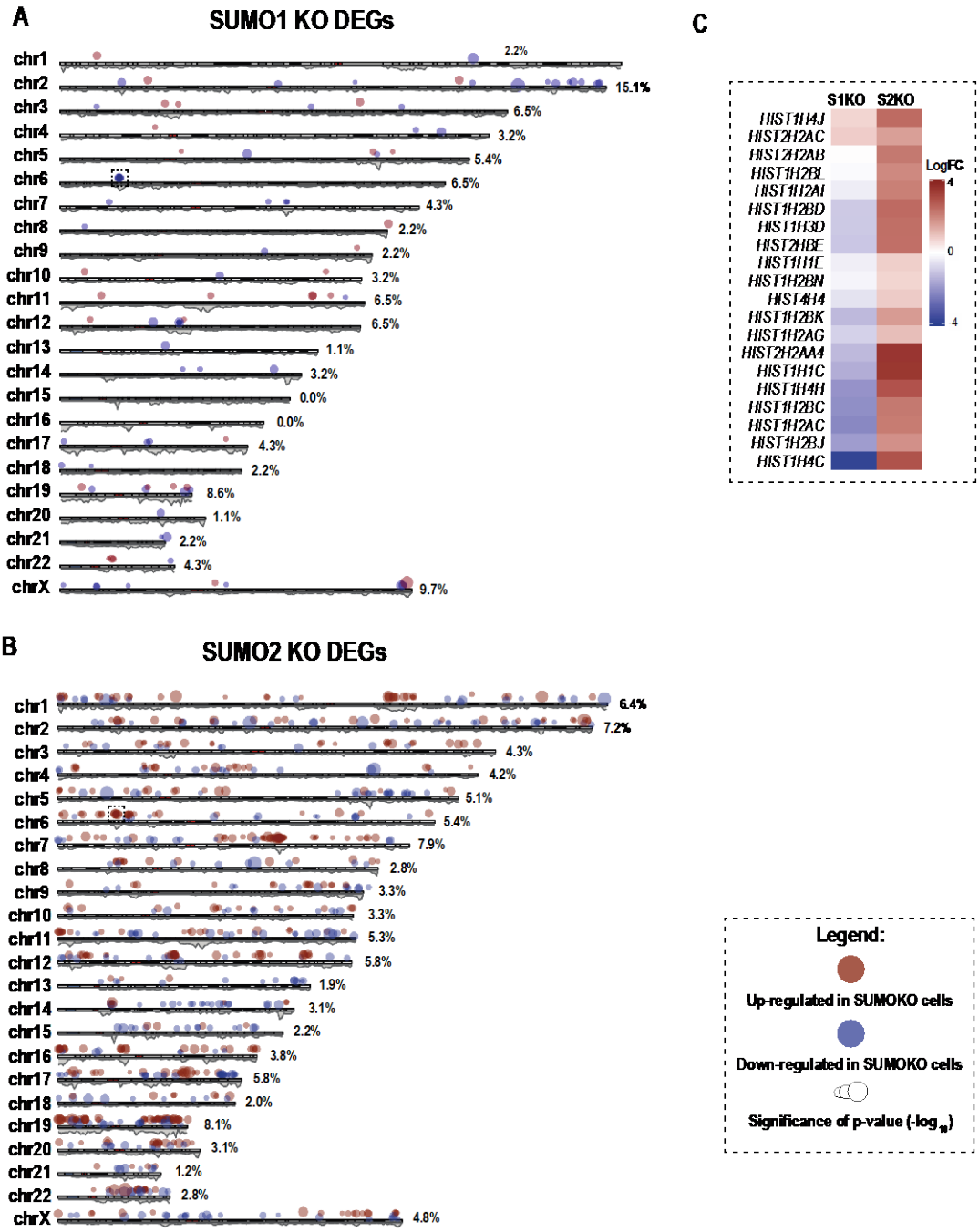


Figure 15. Karyoplot of SUMO KO cell DEGs and histone genes. A and B)The genomic locations of up and down-regulated genes in A) SUMO1 KO cells and B) SUMO2 KO cells. DEGs are represented by red and blue dots, respectively. The size of the dots corresponds to significance, with a bigger dot representing a more significant change in gene expression. The gray plot below each chromosome represents gene density at each loci. Percent of DEGs per chromosome is labeled to the right of each chromosome (# of DEGs on the chromosome / total # of DEGs) x 100. The major histone gene locus on chromosome 6 is highlighted. C) A heatmap of histone gene expression as measured by RNA-seq.

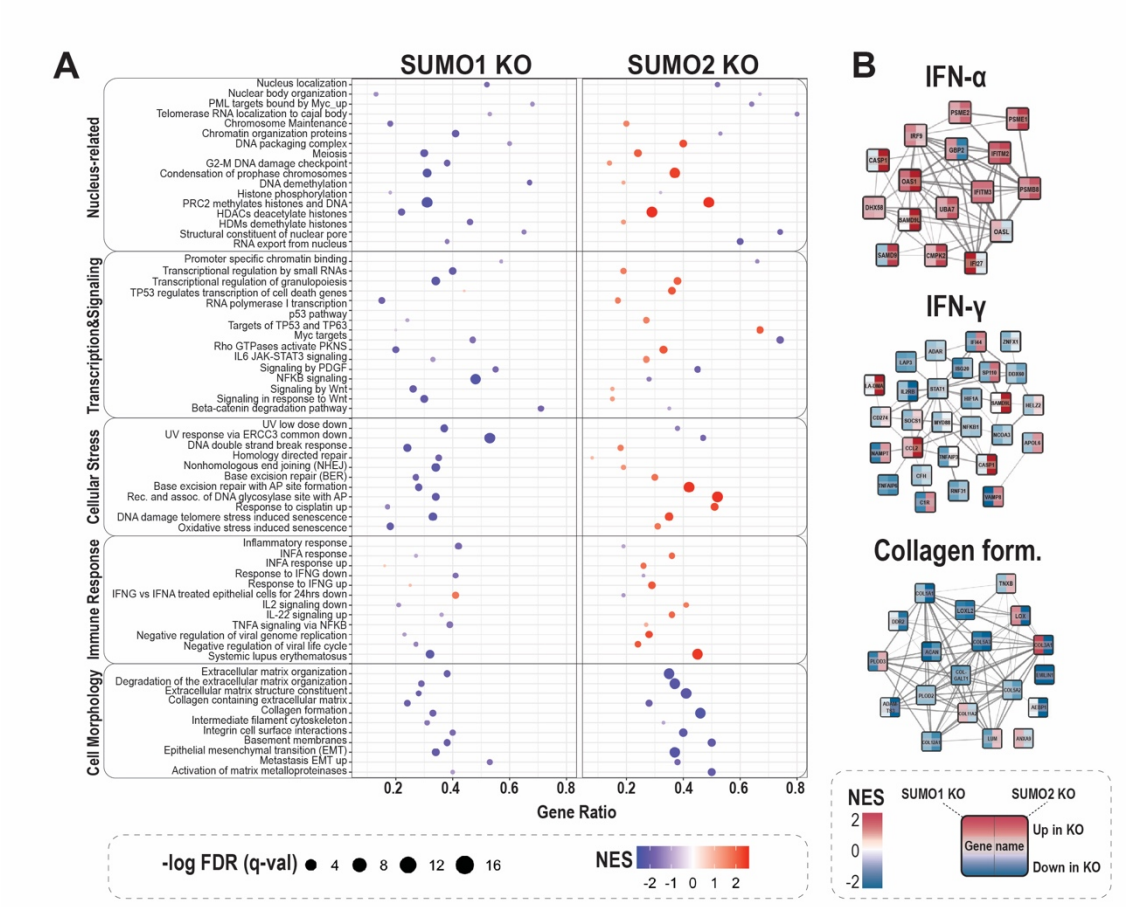


Figure 16. Gene Set Enrichment and Cytoscape Analysis of SUMO KO cells. A) Enriched gene sets for SUMO1 KO and SUMO2 KO cell DEGs. The color of the dot corresponds to the GSEA normalized enrichment score (NES), which quantitatively represents the contribution of down-regulated (blue) and up-regulated (red) genes to each gene set. The size of the dot corresponds to the significance, with a bigger dot representing a more significant enrichment. B) STRING protein interaction networks for IFNα, γ, and collagen formation gene sets.

.

Tables

Table 2. Summary of SUMO KO and rescue cell line phenotypes.

Cellular Process	Cell Line (phenotype as compared to WT)			
	SUMO1 KO	**SUMO2 KO (S2KO)**	**S2KO + SUMO1**	**S2KO + SUMO2**
Cellular morphology	Little change	Fibroblast-like	Fibroblast-like	Little change
Cell cycle	No change	> 2n DNA content	N/A	N/A
PML positive nuclear bodies	Fewer	Fewer	N/A	Fewer
DAXX positive nuclear bodies	Fewer	Fewer	N/A	Fewer
Mitochondrial function/number (MTT)	Similar readout	Lower readout	Higher readout	Lower readout
Proteotoxic stress response (Eerl)	Similar sensitivity	More senstitive	Similar sensitivity	Similar sensitivity
Genotoxic stress response (HU)	Less sensitive	Less sensitive	Less sensitive	Similar sensitivity
Gene expression (# of changed genes)	~2000 unique DEGs	~4000 unique DEGs	N/A	N/A
Gene expression (general trend)	Down-regulated	Up-regulated	Up-regulated	Down-regulated

Table 3. Reagents used for stress assays.

Reagent	Vendor	Part Number	Solvent
Eeyarestatin I	EMD Millipore	324521-25MG	DMSO
Hydroxyurea	Amresco	1B1368-25G	PBS
L-Azetidine-2-carboxylic acid	Sigma	A0760-50MG	PBS
MTT	Molecular Probes	V-13154	PBS

Table 4. Primers list for genomic SUMOs and qRT-PCR assays.

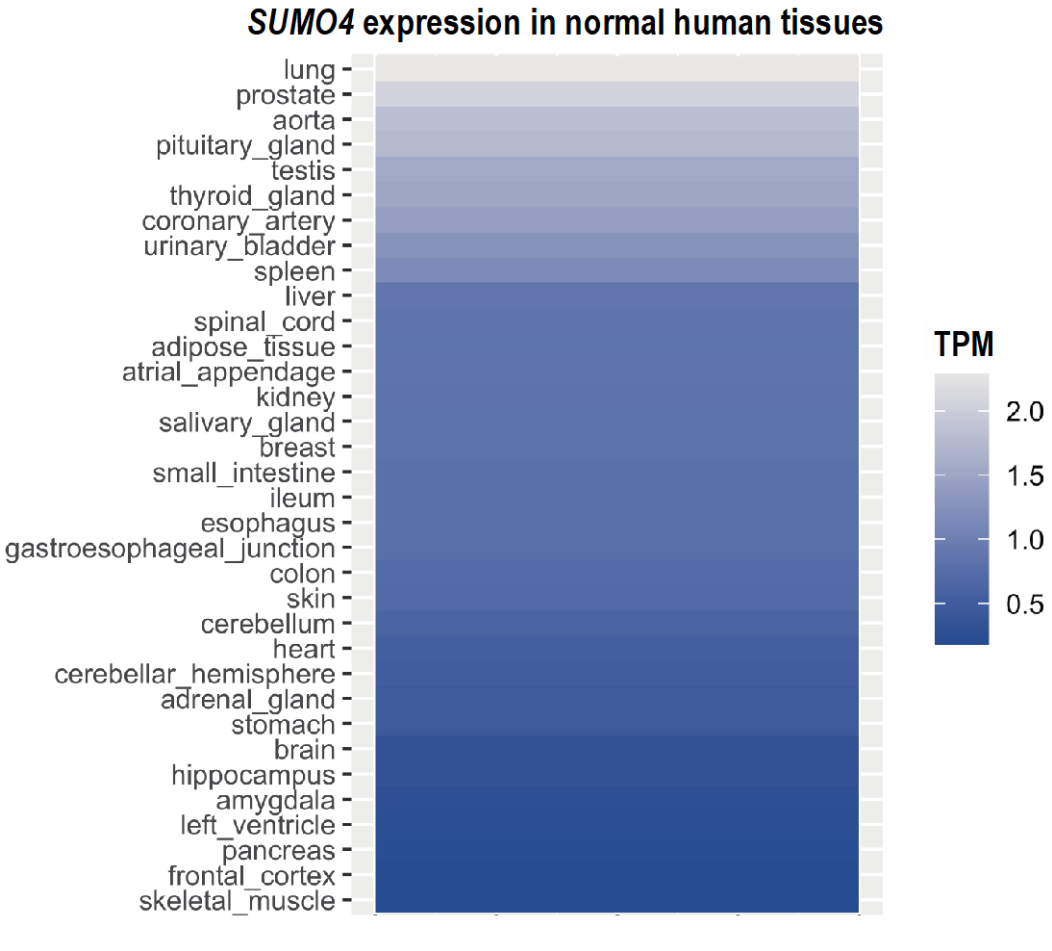
Primer name	Sequence (5'—>3')	Assay used
SUMO1 genomic (F)	CGGTACCCGGGGATCAAGGTGTAACGGAGAGAGCA	Indel ID
SUMO1 genomic (R)	CGACTCTAGAGGATCGAGAAAGAAGTGGGACGACATGA	Indel ID
SUMO2 genomic (F)	CGGTACCCGGGGATCACCAAGTGCACGACAAAACTG	Indel ID
SUMO2 genomic (R)	CGACTCTAGAGGATCATTGCATGTAGGCTTTTCCGT	Indel ID
APOBEC3G (F)	ATGACACCTGGGTCCTGCTGAA	qRT-PCR
APOBEC3G (R)	GAATCACGTCCAGGAAGCACAG	qRT-PCR
HOXD10 (F)	TGGCTGAGGTCTCCGTGTCCA	qRT-PCR
HOXD10 (R)	GCACCTCTTCTTTCTGCCACTC	qRT-PCR
COL3A1 (F)	TGGTCTGCAAGGAATGCCTGGA	qRT-PCR
COL3A1 (R)	TCTTTCCCTGGGACACCATCAG	qRT-PCR
TMEM71 (F)	CAGACTCCTCACCAATGGCTAC	qRT-PCR
TMEM71 (R)	CCATGCAGCCAAGATTTGGAGG	qRT-PCR
FAP (F)	GGAAGTGCCTGTTCCAGCAATG	qRT-PCR
FAP (R)	TGTCTGCCAGTCTTCCCTGAAG	qRT-PCR
OAS1 (F)	AGGAAAGGTGCTTCCGAGGTAG	qRT-PCR
OAS1 (R)	GGACTGAGGAAGACAACCAGGT	qRT-PCR
ISG15 (F)	CTCTGAGCATCCTGGTGAGGAA	qRT-PCR
ISG15 (R)	AAGGTCAGCCAGAACAGGTCGT	qRT-PCR
TLR4 (F)	CCCTGAGGCATTTAGGCAGCTA	qRT-PCR
TLR4 (R)	AGGTAGAGAGGTGGCTTAGGCT	qRT-PCR
HSPB7 (F)	GAGACGCCTATGAGTTTGCGGT	qRT-PCR
HSPB7 (R)	GGCACTTGTGAGCGAAGGTGTT	qRT-PCR
CTAG2 (F)	ATGGCGGTGCCGCTTCTGCG	qRT-PCR
CTAG2 (R)	CGGACCAGCTCCGCTTCCATG	qRT-PCR
TMEM156 (F)	CTTCACCAGGACTTGCCAAGAC	qRT-PCR
TMEM156 (R)	GCTTTCACTTCCATTGATCCTCTC	qRT-PCR
MAP2K6 (F)	GGCTACTTGGTGGACTCTGTTG	qRT-PCR
MAP2K6 (R)	CATCGTGATGCCCAGACTCCAA	qRT-PCR
HIST1H1C (F)	TGCCACTTGTACCCGAGTTT	qRT-PCR
HIST1H1C (R)	CCTTTTTGGCCGCCTTCTTC	qRT-PCR
HIST2H2BE (F)	TCCTCGGCAGAATCAAGGGT	qRT-PCR
HIST2H2BE (R)	AGGCCAAGAGAAGCCTCATT	qRT-PCR
HIST1H4H (F)	AGCAAGCAGGAGCCTTAGC	qRT-PCR
HIST1H4H (R)	GCAATGGATGTGGTCTACGC	qRT-PCR
SUMO1 (F)	TGACCAGGAGGCAAAACCTTC	qRT-PCR
SUMO1 (R)	AATTCATTGGAACACCCTGTCTT	qRT-PCR
SUMO2 (F)	ACGAAAAGCCCAAGGAAGGA	qRT-PCR
SUMO2 (R)	TGCCTCATTGACAATCCCTGT	qRT-PCR
GAPDH (F)	ACGGATTTGGTCGTATTGGG	qRT-PCR
GAPDH (R)	TGGAATTTGCCATGGGTGGA	qRT-PCR

Table 5. Antibodies and dilutions used in SUMO KO studies.

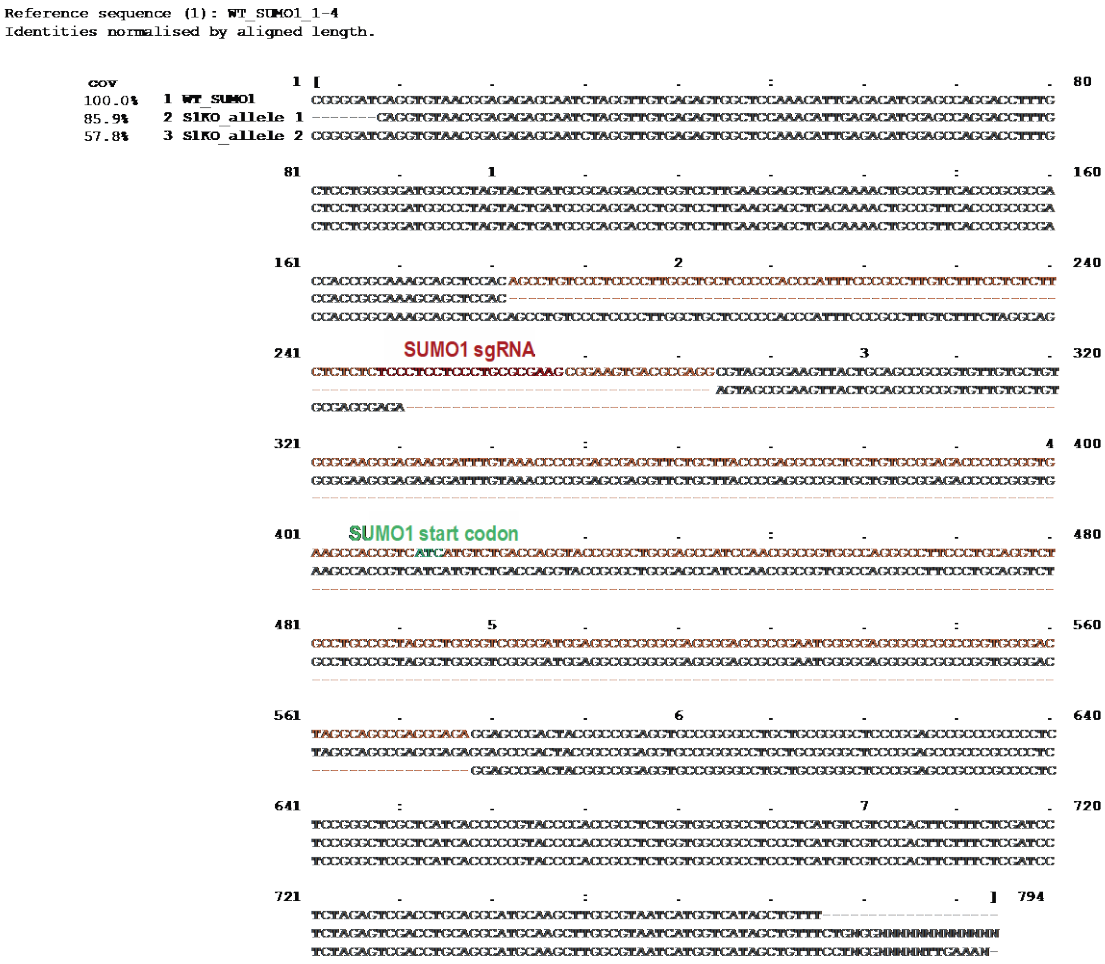
Vendor	Part Number	Antigen – Fluorophore*	Dilution	Species
Matunis Lab / Invitrogen	21C7 (33-2400)	SUMO1 (IF & WB)	[1:1000]	Mouse
Matunis Lab / Abcam	8A2 (ab81371)	SUMO 2/3 (IF & WB)	[1:800]	Mouse
Sigma	T-9026 DM1a	Tubulin (IF) Tubulin (WB)	[1:2000] [1:10,000]	Mouse Mouse
Santa Cruz	SC-966 PG-M3	PML (IF)	[1:100]	Mouse
Atlas	HPA008736	DAXX (IF)	[1:500]	Rabbit
Invitrogen	A11001	Mouse – 488 (IF)	[1:400]	Goat
Invitrogen	35561	Rabbit – 594 (IF)	[1:400]	Goat
Jackson Lab	115-035-003	Mouse – HRP (WB)	[1:10,000]	Goat
Jackson Lab	111-035-144	Rabbit – HRP (WB)	[1:10,000]	Goat

*IF = used for immunofluorescence microscopy. WB = used for western blot assays

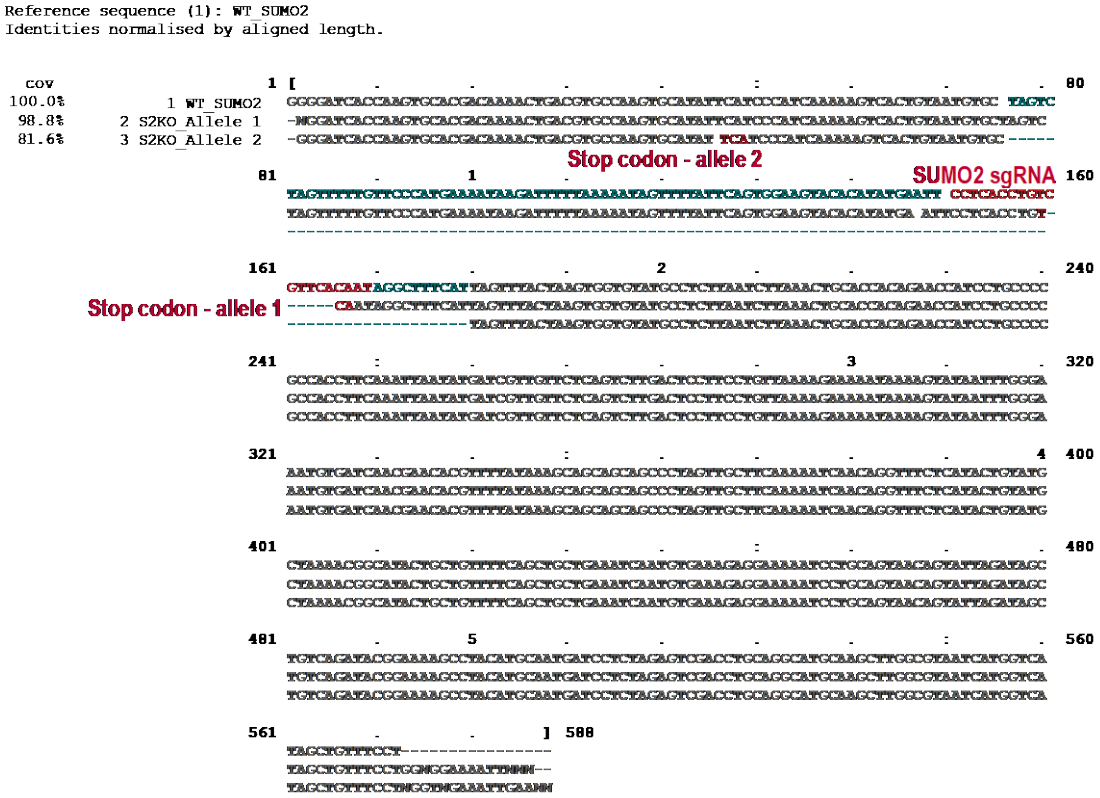
Supplemental Figures



Supplemental Figure 1. SUMO4 gene expression is low in normal human tissues. *SUMO4* gene expression values from GTEx, reported as transcripts per million (TPM). Note the scale, which goes from 0.5 - 2.0, as compared to *SUMO1-3* expression which ranges from 20-100.



Supplemental Figure 2. SUMO1 KO cell CRISPR mutations. The SUMO1 loci in U2OS WT cells is on top, with the two SUMO alleles from the SUMO1 KO cells aligned below. SUMO1 sgRNA is highlighted in red text and the SUMO1 start codon is highlighted in green text. The start codon is deleted in both alleles, resulting in different mutations and therefore heterozygous biallelic mutant *SUMO1* gene knockout.



Supplemental Figure 3. SUMO2 KO cell CRISPR mutations. The SUMO2 loci in U2OS WT cells is on top, with the SUMO2 alleles in the S2KO cells aligned below. SUMO2 sgRNA is highlighted in red text. Stop codons are prematurely created in each allele, as highlighted in red text, thus creating a heterozygous biallelic mutant *SUMO2* gene.

Precursor SUMO1 gene sequence, ordered from TWIST biosciences:

ATGTCTGACCAGGAGGCAAAACCTTCAACTGAGGACTTGGGGGATAAGAAGGAAGGTGAATATATTAAACTCAA
AGTCATTGGACAGGATAGCAGTGAGATTCACTTCAAAGTGAAAATGACAACACATCTCAAGAAACTCAAAGAATC
ATACTGTCAAAGACAGGGTGTTCCAATGAATTCACTCAGGTTTCTCTTTGAGGGTCAGAGAATTGCTGATAATCA
TACTCCAAAAGAACTGGGAATGGAGGAAGAAGATGTGATTGAAGTTTATCAGGAACAAACGGGGGGTCATTCAA
CAGTTTAG

Translated precursor SUMO1 protein sequence:

MSDQEAKPSTEDLGDKKEGEYIKLKVIGQDSSEIHFKVKMTTHLKKLKESYCQRQGVPMNSLRFLFEGQRI
ADNHTPKELGMEEEDVIEVYQEQTGGHSTV

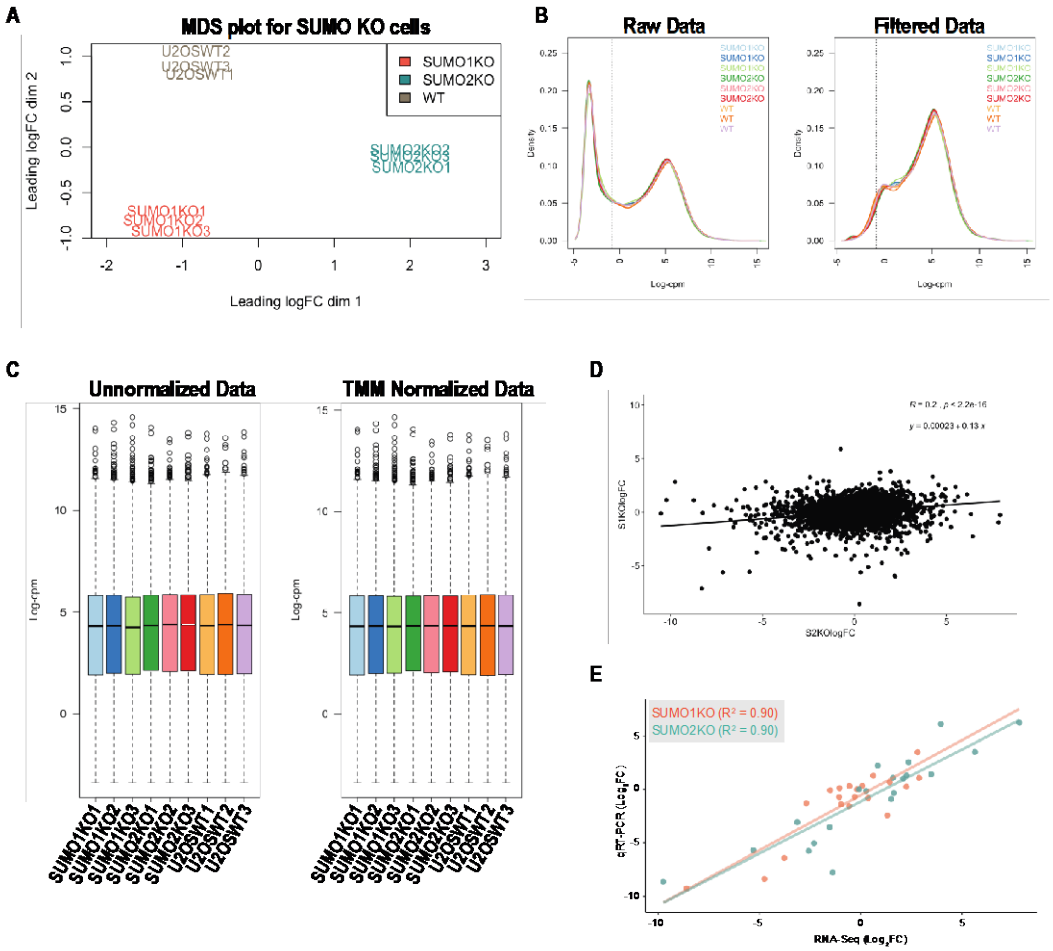
Precursor SUMO2 gene sequence, ordered from TWIST biosciences:

ATGGCCGACGAAAAGCCCAAGGAAGGAGTCAAGACTGAGAACAACGATCATATTAATTTGAAGGTGGCGGGGC
AGGATGGTTCTGTGGTGCAGTTTAAGATTAAGAGGCATACACCACTTAGTAAACTAATGAAAGCCTATTGTGAAC
GACAGGGATTGTCAATGAGGCAGATCAGATTCCGATTTGACGGGCAACCAATCAATGAAACAGACACACCTGCA
CAGTTGGAAATGGAGGATGAAGATACAATTGATGTGTTCCAACAGCAGACGGGAGGTGTCTACTGATGAGCTGC
AGTA

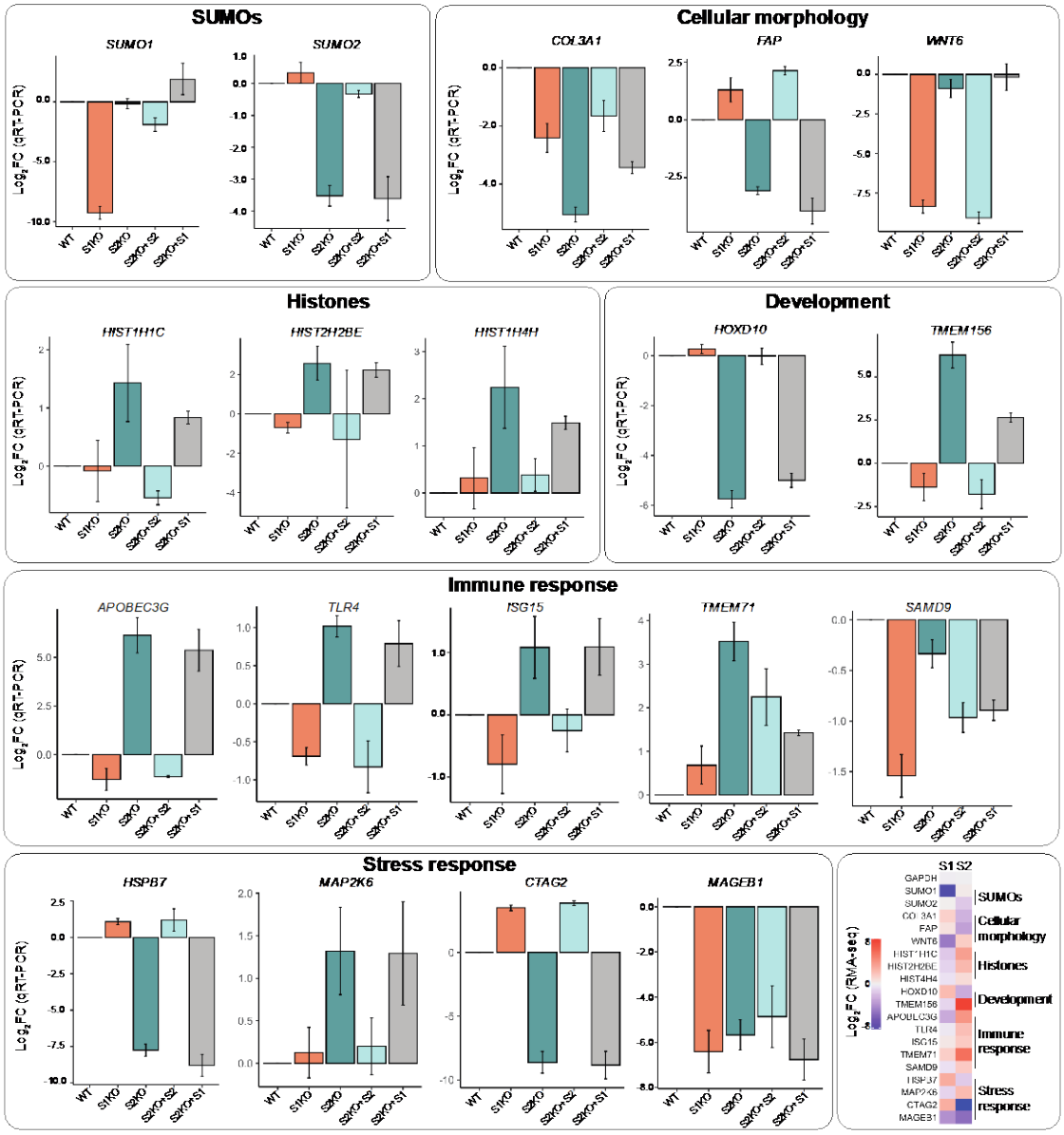
Translated precursor SUMO2 protein sequence:

MADEKPKEGVKTENNDHINLKVAGQDGSVVQFKIKRHTPLSKLMKAYCERQGLSMRQIRFRFDGQPINETD
TPAQLEMEDEDTIDVFQQQTGGVY

Supplemental Figure 4. Precursor SUMO1 and SUMO2 sequences for rescue cell lines. SUMO1 and SUMO2 gene sequences, and translated amino acid sequences are shown. Di-glycine motifs, required for SUMO activation, are highlighted in red.



Supplemental Figure 5. Processing of SUMO KO cell RNA-seq data. A) An MDS plot showing tight clustering of replicates, and differences between cell lines. B) Pre- and post-filtering plots based on low number of reads. C) Plots showing pre- and post-TMM normalization. Library sizes were relatively similar. D) Correlation plot of SUMO1 KO log_2FC values as compared to SUMO2 KO log_2FC values. The low correlation suggests that the noise in our samples are low, and thus the large differences in DEG numbers observed between the SUMO1 KO and SUMO2 KO cells represent a true biological difference. E) Correlation plot of gene expression data from qRT-PCR as compared to RNA-seq. The correlation (R^2>0.9) reveals that results obtained from both methods are similar and therefore that RNA-sequencing values are reproducible.



Supplemental Figure 6. Individual bar plots of SUMO KO cell DEGs assayed by qRT-PCR. Bar plots of log_2FC gene expression values as measured by qRT-PCR. Genes are grouped by their cellular function categories. For reference, log_2FC gene expression values from RNA-seq are summarized in the heatmap in the lower right-hand corner.

Chapter 4: A Cellular and Bioinformatics Analysis of the SENP1 SUMO Isopeptidase in Pancreatic Cancer

This chapter was published in The *Journal of Gastrointestinal Oncology* (2019, Oct 25;10(5): 821-830) and is reproduced here with copyright permission from the AME Publishing Company.

Article DOI: 10.21037/jgo.2019.05.09

Abstract

Sumoylation is an important post-translational modification that involves the conjugation of the Small Ubiquitin-related Modifier (SUMO) onto target proteins. This modification is reversed through the catalytic activity of SUMO isopeptidases, known as SENPs. One of these SENPs, SENP1, was reported to be overexpressed in human pancreatic cancer cells and patient tissues. Since elevated SENP1 expression levels can be used as a prognostic marker for a subset of cancers, we set out to further explore the overexpression of SENP1 in pancreatic cancer. We found that SENP1 protein levels were not significantly different between pancreatic cancer and normal pancreas-derived cell lines. To evaluate SENP1 expression in patient samples, we analyzed large publicly available datasets and found that SENP1 mRNA levels were significantly lower in pancreatic cancer tissue as compared to normal pancreas tissue samples. Furthermore, we found that the *SENP1* gene is amplified in less than 1% of sequenced pancreatic cancer patient samples and that expression levels have no association with patient survival. Based on our analysis, we conclude that SENP1 is not overexpressed in pancreatic cancer and is therefore not likely to be an effective biomarker for this disease. Through this work, we also outline a simple but powerful bioinformatics workflow for the assessment of mRNA expression levels, genomic alterations and survival analysis for putative biomarkers for common human cancers.

Introduction

The Small Ubiquitin-related Modifier (SUMO) is a well conserved 110 amino acid protein that is post-translationally conjugated onto target proteins in a dynamic and reversible process called sumoylation. SUMO is covalently attached to target proteins through an E1, E2 and E3 enzymatic cascade. Deconjugation of SUMO from target proteins is catalyzed by sentrin-specific isopeptidases, or SENPs (22, 269). A wide range of essential cellular functions are regulated by sumoylation, such as transcription, chromatin remodeling, DNA replication and cell division, among many others (76, 77). Many of these essential cellular processes are misregulated in human cancers (75). As such, misregulation of SUMO conjugating and deconjugating enzymes have been implicated as contributing factors in the development and progression of many cancers (197, 270-273). Consequently, individual pathway components have become attractive drug targets and potential biomarkers for cancer therapies (104, 274, 275). For instance, an inhibitor of the SUMO E1 conjugating enzyme has been shown to inhibit sumoylation globally and thereby decrease cancer cell proliferation and viability (64).

Our lab and others have identified unique roles of the SUMO isopeptidase, SENP1, in regulating genes important for cancer-related processes, such as chromosome segregation (24, 200) and cellular proliferation (103, 108, 110, 276). Through misregulation of SENP1 expression, these genes and processes can become misregulated and contribute to cancer development and progression. As such, it has been demonstrated that SENP1 expression levels can be used as a prognostic marker for a molecularly defined subset of prostate cancers (277). These promising findings prompted us to further explore reported SENP1 overexpression in pancreatic cancer

(112), a lethal disease that is projected to become the second leading cause of cancer-related deaths in both men and women by 2030 (278). Lack of effective treatment options render this disease particularly lethal and early diagnosis of the disease is essential in order to optimize treatment effectiveness and patient survival (279). The previous observation that SENP1 is up-regulated in pancreatic cancer raised the intriguing possibility that SUMO inhibitors could be used as an effective treatment option for this disease. We therefore chose to further characterize and validate SENP1 expression levels in pancreatic cancer cells and patient tissues as a step toward further establishing SENP1 as a biomarker for treatment of pancreatic cancer with a SUMO inhibitor.

We found using cell-based assays and analyses of large-scale sequencing studies from pancreatic cancer patients that in contrast to a previous report (112), SENP1 is not significantly overexpressed in pancreatic cancer. Through this work, we also provide the field with a powerful bioinformatics workflow that can be used by researchers to evaluate expression levels and genomic alterations of putative biomarkers for many common human cancers.

Results

Characterizing SENP1 Expression and Localization in Human Cell Lines

To evaluate SENP1 expression in pancreatic cancer, we first looked at SENP1 messenger RNA (mRNA) and protein levels in six human cell lines. We used the HPNE cell line, which is derived from non-cancerous pancreas tissue (280), for comparison to the pancreatic cancer cell lines AsPC-1, BxPC-3, CFPAC-1 and PANC-1, and to

the cervical cancer cell line, HeLa. We found using qRT-PCR that the HPNE cells had significantly lower relative SENP1 mRNA expression as compared to all other tested cell lines (p-values < 0.05) (Figure 17A). Consistent with previous findings, the AsPC-1, BxPC-3 and CFPAC-1 cells had indistinguishable differences in SENP1 expression, whereas the PANC-1 cell line had the highest levels of SENP1 expression (112).

To explore whether the elevated SENP1 mRNA levels in PANC-1 cells were associated with a gene duplication event, we turned to the Broad Institute's Cancer Cell Line Encyclopedia (CCLE, portals.broadinstitute.org/ccle) (239). We found that there were two copies of SENP1 in all four of our tested pancreatic cancer cell lines (Figure 17B), indicating that the elevated SENP1 mRNA levels in PANC-1 were not associated with a SENP1 gene duplication event. Consistent with our findings, RNA sequencing data from CCLE also showed a similar pattern of SENP1 mRNA expression for the tested pancreatic cancer cell lines.

To investigate the relationship between mRNA and protein expression levels, we probed whole cell lysates from our six human cell lines using a validated SENP1 antibody. We found that in contrast to our qPCR results, there were no significant differences in SENP1 protein levels between the tested cell lines (Figure 17C and 1D). Specifically, SENP1 protein levels were not as elevated in the PANC-1 cells as were expected based on our qRT-PCR results and the previous data (112). However, high molecular weight forms of SENP1 varied between PANC-1 cells and the other cell lines, as indicated by the high molecular weight bands marked by an asterisk in Figure 17C.

The nature of detected high molecular weight forms of SENP1 is not known, however the N-terminus of SENP1 contains multiple phosphorylation and acetylation sites, as well as predicted sumoylation and ubiquitination sites (281). Our lab and others have found that signals in the N-terminus of SENP1 determine its subcellular distribution between the nucleus, cytoplasm and nuclear pore complexes (NPCs), and it is predicted that posttranslational modifications could affect localization (24, 282, 283). Since we observed variations in high molecular weight forms of SENP1 in the PANC-1 cells by western blot, we investigated whether these correlate with changes in subcellular localization. We used immunofluorescence microscopy to image HPNE, AsPC-1 and PANC-1 cells co-stained for SENP1 and NPCs. Consistent with previous work from our lab (24), we found that SENP1 colocalizes with NPCs and is detectable at varying levels in small foci throughout the nucleoplasm in the three pancreas-derived cell lines (Figure 18A). More specifically, we observed similar SENP1 levels in the nucleoplasm of AsPC-1 and PANC-1 cells, and elevated levels in the nucleoplasm of HPNE cells, as quantitated in Figure 18B.

Bioinformatics Evaluation of SENP1 in Pancreatic Cancer Patient Samples

To extend our studies beyond cell lines, we next evaluated SENP1 mRNA expression, gene alterations, and survival association in pancreatic cancer patient samples, as outlined in Figure 19A. We first turned to the University of California Santa Cruz (UCSC) Xena Public Data Hub (xena.ucsc.edu) to acquire normalized mRNA data from RNA sequencing studies (284) from The Cancer Genome Atlas (TCGA) (285) and the Genotype-Tissue Expression program (GTex) (286). TCGA is a multicenter effort that profiles data at the molecular level from thousands of cancer patients across

33 cancer types. The GTEx program is another multicenter effort that generates genomic and transcriptomic profiling data from over 50 types of tissues derived from non-cancerous patient biopsies. Importantly, there are approximately equal numbers of pancreatic cancer and non-cancer samples from TCGA and GTEx, respectively, which when normalized by Xena, allows for powerful statistical comparisons between the two sources of data. To that end, we compared *SENP1* mRNA expression levels from 178 pancreatic cancer samples to 165 non-cancerous pancreatic tissue samples and found that *SENP1* was significantly lower in the pancreatic cancer tissues as compared to the non-cancerous tissue (p-value < 0.05, Figure 3B). As a second approach, we also used the Oncomine Platform by Thermo Fisher Scientific (https://www.oncomine.org) (287) as an alternative data source for evaluating *SENP1* mRNA expression in pancreatic ductal adenocarcinoma tumors and matching normal pancreatic tissue samples by microarray (288). Here, we found that there is no significant difference in *SENP1* mRNA expression levels between the paired tissues (Figure 19C).

To complement our SENP1 mRNA expression data, we analyzed *SENP1* gene alterations using the Memorial Sloan Kettering Cancer Center (MSKCC) cBioPortal (cbioportal.org) (289, 290) in 676 human pancreatic cancer samples. We found that *SENP1* was amplified in 4 of the samples, deleted in 4 of the samples, and had a missense mutation of unknown significance in 1 sample (Figure 19D). Thus, the total alteration rate of the *SENP1* gene in pancreatic cancer based on these samples is approximately 1.3%. Of that, only 0.6% (4/676) of the cases had a gene amplification. As a second approach, we also analyzed SENP1 single nucleotide polymorphisms (SNPs) in genome-wide association studies (GWAS) using the joint National Human Genome Research Institute (NHGRI) and European Bioinformatics Institute (EMBL-

EBI) quality controlled and literature-derived catalog of published GWAS studies (https://www.ebi.ac.uk/gwas) (291). Our search identified two SENP1 variants, rs10875742 and rs2408955-T, associated with vital lung function and glycated hemoglobin levels, respectively.

Lastly, to look at the association of SENP1 expression levels and pancreatic cancer patient survival, we used the Kaplan-Meier plotter (kmplot.com) (292) to stratify patient survival data based on calculated high versus low *SENP1* mRNA expression levels. The results showed no statistically significant difference in pancreatic cancer patient survival based on *SENP1* mRNA expression levels (Figure 19E).

Discussion

Sumoylation regulates essential cellular processes, many of which are often misregulated in human cancers. As the SUMO pathway itself is also misregulated in numerous cancers, it has been implicated as a contributing factor in the development and progression of these diseases (64, 76). Researchers have found that expression levels of individual SUMO pathway enzymes can be used as prognostic markers for cancers such as prostate and cervical cancer (275, 277). Here, we used authenticated cell lines, validated reagents and data from large-scale genomics studies to evaluate the utility of SENP1 expression as a biomarker in pancreatic cancer.

To explore reported SENP1 overexpression in pancreatic cancer, we first evaluated SENP1 mRNA levels in pancreas-derived human cell lines. We found that the normal control cells had significantly lower *SENP1* mRNA expression as compared to the cancer cell lines. We also found that AsPC-1, BxPC-3 and CFPAC-1 had

indistinguishable differences in *SENP1* expression, whereas the PANC-1 cell line had significantly higher levels of *SENP1* expression. However, the magnitude of *SENP1* mRNA differences between cell lines did not match previously published findings (112). In this published study, an approximate 6-fold increase in *SENP1* expression in the PANC-1 cell line was observed when compared to AsPC-1 cells, whereas we observed an approximate 2-fold increase. These differences could be due to differences in cell lines (our cell line identities were validated by short tandem repeat profiling), the use of different equipment, reagents or relative expression calculations. For instance, the previous report used the $\Delta\Delta C_T$ method, whereas we used the ΔCq method (293) since evaluating endogenous *SENP1* mRNA levels does not involve the use of a treatment group. Consistent with our findings, we found that our *SENP1* mRNA expression patterns were similar to those obtained by CCLE using RNA sequencing. Furthermore, data from CCLE revealed that elevated *SENP1* mRNA levels in the PANC-1 cells were not associated with a gene duplication event, as all four pancreatic cancer cell lines were found to be diploid at the *SENP1* locus. Surprisingly, we found that SENP1 protein levels were similar across all tested cell lines, despite higher mRNA expression in PANC-1 cells. This indicates that SENP1 protein levels are regulated post-transcriptionally, possibly at the level of translation or protein stability. Interestingly, although SENP1 protein levels did not differ between cell lines, we did observe variations in predicted modified forms of SENP1 by western blot analysis. We also observed variations in the relative distribution of SENP1 within the nucleoplasm of HPNE cells in comparison to AsPC-1 and PANC-1 cells. The prediction that these differences in observed localization reflect differences in posttranslational modifications of the SENP1 N-terminus will require further study.

Using publicly available patient datasets, we compared *SENP1* expression levels from hundreds of pancreatic cancer tissues to non-cancerous pancreas tissues. We found that SENP1 expression is lower in pancreatic cancer tissues when compared to unpaired-normal pancreas tissue, and is unchanged when compared to paired-adjacent normal pancreas tissue. The difference between these two outcomes could be explained by tissue environment, especially considering the strong desmoplastic reaction that occurs in pancreatic cancer (288). It is possible that the normal-adjacent tissues are influenced by the tumor microenvironment (TME) (294), which in turn affects *SENP1* mRNA expression in the surrounding tissues. These results indicate that *SENP1* levels are highest in healthy pancreas tissues and decrease in pancreatic tumor tissues. This finding is in contrast to a previous observation (112) which found elevated *SENP1* mRNA levels in pancreatic ductal adenocarcinoma tissues from 22 patients as compared to adjacent normal tissues when assayed by qRT-PCR. These discordant findings could be explained by the different approaches used to evaluate *SENP1* mRNA levels, the differences in sample sizes, or potential epidemiological variables related to the sources of tissue.

To further explore SENP1 in patient samples, we also looked at *SENP1* gene mutations in over 600 sequenced pancreatic cancer tissues using cBioPortal. Here, we found that *SENP1* was amplified in 4 of the samples, deleted in 4 of the samples, and had a missense mutation of unknown significance in 1 sample. This amounts to a 1.3% *SENP1* gene alteration rate in pancreatic cancer, and furthermore, the observed differences between types of alterations suggests that *SENP1* mutations in pancreatic cancer are not conserved. For comparison, KRAS, a protein well-known to promote pancreatic cancer tumorigenesis, has been found to have an alteration rate of greater than 90% and the alteration is almost always a single nucleotide variant (295).

Our search of the GWAS catalog (291) identified two *SENP1* variants, rs10875742 and rs2408955-T. The rs10875742 variant was associated with vital lung function, and the rs2408955-T variant was associated with hemoglobin A1c (HbA1c) at genome-wide significance (296). Of relevance to our study, HbA1c is used to diagnose and monitor Type 2 diabetes (T2D), which is a risk factor and a prognostic factor for pancreatic cancer (297). This *SENP1* variant was further classified into an erythrocytic group to better define its mode of action on HbA1c, however the specific effects of this variant on the function, structure or lifespan of red blood cells have yet to be explored. Given the utility of HbA1c in diagnosing and monitoring T2D, and the link between T2D and pancreatic cancer, this variant could be of interest for further exploration.

Lastly, we found that there is no association between *SENP1* expression levels and pancreatic cancer patient survival. Taken together, our data provides evidence that *SENP1* is not altered at the genetic level, nor is it overexpressed in pancreatic cancer tissues or associated with patient survival. Thus, although a previous study has suggested a link between SENP1 and pancreatic cancer (112), our results do not support this finding. We therefore conclude that SENP1 is not likely to be an effective biomarker for this disease. Through this work, we have also outlined a powerful and freely-available bioinformatics workflow for the evaluation of potential biomarkers for the most common human cancers.

Conclusions

We used authenticated cell lines, validated reagents and data from large-scale genomics studies to evaluate SENP1 localization, mRNA and protein level expression,

gene mutations and survival association in human pancreas cells and tissue samples. We found that SENP1 is not overexpressed in pancreatic cancer, has no association with patient survival and would therefore not make an effective biomarker. Through this work we have outlined an easy to use and freely available bioinformatics workflow for evaluating putative biomarkers for use in cancer diagnostics and therapies.

Acknowledgements

This work was supported by National Institute of Health [grant number GM060980 to M.J.M] and [grant number T32CA09110 to D.M.B]; and was also supported in part through funding from the Sol Goldman Pancreatic Cancer Research Center. We thank Dr. Christopher Heaphy, Dr. Alan Meeker, Dr. Scott Kern, Dr. Laura Wood and Dr. Robert Anders for their discussion, guidance and reagents.

Author Contributions

D.M.B designed and performed experiments, data mining, data analysis and writing of the manuscript. M.J.M facilitated experimental design and edited the manuscript for clarity.

Materials and Methods

Cell Culture

We received the “normal” hTERT-transformed HPNE cell line (ATCC#: CRL-4023) graciously from Dr. Laura Wood, and three pancreatic cancer cell lines, AsPC-1 (ATCC#: CRL-1682), BxPC-3 (ATCC#: CRL-1687), CFPAC-1 (ATCC#: CRL-1918)

graciously from Dr. Scott Kern. We ordered the PANC-1 cells directly from ATCC (ATCC#: CRL-1469). We also used the cervical cancer cell line, HeLa (ATCC# CCL-2), as a non-pancreas control. All cell lines were cultured in Dulbecco's Modified Eagle's Medium (DMEM, ThermoFisher Scientific PN: 11965118) supplemented with 10% Fetal Bovine Serum (Atlanta Biologicals PN: S11150) and grown in a monolayer at 37^{o}C and 5% CO_2. Cells were passaged one to two times per week, or until cells reached approximately 80% confluence.

All six cell lines used for this study were authenticated by the Johns Hopkins University (JHU) Genetic Resources Core Facility (GRCF) using the Promega GenePrint10 Short Tandem Repeat Profile Kit and had identities with an >80% match against the ATCC database. Additionally, we used the JHU GCRF to confirm that all six cell lines were mycoplasma free using a PCR based MycoDtect kit from Greiner Bio-One.

qRT-PCR Analysis

All six cell lines were seeded at 5.0 x 10^5 cells/well in a 6-well dish and grown at 37^{o}C and 5% CO_2 for approximately 24 hours. Total RNA was extracted using the Sigma GenElute Mammalian MiniPrep kit (Sigma PN: RTN10) following the vendor's protocol. Extracted RNA was analyzed by nanodrop for concentration and purity. cDNA was generated using the New England BioLabs ProtoScript First Strand cDNA Synthesis Kit (NEB PN: E6300S) using 200ng of RNA, $d(T)_{23}$ VN primers and following the vendor's recommended protocol. The qPCR reaction was performed using Bio-Rad iTaq Universal SYBR Green Supermix (Bio-Rad PN: 1725121) and following the vendors recommended protocol. qPCR runs were performed using an Applied Biosystems Quant Studio with Quant Studio v1.3.1 software. Relative SENP1

expression was calculated using average C_T values from three biological replicates, a validated housekeeping gene (GAPDH), and the ΔCq equation: $2^{-(SENP1-GAPDH)}$. Primer sequences:

Table 6. SENP1 and GAPDH primers used for qRT-PCR.

	Forward	Reverse
SENP1	5'- ATCAGGCAGTGAAACGTTGGAC -3'	5'- GCAGGCTTCATTGTTTATCCCA -3'
GAPDH	5'- ACGGATTTGGTCGTATTGGG -3'	5'- CGCTCCTGGAAGATGGTGAT -3'

Immunoblotting Analysis

All six cell lines were seeded at 5.0 x 10^5 cells/well in a 6-well dish and grown at 37°C and 5% CO_2 for approximately 24 hours. Cells were harvested by scraping in 100uL of 2X Laemmli sample buffer with 10% β-mercaptoethanol. Cells were lysed in a water bath sonicator for 3 x 15 second pulses, heated at 95°C for 5 minutes, cooled and spun at 13,000xG for 5 minutes. Samples were loaded onto a 10-well, 10% tris-glycine gel and run at 110V for 1 hour 15 minutes. Samples were transferred to a LF-PVDF membrane (Bio-Rad PN: 1704274) using the Bio-Rad TransBlot Turbo Mixed MW setting. Blots were blocked in 5% non-fat dry milk for 1 hour at room temperature (RT) with gentle shaking. Blots were then rinsed with 1X TBST and incubated overnight at 4^0C with a rabbit monoclonal SENP1 antibody (abcam PN: ab108981 [1:1,000]) and a mouse monoclonal alpha tubulin antibody (abcam PN: ab7291 [1:15,000]) diluted in 2% BSA, 0.02% NaN_3 and 1x PBS. Blots were washed in 1xTBST and incubated in Goat anti-rabbit 800CW (LI-COR PN: 926-32211 [1:10,000]) and goat anti-mouse 680LT (LI-COR PN: 926-68020 [1:10,000]) protected from light for 1 hour at RT with gentle shaking. Blots were imaged using the LiCor Odyssey imaging system and quantitated using ImageStudio v.5.2.5 software.

Immunofluorescence Microscopy

Cells were seeded on coverslips at 2.5 x 105 cells/well in a 6-well dish and grown at 37°C and 5% CO2 for approximately 24 hours. Media was carefully aspirated, cells were washed one time with 1x PBS, fixed in 3.5% paraformaldehyde in 1x PBS for 7min. at RT, washed with 1x PBS and permeabilized in 0.5% Triton-X-100 in 1x PBS for 20min. at RT. Cells were gently washed twice with 1x PBS and primary antibodies were applied (SENP1 abcam PN: ab108981 [1:500]; mAb 414 abcam PN: ab24609 [1:2,000]). Zeiss Observer Z1 fluorescence microscope with an Apotome VH optical sectioning grid was used to acquire images. SENP1 nucleoplasmic fluorescence intensity was measured using ImageJ software and graphed using RStudio.

Statistical Analysis

All statistical analyses were performed using RStudio. Differences in means for qPCR and western blot data were analyzed by ANOVA, followed by a Tukey Honestly Significant Difference test with 95% confidence intervals to identify statistically significant differences between sample pairs. Differences in means for the patient data from Xena was calculated using a Students t-test. P-values < 0.05 were considered significant for all reported data.

Bioinformatics: Cancer Cell Line Encyclopedia (CCLE)

SENP1 mRNA expression and copy number data from human cancer cell lines was downloaded from the Broad Institute's Cancer Cell Line Encyclopedia (https://portals.broadinstitute.org/ccle). Using RStudio, data was subset and graphed to include information only from the AsPC-1, BxPC-3, CFPAC-1 and PANC-1 cell lines.

Bioinformatics: Xena

Normalized mRNA data for pancreatic cancer tissues and normal pancreas tissues were downloaded from the UCSC Xena public data hub (https://xena.ucsc.edu) and opened in RStudio. The pancreatic cancer SENP1 mRNA expression values were obtained from the TCGA Pancreatic Cancer (PAAD) cohort, which had 10 studies for a total of 196 samples. The non-cancerous SENP1 mRNA expression values were downloaded from the GTEX study. In RStudio, GTEx SENP1 data was subset by organ type to include only data from normal pancreas samples, providing a total of 167 samples. Then, both TCGA and GTEx data sets were cleaned to remove samples with missing data, resulting in 178 TCGA pancreatic cancer samples, and 165 GTex normal samples. Lastly, normality assumptions were confirmed using the Shapiro-Wilk normality test and mean expression values were compared using a Students t-test and the results were plotted using ggplot (261).

Bioinformatics: cBioPortal

The web-based cBioPortal for Cancer Genomics (289, 290) (www.cbioportal.org) version 1.18.0 was used to analyze SENP1 alterations in large-scale pancreatic cancer genomic data sets. Three pancreatic adenocarcinoma cancer studies were queried: QCMG, Nature 2016 (298); TCGA, Provisional; and UTSW, Nature Communications 2015 (295). Molecular profiles were selected for Mutations and Copy Number Alterations, resulting in 751 unique patient/case sets, of which 676 were sequenced. The gene symbol “SENP1” was used to run the query. Presented data are from the OncoPrint and Cancer Types Summary tabs.

Bioinformatics: Oncomine

The Thermo Fisher Scientific Oncomine platform version 4.5 (oncomine.org) was used to analyze SENP1 mRNA expression levels in pancreatic cancer patient samples as compared to adjacent normal tissues from the Badea Pancreas study (288).

Bioinformatics: GWAS Catalog

The web-based NHGRI-EBI Catalog of published genome-wide association studies (https://www.ebi.ac.uk/gwas/) was used to analyze SENP1 in GWAS studies (291).

Bioinformatics: Kaplan-Meier Plotter (KM Plot)

Relapse free and overall survival data for pancreatic cancer patients based on SENP1 mRNA expression was analyzed using the Kaplan-Meier Plotter (http://kmplot.com/analysis/) using data from the pan-cancer study. The pancreatic ductal adenocarcinoma (n=177) study was selected and analysis was not restricted by subtypes.

Figures

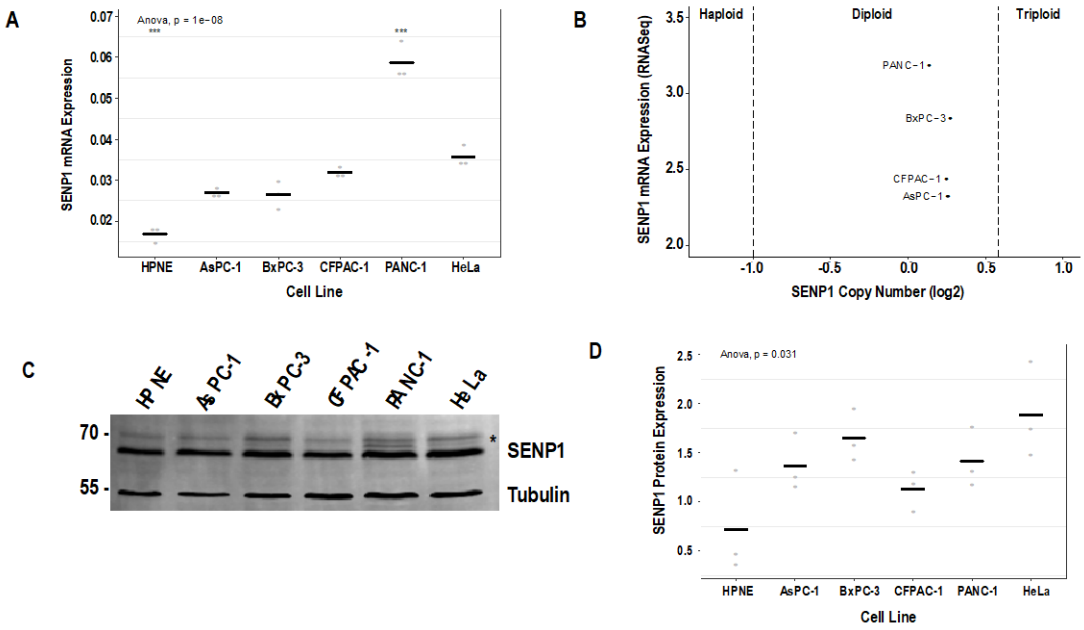


Figure 17. Evaluation of SENP1 expression levels in human cell lines. A) Relative *SENP1* mRNA expression levels measured by qRT-PCR. B) Plot of *SENP1* mRNA expression levels against *SENP1* copy number in four pancreatic cancer cell lines using data from CCLE. C) Representative western blot image of SENP1 and tubulin signal from whole cell lysates. D) Quantitation of normalized SENP1 protein levels from 3 independent western blot assays. Gray dots are individual data points, black lines are mean normalized SENP1 values.

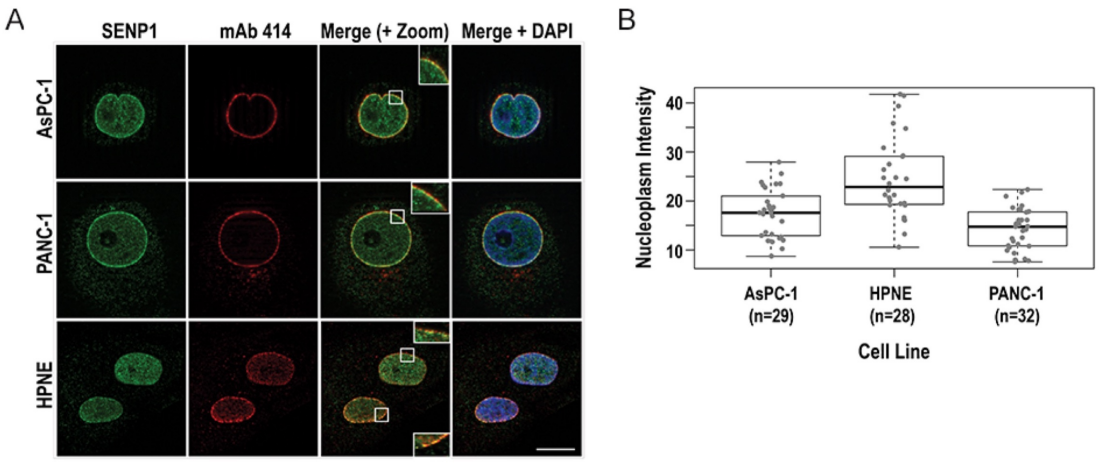


Figure 18. SENP1 localization in pancreas-derived cells. A) Representative immunofluorescence microscopy images of pancreas-derived cells co-stained with antibodies recognizing SENP1 and NPCs (mAb 414). Scale bar is 10μm. B) Quantitation of SENP1 nucleoplasmic signal from immunofluorescence images.

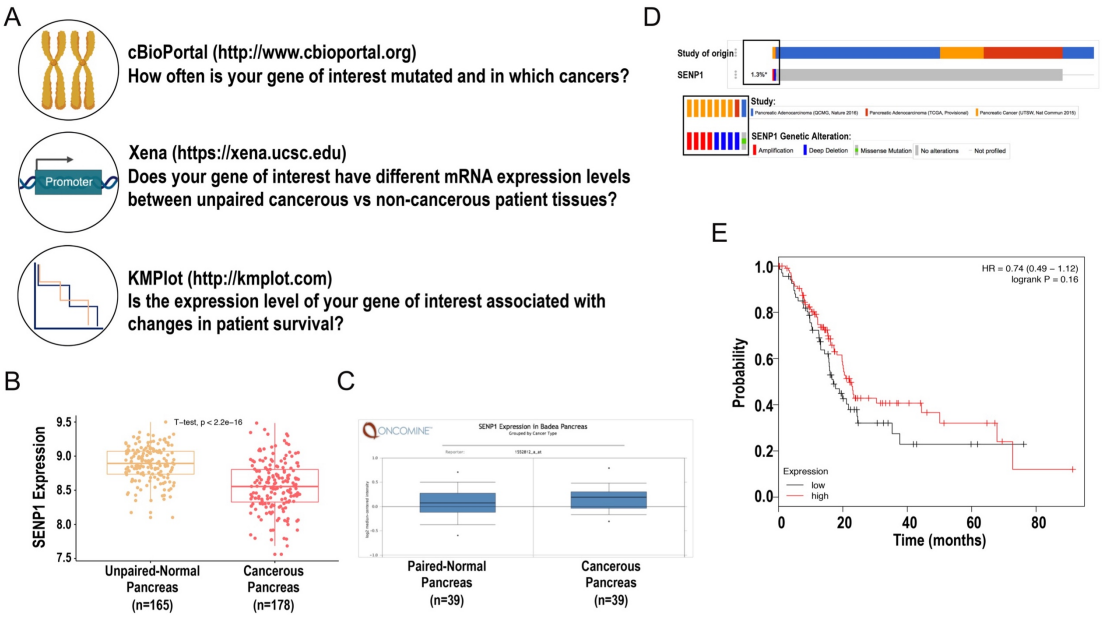


Figure 19. Bioinformatics evaluation of SENP1 in pancreatic cancer patient samples. A) Outline of bioinformatics resources used for SENP1 analysis. B) Quantitation of SENP1 mRNA expression from cancerous and unpaired non-cancerous pancreas tissue using data from the UCSC Xena Public Data Hub. C) SENP1 mRNA expression levels in paired-normal pancreas tissue as compared to cancer tissue samples using Oncomine. D) Oncoprint of SENP1 gene alterations in 676 patient samples using cBioPortal. E) Kaplan-Meier survival analysis of 261 patients with high versus low SENP1 expression analyzed using KMPlot.com. Hazard ratio (HR), 95% confidence intervals, and logrank P-value presented in the graph.

Conclusions and Future Directions

Conclusions

There is growing evidence that SUMO pathway components contribute to many diverse human diseases. The SUMO paralogs in particular regulate hundreds of proteins involved in a wide-range of processes that are associated with cancer, neurodegeneration and auto-immune disorders, among others. Questions surrounding the unique roles of the SUMO paralogs in these diverse cellular processes however, remain unanswered. Do the paralogs have unique and non-redundant roles, which could help explain the vast array of cellular functions regulated by sumoylation? If so, how is this paralog-specificity itself regulated? Moreover, could we therapeutically target different paralogs in the pathway? And if so, how might we identify the best patients for such a treatment? The work presented in this thesis uses a multi-faceted approach to address these questions. We first reviewed the literature for evidence of paralog-specific functions, where we found examples ranging from regulation of individual proteins to entire cellular processes. We then generated SUMO1 and SUMO2 knockout (KO) cells and systematically analyzed them for paralog-specific functions. This analysis revealed non-redundant roles for the paralogs in regulating cellular morphology, the cellular stress response, nuclear body integrity and gene expression. Lastly, we developed a bioinformatics workflow to analyze SUMO pathway enzyme expression levels in cancerous tissues as compared to normal human tissues. We used this workflow and found that the SENP1 SUMO protease is not a predictive biomarker of pancreatic cancer. Collectively, we have identified unique and non-redundant roles for the SUMO paralogs, and have provided a method for identifying SUMO paralog and pathway misregulation in human cancers.

SUMO Paralog-Specific Functions

As presented in Chapter 2, we compiled findings from published studies that explored SUMO paralog-specific functions. Two overarching themes that emerged from this review were that the SUMO paralogs have conserved functions in mediating specific protein-protein interactions, and that biological outcomes of SUMO-modified proteins are complex and often context dependent. Beyond these themes, evidence of paralog-specific differences are abundant, and supported by the unique sets of proteins that are selectively modified by different paralogs. Moreover, evidence of paralog-specific conjugating and deconjugating enzymes suggest mechanisms for differential regulation of paralog-specific functions.

In Chapter 3, we provided evidence that SUMO1 and SUMO2 have non-redundant functions in regulating a host of cellular processes, including control of gene expression. Moving beyond a description of phenotypes and exploring the underlying molecular basis of changes observed in SUMO1 and SUMO2 KO cells will require a more comprehensive analysis of paralog specific substrates, both on and off chromatin. Proteomics experiments have identified a limited number of unique SUMO1 and SUMO2/3 substrates, however, the most recent paralog-specific studies were done almost ten years ago (19, 172, 173). Now that more advanced and sensitive mass spectrometry technologies are available, there is an opportunity to conduct a modern and systematic analysis of paralog-specific substrates (19). An ideal experimental design includes the use of paralog-specific antibodies to identify endogenously modified substrates by mass spectrometry, and then coupling this data with chromatin-immunoprecipitation and sequencing analysis. In conjunction with our

current RNA-sequencing data, this combination of results would provide a comprehensive view of what the SUMO paralogs are doing in cells, at the level of gene and protein regulation. In addition to identifying unique SUMO1 or SUMO2/3 modified substrates, it will also be important to consider overlapping substrates since the effects of paralog-specific modification can differ, as shown with CFTR and IκBα (71, 184, 225).

How is substrate-specificity by the paralogs achieved? Two primary mechanisms were revealed through our analysis of the literature. First, differences in the SUMO interacting motif (SIM) binding domain between SUMO1 and SUMO2/3 play a role in determining paralog-specific modifications (48). For instance, DAXX and BLM are preferentially modified by SUMO1 and SUMO2/3, respectively. This is dictated through paralog-specific non-covalent interactions with these substrates that subsequently promotes covalent modification (188, 189, 191, 195). Since this has been observed for multiple SUMO1 and SUMO2/3 specific substrates, it is likely a common mechanism of paralog-specific modification. Secondly, some of the SUMO pathway enzymes display preferences for specific paralogs, as observed by SENP6 which preferentially hydrolyzes polymeric-SUMO2/3 chains (32). Understanding these mechanisms in greater detail is important given our evidence that the SUMO paralogs have unique functions in regulating a vast array of cellular functions. The next questions to ask include, how can the same SIM binding domains mediate interactions with so many diverse substrates, and is there a way to characterize the functional outcomes? For instance, lysine and arginine residues exist in the SIM binding domain of both SUMO1 and SUMO2/3, however they selectively provide SUMO2/3 with the ability to strongly repress transcription (54). Beyond its role in transcription, the SUMO2/3 SIM binding domain is also involved in mediating protein-protein

interactions between DNA repair factors, and kinetochore-associated proteins during mitosis (188, 198). Thus, it remains to be answered how this single domain can impart paralog-specific interactions with so many functionally diverse proteins. Of note, it's possible that varying polymeric-SUMO2/3 chain lengths may have important roles in mediating this specificity, but this too requires further study.

Future Directions to Evaluate Paralog-Specific Functions

In Chapter 3 we presented our systematic analysis of SUMO KO cell lines, which provided evidence for paralog-specific functions in regulating cellular morphology, the cellular stress response, nuclear body integrity and gene expression. The findings from this study revealed many questions, such as, what are the effects of SUMO2 on the cytoskeleton and regulatory pathways that control cell morphology? How do SUMO1 and SUMO2 individually sensitize cells to treatment with hydroxyurea? What are the targets and consequences of SUMO1 and SUMO2 modification on proteins localized to PML-NBs? Do SUMO1 and SUMO2 mediate association of transcription factors with unique regulatory proteins at gene promoters? How is expression of different genes specifically regulated by SUMO1 or SUMO2? These questions represent exciting new directions for future investigation. For instance, we are intrigued by the finding that a significant number of histone genes were specifically up-regulated in the SUMO2 KO cells, and had a corresponding decrease in SUMO1 KO cells. What is particularly peculiar about this finding is that although histone genes are localized to three histone gene loci, only a subset of the genes in these loci were changed (Figure 20). How are the SUMO paralogs selectively affecting the expression of only some histone genes? How is SUMO2 repressing their activation in WT cells,

and conversely, what is the mechanism behind SUMO1 activation of these histone genes? As similar trends were observed for immune response genes, uncovering this mechanism could contribute to our general understanding of SUMO-regulated gene expression in other cellular contexts.

To begin to understand the paralog-specific mechanisms driving the phenotypes described in Chapter 3, future studies need to be performed in WT cells. For instance, to better understand the effects of SUMO1 and SUMO2 on histone gene expression, it would be helpful to identify sumoylated substrates that are involved in this process in WT cells. Of note, to more generally understand how SUMO1 and SUMO2/3 are uniquely regulating transcription, we have performed Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq) on our U2OS WT and SUMO KO cells. Based on our RNA-seq analysis where we identified thousands of uniquely differentially expressed genes in the SUMO KO cells, we hypothesize that there will be many differences in accessible chromatin regions between the cell lines that correlate with affected genes. This finding would provide evidence that loss of the SUMO paralogs is associated with varying changes in the chromatin landscape, thereby suggesting that the SUMO paralogs uniquely regulate chromatin remodeling factors. Whether this regulation is through paralog-specific modification of unique factors, or through differing effects of the paralogs on shared modifiers will require further study, again in WT cells.

Lastly, we are interested in understanding what molecular attributes of the SUMO paralogs are required for the unique and non-redundant phenotypes reported in Chapter 3. We could explore these attributes by performing rescue experiments with targeted SUMO1 and SUMO2 mutants (Figure 21). For instance, to see whether

effects on transcription are dependent on SUMO conjugation, we could generate SUMO paralogs that do not have the di-glycine motif required for activation and conjugation onto target proteins. Moreover, we could generate mutants with defective SIM binding domains, and for SUMO2/3, mutants that lack the K11 internal sumoylation site that facilitates chain formation. Using our established qRT-PCR method, we could easily assay for effects of such mutants on gene expression. Lastly, it should be mentioned that in addition to doing these studies in U2OS cells for consistency and reproducibility, it will also be advantageous to use other human cell lines to demonstrate conservation of phenotypes. Identification of potentially interesting cell lines could be aided by the use of bioinformatics, as discussed in the following section.

Exploration of the SUMO Pathway Using Bioinformatics

Using bioinformatics, we developed a pipeline for evaluating SUMO pathway enzyme misregulation in human cancers. Beyond using this workflow to identify cancers with misregulated sumoylation, we can also use bioinformatics to learn about sumoylation more generally. For instance, the Genotype Expression (GTEx) project has gene expression data from over 15,000 human samples taken from 54 different non-cancerous tissues (https://www.gtexportal.org). Importantly, the data are annotated with patient attributes and thus it will be interesting to use this data and ask whether there is an association between the SUMO pathway components with age, sex or cause of death overall, and in specific tissues. Moreover, data portals such as cBioPortal (299), COSMIC (300) and TCGA (https://www.cancer.gov/tcga) contain a wealth of mutational information at the gene and protein levels from human cancer patient samples. Thus, it would also be interesting to mine data from these repositories

and ask whether the paralogs or pathway enzymes are misregulated in a particular cancer. Furthermore, we could ask if there are mutations that are associated with poor disease progression or prognosis. We could then study the biological consequences of such mutations in the lab. In summary, the questions that can be both asked and answered with the available data are seemingly endless, thereby opening many doors for further exploration.

The SUMO Pathway as a Therapeutic Target and a Biomarker

The work in this thesis highlights many of the essential biological processes that are regulated by the SUMO pathway. Misregulation of these processes is often reported in many human diseases, thereby implicating sumoylation in their onset and progression. As such, pharmaceutical companies have developed a SUMO E1 inhibitor, TAK-981, that is currently in clinical trials for cancer therapy, and interestingly, also for COVID-19 treatment (140-143). It will be interesting to see the results of these Phase I studies, which are evaluating the safety and tolerability of TAK-981 using dose escalation studies in human patients.

Given that a SUMO-pathway targeting drug has been developed, and hopefully will prove to be safe and effective, it would also be advantageous to have a biomarker for predicting the success of this drug in potential patients. Although we found that the SENP1 SUMO protease is not an effective biomarker for pancreatic cancer (Chapter 4), we are still interested in whether differential expression of SUMO pathway enzymes could be an indicator of success for a SUMO inhibitor. More specifically, are tissues or cells with elevated levels of SUMO proteases more sensitive to treatment with a SUMO inhibitor? To begin to address this question, we have established optimal

concentrations to obtain varying levels of SENP2 using tetracycline-inducible SENP2 expressing cells (Figure 22A). We have obtained a pre-clinical SUMO E1 inhibitor that is available for research use, and now need to determine the EC50 value of this drug in our SENP2 expressing cell line (197). We can then test the hypothesis that cells with the highest levels of the SUMO protease will be more sensitive to the drug, since they have the lowest levels of SUMO-conjugated proteins, as demonstrated through overexpressing SENP1 in HeLa cells (Figure 22B). Results from this study would provide evidence for or against the use of SUMO proteases as biomarkers for treatment with a SUMO inhibitor.

In summary, we have provided evidence that SUMO1 and SUMO2 have unique roles in regulating proteins associated with many human diseases. Our work provides a foundation for future medical and scientific SUMO-focused endeavors aimed at better understanding the contributions of sumoylation to health and disease and providing new avenues for targeted therapies.

Figures

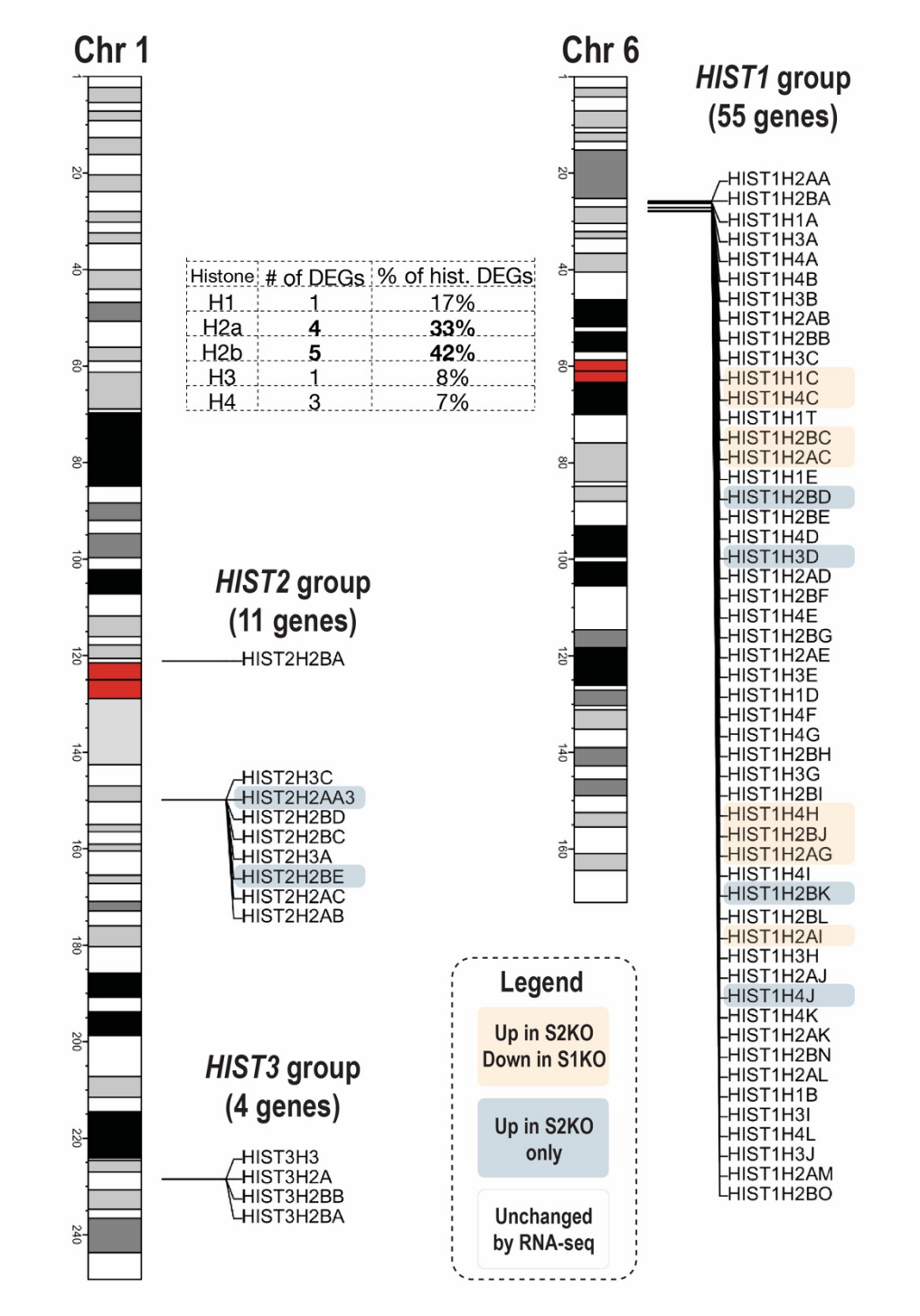


Figure 20. Labeled karyoplot of all histone genes. Histone genes are categorized into three distinct groups, as labeled, and localized to 3 predominant loci on human Chromosomes (Chr) 1 and 6. Genes that were changed in SUMO KO cells (from our RNA-seq data) are highlighted. The question remains to be answered about how only select and seemingly random histone genes are mis-regulated upon loss of SUMO1 and SUMO2.

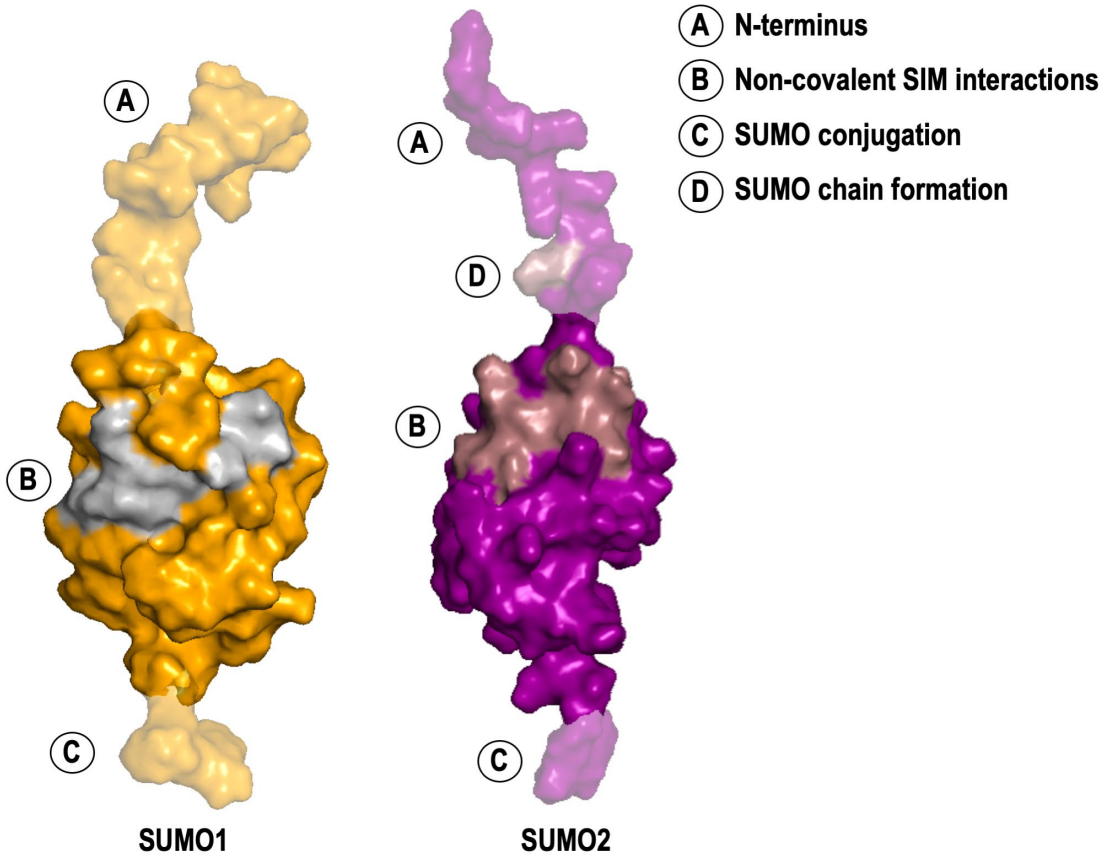


Figure 21. Targeted SUMO mutants for future mechanistic studies. Surface models of SUMO1 (PDB: 1A5R.A) and SUMO2 (PDB: 2N9E.B) showing targeted mutations (labeled), that could be interesting to use for future rescue experiments.

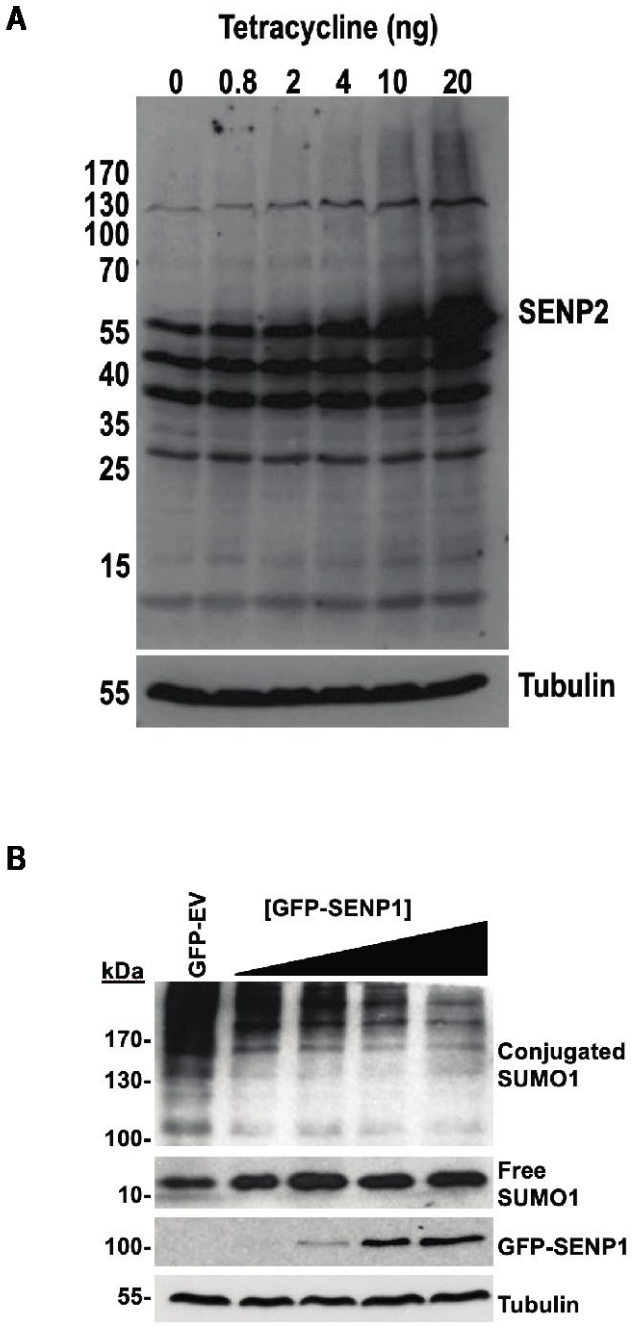


Figure 22. Varying levels of SUMO proteases in cell culture. A) Varying levels of SENP2 can be induced in HEK-293 cells, demonstrating a proof of concept method for directly testing the sensitivity of cancer cell lines with varying levels of SUMO proteases to treatment with sumoylation inhibitors. B) Preliminary data showing that increasing levels of SENP1 lead to a corresponding decrease in conjugated SUMO1 proteins in PANC-1 cells.

References

1. A. Varshavsky, The early history of the ubiquitin field. *Protein Sci* **15**, 647-654 (2006).
2. C. M. Pickart, Back to the future with ubiquitin. *Cell* **116**, 181-190 (2004).
3. M. Hochstrasser, Origin and function of ubiquitin-like proteins. *Nature* **458**, 422-429 (2009).
4. P. Beltrao, P. Bork, N. J. Krogan, V. van Noort, Evolution and functional cross-talk of protein post-translational modifications. *Mol Syst Biol* **9**, 714 (2013).
5. O. Kerscher, R. Felberbaum, M. Hochstrasser, Modification of proteins by ubiquitin and ubiquitin-like proteins. *Annu Rev Cell Dev Biol* **22**, 159-180 (2006).
6. M. J. Matunis, E. Coutavas, 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. *The Journal of Cell Biology* **135**, 1457-1470 (1996).
7. R. Mahajan, C. Delphin, T. Guan, L. Gerace, F. Melchior, A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* **88**, 97-107 (1997).
8. B. A. Schulman, J. W. Harper, Ubiquitin-like protein activation by E1 enzymes: the apex for downstream signalling pathways. *Nat Rev Mol Cell Biol* **10**, 319-331 (2009).
9. J. S. Seeler, A. Dejean, SUMO: of branched proteins and nuclear bodies. *Oncogene* **20**, 7243-7249 (2001).
10. N. Eisenhardt *et al.*, A new vertebrate SUMO enzyme family reveals insights into SUMO-chain assembly. *Nat Struct Mol Biol* **22**, 959-967 (2015).
11. M. J. Clague, C. Heride, S. Urbe, The demographics of the ubiquitin system. *Trends Cell Biol* **25**, 417-426 (2015).
12. A. J. George, Y. C. Hoffiz, A. J. Charles, Y. Zhu, A. M. Mabb, A Comprehensive Atlas of E3 Ubiquitin Ligase Mutations in Neurological Disorders. *Front Genet* **9**, 29 (2018).
13. F. Melchior, M. Schergaut, A. Pichler, SUMO: ligases, isopeptidases and nuclear pores. *Trends Biochem Sci* **28**, 612-618 (2003).
14. A. Rabellino, C. Andreani, P. P. Scaglioni, The Role of PIAS SUMO E3-Ligases in Cancer. *Cancer Res* **77**, 1542-1547 (2017).
15. H. M. Chang, E. T. H. Yeh, SUMO: From Bench to Bedside. *Physiol Rev* **100**, 1599-1619 (2020).
16. P. Radivojac *et al.*, Identification, analysis, and prediction of protein ubiquitination sites. *Proteins* **78**, 365-380 (2010).
17. S. H. Yang, A. Galanis, J. Witty, A. D. Sharrocks, An extended consensus motif enhances the specificity of substrate modification by SUMO. *EMBO J* **25**, 5083-5093 (2006).

18. J. Schimmel *et al.*, The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. *Mol Cell Proteomics* **7**, 2107-2122 (2008).
19. I. A. Hendriks, A. C. Vertegaal, A comprehensive compilation of SUMO proteomics. *Nat Rev Mol Cell Biol* **17**, 581-595 (2016).
20. A. Nayak, S. Muller, SUMO-specific proteases/isopeptidases: SENPs and beyond. *Genome Biol* **15**, 422 (2014).
21. T. Okura *et al.*, Protection against Fas/APO-1- and tumor necrosis factor-mediated cell death by a novel protein, sentrin. *J Immunol* **157**, 4277-4281 (1996).
22. K. Kunz, T. Piller, S. Müller, SUMO-specific proteases and isopeptidases of the SENP family at a glance. *Journal of Cell Science* **131**, jcs211904 (2018).
23. C. Cubenas-Potts *et al.*, Identification of SUMO-2/3-modified proteins associated with mitotic chromosomes. *Proteomics* **15**, 763-772 (2015).
24. C. Cubenas-Potts, J. D. Goeres, M. J. Matunis, SENP1 and SENP2 affect spatial and temporal control of sumoylation in mitosis. *Mol Biol Cell* **24**, 3483-3495 (2013).
25. N. Kolli *et al.*, Distribution and paralogue specificity of mammalian deSUMOylating enzymes. *Biochem J* **430**, 335-344 (2010).
26. D. Bailey, P. O'Hare, Characterization of the localization and proteolytic activity of the SUMO-specific protease, SENP1. *J Biol Chem* **279**, 692-703 (2004).
27. J. Goeres *et al.*, The SUMO-specific isopeptidase SENP2 associates dynamically with nuclear pore complexes through interactions with karyopherins and the Nup107-160 nucleoporin subcomplex. *Mol Biol Cell* **22**, 4868-4882 (2011).
28. K.-H. Chow, S. Elgort, M. Dasso, M. A. Powers, K. S. Ullman, The SUMO proteases SENP1 and SENP2 play a critical role in nucleoporin homeostasis and nuclear pore complex function. *Molecular Biology of the Cell* **25**, 160-168 (2014).
29. H. M. Odeh, E. Coyaud, B. Raught, M. J. Matunis, The SUMO-specific isopeptidase SENP2 is targeted to intracellular membranes via a predicted N-terminal amphipathic alpha-helix. *Mol Biol Cell* **29**, 1878-1890 (2018).
30. R. Zunino, A. Schauss, P. Rippstein, M. Andrade-Navarro, H. M. McBride, The SUMO protease SENP5 is required to maintain mitochondrial morphology and function. *J Cell Sci* **120**, 1178-1188 (2007).
31. R. Zunino, E. Braschi, L. Xu, H. M. McBride, Translocation of SenP5 from the nucleoli to the mitochondria modulates DRP1-dependent fission during mitosis. *J Biol Chem* **284**, 17783-17795 (2009).
32. D. Mukhopadhyay, A. Arnaoutov, M. Dasso, The SUMO protease SENP6 is essential for inner kinetochore assembly. *J Cell Biol* **188**, 681-692 (2010).
33. A. V. Mendes, C. P. Grou, J. E. Azevedo, M. P. Pinto, Evaluation of the activity and substrate specificity of the human SENP family of SUMO proteases. *Biochim Biophys Acta* **1863**, 139-147 (2016).
34. C. Burdelski *et al.*, The prognostic value of SUMO1/Sentrin specific peptidase 1 (SENP1) in prostate cancer is limited to ERG-fusion positive tumors lacking PTEN deletion. *BMC Cancer* **15**, 538 (2015).

35. D. M. Bouchard, M. J. Matunis, A cellular and bioinformatics analysis of the SENP1 SUMO isopeptidase in pancreatic cancer. *J Gastrointest Oncol* **10**, 821-830 (2019).
36. I. Matic, B. Macek, M. Hilger, T. C. Walther, M. Mann, Phosphorylation of SUMO-1 occurs in vivo and is conserved through evolution. *J Proteome Res* **7**, 4050-4057 (2008).
37. J. Bouchenna *et al.*, The Role of the Conserved SUMO-2/3 Cysteine Residue on Domain Structure Investigated Using Protein Chemical Synthesis. *Bioconjug Chem* **30**, 2684-2696 (2019).
38. M. H. Tatham *et al.*, Polymeric chains of SUMO-2 and SUMO-3 are conjugated to protein substrates by SAE1/SAE2 and Ubc9. *J Biol Chem* **276**, 35368-35374 (2001).
39. A. M. Sriramachandran *et al.*, Arkadia/RNF111 is a SUMO-targeted ubiquitin ligase with preference for substrates marked with SUMO1-capped SUMO2/3 chain. *Nat Commun* **10**, 3678 (2019).
40. H. Sun, J. D. Leverson, T. Hunter, Conserved function of RNF4 family proteins in eukaryotes: targeting a ubiquitin ligase to SUMOylated proteins. *EMBO J* **26**, 4102-4112 (2007).
41. I. Matic *et al.*, In vivo identification of human small ubiquitin-like modifier polymerization sites by high accuracy mass spectrometry and an in vitro to in vivo strategy. *Mol Cell Proteomics* **7**, 132-144 (2008).
42. D. Owerbach, E. M. McKay, E. T. Yeh, K. H. Gabbay, K. M. Bohren, A proline-90 residue unique to SUMO-4 prevents maturation and sumoylation. *Biochem Biophys Res Commun* **337**, 517-520 (2005).
43. W. Wei *et al.*, A stress-dependent SUMO4 sumoylation of its substrate proteins. *Biochem Biophys Res Commun* **375**, 454-459 (2008).
44. D. Guo *et al.*, Proteomic analysis of SUMO4 substrates in HEK293 cells under serum starvation-induced stress. *Biochem Biophys Res Commun* **337**, 1308-1318 (2005).
45. Y. Y. Li *et al.*, Small Ubiquitin-Like Modifier 4 (SUMO4) Gene M55V Polymorphism and Type 2 Diabetes Mellitus: A Meta-analysis Including 6,823 Subjects. *Front Endocrinol (Lausanne)* **8**, 303 (2017).
46. Y. C. Liang *et al.*, SUMO5, a Novel Poly-SUMO Isoform, Regulates PML Nuclear Bodies. *Sci Rep* **6**, 26509 (2016).
47. E. Meulmeester, M. Kunze, H. H. Hsiao, H. Urlaub, F. Melchior, Mechanism and consequences for paralog-specific sumoylation of ubiquitin-specific protease 25. *Mol Cell* **30**, 610-619 (2008).
48. J. Song, L. K. Durrin, T. A. Wilkinson, T. G. Krontiris, Y. Chen, Identification of a SUMO-binding motif that recognizes SUMO-modified proteins. *Proc Natl Acad Sci U S A* **101**, 14373-14378 (2004).
49. C. M. Hecker, M. Rabiller, K. Haglund, P. Bayer, I. Dikic, Specification of SUMO1- and SUMO2-interacting motifs. *J Biol Chem* **281**, 16117-16127 (2006).
50. M. H. Tatham *et al.*, Unique binding interactions among Ubc9, SUMO and RanBP2 reveal a mechanism for SUMO paralog selection. *Nat Struct Mol Biol* **12**, 67-74 (2005).

51. J. Song, Z. Zhang, W. Hu, Y. Chen, Small ubiquitin-like modifier (SUMO) recognition of a SUMO binding motif: a reversal of the bound orientation. *J Biol Chem* **280**, 40122-40129 (2005).
52. S. Citro, S. Chiocca, Sumo paralogs: redundancy and divergencies. *Front Biosci (Schol Ed)* **5**, 544-553 (2013).
53. W. C. Huang, T. P. Ko, S. S. Li, A. H. Wang, Crystal structures of the human SUMO-2 protein at 1.6 A and 1.2 A resolution: implication on the functional differences of SUMO proteins. *Eur J Biochem* **271**, 4114-4122 (2004).
54. S. Chupreta, S. Holmstrom, L. Subramanian, J. A. Iniguez-Lluhi, A small conserved surface in SUMO is the critical structural determinant of its transcriptional inhibitory properties. *Mol Cell Biol* **25**, 4272-4282 (2005).
55. T. Hoppe, Multiubiquitylation by E4 enzymes: 'one size' doesn't fit all. *Trends Biochem Sci* **30**, 183-187 (2005).
56. S. Zhong *et al.*, Role of SUMO-1-modified PML in nuclear body formation. *Blood* **95**, 2748-2752 (2000).
57. T. H. Shen, H. K. Lin, P. P. Scaglioni, T. M. Yung, P. P. Pandolfi, The Mechanisms of PML-Nuclear Body Formation. *Mol Cell* **24**, 805 (2006).
58. V. Lallemand-Breitenbach, H. de The, PML nuclear bodies. *Cold Spring Harb Perspect Biol* **2**, a000661 (2010).
59. R. Bernardi, P. P. Pandolfi, Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat Rev Mol Cell Biol* **8**, 1006-1016 (2007).
60. S. Zhong, P. Salomoni, P. P. Pandolfi, The transcriptional role of PML and the nuclear body. *Nat Cell Biol* **2**, E85-90 (2000).
61. D. Torok, R. W. Ching, D. P. Bazett-Jones, PML nuclear bodies as sites of epigenetic regulation. *Front Biosci (Landmark Ed)* **14**, 1325-1336 (2009).
62. M. Scherer, T. Stamminger, Emerging Role of PML Nuclear Bodies in Innate Immune Signaling. *J Virol* **90**, 5850-5854 (2016).
63. K. Nacerddine *et al.*, The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev Cell* **9**, 769-779 (2005).
64. X. He *et al.*, Characterization of the loss of SUMO pathway function on cancer cells and tumor proliferation. *PLoS One* **10**, e0123882 (2015).
65. F. El-Asmi, F. P. McManus, P. Thibault, M. K. Chelbi-Alix, Interferon, restriction factors and SUMO pathways. *Cytokine Growth Factor Rev* 10.1016/j.cytogfr.2020.03.001 (2020).
66. X. Cheng, H. Y. Kao, Post-translational modifications of PML: consequences and implications. *Front Oncol* **2**, 210 (2012).
67. J. A. Brosnan-Cashman *et al.*, ATRX loss induces multiple hallmarks of the alternative lengthening of telomeres (ALT) phenotype in human glioma cell lines in a cell line-specific manner. *PLoS One* **13**, e0204159 (2018).
68. J. Schimmel *et al.*, Uncovering SUMOylation dynamics during cell-cycle progression reveals FoxM1 as a key mitotic SUMO target protein. *Mol Cell* **53**, 1053-1066 (2014).
69. Y. Galanty, R. Belotserkovskaya, J. Coates, S. P. Jackson, RNF4, a SUMO-targeted ubiquitin E3 ligase, promotes DNA double-strand break repair. *Genes Dev* **26**, 1179-1195 (2012).

70. S. Muller *et al.*, c-Jun and p53 activity is modulated by SUMO-1 modification. *J Biol Chem* **275**, 13321-13329 (2000).
71. F. Aillet *et al.*, Heterologous SUMO-2/3-ubiquitin chains optimize IkappaBalpha degradation and NF-kappaB activity. *PLoS One* **7**, e51672 (2012).
72. C. Smet-Nocca, J. M. Wieruszeski, H. Leger, S. Eilebrecht, A. Benecke, SUMO-1 regulates the conformational dynamics of thymine-DNA Glycosylase regulatory domain and competes with its DNA binding activity. *BMC Biochem* **12**, 4 (2011).
73. D. Baba *et al.*, Crystal structure of thymine DNA glycosylase conjugated to SUMO-1. *Nature* **435**, 979-982 (2005).
74. D. McLaughlin, C. T. Coey, W.-C. Yang, A. C. Drohat, M. J. Matunis, Characterizing Requirements for SUMO Modification and Binding on Base Excision Repair Activity of Thymine DNA Glycosylase in Vivo. *Journal of Biological Chemistry* 10.1074/jbc.M115.706325 (2016).
75. B. Vogelstein, K. W. Kinzler, Cancer genes and the pathways they control. *Nature Medicine* **10**, 789 (2004).
76. J. S. Seeler, A. Dejean, SUMO and the robustness of cancer. *Nat Rev Cancer* **17**, 184-197 (2017).
77. K. Eifler, A. C. Vertegaal, SUMOylation-Mediated Regulation of Cell Cycle Progression and Cancer. *Trends Biochem Sci* **40**, 779-793 (2015).
78. V. Dorval, P. E. Fraser, SUMO on the road to neurodegeneration. *Biochim Biophys Acta* **1773**, 694-706 (2007).
79. J. Wang *et al.*, Defective sumoylation pathway directs congenital heart disease. *Birth Defects Res A Clin Mol Teratol* **91**, 468-476 (2011).
80. W. Yang, H. Sheng, H. Wang, Targeting the SUMO pathway for neuroprotection in brain ischaemia. *Stroke Vasc Neurol* **1**, 101-107 (2016).
81. D. B. Anderson, C. A. Zanella, J. M. Henley, H. Cimarosti, Sumoylation: Implications for Neurodegenerative Diseases. *Adv Exp Med Biol* **963**, 261-281 (2017).
82. L. Mendler, T. Braun, S. Muller, The Ubiquitin-Like SUMO System and Heart Function: From Development to Disease. *Circ Res* **118**, 132-144 (2016).
83. G. B. D. C. o. D. Collaborators, Global, regional, and national age-sex-specific mortality for 282 causes of death in 195 countries and territories, 1980-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet* **392**, 1736-1788 (2018).
84. L. Schorova, S. Martin, Sumoylation in Synaptic Function and Dysfunction. *Front Synaptic Neurosci* **8**, 9 (2016).
85. V. Lallemand-Breitenbach *et al.*, Arsenic degrades PML or PML-RARalpha through a SUMO-triggered RNF4/ubiquitin-mediated pathway. *Nat Cell Biol* **10**, 547-555 (2008).
86. H. de The, M. Le Bras, V. Lallemand-Breitenbach, The cell biology of disease: Acute promyelocytic leukemia, arsenic, and PML bodies. *J Cell Biol* **198**, 11-21 (2012).
87. J. Zhu *et al.*, A sumoylation site in PML/RARA is essential for leukemic transformation. *Cancer Cell* **7**, 143-153 (2005).

88. C. S. Grove, G. S. Vassiliou, Acute myeloid leukaemia: a paradigm for the clonal evolution of cancer? *Dis Model Mech* **7**, 941-951 (2014).
89. K. Alimoghaddam, A review of arsenic trioxide and acute promyelocytic leukemia. *Int J Hematol Oncol Stem Cell Res* **8**, 44-54 (2014).
90. S. J. Moschos *et al.*, Expression analysis of Ubc9, the single small ubiquitin-like modifier (SUMO) E2 conjugating enzyme, in normal and malignant tissues. *Hum Pathol* **41**, 1286-1298 (2010).
91. F. Wu, S. Zhu, Y. Ding, W. T. Beck, Y. Y. Mo, MicroRNA-mediated regulation of Ubc9 expression in cancer cells. *Clin Cancer Res* **15**, 1550-1557 (2009).
92. H. Li, H. Niu, Y. Peng, J. Wang, P. He, Ubc9 promotes invasion and metastasis of lung cancer cells. *Oncol Rep* **29**, 1588-1594 (2013).
93. S. F. Chen *et al.*, Ubc9 expression predicts chemoresistance in breast cancer. *Chin J Cancer* **30**, 638-644 (2011).
94. M. Taheri, V. K. Oskooei, S. Ghafouri-Fard, Protein inhibitor of activated STAT genes are differentially expressed in breast tumor tissues. *Per Med* **16**, 277-285 (2019).
95. N. Kotaja, U. Karvonen, O. A. Janne, J. J. Palvimo, PIAS proteins modulate transcription factors by functioning as SUMO-1 ligases. *Mol Cell Biol* **22**, 5222-5234 (2002).
96. M. Shiota, N. Fujimoto, E. Kashiwagi, M. Eto, The Role of Nuclear Receptors in Prostate Cancer. *Cells* **8** (2019).
97. M. Puhr *et al.*, PIAS1 is a crucial factor for prostate cancer cell survival and a valid target in docetaxel resistant cells. *Oncotarget* **5**, 12043-12056 (2014).
98. J. Hoefer *et al.*, PIAS1 is increased in human prostate cancer and enhances proliferation through inhibition of p21. *Am J Pathol* **180**, 2097-2107 (2012).
99. M. Puhr *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* **35**, 2322-2332 (2016).
100. Z. Zheng *et al.*, SUMO-3 enhances androgen receptor transcriptional activity through a sumoylation-independent mechanism in prostate cancer cells. *J Biol Chem* **281**, 4002-4012 (2006).
101. J. S. Rawlings, K. M. Rosler, D. A. Harrison, The JAK/STAT signaling pathway. *J Cell Sci* **117**, 1281-1283 (2004).
102. P. E. Lonergan, D. J. Tindall, Androgen receptor signaling in prostate cancer development and progression. *J Carcinog* **10**, 20 (2011).
103. Z. Wang, J. Jin, J. Zhang, L. Wang, J. Cao, Depletion of SENP1 suppresses the proliferation and invasion of triple-negative breast cancer cells. *Oncol Rep* **36**, 2071-2078 (2016).
104. J. Cheng, T. Bawa, P. Lee, L. Gong, E. T. Yeh, Role of desumoylation in the development of prostate cancer. *Neoplasia* **8**, 667-676 (2006).
105. C. P. Cui *et al.*, SENP1 promotes hypoxia-induced cancer stemness by HIF-1alpha deSUMOylation and SENP1/HIF-1alpha positive feedback loop. *Gut* 10.1136/gutjnl-2016-313264 (2017).
106. M. e. a. Tan, SENP2 regulates MMP13 expression in a bladder cancer cell line through SUMOylation of TBL1/TBLR1. *Sci Rep* **5** (2015).

107. K. Liu, J. Zhang, H. Wang, Small ubiquitin-like modifier/sentrin-specific peptidase 1 associates with chemotherapy and is a risk factor for poor prognosis of non-small cell lung cancer. *J Clin Lab Anal* 10.1002/jcla.22611, e22611 (2018).
108. X. Zhang *et al.*, SUMO-Specific Cysteine Protease 1 Promotes Epithelial Mesenchymal Transition of Prostate Cancer Cells via Regulating SMAD4 deSUMOylation. *International Journal of Molecular Sciences* **18**, 808 (2017).
109. Q. Wang *et al.*, SUMO-specific protease 1 promotes prostate cancer progression and metastasis. *Oncogene* **32**, 2493-2498 (2013).
110. Y. Xu *et al.*, SUMO-specific protease 1 regulates the in vitro and in vivo growth of colon cancer cells with the upregulated expression of CDK inhibitors. *Cancer Lett* **309**, 78-84 (2011).
111. T. Bawa-Khalfe, E. T. Yeh, SUMO Losing Balance: SUMO Proteases Disrupt SUMO Homeostasis to Facilitate Cancer Development and Progression. *Genes Cancer* **1**, 748-752 (2010).
112. C. Ma *et al.*, SUMO-specific protease 1 regulates pancreatic cancer cell proliferation and invasion by targeting MMP-9. *Tumour Biol* **35**, 12729-12735 (2014).
113. Y. H. Ren *et al.*, De-SUMOylation of FOXC2 by SENP3 promotes the epithelial-mesenchymal transition in gastric cancer cells. *Oncotarget* **5**, 7093-7104 (2014).
114. X. L. Chen *et al.*, SENP2 exerts an antitumor effect on chronic lymphocytic leukemia cells through the inhibition of the Notch and NFkappaB signaling pathways. *Int J Oncol* **54**, 455-466 (2019).
115. H. Pei *et al.*, SUMO-specific protease 2 (SENP2) functions as a tumor suppressor in osteosarcoma via SOX9 degradation. *Exp Ther Med* **16**, 5359-5365 (2018).
116. A. Lee, Y. Zhu, Y. Sabo, S. P. Goff, Embryonic Cells Redistribute SUMO1 upon Forced SUMO1 Overexpression. *mBio* **10** (2019).
117. J. R. Morris *et al.*, The SUMO modification pathway is involved in the BRCA1 response to genotoxic stress. *Nature* **462**, 886-890 (2009).
118. C. M. Guzzo *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* **5**, ra88 (2012).
119. W. Wu, A. Koike, T. Takeshita, T. Ohta, The ubiquitin E3 ligase activity of BRCA1 and its biological functions. *Cell Div* **3**, 1 (2008).
120. C. Wang, M. P. Lisanti, D. J. Liao, Reviewing once more the c-myc and Ras collaboration: converging at the cyclin D1-CDK4 complex and challenging basic concepts of cancer biology. *Cell Cycle* **10**, 57-67 (2011).
121. S. Jancik, J. Drabek, D. Radzioch, M. Hajduch, Clinical relevance of KRAS in human cancers. *J Biomed Biotechnol* **2010**, 150960 (2010).
122. J. Luo *et al.*, A genome-wide RNAi screen identifies multiple synthetic lethal interactions with the Ras oncogene. *Cell* **137**, 835-848 (2009).
123. B. Yu *et al.*, Oncogenesis driven by the Ras/Raf pathway requires the SUMO E2 ligase Ubc9. *Proc Natl Acad Sci U S A* **112**, E1724-1733 (2015).
124. S. Mori *et al.*, Anchorage-independent cell growth signature identifies tumors with metastatic potential. *Oncogene* **28**, 2796-2805 (2009).

125. J. D. Kessler *et al.*, A SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. *Science* **335**, 348-353 (2012).
126. J. S. Steffan *et al.*, SUMO modification of Huntingtin and Huntington's disease pathology. *Science* **304**, 100-104 (2004).
127. L. Ford, L. Fioriti, E. R. Kandel, Ubiquitination and SUMOylation of Amyloid and Amyloid-like Proteins in Health and Disease. *Curr Issues Mol Biol* **35**, 195-230 (2020).
128. J. G. O'Rourke *et al.*, SUMO-2 and PIAS1 modulate insoluble mutant huntingtin protein accumulation. *Cell Rep* **4**, 362-375 (2013).
129. M. Marinello *et al.*, SUMOylation by SUMO2 is implicated in the degradation of misfolded ataxin-7 via RNF4 in SCA7 models. *Dis Model Mech* **12** (2019).
130. A. Decque *et al.*, Sumoylation coordinates the repression of inflammatory and anti-viral gene-expression programs during innate sensing. *Nat Immunol* **17**, 140-149 (2016).
131. J. M. Kahlenberg, M. J. Kaplan, The inflammasome and lupus: another innate immune mechanism contributing to disease pathogenesis? *Curr Opin Rheumatol* **26**, 475-481 (2014).
132. J. T. Crowl, D. B. Stetson, SUMO2 and SUMO3 redundantly prevent a noncanonical type I interferon response. *Proc Natl Acad Sci U S A* **115**, 6798-6803 (2018).
133. A. El Motiam *et al.*, SUMO and Cytoplasmic RNA Viruses: From Enemies to Best Friends. *Adv Exp Med Biol* **1233**, 263-277 (2020).
134. P. Domingues *et al.*, Global Reprogramming of Host SUMOylation during Influenza Virus Infection. *Cell Rep* **13**, 1467-1480 (2015).
135. A. M. Cole, N. Petousi, G. L. Cavalleri, P. A. Robbins, Genetic variation in SENP1 and ANP32D as predictors of chronic mountain sickness. *High Alt Med Biol* **15**, 497-499 (2014).
136. M. M. Hsieh *et al.*, SENP1, but not fetal hemoglobin, differentiates Andean highlanders with chronic mountain sickness from healthy individuals among Andean highlanders. *Exp Hematol* **44**, 483-490 e482 (2016).
137. D. Zhou *et al.*, Whole-genome sequencing uncovers the genetic basis of chronic mountain sickness in Andean highlanders. *Am J Hum Genet* **93**, 452-462 (2013).
138. P. Azad *et al.*, Senp1 drives hypoxia-induced polycythemia via GATA1 and Bcl-xL in subjects with Monge's disease. *J Exp Med* **213**, 2729-2744 (2016).
139. A. Y. Alontaga *et al.*, Design of high-throughput screening assays and identification of a SUMO1-specific small molecule chemotype targeting the SUMO-interacting motif-binding surface. *ACS Comb Sci* **17**, 239-246 (2015).
140. P. Biosciences (Intratumoral Microdosing of TAK-981 in Head and Neck Cancer.
141. A. M. Sarit Assouline, Paolo F Caimi, Bingxia Wang, Chirag Patel, Mi-Sook Kim, Dennis Huszar, Allison J Berger, Sharon Friedlander, Alejandro Gomez-Pinillos, Igor Proscurshim, A Phase 1b/2 Study of TAK-981, a First-in-Class Sumoylation Inhibitor, in Combination with Rituximab in Patients with Relapsed/Refractory (r/r) CD20-Positive Non-Hodgkin Lymphoma (NHL). *Blood* **134** (2019).

142. Takeda (A Safety and Efficacy Study of TAK-981 Plus Pembrolizumab in Participants With Select Advanced or Metastatic Solid Tumors.
143. I. Millennium Pharmaceuticals (A Study of TAK-981 in Combination With Rituximab in Participants With Relapsed/Refractory (r/r) CD20-positive (CD20+) Non-Hodgkin Lymphoma (NHL).
144. D. N. Valencia, Brief Review on COVID-19: The 2020 Pandemic Caused by SARS-CoV-2. *Cureus* **12**, e7386 (2020).
145. R. N. Gilbreth *et al.*, Isoform-specific monobody inhibitors of small ubiquitin-related modifiers engineered using structure-guided library design. *Proc Natl Acad Sci U S A* **108**, 7751-7756 (2011).
146. J. Wu *et al.*, Momordin Ic, a new natural SENP1 inhibitor, inhibits prostate cancer cell proliferation. *Oncotarget* **7**, 58995-59005 (2016).
147. H. A. Newman *et al.*, A high throughput mutagenic analysis of yeast sumo structure and function. *PLoS Genet* **13**, e1006612 (2017).
148. S. A. Saracco, M. J. Miller, J. Kurepa, R. D. Vierstra, Genetic analysis of SUMOylation in Arabidopsis: conjugation of SUMO1 and SUMO2 to nuclear proteins is essential. *Plant Physiol* **145**, 119-134 (2007).
149. M. Smith, W. Turki-Judeh, A. J. Courey, SUMOylation in Drosophila Development. *Biomolecules* **2**, 331-349 (2012).
150. K. Drabikowski *et al.*, Comprehensive list of SUMO targets in Caenorhabditis elegans and its implication for evolutionary conservation of SUMO signaling. *Sci Rep* **8**, 1139 (2018).
151. H. A. van den Burg, R. K. Kini, R. C. Schuurink, F. L. Takken, Arabidopsis small ubiquitin-like modifier paralogs have distinct functions in development and defense. *Plant Cell* **22**, 1998-2016 (2010).
152. L. Castano-Miquel, J. Segui, L. M. Lois, Distinctive properties of Arabidopsis SUMO paralogues support the in vivo predominant role of AtSUMO1/2 isoforms. *Biochem J* **436**, 581-590 (2011).
153. T. Colby, A. Matthai, A. Boeckelmann, H. P. Stuible, SUMO-conjugating and SUMO-deconjugating enzymes from Arabidopsis. *Plant Physiol* **142**, 318-332 (2006).
154. V. Abad-Morales, E. B. Domenech, A. Garanto, G. Marfany, mRNA expression analysis of the SUMO pathway genes in the adult mouse retina. *Biol Open* **4**, 224-232 (2015).
155. L. Wang *et al.*, SUMO2 is essential while SUMO3 is dispensable for mouse embryonic development. *EMBO Rep* **15**, 878-885 (2014).
156. E. Pauws, P. Stanier, FGF signalling and SUMO modification: new players in the aetiology of cleft lip and/or palate. *Trends Genet* **23**, 631-640 (2007).
157. F. S. Alkuraya *et al.*, SUMO1 haploinsufficiency leads to cleft lip and palate. *Science* **313**, 1751 (2006).
158. E. Evdokimov, P. Sharma, S. J. Lockett, M. Lualdi, M. R. Kuehn, Loss of SUMO1 in mice affects RanGAP1 localization and formation of PML nuclear bodies, but is not lethal as it can be compensated by SUMO2 or SUMO3. *J Cell Sci* **121**, 4106-4113 (2008).
159. F. P. Zhang *et al.*, Sumo-1 function is dispensable in normal mouse development. *Mol Cell Biol* **28**, 5381-5390 (2008).

160. L. Mikkonen, J. Hirvonen, O. A. Janne, SUMO-1 regulates body weight and adipogenesis via PPARgamma in male and female mice. *Endocrinology* **154**, 698-708 (2013).

161. N. A. de Assis *et al.*, SUMO1 as a candidate gene for non-syndromic cleft lip with or without cleft palate: no evidence for the involvement of common or rare variants in Central European patients. *Int J Pediatr Otorhinolaryngol* **75**, 49-52 (2011).

162. J.-S. Seeler, A. Dejean, Nuclear and unclear functions of SUMO. *Nature Reviews Molecular Cell Biology* **4**, 690 (2003).

163. V. G. Wilson, "Introduction to Sumoylation" in SUMO Regulation of Cellular Processes, V. G. Wilson, Ed. (Springer International Publishing, Cham, 2017), 10.1007/978-3-319-50044-7_1, pp. 1-12.

164. N. Schmidt *et al.*, An influenza virus-triggered SUMO switch orchestrates co-opted endogenous retroviruses to stimulate host antiviral immunity. *Proc Natl Acad Sci U S A* **116**, 17399-17408 (2019).

165. J. I. Abe *et al.*, Coordination of Cellular Localization-Dependent Effects of Sumoylation in Regulating Cardiovascular and Neurological Diseases. *Adv Exp Med Biol* **963**, 337-358 (2017).

166. A. Pichler, A. Gast, J. S. Seeler, A. Dejean, F. Melchior, The nucleoporin RanBP2 has SUMO1 E3 ligase activity. *Cell* **108**, 109-120 (2002).

167. O. Kirsh *et al.*, The SUMO E3 ligase RanBP2 promotes modification of the HDAC4 deacetylase. *EMBO J* **21**, 2682-2691 (2002).

168. S. Zhu *et al.*, Protection from isopeptidase-mediated deconjugation regulates paralog-selective sumoylation of RanGAP1. *Mol Cell* **33**, 570-580 (2009).

169. L. N. Shen, C. Dong, H. Liu, J. H. Naismith, R. T. Hay, The structure of SENP1-SUMO-2 complex suggests a structural basis for discrimination between SUMO paralogues during processing. *Biochem J* **397**, 279-288 (2006).

170. K. O. Alegre, D. Reverter, Swapping small ubiquitin-like modifier (SUMO) isoform specificity of SUMO proteases SENP6 and SENP7. *J Biol Chem* **286**, 36142-36151 (2011).

171. G. Rosas-Acosta, W. K. Russell, A. Deyrieux, D. H. Russell, V. G. Wilson, A universal strategy for proteomic studies of SUMO and other ubiquitin-like modifiers. *Mol Cell Proteomics* **4**, 56-72 (2005).

172. J. Becker *et al.*, Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat Struct Mol Biol* **20**, 525-531 (2013).

173. F. Impens, L. Radoshevich, P. Cossart, D. Ribet, Mapping of SUMO sites and analysis of SUMOylation changes induced by external stimuli. *Proc Natl Acad Sci U S A* **111**, 12432-12437 (2014).

174. A. C. Vertegaal *et al.*, Distinct and overlapping sets of SUMO-1 and SUMO-2 target proteins revealed by quantitative proteomics. *Mol Cell Proteomics* **5**, 2298-2310 (2006).

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

176. Z. Xiao *et al.*, System-wide Analysis of SUMOylation Dynamics in Response to Replication Stress Reveals Novel Small Ubiquitin-like Modified Target

Proteins and Acceptor Lysines Relevant for Genome Stability. *Mol Cell Proteomics* **14**, 1419-1434 (2015).

177. I. A. Hendriks, L. W. Treffers, M. Verlaan-de Vries, J. V. Olsen, A. C. O. Vertegaal, SUMO-2 Orchestrates Chromatin Modifiers in Response to DNA Damage. *Cell Rep* **10**, 1778-1791 (2015).
178. H. Saitoh, J. Hinchey, Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J Biol Chem* **275**, 6252-6258 (2000).
179. F. Golebiowski *et al.*, System-wide changes to SUMO modifications in response to heat shock. *Sci Signal* **2**, ra24 (2009).
180. A. Seifert, P. Schofield, G. J. Barton, R. T. Hay, Proteotoxic stress reprograms the chromatin landscape of SUMO modification. *Sci Signal* **8**, rs7 (2015).
181. M. H. Z. Guang *et al.*, Targeting Proteotoxic Stress in Cancer: A Review of the Role that Protein Quality Control Pathways Play in Oncogenesis. *Cancers (Basel)* **11** (2019).
182. J. R. Blount, A. A. Burr, A. Denuc, G. Marfany, S. V. Todi, Ubiquitin-specific protease 25 functions in Endoplasmic Reticulum-associated degradation. *PLoS One* **7**, e36542 (2012).
183. A. Ahner, X. Gong, R. A. Frizzell, Cystic fibrosis transmembrane conductance regulator degradation: cross-talk between the ubiquitylation and SUMOylation pathways. *FEBS J* **280**, 4430-4438 (2013).
184. A. Ahner, X. Gong, R. A. Frizzell, Divergent signaling via SUMO modification: potential for CFTR modulation. *Am J Physiol Cell Physiol* **310**, C175-180 (2016).
185. C. S. Wu *et al.*, SUMOylation of ATRIP potentiates DNA damage signaling by boosting multiple protein interactions in the ATR pathway. *Genes Dev* **28**, 1472-1484 (2014).
186. M. Nie, M. N. Boddy, Cooperativity of the SUMO and Ubiquitin Pathways in Genome Stability. *Biomolecules* **6**, 14 (2016).
187. K. J. Ouyang *et al.*, SUMO modification regulates BLM and RAD51 interaction at damaged replication forks. *PLoS Biol* **7**, e1000252 (2009).
188. J. Zhu *et al.*, Small ubiquitin-related modifier (SUMO) binding determines substrate recognition and paralog-selective SUMO modification. *J Biol Chem* **283**, 29405-29415 (2008).
189. S. Eladad *et al.*, Intra-nuclear trafficking of the BLM helicase to DNA damage-induced foci is regulated by SUMO modification. *Hum Mol Genet* **14**, 1351-1365 (2005).
190. I. A. Sawyer, J. Bartek, M. Dundr, Phase separated microenvironments inside the cell nucleus are linked to disease and regulate epigenetic state, transcription and RNA processing. *Semin Cell Dev Biol* **90**, 94-103 (2019).
191. M. S. Jang, S. W. Ryu, E. Kim, Modification of Daxx by small ubiquitin-related modifier-1. *Biochem Biophys Res Commun* **295**, 495-500 (2002).
192. Y. S. Mao, B. Zhang, D. L. Spector, Biogenesis and function of nuclear bodies. *Trends Genet* **27**, 295-306 (2011).
193. D. Fasci, V. G. Anania, J. R. Lill, G. S. Salvesen, SUMO deconjugation is required for arsenic-triggered ubiquitylation of PML. *Sci Signal* **8**, ra56 (2015).

194. M. T. Naik *et al.*, NMR chemical shift assignments of a complex between SUMO-1 and SIM peptide derived from the C-terminus of Daxx. *Biomol NMR Assign* **5**, 75-77 (2011).
195. D. Y. Lin *et al.*, Role of SUMO-interacting motif in Daxx SUMO modification, subnuclear localization, and repression of sumoylated transcription factors. *Mol Cell* **24**, 341-354 (2006).
196. C. C. Chang *et al.*, Structural and functional roles of Daxx SIM phosphorylation in SUMO paralog-selective binding and apoptosis modulation. *Mol Cell* **42**, 62-74 (2011).
197. X. He *et al.*, Probing the roles of SUMOylation in cancer cell biology by using a selective SAE inhibitor. **13**, 1164 (2017).
198. X. D. Zhang *et al.*, SUMO-2/3 modification and binding regulate the association of CENP-E with kinetochores and progression through mitosis. *Mol Cell* **29**, 729-741 (2008).
199. J. Joseph, S. H. Tan, T. S. Karpova, J. G. McNally, M. Dasso, SUMO-1 targets RanGAP1 to kinetochores and mitotic spindles. *J Cell Biol* **156**, 595-602 (2002).
200. C. C. Lee, B. Li, H. Yu, M. J. Matunis, Sumoylation promotes optimal APC/C Activation and Timely Anaphase. *Elife* **7** (2018).
201. C. Cubenas-Potts, M. J. Matunis, SUMO: a multifaceted modifier of chromatin structure and function. *Dev Cell* **24**, 1-12 (2013).
202. H. Neyret-Kahn *et al.*, Sumoylation at chromatin governs coordinated repression of a transcriptional program essential for cell growth and proliferation. *Genome Res* **23**, 1563-1579 (2013).
203. J. C. Cossec *et al.*, SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States. *Cell Stem Cell* **23**, 742-757 e748 (2018).
204. Y. Shiio, R. N. Eisenman, Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A* **100**, 13225-13230 (2003).
205. H. W. Liu *et al.*, Chromatin modification by SUMO-1 stimulates the promoters of translation machinery genes. *Nucleic Acids Res* **40**, 10172-10186 (2012).
206. H. W. Liu, T. Banerjee, X. Guan, M. A. Freitas, J. D. Parvin, The chromatin scaffold protein SAFB1 localizes SUMO-1 to the promoters of ribosomal protein genes to facilitate transcription initiation and splicing. *Nucleic Acids Res* **43**, 3605-3613 (2015).
207. J. Ouyang, G. Gill, SUMO engages multiple corepressors to regulate chromatin structure and transcription. *Epigenetics* **4**, 440-444 (2009).
208. G. Gill, Post-translational modification by the small ubiquitin-related modifier SUMO has big effects on transcription factor activity. *Curr Opin Genet Dev* **13**, 108-113 (2003).
209. H. Hofmann, S. Floss, T. Stamminger, Covalent modification of the transactivator protein IE2-p86 of human cytomegalovirus by conjugation to the ubiquitin-homologous proteins SUMO-1 and hSMT3b. *J Virol* **74**, 2510-2524 (2000).
210. M. J. Lyst, I. Stancheva, A role for SUMO modification in transcriptional repression and activation. *Biochem Soc Trans* **35**, 1389-1392 (2007).

211. S. Ross, J. L. Best, L. I. Zon, G. Gill, SUMO-1 modification represses Sp3 transcriptional activation and modulates its subnuclear localization. *Mol Cell* **10**, 831-842 (2002).

212. S. H. Yang, A. D. Sharrocks, SUMO promotes HDAC-mediated transcriptional repression. *Mol Cell* **13**, 611-617 (2004).

213. H. Poukka, U. Karvonen, O. A. Janne, J. J. Palvimo, Covalent modification of the androgen receptor by small ubiquitin-like modifier 1 (SUMO-1). *Proc Natl Acad Sci U S A* **97**, 14145-14150 (2000).

214. J. A. Iniguez-Lluhi, D. Pearce, A common motif within the negative regulatory regions of multiple factors inhibits their transcriptional synergy. *Mol Cell Biol* **20**, 6040-6050 (2000).

215. S. Holmstrom, M. E. Van Antwerp, J. A. Iniguez-Lluhi, Direct and distinguishable inhibitory roles for SUMO isoforms in the control of transcriptional synergy. *Proc Natl Acad Sci U S A* **100**, 15758-15763 (2003).

216. E. Rosonina, A. Akhter, Y. Dou, J. Babu, V. S. Sri Theivakadadcham, Regulation of transcription factors by sumoylation. *Transcription* **8**, 220-231 (2017).

217. E. Rosonina, A conserved role for transcription factor sumoylation in binding-site selection. *Curr Genet* **65**, 1307-1312 (2019).

218. J. Ouyang, Y. Shi, A. Valin, Y. Xuan, G. Gill, Direct binding of CoREST1 to SUMO-2/3 contributes to gene-specific repression by the LSD1/CoREST1/HDAC complex. *Mol Cell* **34**, 145-154 (2009).

219. L. Subramanian, M. D. Benson, J. A. Iniguez-Lluhi, A synergy control motif within the attenuator domain of CCAAT/enhancer-binding protein alpha inhibits transcriptional synergy through its PIASy-enhanced modification by SUMO-1 or SUMO-3. *J Biol Chem* **278**, 9134-9141 (2003).

220. P. Sutinen, M. Malinen, S. Heikkinen, J. J. Palvimo, SUMOylation modulates the transcriptional activity of androgen receptor in a target gene and pathway selective manner. *Nucleic Acids Res* **42**, 8310-8319 (2014).

221. D. Girdwood *et al.*, P300 transcriptional repression is mediated by SUMO modification. *Mol Cell* **11**, 1043-1054 (2003).

222. F. Lehembre, S. Muller, P. P. Pandolfi, A. Dejean, Regulation of Pax3 transcriptional activity by SUMO-1-modified PML. *Oncogene* **20**, 1-9 (2001).

223. M. Gostissa *et al.*, Activation of p53 by conjugation to the ubiquitin-like protein SUMO-1. *EMBO J* **18**, 6462-6471 (1999).

224. M. S. Rodriguez *et al.*, SUMO-1 modification activates the transcriptional response of p53. *EMBO J* **18**, 6455-6461 (1999).

225. J. M. Desterro, M. S. Rodriguez, R. T. Hay, SUMO-1 modification of IkappaBalpha inhibits NF-kappaB activation. *Mol Cell* **2**, 233-239 (1998).

226. S. Tian, H. Poukka, J. J. Palvimo, O. A. Janne, Small ubiquitin-related modifier-1 (SUMO-1) modification of the glucocorticoid receptor. *Biochem J* **367**, 907-911 (2002).

227. S. R. Holmstrom, S. Chupreta, A. Y. So, J. A. Iniguez-Lluhi, SUMO-mediated inhibition of glucocorticoid receptor synergistic activity depends on stable assembly at the promoter but not on DAXX. *Mol Endocrinol* **22**, 2061-2075 (2008).

228. A. Khanna-Gupta, Sumoylation and the function of CCAAT enhancer binding protein alpha (C/EBP alpha). *Blood Cells Mol Dis* **41**, 77-81 (2008).
229. M. L. Goodson *et al.*, Sumo-1 modification regulates the DNA binding activity of heat shock transcription factor 2, a promyelocytic leukemia nuclear body associated transcription factor. *J Biol Chem* **276**, 18513-18518 (2001).
230. S. H. Yang, E. Jaffray, R. T. Hay, A. D. Sharrocks, Dynamic interplay of the SUMO and ERK pathways in regulating Elk-1 transcriptional activity. *Mol Cell* **12**, 63-74 (2003).
231. S. Salinas *et al.*, SUMOylation regulates nucleo-cytoplasmic shuttling of Elk-1. *J Cell Biol* **165**, 767-773 (2004).
232. T. E. Danciu *et al.*, Small ubiquitin-like modifier (SUMO) modification mediates function of the inhibitory domains of developmental regulators FOXC1 and FOXC2. *J Biol Chem* **287**, 18318-18329 (2012).
233. P. Gomez-del Arco, J. Koipally, K. Georgopoulos, Ikaros SUMOylation: switching out of repression. *Mol Cell Biol* **25**, 2688-2697 (2005).
234. R. Geiss-Friedlander, F. Melchior, Concepts in sumoylation: a decade on. *Nat Rev Mol Cell Biol* **8**, 947-956 (2007).
235. E. Urena *et al.*, Evolution of SUMO Function and Chain Formation in Insects. *Mol Biol Evol* **33**, 568-584 (2016).
236. H. Yuan *et al.*, Small ubiquitin-related modifier paralogs are indispensable but functionally redundant during early development of zebrafish. *Cell Res* **20**, 185-196 (2010).
237. D. Baczyk, S. Drewlo, J. C. Kingdom, Emerging role of SUMOylation in placental pathology. *Placenta* **34**, 606-612 (2013).
238. D. Baczyk, M. C. Audette, E. Coyaud, B. Raught, J. C. Kingdom, Spatiotemporal distribution of small ubiquitin-like modifiers during human placental development and in response to oxidative and inflammatory stress. *J Physiol* **596**, 1587-1600 (2018).
239. J. Barretina *et al.*, The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* **483**, 603-607 (2012).
240. J. Schindelin *et al.*, Fiji: an open-source platform for biological-image analysis. *Nature methods* **9**, 676-682 (2012).
241. M. Dasso, Emerging roles of the SUMO pathway in mitosis. *Cell Div* **3**, 5 (2008).
242. E. Van Damme, K. Laukens, T. H. Dang, X. Van Ostade, A manually curated network of the PML nuclear body interactome reveals an important role for PML-NBs in SUMOylation dynamics. *Int J Biol Sci* **6**, 51-67 (2010).
243. S. F. Banani *et al.*, Compositional Control of Phase-Separated Cellular Bodies. *Cell* **166**, 651-663 (2016).
244. J. M. Enserink, Sumo and the cellular stress response. *Cell Div* **10**, 4 (2015).
245. J. van Meerloo, G. J. Kaspers, J. Cloos, Cell sensitivity assays: the MTT assay. *Methods Mol Biol* **731**, 237-245 (2011).
246. Y. Song *et al.*, Double mimicry evades tRNA synthetase editing by toxic vegetable-sourced non-proteinogenic amino acid. *Nat Commun* **8**, 2281 (2017).

247. E. Fiebiger *et al.*, Dissection of the dislocation pathway for type I membrane proteins with a new small molecule inhibitor, eeyarestatin. *Mol Biol Cell* **15**, 1635-1646 (2004).
248. A. Singh, Y. J. Xu, The Cell Killing Mechanisms of Hydroxyurea. *Genes (Basel)* **7** (2016).
249. G. Gill, Something about SUMO inhibits transcription. *Curr Opin Genet Dev* **15**, 536-541 (2005).
250. A. Subramanian *et al.*, Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **102**, 15545-15550 (2005).
251. Z. Hannoun, G. Maarifi, M. K. Chelbi-Alix, The implication of SUMO in intrinsic and innate immunity. *Cytokine Growth Factor Rev* **29**, 3-16 (2016).
252. F. El-Asmi *et al.*, Cross-talk between SUMOylation and ISGylation in response to interferon. *Cytokine* **129**, 155025 (2020).
253. A. Alonso *et al.*, Emerging roles of sumoylation in the regulation of actin, microtubules, intermediate filaments, and septins. *Cytoskeleton (Hoboken)* **72**, 305-339 (2015).
254. M. L. Whitfield *et al.*, Stem-loop binding protein, the protein that binds the 3' end of histone mRNA, is cell cycle regulated by both translational and posttranslational mechanisms. *Mol Cell Biol* **20**, 4188-4198 (2000).
255. W. F. Marzluff, K. P. Koreski, Birth and Death of Histone mRNAs. *Trends Genet* **33**, 745-759 (2017).
256. T. Kubota *et al.*, Virus infection triggers SUMOylation of IRF3 and IRF7, leading to the negative regulation of type I interferon gene expression. *J Biol Chem* **283**, 25660-25670 (2008).
257. S. Citro, S. Chiocca, Assessing the Role of Paralog-Specific Sumoylation of HDAC1. *Methods Mol Biol* **1510**, 329-337 (2017).
258. A. M. Ishov *et al.*, PML is critical for ND10 formation and recruits the PML-interacting protein daxx to this nuclear structure when modified by SUMO-1. *J Cell Biol* **147**, 221-234 (1999).
259. X. Xu *et al.*, SUMO-1 modification of FEN1 facilitates its interaction with Rad9-Rad1-Hus1 to counteract DNA replication stress. *Journal of Molecular Cell Biology* 10.1093/jmcb/mjy047, mjy047-mjy047 (2018).
260. M. Li *et al.*, SUMO2 conjugation of PCNA facilitates chromatin remodeling to resolve transcription-replication conflicts. *Nat Commun* **9**, 2706 (2018).
261. W. H, *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag New York, 2016).
262. F. A. Ran *et al.*, Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* **8**, 2281-2308 (2013).
263. Z. Darzynkiewicz, G. Juan, E. Bedner, Determining cell cycle stages by flow cytometry. *Curr Protoc Cell Biol* **Chapter 8**, Unit 8 4 (2001).
264. J. N. Kapur, P. K. Sahoo, A. K. C. Wong, A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram. *Comput Vision Graph* **29**, 273-285 (1985).

265. J. Brocher, Qualitative and Quantitative Evaluation of Two New Histogram Limiting Binarization Algorithms. *International Journal of Image Processing (IJIP)* **8**, 30-48 (2014).
266. C. W. Law *et al.*, RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* **5** (2016).
267. B. Gel, E. Serra, karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* **33**, 3088-3090 (2017).
268. J. Reimand *et al.*, Pathway enrichment analysis and visualization of omics data using g:Profiler, GSEA, Cytoscape and EnrichmentMap. *Nat Protoc* **14**, 482-517 (2019).
269. X. Zhao, SUMO-Mediated Regulation of Nuclear Functions and Signaling Processes. *Mol Cell* **71**, 409-418 (2018).
270. J. D. Bernstock *et al.*, Topotecan is a potent inhibitor of SUMOylation in glioblastoma multiforme and alters both cellular replication and metabolic programming. *Sci Rep* **7**, 7425 (2017).
271. M. V. Bogachek *et al.*, Inhibiting the SUMO Pathway Represses the Cancer Stem Cell Population in Breast and Colorectal Carcinomas. *Stem Cell Reports* **7**, 1140-1151 (2016).
272. I. A. Hendriks *et al.*, Uncovering global SUMOylation signaling networks in a site-specific manner. *Nat Struct Mol Biol* **21**, 927-936 (2014).
273. R. Li *et al.*, Akt SUMOylation regulates cell proliferation and tumorigenesis. *Cancer Res* **73**, 5742-5753 (2013).
274. A. Kumar, K. Y. Zhang, Advances in the development of SUMO specific protease (SENP) inhibitors. *Comput Struct Biotechnol J* **13**, 204-211 (2015).
275. D. Mattoscio, C. Casadio, M. Fumagalli, M. Sideri, S. Chiocca, The SUMO conjugating enzyme UBC9 as a biomarker for cervical HPV infections. *Ecancermedicalscience* **9**, 534 (2015).
276. Q. S. Zhang, M. Zhang, X. J. Huang, X. J. Liu, W. P. Li, Downregulation of SENP1 inhibits cell proliferation, migration and promotes apoptosis in human glioma cells. *Oncol Lett* **12**, 217-221 (2016).
277. C. Burdelski *et al.*, The prognostic value of SUMO1/Sentrin specific peptidase 1 (SENP1) in prostate cancer is limited to ERG-fusion positive tumors lacking PTEN deletion. *BMC cancer* **15**, 538-538 (2015).
278. L. Rahib *et al.*, Projecting Cancer Incidence and Deaths to 2030: The Unexpected Burden of Thyroid, Liver, and Pancreas Cancers in the United States. *Cancer Research* **74**, 2913-2921 (2014).
279. S. H. Loosen, U. P. Neumann, C. Trautwein, C. Roderburg, T. Luedde, Current and future biomarkers for pancreatic adenocarcinoma. *Tumor Biology* **39**, 1010428317692231 (2017).
280. K. M. Lee, C. Nguyen, A. B. Ulrich, P. M. Pour, M. M. Ouellette, Immortalization with telomerase of the Nestin-positive cells of the human pancreas. *Biochemical and Biophysical Research Communications* **301**, 1038-1044 (2003).
281. B. Saunders *et al.*, The Molecule Pages database. *Nucleic Acids Res* **36**, D700-706 (2008).

282. C. M. Hickey, N. R. Wilson, M. Hochstrasser, Function and regulation of SUMO proteases. *Nat Rev Mol Cell Biol* **13**, 755-766 (2012).
283. S. J. Li, M. Hochstrasser, The Ulp1 SUMO isopeptidase: distinct domains required for viability, nuclear envelope localization, and substrate specificity. *J Cell Biol* **160**, 1069-1081 (2003).
284. J. Vivian *et al.*, Toil enables reproducible, open source, big biomedical data analyses. *Nature Biotechnology* **35**, 314 (2017).
285. X. Li, W. Wang, J. Chen, From pathways to networks: connecting dots by establishing protein-protein interaction networks in signaling pathways using affinity purification and mass spectrometry. *Proteomics* **15**, 188-202 (2015).
286. J. Lonsdale *et al.*, The Genotype-Tissue Expression (GTEx) project. *Nature Genetics* **45**, 580 (2013).
287. D. R. Rhodes *et al.*, Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166-180 (2007).
288. L. Badea, V. Herlea, S. O. Dima, T. Dumitrascu, I. Popescu, Combined gene expression analysis of whole-tissue and microdissected pancreatic ductal adenocarcinoma identifies genes specifically overexpressed in tumor epithelia. *Hepatogastroenterology* **55**, 2016-2027 (2008).
289. E. Cerami *et al.*, The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401-404 (2012).
290. B. A. A. Jianjiong Gao, Ugur Dogrusov, Gideon Dresdner, Benjamin Gross, S. Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, Erik Larsson, Ethan Cerami, Chris Sander, Nikolaus Schultz, Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. *Science Signaling* **6** (2013).
291. A. Klemm *et al.*, The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Research* **42**, D1001-D1006 (2013).
292. B. Győrffy *et al.*, An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Research and Treatment* **123**, 725-731 (2010).
293. M. W. Pfaffl, A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45 (2001).
294. D. Aran *et al.*, Comprehensive analysis of normal adjacent to tumor transcriptomes. *Nature Communications* **8**, 1077 (2017).
295. A. K. Witkiewicz *et al.*, Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nature Communications* **6**, 6744 (2015).
296. E. Wheeler *et al.*, Impact of common genetic determinants of Hemoglobin A1c on type 2 diabetes risk and diagnosis in ancestrally diverse populations: A transethnic genome-wide meta-analysis. *PLOS Medicine* **14**, e1002383 (2017).
297. A. De Souza, K. Irfan, F. Masud, M. W. Saif, Diabetes Type 2 and Pancreatic Cancer: A History Unfolding. *JOP : Journal of the pancreas* **17**, 144-148 (2016).

298. P. Bailey *et al.*, Genomic analyses identify molecular subtypes of pancreatic cancer. *Nature* **531**, 47-52 (2016).

299. E. Cerami *et al.*, The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data: Figure 1. *Cancer Discovery* **2**, 401-404 (2012).

300. J. G. Tate *et al.*, COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res* **47**, D941-D947 (2019).

Curriculum Vitae

Danielle Marie Bouchard
PhD Candidate

Johns Hopkins University School of Public Health
Department of Biochemistry and Molecular Biology
615 N Wolfe St, Room W8104-2
Baltimore, MD 21205
775-830-3374
dboucha2@jhu.edu

PRESENT POSITION

PhD Candidate August 2015 – Present
Department of Biochemistry and Molecular Biology, Johns Hopkins University, School of Public Health Advisor: Dr. Michael J. Matunis

EDUCATION

Ph.D. in Biochemistry and Molecular Biology 2015 – 2020
Department of Biochemistry and Molecular Biology, Johns Hopkins University, School of Public Health
Advisor: Dr. Michael J. Matunis

B.S. in Biotechnology 2006 – 2010
College of Agriculture, Montana State University, Bozeman, MT
Distinction: Cum Laude

RESEARCH EXPERIENCE

Johns Hopkins University May 2016 – Present
Graduate Research Assistant; Advisor: Dr. Michael J. Matunis

Discovering paralog-specific functions of the SUMO1 and SUMO2 protein modifiers. Previously, I explored the role of the SUMO isopeptidase, SENP1, in the development and progression of pancreatic ductal adenocarcinoma which culminated in a first-author publication.

PROFESSIONAL EXPERIENCE

Grifols Diagnostic Solutions (Prev. Novartis V&D) November 2013 – June 2015
Analytical Assistant

Developed quantitative SDS-PAGE Coomassie and western blot assays and authored development and qualification reports for method transfer into QC for release testing.

Novartis Vaccines & Diagnostics August 2012 – November 2013
Associate Scientist I

Performed routine testing of HIV and HCV antigens using SDS-PAGE coomassie, silver stain, western blot, size exclusion HPLC and BCA assays for in-process product release.

Sepax Technologies November 2010 – August 2012
Field Sales Account Manager, Northern CA, OR, NV, UT, CO

Directly represented Sepax and their HPLC resin technology to over 250 academic, government and industry accounts while meeting and exceeding quarterly sales goals

AWARDS

Roger McMacken Scholarship Award *June 2020*
Biochemistry and Molecular Biology Department, JHSPH

Best Scientific Talk *April 2019*
Biochemistry and Molecular Biology Annual Retreat

Honorable Mention for the 3 Minute-Thesis Competition *April 2019*
Johns Hopkins University

STAR Award for collaborative leadership *July 2013*
Novartis Vaccines & Diagnostics

Fisher Scientific 2010 Outstanding Graduate in Animal Biotechnology Recipient *May 2010*
Montana State University

PUBLICATIONS

Bouchard DM, Wang W, Yang WC, He S, and Matunis MJ. (2020). *Discovery of Paralog-Specific SUMO1 and SUMO2 Functions using Human Knock-out Cells.* Manuscript in preparation.

Bouchard DM and Matunis MJ. *A Cellular and Bioinformatics Analysis of the SENP1 SUMO Isopeptidase in Pancreatic Cancer.* J Gastrointest Oncol 2019. Doi:10.21037/jgo.2019.05.09

ORAL PRESENTATIONS

Bouchard DM, Wang W, Yang WC, He S, and Matunis MJ. Discovery of Paralog-Specific SUMO1 and SUMO2 Functions using Human Knock-out Cells. *Biochemistry and Molecular Biology 2020 Achievement Awards Ceremony*, Virtual 5-minute presentation; September 2020.

Bouchard DM, Yang WC and Matunis MJ. Discovering SUMO1 and SUMO2 Paralog-Specific Functions using Human Knockout Cells. *Biochemistry and Molecular Biology Annual Retreat*, Baltimore, MD; April 2019.

Bouchard DM, Matunis MJ. An Evaluation of the SUMO Isopeptidase (SENP1) in Human Cancer. *Hopkins Ubiquitin Club*, Baltimore, MD; December 2018.

Bouchard DM, Matunis MJ. Understanding SENP1 Regulation and Function in Pancreatic Cancer. *Biochemistry and Molecular Biology Annual Retreat*, Towson, MD; April 2018.

Bouchard DM, Matunis MJ. Evaluation of the SUMO Pathway as a Target for Pancreatic Cancer Therapy. *Sol Goldman Pancreatic Cancer Research Center 2016 Grant Update Meeting*, Baltimore, MD; December 2016.

POSTER PRESENTATIONS

Bouchard DM, Yang WC and Matunis MJ. Discovery of Paralog-Specific SUMO1 and SUMO2 Functions using Human Cancer Knockout Cells. *American Association of Cancer Research Virtual Annual Meeting II*, Virtual e-poster presentation; June 2020.

Bouchard DM, Yang WC and Matunis MJ. Discovery of Paralog-Specific SUMO1 and SUMO2 Functions using Human Cancer Knockout Cells. *The Sidney Kimmel Comprehensive Cancer Center at Johns Hopkins Fellow Research Day 2019*, Baltimore, MD; June 2019.

Bouchard DM, Matunis MJ. Development of SENP1 as a Novel Biomarker for Pancreatic Cancer Therapy. *Baltimore Area Repair Symposium (BARS) Meeting*, Baltimore, MD; March 2018

Bouchard DM, Matunis MJ. Development of SENP1 as a Novel Biomarker for Pancreatic Cancer Therapy. *ASCB-EMBO 2017 Meeting*, Philadelphia, PA; December 2017.

Bouchard DM, Matunis MJ. Understanding SENP1 Regulation and Function in Pancreatic Cancer. *ASCB 2017 M3 Meeting*, Bethesda, MD; May 2017.

Bouchard DM, Matunis MJ. Understanding SENP1 Regulation and Function in Pancreatic Cancer. *Biochemistry and Molecular Biology Annual Retreat*, Gettysburg, PA; March 2017.

TRAINING & LEADERSHIP EXPERIENCE

Summer Academic Research Experience (SARE) June - August, 2019

Mentor

Mentored a rising sophomore summer high school student in the laboratory. I taught my trainee how to pipette, work with mammalian cell culture, keep a detailed and organized lab notebook and how to make and present a poster. She later went on to win 1st place in the 4th Annual Excellence in Diversity Symposium for her poster presentation of her project.

Grifols Diagnostic Solutions, USA Headquarters August 14-15, 2015

Project Management Essentials

Traveled to Grifols Headquarters and completed a two-day training on Project Management Essentials lead by Franklin Covery. Used skills acquired to schedule the full laboratory schedule and ordering for the Analytical Development group.

Novartis Vaccines & Diagnostics/Grifols August 2013 – June 2015

Qualified Trainer

Completed the full day Novartis Vaccines & Diagnostics "Train the Trainer" certificate program which allowed me to train eight new associates on all laboratory skills required for them to successfully do their job over the course of two years.

PROFESSIONAL AFFILIATIONS

American Association for Cancer Research (AACR)
The American Society for Cell Biology (ASCB)
The American Society for Biochemistry and Molecular Biology (ASBMB)

VOLUNTEER EXPERIENCE

Hopkins Biotech Network (HBN) August 2017 – June 2020

Executive Director of Alumni Relations

Maintained and promoted relationships between current Hopkins students with an interest in the Biotechnology and Healthcare industry and Alumni through the development of an alumni database and organization of career panels and networking events.

JHSPH Academic Ethics Review Committee April 2017 – April 2018

Student Board Member

Served on a board for academic hearings that involving suspected ethical issues related to course or lab work. Served on one committee regarding cheating and plagiarism during an online exam.

JHSPH Insoluble Fraction November 2016 – March 2020

President

Organized and ran a weekly networking event for the JHSPH students, staff and faculty.

2019 Girl Power STEM Event at the Johns Hopkins APL March 10, 2019

Volunteer with BioEYES

Was one of four volunteers at the BioEyes exhibition table, where we helped girls aged 5-15 understand the developmental lifecycle of a Zebrafish using light microscopy and live embryos.

2018 ASCB M4 Meeting May 2017 – April 2018

Lead Organizer

Organized the 2018 Mid-Atlantic Mitosis and Meiosis Meeting (M4), a regional conference run by 5 trainees with over 100 attendees from over 15 institutions. Raised $7 in event funds.

Johns Hopkins Biomedical Careers Initiative January 2017 – January 2018

Member of organizing committee

Active member of the JHU BCI-EDGE committee which discusses alternative scientific careers for PhD students outside of academia. Positions are held for one year.

JHSPH Student Assembly April 2016 – May 2018

Member at Large

TECHNICAL COMPUTER SKILLS

R, RStudio, Adobe Illustrator, Adobe Photoshop, FIJI/ImageJ, Apple Keynote, Microsoft Office Suite

BIBLIOGRAPHY

Bouchard DM and Matunis MJ. *A Cellular and Bioinformatics Analysis of the SENP1 SUMO Isopeptidase in Pancreatic Cancer.* J Gastrointest Oncol 2019. Doi:10.21037/jgo.2019.05.09