The Contribution of Alternative Splicing and Post-Transcriptional Mechanisms Toward the Global Control of SUMO1/2/3 SUMOylation

Myriah Lorraine Acuna
The University of Texas at El Paso

Follow this and additional works at: https://scholarworks.utep.edu/open_etd

Part of the Cell Biology Commons

Recommended Citation
https://scholarworks.utep.edu/open_etd/3467

This is brought to you for free and open access by ScholarWorks@UTEP. It has been accepted for inclusion in Open Access Theses & Dissertations by an authorized administrator of ScholarWorks@UTEP. For more information, please contact lweber@utep.edu.
THE CONTRIBUTION OF ALTERNATIVE SPLICING AND POST-TRANSCRIPTIONAL MECHANISMS TOWARD THE GLOBAL CONTROL OF SUMO1/2/3 SUMOYLATION

MYRIAH LORRAINE ACUÑA

Doctoral Program in Biology

APPROVED:

_________________________________________
German Rosas-Acosta, Ph. D., Chair

_________________________________________
Anita Quintana, Ph.D.

_________________________________________
Manuel Llano, Ph.D.

_________________________________________
Chuan Xiao, Ph.D.

_________________________________________
Kristine Garza, Ph.D. Advocate

_________________________________________
Stephen L. Crites, Jr., Ph.D.
Dean of the Graduate School
THE CONTRIBUTION OF ALTERNATIVE SPLICING AND POST-TRANSCRIPTIONAL MECHANISMS TOWARD THE GLOBAL CONTROL OF SUMO1/2/3 SUMOYLATION

by

MYRIAH LORRAINE ACUÑA, B.S., M.Sc.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

DEPARTMENT OF BIOLOGICAL SCIENCES

THE UNIVERSITY OF TEXAS AT EL PASO

May 2022
Acknowledgments

Through the course of the writing of this study, I have received an abundance of encouragement.

I would first like to thank my mentor, Dr. German Rosas-Acosta, whose expertise was invaluable in formulating the research questions and methodology. Your insightful feedback pushed me to sharpen my thinking and brought my work to a higher level.

I would like to acknowledge my colleagues from the University of Texas-El Paso Bioresearch facility. I would particularly like to single out my assistant Andrea Garcia; I want to thank you for your patient support and opportunities to further my research. It has been a pleasure to mentor you through your undergraduate and see the person you have become now entering your own Ph.D. journey. I would also like to thank my colleague, Miguel Beltran, for his valuable guidance throughout my experiments. I would also like to thank my committee: Dr. Manuel Llano, Dr. Chuan Xiao, Dr. Anita Quintana, and Dr. Garza; you provided me with the tools I needed to choose the right direction and complete my dissertation. Luis Alvarado of Texas Tech contributed a major portion of help through the proofing and generating of the proper statistical tests for these experiments. As well as Dr. Amy Wagler of the University of Texas at El Paso Data Science Laboratory.

In addition, I would like to thank my parents for their wise counsel and sympathetic ear. You are always there for me. Finally, I could not have completed this dissertation without the support of my husband and daughter, Anthony, and Scarlett Mock. They provided happy distractions to rest my mind outside of my research.
Abstract

Evolutionary pressures have caused eukaryotic cells to develop various response mechanisms to changing environments, particularly stress adaptations that have been inherited over time. SUMOylation, the post-translational conjugation of the Small Ubiquitin-like Modifiers to a lysine residue on target proteins, is well-documented to regulate essential processes within the cell, including nucleocytoplasmic transport, transcription, apoptosis, protein stability, and progression through the cell cycle (1). SUMO protein levels have been reported to increase upon diverse stress states, including viral infection (2), cardiovascular ischemic events (3), temperature shock (4), and cancer (5). However, very little is known about how SUMOylation is regulated within the cell, particularly how SUMO protein levels are rapidly increased upon stress. Analysis of human SUMO mRNA sequences obtained from the NCBI database indicates the existence of three different transcripts, or variants (var), for the SUMO1 gene and two distinct mature mRNA transcript variants for both the SUMO2 and SUMO3 genes. For each SUMO paralog, one of the variants described codes for a previously unidentified protein isoform displaying a slightly different amino acid sequence from that of the prototype paralog, hereafter referred to as the α-isoform. To achieve a more thorough understanding of the molecular biology and the mechanisms governing SUMO protein levels within the cell, in these studies we establish the relative abundance of both the regular and alternative transcripts coding for the main SUMO paralogs in human cells, SUMO1, SUMO2, and SUMO3, under normal and stress conditions. For all three SUMOs, the variant transcripts that code for the well-characterized protein isoforms appears to be the most abundant transcripts. In contrast, the one coding for the α-isoforms tends to change upon stress. Given the known cytoprotective properties of protein SUMOylation (6), an improved understanding of the contribution of such post-transcriptional mechanisms to the regulation of
SUMOylation may lead to the development of innovative therapies for conditions in which the role of protein SUMOylation plays a critical function in host survival, including infectious diseases such as Influenza A virus infection, and cardiovascular ischemic events such as strokes and heart attacks.
# Table of Contents

Acknowledgments........................................................................................................ iv
Abstract......................................................................................................................... v
Table of Contents.......................................................................................................... vii
List of Figures............................................................................................................... ix

Chapter 1: Introduction.................................................................................................. 1
  1.1 The Central Dogma of Gene Expression ................................................................. 1
  1.2 Post-Translational Modifications............................................................................. 6
  1.3 Small Ubiquitin-Like Modifiers ............................................................................... 7
  1.4 Processing of SUMO1/2/3 Proteins......................................................................... 10
  1.5 SUMO, The Regulator ............................................................................................ 13
  1.6 SUMOylation and Disease....................................................................................... 17
  1.7 SUMOylation and the Cellular Stress Response....................................................... 18
  1.8 SUMOylation Regulation......................................................................................... 20

Chapter 2: Aims.............................................................................................................. 23

Chapter 3: Materials and Methods.............................................................................. 26
  3.1 Primer Design Strategy.......................................................................................... 26
  3.2 Cell Culture............................................................................................................ 26
  3.3 RNA Purification..................................................................................................... 27
  3.4 Assessment of Purified RNA Quality and Quantity............................................... 28
  3.5 cDNA Synthesis..................................................................................................... 29
  3.6 qRT-PCR............................................................................................................... 29
  3.7 Cloning of SUMO1/2/3 Variants............................................................................ 30
  3.8 Transformation....................................................................................................... 30
  3.9 Maxi Prep.............................................................................................................. 31
  3.10 T7 RNA Polymerase In Vivo Transcription........................................................... 32
  3.11 Generating Calibration Curves.............................................................................. 33
  3.12 Stress Treatments................................................................................................ 33
  3.13 Nuclear vs. Cytosolic Fractionation...................................................................... 34
  3.14 His and YFP Tag SUMO1/2/3 Plasmid Cloning.................................................... 35
  3.15 Transient Transfections....................................................................................... 35
  3.16 Western Blot Analyses.......................................................................................... 36
  3.17 Confocal Microscopy............................................................................................ 37
  3.18 Tertiary Structure Prediction Models.................................................................... 37
  3.19 Statistical Reports................................................................................................ 38

Chapter 4: Results......................................................................................................... 39

Chapter 5: Conclusions............................................................................................... 67

Chapter 6: Limitations of the Study............................................................................ 74

Chapter 7: Future Studies............................................................................................ 76
List of Figures

Figure 1: Dynamic Points in the Central Dogma Affect Gene Expression Levels.......................... 1
Figure 2: Splicing is an Intricately Regulated Process that Contributes to Gene Expression......... 3
Figure 3: Small Ubiquitin Like Modifiers Protein Structures.................................................... 8
Figure 4: SUMO Proteins Must be Processed to Achieve Conjugation......................................... 11
Figure 5: The Functional Properties of SUMO Affects Many Biological Processes................... 12
Figure 6: SUMO1/2/3 Alternative mRNA Transcripts Informatic Predicted Properties............. 40
Figure 7: Primer-Specific Targeting of the SUMO1/2/3 Variants................................................. 43
Figure 8: SUMO1/2/3 mRNA Normalcy Levels Across Four Mammalian Cell Lines................. 45
Figure 9: SUMO1/2/3 Variants Transcriptomic Profile Upon Stress ............................................. 48
Figure 10: SUMO1/2/3 Variants Transcriptomic Profile Upon SARS-CoV-2 Infection ............. 52
Figure 11: Detection of SUMO1/2/3 Variants in Cytoplasmic Fractions................................. 55
Figure 12: The SUMO1/2/3 Variants’ Nucleo-Cytosolic Profiles Shift at 27°C............................. 57
Figure 13: Ribosome- Sequencing Confirmation of SUMO1/2/3’s Translation.............................. 59
Figure 14: Artificial Intelligence Predicted Functional Properties................................................. 61
Figure 15: Characterized Functional Properties of the SUMO1/2/3 Isoforms............................... 63
Figure 15: SUMO1/2/3 His-Tagged Plasmids Affect SUMO1/2/3 Localization............................. 65

Supplementary Data

Figure S1: NCBI Accession Numbers for each SUMO1/2/3 Variant........................................... 79
Figure S2: Primer Sequences for Amplification of SUMO1/2/3 Variants..................................... 80
Figure S3: SUMO1/2/3 Stress-Induced Responses in the A549 Cell Line.................................... 81
Figure S4: SUMO1/2/3 Stress-Induced Responses in the HEK293A Cell Line............................. 82
Figure S5: Graphical Representation of Subcellular Fractionation at 27°C.................................. 83
1. Introduction

1.1 The Central Dogma of Gene Expression

“The significance of deoxyribonucleic acid within living cells is undisputed” (7, 8). All life essentially relies on endless cellular replication and adaptations to stress that ensure survival. While it is presently known many of these adaptations initiate at the gene level, there is leaves much left to discover about the regulation of these changes. Cellular instructions come from genes, whose expression begins at the transcriptional level but can also undergo significant changes to mitigate proper responses to external signals. Transcription generates highly specific RNA molecules containing information within the nucleotide sequence, is the beginning of the central dogma of molecular biology, which can be naturally and artificially manipulated at multiple levels namely histone modification, alternative splicing, and protein biosynthesis (Fig 1)(9).

![Diagram](image)

**Figure 1-Dynamic Points in the Central Dogma Affect Gene Expression Levels**

Instructional material contained within the DNA sequence to be transcribed into RNA. After modification, including capping, splicing, and poly-A tailing, it is exported into the cytoplasm, where the RNA sequence is translated into a polypeptide sequence by the ribosome (9).

Splicing is a sequence-specific post-transcriptional process in which sections of an RNA transcript called introns are excised, and the remaining sequences called exons are merged to construct a mature mRNA transcript. Though introns constitute a significant amount of the human
genome, their immense majority is eliminated from transcripts via the splicing of RNA occurring within a ribonucleoprotein splicing machine called the suprasplicosome (10). The suprasplicosome contains five spliceosomal small nuclear ribonucleoprotein particles, snRNPs, and non-snRNP proteins, including regulatory splicing factors such as hnRNPs (11) and phosphorylated Serine/Arginine-rich-proteins (Fig. 2A)(12). The five primary snRNAs (U1, U2, U4, U5, and U6) begin the splicing process by cleaving the intron from exon one at the 5’ splice junction to produce a bond to an adenine nucleotide in the branch-point sequence, creating what is referred to as a lariat structure (13). This lariat formation contains an irregular 2’-5’ phosphodiester bond between the adenine and guanine nucleotides towards the end of the intron (13). The adenine nucleotide maintains its 3’-5’ linkage, and, finally, the spliceosome complex cleaves at the 3’ splice junction, ligating exons 1 and 2 together (13). The snRNPs then detach and repeat for each intron.

Alternative splicing is an RNA editing method where a single gene encodes multiple protein isoforms via exons or parts of exons and introns that are joined or skipped resulting in diverse conformations. This editing procedure not only serves as another point of gene expression regulation but also generates immense biodiversity by amplifying the genome causing multiple messages to arise from a single DNA gene. The human genome contains numerous genes known to be under the regulation of alternative splicing comprising significant housekeeping genes such as telomerase (14), actin (15), and tubulin (16); it is estimated that about 92-94% of genes undergo alternative splicing in which 86% produce an isoform that can vary during different developmental stages, as well as among other tissues and individuals (17). While introns were formerly considered as nothing more than “junk sequences”, they now appear to also contain regulatory functions by signaling additional transcript editing machinery such as where to join and where to cut, ultimately
serving as a switch to generate coding or non-coding RNAs from the same transcription entity (Fig. 2B) (18) (19).

Figure 2- Splicing is an Intricately Regulated Process that Contributes to Gene Expression
A: The native spliceosome houses all five snRNPs and other spliceosomal constituents providing the ultimate platform for conformational changes. Additionally, the functional domains of the snRNPs shape an uninterrupted segment along the surface of the large spliceosomal cavity generating a protected space for the pre-mRNA binding and the splicing reaction to occur (20).

B: These standard forms can be combined to generate more complicated alternative splicing events in many cases (21).

C: Splicing is Regulated by SR Proteins and hnRNPs. (Open boxes) Exons, (jagged lines) introns, (brackets) splice sites (ss). The consensus motifs of ss are shown, and the branch point adenosine is indicated. (Dashed lines) Two alternative splicing pathways with the middle exon are either included or excluded. Splicing is regulated by cis-elements (ESE, ESS, ISS, and ISE) and trans-acting splicing factors (SR proteins, hnRNPs, and unknown factors) (21).

Conversely, the splicing of introns also produces small regulatory RNAs that emerge to add plasticity to the nature of the protein produced, its distribution in each cell type, and the timing of its translation all of which become further points of expression regulation (22). Despite a significant potential for error, the splicing process is conducted with extreme fidelity implying the commitment of additional transcript qualities such as cis-regulatory elements who act as core
splice signals in the splice selection site that can serve as either splicing enhancers or silencers. (21). Generally, these splicing regulatory elements function through the enlistment of trans-acting splicing factors to either activate or suppress splice site detection or spliceosome assembly via diverse mechanisms (Fig. 2C) (23). Further RNA handling steps ensue, such as 5’end-capping (24), 3’end polyadenylation (25), and RNA editing enzymes, ADAR1 and ADAR2 (26), still occurring within the supraspliceosome located inside the nucleus.

In eukaryotes, mRNAs must first be transported to the cytoplasm before translation, however, not all transcripts are translated, and some mRNA transcripts are stored for later while others are degraded. Surveillance quality control mechanisms such as non-sense mediated decay find and degrade mRNA sequences that lack whole open reading frames and premature stop codons to prevent the manufacture of truncated proteins that can be fatal to the organism (27). Widespread analyses corroborate that as many as half of all fluxes in mRNA quantities in response to stress are caused by fluctuating rates of decay (28). Moreover, microarray experiments have revealed that 40–50% of gene expression fluctuations in response to cellular signals occur at the level of the mRNA stability (29). miRNAs are endogenous non-coding RNAs and are considered one of the amplest groups of regulatory genes encompassing 1% of all genes (30). These RNA species provide an additional regulatory pathway capable of gene silencing via complementary base pairing among miRNAs within the 3’ untranslated regions of the target mRNAs (31). This post-transcriptional tweaking of cellular physiology in reaction to both intracellular and environmental cues commands not only explicit temporal but also spatial control of gene expression. Another method of influence on gene expression is the precise mRNA localization with direct implications in cell migration (32) and differentiation (33). Hence it is now considered more evolutionarily efficient for the cell to dictate mRNA localization because these transcripts are capable of being
translated tens to hundreds of times in response to local stimuli rather than transporting individual proteins back and forth to different cellular compartments (34). This is an innovative notion branded as the “RNA Signature” hypothesis, in which “each mRNA transcript contains an individualized set of regulatory features that regulate its transport, localization, and translational control” (35).

Whereas the translation of mRNA is considered a concluding point in gene expression, it is only the initial phase in spawning a functional protein. Protein synthesis can be arranged into four points: initiation, elongation, termination, and ribosome recycling (36). This procedure is conducted in ribosomes located outside the nucleus within the cytosol with the aid of two further RNA species: tRNAs, and rRNAs, as well as various proteins vital for translation. A single mammalian cell is estimated to contain nearly 10 million ribosomes with a synthesis rate of 7500 subunits per minute (37). Initiation commences through the binding of methionyl tRNA and mRNA to the small ribosomal subunit trailed by the large subunit who together scan the mRNA transcript in the 5’ to 3’ direction in a cap-dependent method (38). A polypeptide chain proceeds the elongation towards the carboxy terminus resulting with each amino acid being specified by three bases as per the genetic code within the transcript advancing not at a continuous pace but rather in a stop-and-go method until arriving at a stop codon (39). Termination occurs via the concerted action of release factors eRF1 and eRF3, however, a polypeptide is not equivalent to the production of functional proteins. The final step comprises folding into an appropriate confirmation and association into functional complexes further sustained by molecular chaperones who bind and stabilize unfolded or incompletely folded polypeptide chains. Further vital modifications ensue such as cleavage and the covalent attachment of carbohydrates and lipids that can occur on the new protein that affect the function and precise localization within the cell.
Translation of particular mRNAs can be regulated not only by repressor proteins but also by noncoding miRNAs that temper the global translational function of cells in response to stress and further physiological conditions (38). Although the level of transcription is fundamental in facilitating the magnitude of stress responses, translational control regularly feeds urgent and efficient fluctuations in protein levels (40). Nevertheless, because protein synthesis is viewed as a rather costly cellular procedure, translation is chiefly repressed under most if not all types of stress conditions. However, subsets of mRNAs that encode stress response proteins exist and facilitate post-stress recovery and can circumvent this translational inhibition by engaging in what is referred to as the selective translation (41). Cells can employ an assortment of mechanisms to achieve selective translation that also contributes to the variations of elongation speed of the protein synthesis with many of these cis-elements residing in the untranslated region of mRNAs, including internal ribosome entry sites, upstream open reading frames, motifs with special sequences or secondary structures, and miRNA binding sites (42, 43).

1.2 Post-Translational Modifications

Post-translational modifications are chemical reactions that can alter states of charge, hydrophobicity, conformation, stability, and even protein function (44). These dynamic shifts are reversible in different organelles affecting a broad range of cellular activities such as enzyme regulation, signal transduction, and mediation of protein localization (45). In addition, protein levels are coordinated by different rates of degradation indicating that the control of both the quantities and actions of proteins ultimately regulates countless facets of cell behavior (46). Presently, more than 400 distinctive protein modifications have been acknowledged such as phosphorylation, acetylation, ubiquitination, and SUMOylation, all of which have been
documented to affect activity, intracellular distribution, and protein interactions, and protein longevity (47).

Phosphorylation stands as the most common and extensively studied post-translational modification with more than 500 different kinases in mammals who conduct their reactions mainly on the serine, threonine, and tyrosine residues of target substrate proteins (48) (49). Ubiquitination, also a broadly studied modification regulates numerous biological processes, including immune responses (50), apoptosis (51), and cancer (52). Contributing to the intricacy of the proteome, ubiquitination is capable of modifying the roles of these altered proteins and thereby the cell’s ability to counter developmental and environmental deviations where protein degradation is needed to be sped up or slowed down (53). The process, or, conjugation of ubiquitination is conducted by three enzymes, the ubiquitin-activating enzyme E1, the ubiquitin-conjugating enzyme E2, and the ubiquitin ligase E3 (54). Moreover, these ligases are capable of conjugating more than one ubiquitin molecule to a substrate thus forming polyubiquitin chains and acts as a molecular signal for degradation by the 26S proteasome (54). Furthermore, the explicit regulatory mechanism mediated by ubiquitination fluctuates with the different structures of the poly-ubiquitin chain (55). Contending with ubiquitination for target substrates, the term SUMO refers to a dynamic familial group of 100 amino acids called the ubiquitin-like modifiers (Ubls) also capable of deconjugation by cysteine proteases of the sentrin-specific proteases (SENPs) (56).

1.3 Small Ubiquitin Modifiers
Since its discovery in 1996, SUMO has achieved a high degree of attention due to intriguing and essential functions including a variety of biomedically indispensable protein substrates such as tumor suppressor p53, C-Jun, PML, and huntingtin (57). Although SUMO and Ubiquitin proteins exhibit similar 3-D structures and undergo similar cascades of events, ending in their post-
translational conjugation to a plethora of protein targets; unlike Ubiquitination, which simply targets proteins for degradation, SUMOylation regulates many other vital cellular processes such as chromosome segregation, nucleocytoplasmic transport, DNA replication, nuclear protein import, formation of specific subnuclear structures. transcription, apoptosis, repair of DNA damage and the inflammatory response in mammals (58). All SUMO proteins from yeast to humans share the conserved Ubiquitin domain and the C-terminal di-glycine cleavage/attachment site (Fig. 3) (59).

![Figure 3- Small Ubiquitin Like Modifiers Protein Structures](image)

A: Ubiquitin protein B: Neeed8 Protein and C: SUMO proteins are all similar in structure, however, SUMO proteins can be distinguished by long-unstructured N-terminus tails. SUMO proteins are considered small, weighing 12kDa in size. To perform post-translational modifications, SUMO proteins utilize a reversible conjugation pathway involving three enzymes, E1, E2, and E3 attaching proteins on lysine residues and can also form polymer chains of SUMO2 and SUMO3 proteins.

However, the most blatant difference between the SUMO proteins and Ubiquitin is a large unstructured N-terminal extension in the SUMO proteins stretching in length between 11 and 35 amino acids. The unstructured nature suggests that it likely contributes to mediating specific
protein-protein interactions (PPI) (60). SUMO proteins are intriguingly among the most abundant fraction of proteins with nearly $10^7$ molecules/cell in the human cervix cancer cell line HeLa, only marginally less than ubiquitin levels (61). There are three diverse well-characterized proteins SUMO protein orthologs existent within mammals: SUMO1, SUMO2, and SUMO3, and two less-studied paralogs: SUMO4 and SUMO5 who appear to encompass tissue constrained functions. It was revealed that SUMO4 is encoded by an intron within the human TAB2 gene with potent expression in kidney cells (62). However, the latter two paralogs will not be considered here. The SUMO1 proteins display a mere 44% sequence identity with the SUMO2/3 proteins contrasted to SUMO2 and SUMO3 who share an 86% sequence identity (63). However, distinctions between SUMO2 and SUMO3 have not been as thoroughly investigated due to 1) high sequence conservation and 2) a lack of means to distinguish them functionally and biochemically in vivo. Therefore, SUMO2 and SUMO3 are almost always studied and grouped together as SUMO2/3 (63). It was shown that certain target substrates have preferential binding for the diverse SUMO1/2/3 paralogs as described in large-scale proteomic studies which may be attributed to the disparities in abundances of the different SUMO1/2/3 proteins (64). While most SUMO1/2/3 proteins are mostly found conjugated to their protein substrates, free SUMO2 is present in the highest amounts followed by free SUMO3 and lastly, SUMO1 has the lowest amount of free SUMO (65). Interestingly, conjugated SUMO1/2/3 proteins localize on the nuclear membrane, nuclear bodies, and cytoplasm, respectively (63), while free forms of SUMO1/2/3 proteins exhibit an even diffusion within the cell, however, there have also been documented occasional instances in which their distribution was dynamically affected by signaling and/or stresses (66).
1.4 Processing of SUMO Proteins

As with many Ubiquitin and Ubiquitin-like proteins, all eukaryotic SUMO proteins are firstly translated as immature precursors that require processing by sentrin-specific proteases (67), either SENP1, SENP2 or SENP5 to yield a mature form with an exposed C-terminal diglycine motif that is then necessary for adenylation by a SUMO E1(activating) enzyme (68). It should be noted there are six SENPs who share a conserved catalytic domain in the COOH terminus that contains the active-site pocket of His-Asp-Cys (69). The activated adenylate is then attached by an E1(activating) enzyme and reassigned to a SUMO E2 (conjugating) enzyme also known as UBC9 (68). Ubc9 can either collaborate straightforwardly with substrates to complete the transfer of SUMO to acceptor Lysine residues, or E3 SUMO (ligases) proteins can facilitate the conjugation (Fig. 4) (68). The conjugation process frequently occurs through a SUMOylation consensus site (CS) consisting of a motif of residues adjacent to the modified lysine within the substrate. This motif corresponds to the sequence “ΨKXE,” where Ψ is a hydrophobic residue, K is the lysine site in which the actual conjugation occurs, X is any amino acid, and E is glutamic acid (63). While it has been shown the hydrophobic and acidic residues promote the stability of the interaction between the substrate and the E2 enzyme (70), various forms of the SUMOylation motif have been characterized such as negatively charged amino acid-dependent SUMO motifs (NDSMs) and phosphorylation-dependent SUMO motifs (PDSMs) (71). PDSMs and NDSMs are speculated to promote SUMOylation efficiency via the increasing of the interaction stability between Ubc9 and the substrate (72). It should be noted that SUMOylation can also occur on substrate lysine residues that don’t always conform to these known consensus sites as is the case for the well-studied K164 SUMO site in PCNA (73) (74). While data from high-throughput studies indicate that non-
consensus SUMOylation may be a relatively common event, the mechanisms in which these sites are recognized by the SUMOylation machinery endure to be determined (75).

**Figure 4- SUMO Proteins Must be Processed to Achieve Conjugation**

The C-terminal extensions of SUMO1/2/3 precursors are removed by SENP1/2/3/5 isopeptidases to expose a C-terminal di-glycine motif essential for SUMO activation and conjugation in an ATP-dependent manner. SUMO1/2/3 can be removed from its substrates and SUMO chains can be depolymerized by proteases (65).

The conjugation of SUMO proteins characteristically stimulates protein-protein interactions through the binding of SUMO conjugates to specific substrates containing SUMO interaction motifs, also called SIMs(76). Much like ubiquitin, SUMO proteins attach as a monomer on single or multiple lysine sites of target substrates generating mono- or multi-SUMOylated proteins, and are habitually induced in response to cellular tension preferentially via lysine residue 11 in SUMO2/3 (Fig. 5) (77). While SUMO2 and SUMO3 share this attachment consensus site,
the necessary sequence for conjugation is lacking within SUMO1 (58), enabling SUMO2/3 proteins to form and terminate chains in vitro and in vivo by the attachment of one SUMO molecule to internal lysine residues of another SUMO moiety (58). One signaling process initiated by SUMO chains is the evolutionarily conserved SUMO-targeted Ubiquitin ligase (StUbL) pathway in which poly and multi SUMOylated proteins, are recognized by E3 ubiquitin ligases containing specific binding modules for these structures.

Figure 5- SUMO’s Characterized Functional Properties Affect Many Biological Processes
SUMO2/3 proteins can form chains to each other or substrates at lysine 11 in different conformations to aid in specific cellular processes such as protein trafficking, chromatin regulation, stress response, cell cycle progression and genome stability.

Such is the case for the RING-type ubiquitin ligases that can also synthesize hybrid SUMO-ubiquitin chains by ubiquitylating lysine residues on SUMO. Tryptic digests and MALDI-TOF mass spectrometry analysis of SUMO2 conjugation assays show that this self-conjugation occurs via Lys11 that exists within a SUMO modification motif (78). Moreover, mutation of lysine 11 in SUMO2 and SUMO3 blocks the assembly of polymeric chains in reactions either lacking or in the presence of substrate (78). As mentioned earlier, while SUMO2 and SUMO3 are 96% identical,
these studies postulate SUMO2 is the better substrate for chain formation than does SUMO3, thus the 3-amino acid variance in the N-terminal region functions to partially inhibit the conjugation of SUMO3 (78). However, whether this inhibition is due to different secondary structures of the two modifiers in this region, the disruption of specific amino acid interactions with Ubc9, or if the SUMO system has the capability of regulating itself remains to be determined.

The SUMO proteases also have preferential affinities for the diverse SUMO1/2/3 paralogs (56). SENP1 and SENP2 can deconjugate all SUMO1-modified substrates with implications in the hypoxic (79), androgen (80), and interleukin-6 (81) response arrangements. In contrast to SENP3, SENP5, SENP6, and SENP7 all prefer SUMO2/3 as substrates (56). SENP6 and SENP7 also have chain-editing activity, which removes SUMO2/3 from poly-SUMO chains often implicated in homologous recombination (82, 83). In summary, although the enzymatic mechanism of SUMOylation is like ubiquitination, their pathway enzymes perform in entirely distinct ways. Moreover, while the SUMO enzymes are present at lower levels than their counterpart in the ubiquitin pathway, the number of SUMO substrates remain vast. This then raises the question of how does SUMOylation exactly achieve specificity through a limited number of conjugating and deconjugating enzymes? Furthermore, how does such a minute amount of SUMOylation of a given substrate lead to such amplificatory biological effects?

1.5 SUMO, The Regulator

Current mass-spectrometry experiments posit that as many as, if not more, 6000 proteins are SUMOylated (84). Many of these substrates have been implicated in crucial physiological processes, such as early embryonic death in mice due to mitotic defects because of Ubc9 knockdowns (85, 86). A very well-characterized role for SUMOylation in cellular health is the
repair of DNA breaks. Protein SUMOylation occurs at the lesion sites to recruit the E3 ligase that further aids the stabilization of protein/protein interactions within the supramolecular complexes responsible for DNA repair (87). Finally, further recruitment of SUMO-binding proteins such as the SUMO-targeted Ubiquitin Ligase RNF4 occurs who then bind to MDC1, a mediator of DNA damage checkpoint 1, which becomes ubiquitinated to signal it for degradation thus finalizing an efficient repair illuminating a complex orchestrated pathway carried out by the Ubiquitin-like family for carrying out and regulating cellular processes (87).

The other best-studied cellular function that is widely controlled by SUMOylation is transcription (88-90). The initiation of transcription is crucial to regulating gene expression and is a collaborative effort between the structural properties of DNA and specific transcription factors (91). Transcription factors are essential regulatory proteins that affect gene expression by binding to exposed DNA sequences. SUMOylation plays an extensive role in gene regulation considering the sizeable amount of SUMO conjugate targets is transcription factors. While these transcription factors symbolize a major group of proteins regulated by SUMO modification (92), the blocking of SUMOylation modification on gene-specific transcription factors ELK-1 (93), C/EBP (94), c-Jun (95), generates an increased amount of transcription on target genes. However, gene-specific transcription factors are not the sole SUMO targets, recent wide-range proteomic reports have also cataloged general transcription machinery including multiple transcription factor subunits as other SUMO foci such as TFIIA, TFIIF, TFIID, TBP, several TATA-binding protein factors, and more notably subunits of RNA polymerase II (96). While these previous studies demonstrate a strong SUMO connection with transcriptional repression, others have shown SUMO is also existent at transcriptionally active genes (65). Ultimately, a transcription complex is formed to facilitate the binding and transcription by an RNA polymerase that can’t proceed with transcription without the
factors such as activators, regulators, repressors, and enhancer sequences which have all been implicated as substrates for SUMOylation (91) (97). Furthermore, even when transcription factors are present in a cell, transcription does not always occur- because often, the transcription factors cannot reach their target sequences.

Chromatin complexes result from the magnetism of the negatively charged DNA to positively charged histones arrangements into nucleosomes and must be unwound for expression. Thus, the state of chromatin is a major player in gene expression because it limits the entry of transcription factors and RNA polymerase into the DNA double helix (91). These chromatic arrangements ultimately reflect a restrictive ground state of a gene’s expression serving as a means for allowing the concurrent regulation of functionally and/or structurally related genes that also contributes to the shifting levels of intricacy in gene regulation (98). Further findings postulate observations of transcriptional co-repressors, HDACs, preferentially bind to SUMOylated transcription factors (99). Interestingly it was reported that reducing SUMOylation at the induced ARG1 promoter, a critical gene for tissue repair in M2 macrophages, hinders the cell’s ability to stop transcription upon activating the signal disappearance (100). This illustrates a complex inverse mechanism indicating while SUMO is upregulated upon the induction of active genes, modification occurs at the promoters but regulates transcription levels by aiding in switching off the transcription (65).

Once the helical orientation of DNA is unwound, the sequence functions as the archetype for complementary base-pairing transcribed by RNA polymerase II, the enzyme responsible for catalyzing the configuration of a pre-mRNA molecule (101). Genome-wide analysis of SUMO1 conjugation to chromatin-associated proteins revealed a striking enrichment of the SUMO1 mark on exons and no enrichment on introns (102). The mechanism in which the exons are recognized at the DNA level, however, is unclear. It is therefore likely that the SUMO1-conjugated proteins
associated with exons may act as splicing factors (103). It is known that overall splicing patterns are sensitive to changes in specific splicing factors upon deviations from normal environmental conditions, causing swift genome-wide splicing profiles that are unique to the stimuli (104, 105). This introduces an intriguing model with responses to environmental changes, prompt post-translational modifications adjust precise splicing factors to control their roles in the excision of specific introns (106). Due to its large size, the spliceosome encompasses both cis-elements that directly interact with the nucleic material, as well as trans-acting elements acting directly on the cis-acting proteins. Although the RNA is not in direct contact, the trans-acting elements control the arrangement of the cis-acting elements such as the snRNPs: U1 and U2 (107). Proteomic methods revealed RNA-related proteins are the principal group among SUMO substrates as well as essential splicing proteins: U2 snRNPs that aid in the pre-mRNA branch site recruitment at the donor splice site, the U4/U6 proteins required for tri-snRNP formation, the functionally important U5 proteins, the RNA helicase required for spliceosome activation, and the catalytic core scaffold proteins were all identified as SUMO substrates (108). Other proteomic studies have classified SUMOylation targets in mRNA metabolism including capping, splicing, and polyadenylation (109), as well as regulatory roles in RNA editing/binding by hnRNPs (110). Essential splicing factors, such as serine-arginine rich (SR) proteins have been identified as SUMO substrates (111). The SR protein SF2/ASF interacts with UBC9 and enhances SUMOylation of specific substrates as well as promotes SUMO conjugation to RNA processing factors (111). SR proteins ensure the coupling of splicing to transcription as they are recruited together with U1 snRNP by RNA polymerase II and co-transcriptionally deposited on the exons 5’ donor splice sites of nascent transcripts (112). Another RNA editing enzyme, ADAR1, binds to double-stranded RNA and is also modified by both SUMO1 and SUMO2, as well as containing capabilities of engaging as a
polySUMO-modified protein (113). Ultimately, SUMO modification edits activity by repressing the dimerization domain of the protein, demonstrating its role in regulating the RNA editing (114). Thus, the splicing of the primary SUMO1/2/3 transcripts may serve as an important regulatory mechanism, and if splicing can be regulated then it may serve as a potential target for drug therapy.

1.6 SUMOylation and Disease

The SUMOylation system has profound effects on many pathophysiological processes such as cancer, diabetes, heart failure, autoimmune diseases, and neurological disorders. The deregulation of both the SUMO ligases and de-conjugation enzymes have been linked to cancer pathogenesis as well as response therapies which is accredited to the delicate homeostasis of the SUMO1/2/3 proteins involved in DNA repair, cell division, and cellular signaling in normally dividing cells (115). A prime instance of this occurs upon Acute Promyelocytic Leukemias (APL) where SUMOylation inhibition of the PML-RARα fusion oncoprotein is prompted by arsenic trioxide, a therapeutic treatment in conjunction with retinoic acid (116). An analogous circumstance ensues for arsenic-induced SUMO-dependent deprivation of the human T-cell lymphotropic virus type I (HTLV1) viral protein associated with Adult T-cell Leukemogenesis (117). Furthermore, SUMOylation also partakes in Acute Myeloid Leukemia (AML) therapy responses through its facility to regulate gene expression (118).

The SUMOylation pathway enzymes are overexpressed in Multiple Myeloma which results in a reduced response to melphalan-based chemotherapies (119-121). This dysregulation of SUMO enzymes is also seen in microphthalmia-associated transcription factor (E318K) which is not only associated with melanoma but also with renal cell carcinoma (122). Genome-wide shRNA screenings classified SUMO E1 and E2 conjugating enzymes as KRAS synthetic lethal partners, however, this was dependent only on SUMO1, not SUMO2/3 modification (123). Similar
screenings of Myc-driven tumors identified a component of the SUMO E1 ligase, SAE2, as the most substantial artificial lethal partner owed to the SUMOylation requisite for proteins involved in the formation of the mitotic spindle (122). Moreover, in human breast cancer tumors with abnormal Myc levels, low SAE1 and SAE2 conjugation were linked with longer metastasis-free survival (122). Thus, targeting the SUMO pathway in conjunction with recently discovered pharmacological inhibitors and current therapies could potentially represent a potent stratagem to improve the outcome of patients (5).

1.7 SUMOylation and the Cellular Stress Response

The SUMOylation system also has immense roles on cellular stress impressions such as the heat shock response, a transcriptional platform to offset protein homeostasis imbalances orchestrated by Heat Shock Transcription Factor 1 (Hsf1) (124). Post-translational modifications, particularly phosphorylation at multiple sites, are one of the vital mechanisms proposed for the adaptation of Hsf1 activity to the needs of individual cells (124). However, according to more recent proteomics data, Hsf1 is also modified by SUMO at several sites resulting in rapid conjugation of SUMO2 and SUMO3 to protein substrates (125, 126). This morphological modification is postulated to provide prompt protection of partially misfolded proteins by increasing their solubility through the addition of SUMO2/3 chains (127). Consistently, cells depleted for SUMO2/3 are more sensitive to the heat stress (128). While heat stress has been proven to upregulate a global increase in SUMO conjugation, this also concurrently leads to SUMO deconjugation of specific proteins (128). However, this is not always true because SUMO also activates transcription factors HSF1 and HSF2 SUMOylation which elevates their DNA binding activity causing an increase in expression, therefore providing more protection against protein-damaging stress (128).
To further indulge on SUMOs role in the cellular stress response, the promoter region of the SENP1(SUMO protease) gene contains androgen (hormone response element receptor signaling) and hypoxia response elements (79). Oxygen deprivation elevates SENP1 transcription (now identified as a hypoxic-response gene in humans) via HIF-1α dependent pathways increases levels of SENP1 protein levels in hippocampal neurons via attachment of SUMO peptides to promote association with interacting proteins, or conversely, the SUMO moiety inhibits interactions by blocking the sites (129, 130). One candidate capable of coordinating the multimodal molecular mechanisms that underlie the cytoprotection induced in the brain as a consequence of cooling is that of global SUMOylation (131) who previously reported on the

The cellular SUMOylation system has been established as an essential factor for the Influenza A life cycle for viral replication regulation and gene expression during infection. Due to their way of replication and a fragmented genome, Influenza viruses mutate quickly often undergoing antigenic shift rendering seasonal vaccines useless with and sometimes even without the antigenic shifts. Our laboratory has not only shown that several influenza viral proteins namely: NS1, PB1, NP and M1 are modified and regulated by the host cell’s SUMOylation machinery (2), but also large increases in global cellular SUMOylation can obstruct the expression of the late influenza gene, M1(Rosas-Acosta, manuscript under review). Animal studies in which mutagenic mice containing SUMO1/2/3 knockouts were infected with WSN Influenza showed dramatic lethality in the alveolar tissue surrounding the lungs not only causing dramatic morphological differences but functional differences as well (Rosas-Acosta manuscript under review). In total, increases in cellular SUMOylation may play a protective role during influenza infection implied by the inverse relationship between the virus’ ability to increase SUMOylation and pathogenicity. It is therefore
critical to consider broad-spectrum therapies employing diverse approaches to fight infections of this sort, for example by targeting the different *cellular systems* crucial for viral multiplication.

Although our lab’s previous data clearly identifies the viral components responsible for triggering GICS during influenza infection, the molecular mechanisms that mediate the net increase in total SUMO modifiers required to allow *global increases* in cellular SUMOylation are far from clear. Thus, a better understanding of the changes in global SUMOylation occurring during infection and their correlation with viral infection progression could lead to molecules/compounds that modulate the cellular SUMO system and could provide potentially effective Influenza antivirals as well as cellular level insult therapies. The relevance of our laboratories findings indicates potential increases in cellular SUMOylation may play a protective role by other types of stress including oxygen and glucose deprivation (OGD) (3), hypo/hyperthermia (4), and cancer (5).

### 1.8 SUMOylation Regulation

While the SUMOylation system has a far reach into both normal cellular physiological processes as well as stress responses, very little is known about how the SUMOylation system itself is regulated. Given the numerous potential applications for disease it is therefore critical to develop a deeper molecular understanding with a further goal to artificially manipulate the SUMO system for therapeutic purposes. One candidate for SUMOylation regulation is SRSF1, a prototypical SR protein, with roles in constitutive and alternative splicing (132), is implicated in nonsense-mediated mRNA decay (133), mRNA export (134), translation (135), and recently reported as a modulator of protein SUMOylation, by connection with Ubc9 to enhance SUMOylation of specific targets (136). SRSF1 also contributes regulatory functions in *heat-shock-induced SUMOylation*
by promoting SUMO conjugation to RNA-processing factors (137), as well as assisting in PIAS1 (protein inhibitor of activated STAT1) regulation, provoking PIAS1-induced overall protein SUMOylation. It should be noted, PIAS1 can be found co-localized with other SUMO targets in nuclear bodies with the E3 ligase conjugating enzyme of the SUMO pathway, components of pre-mRNA splicing factors, and the RNA processing machinery (138).

Regulation of SUMOylation through target modification has also been posited to occur on SUMO substrates that can enhance or block SUMOylation. Competition with other lysine-directed modifications such as ubiquitination of acetylation substrates can abolish SUMOylation while phosphorylation has the potential to act as both a positive and a negative regulator (139). Enhanced SUMOylation has been documented upon phosphorylation of specific residues that are attached to branched SUMO acceptor sites also previously referred to as Phosphorylation Dependent SUMO Motifs (PDSM) (139). In addition to these outcomes, the ascertainment of SUMO isopeptidases with seemingly contrary specificities for SUMO1 and SUMO2/3 insinuates that the global conjugation capacity of SUMO1 and SUMO2/3-modified proteins may, in fact, be regulated at the level of removal rather than conjugation (140, 141). While little is comprehended about the functional discrepancy of SUMO1 in contrast with SUMO2/3, the unearthing of SUMO2 and SUMO3 multimers generates a novel dimension of complexity to SUMO conjugation coordination. While SUMO2/3 chains habitually form in response to genotoxic or proteotoxic stress and are preferentially linked via lysine 11 of SUMO2/3 (125), it was further reported that K11 of SUMO2/3 undergoes reversible acetylation with SIRT1 being the K11 deacetylase in a purified in vitro system (142). It was moreover reported that K11 acetyl mimicking SUMO2 did not impact the StUbL pathway, signifying these non-canonical chains are more prevalent in cells and shown to form under both basal and stress conditions using mass spectrophotometry-based
SUMO proteomics (142). Thus, if acetylation of SUMO2/3 can impair chain formation and restrict chain length this could serve as a potential regulator of the SUMO stress response (142).
2. Aims

While substantial progress has been achieved in characterizing the functions and effects associated with SUMO modification, our knowledge of the mechanisms regulating the activity of the SUMOylation system remains limited. One area that remains unexplored is the potential contribution that post-transcriptional processing may play in regulating cellular SUMOylation. Therefore, a more detailed understanding of the molecular mechanisms contributing to SUMO gene expression regulation is critical to achieving our long-term goal of using the cytoprotective functions of the SUMO system to develop novel therapeutics against cellular stress. The experiments proposed here will provide essential data on untapped areas of SUMO research and could lead to the development of innovative approaches to combat influenza viral infections while also improving patient outcomes following ischemic events. To test our hypothesis of alternative splicing regulating the SUMO1/2/3 genes in the pursuit of the long-term goal, two specific aims will be carried out:

**Specific Aim 1: Measure Alternative Splicing Contribution in Cellular SUMO Level Regulation.** Evidence of alternative splice products for the *SUMO1/2/3* genes indicates alternative splicing may contribute to the regulation of SUMO levels in the cell. Thus, a core methodology will be needed to quantify each variant within the cell by qRT-PCR.

Sub Aim 1.1: Establish the normal proportion of the different SUMO variants in the cell. Total RNA will be purified from cells maintained under normal culture conditions and used to perform qRT-PCR analyses to determine the abundance of each variant transcript for the *SUMO1/2/3* genes. To ensure a proper representation of various cell types, we will establish these values for at least two different immortalized human cell lines as well as for peripheral
blood mononuclear cells (PBMCs) derived from healthy donors. These cell lines were selected because of their human origin, ease of growth, responsiveness to stress, ease of liposome-mediated transient transfection, and their ability to support influenza virus multiplication.

**Sub Aim 1.2: Observe if the copy number of each variant and its proportional abundance to other variants are affected by stress.** Once copy numbers/normal proportions for each SUMO variant have been established, we will determine whether these values are affected by stressors including **heat-shock** (43°C for 1 hour), **influenza virus infection** (Multiplicity of Infection [MOI] of 10, 12-hour post-infection, and two different viral strains, A/WSN/1933 H1N1 and A/Memphis/31/1998 H3N2), and **cold-shock** (30°C for 16 hours). To ensure that the stressors used are triggering the expected responses, each stress treatment will include the quantitative assessment of a gene transcript known to be affected by the specific stress condition being studied. Specifically, we will measure the transcript levels of HSP70 for heat shock, Influenza gene segment M1 for H1N1 infection, and RBM3 for cold shock.

**Specific Aim 2: Characterize Functional Properties of the SUMOα Isoforms.** Based on the quantitative assessments performed to date, out of all the SUMOα-coding variants the one coding for SUMO3α appears to be by far the most abundant in proportion to its prototype as well as being the only alternatively spliced variant with an intronic inclusion; we, therefore, hypothesize, SUMO3α is likely to have the largest effect on cellular SUMOylation. According to informatic prediction analyses, SUMO3α could also be relevant as it is predicted to have an extra intrinsically disordered domain absent in the SUMO3 prototype protein. To characterize the functional properties of the SUMOα isoforms, we will pursue the questions presented below.
Sub Aim 2.1: **Probe for translational implications of the SUMO1/2/3 variants.** This will be conducted through the fractionation of purified RNA in A549 and Hek293A cell lines under both normal and cold shock conditions, deemed to be the one stress condition to have the largest effect on all SUMO1/2/3 variants. The presence of all the SUMO1/2/3 variants in cytosolic extracts could strongly indicate that all SUMOα isoforms (alternate spliced products) are translated, thus justifying a further need for functional assessment.

Sub Aim 2.2: **Describe conjugation abilities of SUMOαs in vivo.** Overexpression of a His-tagged form of SUMO1α in its processed form (with the diglycine motif at its C-end) by transfection in HEK293A cells will be useful in determining if isoforms are conjugatable in vivo. The use of an N-terminal His-tag allows sensitive detection of the protein using anti-His antibodies, easy differentiation from the endogenous pool of SUMO proteins, and could allow the proteomic identification of SUMOylated targets if any of the SUMOα were to be conjugatable.

Sub Aim 2.3: **Determine the SUMOα differences on cellular SUMO1/2/3 localization.** If our previous predicted results all hold true, then the final test in characterizing the SUMO1/2/3α isoforms would be to determine differences in localization. This will be carried out through the transfections using the respective SUMO1/2/3 His-S and YFP-tagged clones in 96 well plates to use in fluorescence microscopy.
3. Methods and Materials

3.1 Primer Design Approach

An overall map of the RNA-seq data obtained from the NCBI database was first constructed using the SUMO1/2/3 genes and their respective alternative variant sequences to infer their distinctions. The map was drawn based on a series of alignments performed by ClustalW between the full genomic sequences for the SUMO1 gene located within Chromosome 2, the SUMO2 gene located within chromosome 17, the SUMO3 gene located within chromosome 21, as well as the different mature mRNA transcripts for all SUMO1/2/3 variants. Afterward, a more detailed map of each transcript was assembled to focus on the boundaries between exons to simplify the design of primers targeting exon/exon junctions. The primers used to quantify all different mature transcripts were designed using the following guidelines: 1) Paired primers should have similar TM°s (melting temperatures) 2) the expected PCR products produced should be between 200-300bp in length 3) Primers should target exon/exon junctions 4) There should be specific primer sets for every mature mRNA 5) Any given primer pair should not amplify more than one mature mRNA 6) Primers should contain as close as possible 50:50 [GC] : [AT] content with no more than 40:60 content either way 7) Primers should end on a “clamping sequence” (CG, GC, GG, CC) and lastly, 8) primers should be free of sequences likely to form strong secondary structures. All primers were ordered from IDT (integrated DNA technologies) and reconstituted in sterile TE at a 100μM concentration. The 100μM primer stock solutions were further diluted down to 10μM in TE to produce working for direct use in RT-PCR and qRT-PCR reactions.

3.2 Cell Culture

Human embryonic kidney cells (HEK293A) are a suitable cell model due to their ease of growth and high transfection rates, while A549 cells are adenocarcinoma human alveolar basal epithelial
cells are an ideal model that can form an abundance of intercellular connections. The A549 cell lines are often used as an in vitro model for Type II pulmonary epithelial cell model in drug metabolism assays but also serve as an efficient transfection host. To avoid the likelihood that our data may be biased to events that are specific to immortalized cell lines, we included Peripheral Blood Mononuclear Cells (PBMC) in our analyses. These cells include both lymphocytes (T cells, B cells, NK cells) and monocytes and are obtained from healthy volunteers. HEK293A and A549 cells were grown in tissue culture in DMEM media containing 10% Fetal Bovine Serum, plated in 6-well plates at a concentration of $3 \times 10^5$ cells per well, and became confluent roughly 24-34 hours post-plating. The A549 and HEK293A cell lines were collected in groups of 3 wells per each sample for subsequent RNA purification. PBMCs were directly isolated from patients and stored in cryovials awaiting RNA purification.

### 3.3 RNA Purification

Cells were washed in 1mL PBS and total RNA was purified using the Qiagen RNeasy Mini Kit® via the Qiashredder® method. To begin, 200µL of buffer RLT was added to each well of a 6-well plate containing approximately $1 \times 10^6$ cells per well; the resulting mix was incubated for five minutes at 23°C. The lysates from each of the three wells were combined into a single well, thus resulting in two samples of each cell line per plate. The samples were then transferred into a microfuge tube containing a Qiashredder® cartridge, mixed until there were no visible clumps, and centrifuged at full speed for two minutes at 23°C. Afterward, one volume (roughly 600µL) of 70% ethanol (EtOH) was added to the lysate followed by careful pipette mixing to ensure proper resuspension and homogenization of the lysate. Up to 700µL of each sample was added to an RNeasy® spin column and centrifuged for thirty seconds at 8000 x g. The flow-through was discarded and 700µL of Buffer RW1 was added to the RNeasy spin column, and again centrifuged
for thirty seconds at a speed of 8000 x g and 23°C. Next, 500µL of Buffer RPE was added and centrifuged for thirty seconds at 8000 x g and 23°C. Another 500µL of Buffer RPE was added but centrifuged for two minutes at 8000 x g to dry the cartridge. Subsequently, the cartridge was placed in a new 2mL collection tube and centrifuged for one minute at full speed, then placed into a new 1.5mL collection tube, and RNA was eluted in 50µL of RNase-free DEPC-treated H2O and centrifuged for 1 minute at 8000 x g. The eluted RNA was finally then aliquoted in 9µL and stored at -80°C.

3.4 Assessment of Purified RNA Quality and Quantity

To assess the quantity of the purified RNA a Qubit Fluorometer 3.0® was used and formaldehyde-agarose gel electrophoresis was used to determine the quality of the RNA obtained, according to the following procedure: 10xFA gel buffer (consisting of 200mM 3-N-morpholino propane sulfonic acid (MOPS), 50mM Sodium Acetate, and 10mM EDTA), the pH was then adjusted to 7.0 using 68mL of 1M-NaOH per one liter of 10x FA Buffer. 1xFA gel running buffer was then be made using 100mL of the 10xFA gel buffer previously described, 74mL of 10% Formaldehyde and 826mL of DEPC-H2O. To load the gel while ensuring proper loading of the total purified RNA samples, we made a 4xRNA loading buffer consisting of 20µL saturated bromophenol blue, 800µL of 500mM EDTA, pH 8.0, 2.664mL of 10% Formaldehyde, 2mL of 100% glycerol, 3,080 mL formamide, 4mL of 10xFA gel buffer and lastly 656µL of DEPC-H2O. To produce an FA gel mix, we mixed 3.6g Agarose, 30mL 10xFA gel buffer, and 280mL of DEPC-H2O. The contents were then heated up to a boil and the agarose was allowed to solubilize by cooling to 65°C. 20mL of 10% Formaldehyde was added to the mixture and finally stored at 65°C. It should be noted that problems arose using DEPC-H2O in both gel electrophoresis and qPCR reactions, we then switched to using QH20. Qubit Fluorometer quantification began by initially making two standards
that generate a curve for the purified RNA to measure against. A working solution must first be made for both standards and samples by using 199µL of Qubit buffer with 1µL of Qubit Reagent. Both standard one and standard two use 190µL of the working solution combined with 10uL of Standard 1 and Standard 2. Once the standards were properly vortexed for two to three seconds, the samples must sit for two minutes before the reading commences. To begin the assembly of reading the actual samples, 198µL of the working solution is combined with 2µL of the purified RNA followed by the mixing and settling of the samples.

3.5 cDNA Synthesis

Following the purified RNA extraction from both HEK293A and A549 cell lines, cDNA synthesis was performed using an NEB-M-MuLv® kit. This system required a total of 2µg of RNA added to 2µL of 50mM Poly-T primer, 1µL of 10mM dNTPs, and lastly qH20 to bring the reaction volume up to 10µL. The reaction was then heated for five minutes at 65°C and kept on ice while the other components were added: .5µL RNAse inhibitor, 2µL 10x M-MuLv Buffer, 1µL M-MuLv Reverse transcriptase, and lastly qH20 to bring the reaction volume up to 20µL. The final steps of the cDNA synthesis included incubations at 42°C for one hour, followed by a cool down to 4°C. Negative control samples were produced using all the ingredients minus the Poly-T enzyme and the Reverse transcriptase, qH20 was used in place of these components for the negative control reactions.

3.6 qRT-PCR

To perform the initial PCR reactions, the Bio-Rad iTaq™ Universal SYBR® Green One-Step Kit was used completely avoiding the need to synthesize cDNA altogether. Instead of using the typical 2ug of RNA, this new method was able to carry out the reactions with a mere 100ng RNA coupled
with 10µL reaction mix, 2uL of both forward and reverse primers, .5µL of reverse transcriptase, and lastly, water to bring the reaction volume up to 20µL. Negative controls were assembled using all components minus RNA template. Subsequent runs were carried out using a MyGo thermocycler, along with MyGo programming ran on MAC software.

3.7 Cloning of SUMO1/2/3 Variants

Clones of each of the SUMO1/2/3 variants were generated first by using a Taq Polymerase PCR reaction whose products were necessary to ultimately produce calibration curves in hopes of detecting the exact copy number of each of the SUMO1/2/3 variants produced within the cell. The PCR reaction consisted of 5µL cDNA, 2.5µL 10mM DNTP, 5µL Q5 High-Fidelity Buffer, 2µL of each primer, 1µL of Q5 High-Fidelity DNA polymerase, and 32.5µL QH20 to bring the reaction volume up to 50µL. The PCR products generated were then added to a Thermo-Scientific CloneJET PCR Cloning Kit®: 10µL of 2x Reaction Buffer, 1µL of each PCR product, and 7µL of QH20 for a total of an 18µL reaction volume. The reaction was then vortexed briefly and centrifuged for three to five seconds, followed by incubation at 70°C for five minutes. Finally, ligation reactions were compiled on ice by adding 1µL of the pJET1.2/blunt cloning vector® and 1µL of T4 DNA ligase to the 18µL blunt-end restriction digest reactions.

3.8 Transformation of SUMO1/2/3 Clones into NEB 10-Beta Competent E. Coli

Upon the ligation reaction setup, the transformation of NEB 10-beta competent E. coli cells was then necessary. The transformation entailed thawing of the cells on ice for ten minutes followed by splitting the cell mixture into 2x 25µL aliquots. From this reaction, 1.5µL of each ligation reaction was then added to the cell mixture, carefully mixed, and incubated on ice for thirty minutes. Heat shock was performed for exactly thirty seconds at 42°C and the samples were then
placed on ice again for five minutes. The NEB 10-beta/ Stable Outgrowth Media (SOC) was added to the mixture (450µL) and incubated at 37°C for sixty minutes at 250rpm. Two different volumes of each SUMO1/2/3 variant mixtures were plated (75µL and 300µL, respectively) on 2xYT- Agar plates and incubated overnight at 37°C. To screen for the colonies with the correct SUMO1/2/3 variant sequences, a P200 pipette tip was used to pick five colonies of each variant and placed into 300µL of H2O. Of the 300µL bacterial solution: 5µL was used to set up a replication plate, and another 5µL was used to set up another Taq Polymerase PCR reaction (Figure 7) for the screening process along with 2.5µL dNTP, 5µL Taq Polymerase Buffer, 2µL of the forward primer provided by the CloneJET PCR Cloning Kit, 2µL of the respective SUMO12/3 variant reverse primers, 1µL Taq Polymerase and 32.5µL of QH20.

3.9 Maxi- Prep Procedures

Sequence samples were then subjected to Maxi- Prep procedures using a ZymoPURE II Plasmid Maxi Prep kit to continue with the goal of producing the respective SUMO1/2/3 variant calibration curves. The maxi-prep protocol began with the inoculation of the three selected sequenced samples initially stored in cryovials (-80°C) into 500uL 2xYT. The bacterial cells were then harvested by centrifugation at 3400 x g for ten minutes at 4°C. The supernatant was pipetted out and 14mL of ZymoPure P1 containing RNAse A as well as LyseBlue was added to the pellet followed by complete mixing until no cell clumps remained. 14mL of Buffer P2 was then added and mixed thoroughly by inverting the sealed tube six times and incubated at room temperature for 3 minutes. 14mL of chilled buffer P3 was added to the lysate and mixed immediately by inverting four to six times. The lysate was loaded into the ZymoPURE Syringe Filter followed by a 5-minute incubation allowing precipitation to float to the top. The syringe was then added into a clean 50mL conical tube and pushed through until 35mL of the cleared lysate was recovered. 14mL of Binding Buffer
was added to the cleared lysate and mixed thoroughly by inversion 10 times. The mixture was then added to a Zymo-Spon V-P Column Assembly and vacuumed until all liquid had passed. 5mL of Wash 1 was added and vacuumed. 5mL of wash 2 was added and repeated once more. The samples were then centrifuged at 10,000 g for 1 minute. The columns were then transferred to a clean 1.5mL microfuge tube with 400ul of elution buffer added directly to the column matrix. The samples were incubated for 2 minutes and centrifuged at 10,000 g for 1 minute. Lastly, for Endo Zero Plasmid DNA, a final spin was conducted. The maxi-prep samples were quantified using a Qubit Fluorometer and calculations were performed to measure precisely 1µg/µL of each of the SUMO1/2/3 variants plasmid DNA.

3.10 **T7 RNA Polymerase In Vivo Transcription**

Using a Mega script RNA synthesis kit, it was necessary to generate an *in vivo* transcription using the plasmid DNA of each SUMO1/2/3 variant beginning with a HindIII restriction digest, an enzyme that doesn’t cut SUMO itself but only has a cut site within the plasmid itself. This 50µL reaction resulted in 5µL being used for quality verification by gel electrophoresis, and the rest precipitated by adding 4.5µL Ammonium Acetate and 90µL of chilled 100% EtOH. The SUMO variant samples were centrifuged for thirty minutes at full speed and the pellet was re-suspended in 20µL of sterile TE. The precipitated DNA was quantified again using the Qubit to ensure exactly 1µg of DNA would be used for the following *in vitro* transcription which consisted of assembling an NTP mixture (2µL of each), 2µL of 10X Reaction Buffer, 2µL of enzyme mix, and the respective SUMO1 variant 1ug amount at 23°C. The mixes were then incubated for four hours at 37°C and subjected again to gel electrophoresis to ensure the proper product sizes.
3.11 Generating Calibration Curves

Calibration curves were necessary to obtain an exact copy number of each SUMO1/2/3 transcript within the cell to measure true fluctuations upon different stress conditions. To begin, the sequences for each of the SUMO1/2/3 variants were mapped out and inserted into the pJET1.2 plasmid sequence in between the T7 promoter site and the HindII cut site (AAGCTT). The exact number of nucleotides was entered into an oligonucleotide calculator to determine the exact grams/mol of each transcript. The transcript weight was then multiplied by $10^{11}$ (initial copy number dilution) and divided by Avogadro’s number (6.02 x $10^{23}$ molecules) to generate the needed amount of transcribed RNA in nanograms based on the molecular weight of each SUMO1/2/3 gene variant transcript. Based on the nanogram amount required for each RNA sample, dilutions were made following the initial Qubit quantifications. From the first dilution which contained $10^{11}$ copies, serial dilutions loading 10uL of each previous dilution were performed all the way down to $10^{0}$ copies serving as a negative control. The concentration of the SUMO1/2/3 variants transcribed RNA was the amount of each serial dilution loaded into subsequent qRT-PCR triplicate reactions using the one-step method. However, the qRT-PCR reactions started with the $10^{7}$ copy number dilution all the way down to $10^{0}$ serial dilutions being loaded. Statistical analysis involved using the Cq values ran in triplicates with R-values calculated and compared amongst each other using excel spreadsheets.

3.12 Stress Treatments

Cells plated in 6-wells at a concentration of $3 \times 10^5$ cells per well demonstrated proper confluency 36 hours post-plating. Stressors include heat shock (43°C for 1h), influenza virus infection (Multiplicity of Infection [MOI] of 10, 12h post-infection with the viral strain
(A/Memphis/31/1998 H3N2), and cold shock (27°C for 24h) were all conducted three separate times and collected for subsequent RNA purification beginning with lysis buffer also placed in respective treatments to prevent the least amount of change to the cells post-treatment. To ensure all stressors triggered the expected cellular responses, each stress treatment consisted of also quantifying a gene transcript known to be affected by the specific stress condition being studied. Specifically, the transcript levels of Hs p70 (heat shock), Influenza M1 (influenza infection), and Rbm3 (cold shock). Additionally, to verify that the cellular stressor triggered the expected change in global cellular SUMOylation levels, a set of examples exposed to identical stress conditions will also be collected for immunoblot analyses using anti-SUMO and anti-GAPDH antibodies.

3.13 Nuclear vs Cytosolic Fractionation

Media was aspirated and the cellular monolayer was washed with 2mL of PBS. 200ul of ice-cold Lysis Buffer J directly to the culture plate. Cell lysis was carried out by gently tapping the culture dish and swirling the buffer around the plate surface for five mins while keeping the cell culture plate on ice. Using a pipette, the lysate was transferred to an RNase-free microcentrifuge tube and centrifuged for 10 minutes at maximum speed. Using a pipette with a small tup, transfer the supernatant containing cytoplasmic RNA to another RNase-free tube. The nuclear pellet was obtained and mixed into 400ul of Buffer SK while the cytoplasmic fraction was mixed into 200ul SK. A 200ul of 100% ethanol was then added to each respective fraction, transferred into a spin column, and centrifuged for 1 minute at 3500 g and the flow-through was discarded. The columns were then washed with 400ul of Wash Solution A to the column and centrifuged for 1 minute with the flow-through being discarded and repeated three times. The columns were then centrifuged for an additional 2 minutes to thoroughly dry the resin. The columns were then placed into fresh tubes and eluted in 50ul of Elution Buffer E. The samples were spun down for 2 minutes at 200 g
followed by 1 minute at 14000 g. Lastly, the samples were stored at -80°C awaiting qRT-PCR ran in triplicate using 100ng of mRNA. To calculate the % of RNA in each fraction, we began by calculating the change in ΔCT through the subtraction of the CT for the highest minus CT for the lowest. The fold difference was then calculated by $2^{\Delta CT}$. Then calculate A: % for the one present in the lowest amount 100/ (fold difference + 1). Then calculate B: 100-A. Check calculations by A*FD=B.

3.14 **His-Tag SUMO1/2/3 Plasmid Cloning**

To assess whether SUMO1/2/3α can be processed to their GG-C end mature forms in vivo, we over-expressed the His-tagged SUMO1/2/3α by transfection by using constructs coding for a C-terminal extension. The C-terminal extension adds 39 amino acid residues, including two T7 tag epitopes, after the diglycine motif to allow rapid differentiation of the mature and precursor forms by immunoblotting analysis of whole-cell extracts using anti-His and anti-T7 antibodies. The cones were then transformed into E. Coli and screened for colonies using RT-PCR followed by a BamHI digestion. Upon sequence confirmation, the samples were then mini prepped to amplify DNA quantity and ready for transfections followed by immuno-blotting.

3.15 **Transfections**

Cells were plated at a low density of 2 x 10^4 and observed for 70% confluence. At the time of this targeted confluence, remove the media from the wells and wash 2x w/2mL PBS. Aspirate PBS and add 500uL of pre-warmed Opti-MEM per well and incubate at 37°C, 5% CO2. Prepare transfectant as follows: Label a tube as Trans-IT Master Mix. (Ensure this tube can hold 250uL x Number of Wells). Add (250uL x Number of Wells) of prewarmed Opti-MEM into this tube. Add 7.5uL of Lipofectamine 2000 per well into this tube. Vortex the tube and incubate at room
temperature for 10 minutes. Create a tube per well. Add 250uL of pre-warmed Opti-MEM into each tube. Add the respective DNA to each of these tubes; 5ug/well. Add 250uL from the Master Mix to each of the labeled tubes containing DNA carefully avoiding cross-contamination between respective tubes. Vortex each tube for about 15 seconds and incubate at RT for 30 minutes. After the liposome incubation, add transfectant into each well “dropper-style”. Let incubate for 3 hours, after this incubation, add 1mL of complete media into each well.

3.16 Western Blot Analyses

Treated cells were lysed using boiling 500ul of 4x Laemmli Sample Buffer, passed through a syringe twenty times to ensure DNA was sheared and 40ul of B-Mercaptoethanol was added followed by one last boil for 3 minutes. Once sample preparation was complete, the samples were run on a 10% SDS-Page gel using Bio-Rad System at 50V overnight. A large wet transfer unit was then set up and ran for 3 hours using Bio-Rad System and the membrane was dried for 10 minutes. The PVDF membrane was re-wet by soaking in 5mL of absolute methanol for 1 minute, discarded, and rehydrated in QH20 by shaking for two minutes. The water was discarded, replaced, and shook for another 2 minutes at room temperature. The membranes were then washed for 2 minutes three times with 1x PBS at room temperature on a rocker. Lastly, the membrane was blocked by adding 10mL Licor Blocking Buffer for 60 minutes. Primary antibodies were added in solution with 100ul of 10% Tween-20 to 10mL Blocking Buffer and mixed well. Primary incubated overnight at 4C on a rocker, washed 5 minutes 3 times with 40mL 1xPBS + 0.05% Tween-20. Secondary antibodies were added in solution with 100ul of 10% Tween-20 to 10mL Blocking Buffer and mixed well. The secondary was left to incubate at room temperature for 1 hour, washed 5 minutes 3 times with 40mL 1xPBS + 0.05% Tween-20 followed by one final wash with 1xPBS.
3.17 Confocal Microscopy

Hek293A cells were transfected and fixed with Hoechst 33342 staining solution for deep nuclei visualization on the 460-490 nm channel in a 96-well plate with the generated SUMO1/2/3 and SUMO1/2/3\(\alpha\) His-tagged and YFP-tagged clones using the above-mentioned transfection protocol. Upon the desired observed confluence, the cells were subjected to fluorescence confocal microscopy with detection of the YFP-fusions of SUMO on the green channel, phalloidin to detect actin filaments on the red channel, and DAPI to stain for DNA from three representative fields of cells.

3.18 Tertiary Structure Prediction Analyses

An important tool for structural differences used in these studies was RaptorX prediction software. RaptorX is a homology-based prediction software based on the structural properties of a protein sequence without using any templates. Although it is sometimes considered fast and sloppy, it outperforms other servers, especially for proteins without close homologs or with very sparse sequence profiles (143). Quantitative measurements predict an \(\sim\)84\% Q3 accuracy for 3-state SS, \(\sim\)72\% Q8 accuracy for 8-state SS, \(\sim\)66\% Q3 accuracy for 3-state solvent accessibility, and \(\sim\)0.89 area under the ROC curve for the disorder prediction (143). However, this template-based modeling doesn’t answer how and why proteins adopt their specific structures; and if structural homologs do not exist or cannot be identified, models then must be constructed from scratch. This procedure, called ab initio modeling, is essential for a complete solution to this problem; and can help us understand the physicochemical principle of how proteins fold in nature (144). Currently, the accuracy of ab initio modeling is low, and the success is generally limited to small proteins (<120 residues). A more recent innovative approach to protein prediction structures is Alpha Fold which is an artificial intelligence program developed with a system of subnetworks coupled...
together into a single differentiable end-to-end model based on pattern recognition. All six amino acid sequences coding for the prototype SUMO1/2/3 proteins and their encoded isoforms were entered in both the RaptorX and Alpha-Fold working notebook prediction servers to generate the respective alignments of each SUMO paralog group. The scripts were then run in each respective server to generate the tertiary structure of each SUMO protein of interest which were then compared resulting in no apparent structural differences between the servers. For these studies, the Alpha-Fold predictions were chosen.

3.19 Statistical Reports

To estimate the CNest (copy number estimates) using a linear regression analysis based on the average CQ values of each variant and was applied under various experiments (H1N1, Cold Shock, Heat Shock). The copy number was then calculated by taking $10^{\Delta CQ}$. The average comparison across positive and negative treatments were compared using an unpaired students t-test. A Bonferroni correction was used to adjust for the number of multiple comparisons within each treatment ($sig: p<0.008$). Nuclear and Cytosolic cellular fractions were compared using the log2 scale of the $2^{-\Delta CT}$ method. Analysis was conducted using Stata v.17 and GraphPad Prism V.6.0.
4. Results

The extensive RNA sequencing data in the NCBI server has made it possible to catalog virtually all mature mRNA transcripts present within a given organism thus facilitating our investigation of the regulation of the SUMOylation system. ClustalW alignments of *H. sapiens* SUMO1/2/3 mRNA sequences obtained from the NCBI database indicate there are three different forms of mature mRNA transcripts for SUMO1 (Fig. 6); variant1 (S1var1) contains all exons while variant2 (S1var2) undergoes a splicing event which is proposed here to correspond to a cryptic intron that is eliminated in exon 5. However, due to this cryptic introns’ location within the 5’UTR, both SUMO1var1 and SUMO1var2 transcripts code for what is considered the prototypical SUMO1 protein. SUMO1var3 is of particular interest because the second exon is spliced out, thus the mature mRNA codes for a substantially shorter version of the protein here referred to as SUMO1α with a 2.7kDa size difference compared to the familiar SUMO1 prototype protein. Both SUMO2 and SUMO3 also contain alternatively spliced transcripts, each of which has two variants, one corresponding to the normally spliced transcript and one corresponding to an alternatively spliced one: SUMO2α and SUMO3α (Fig. 6). In the case of SUMO2, the alternatively spliced transcript completely skips exon 3 thus coding for a shorter protein isoform with a 2.8kDa difference in size. Similarly, SUMO3 also contains two distinct transcripts, however, instead of deletion, SUMO3 is the only SUMO gene with an intronic inclusion to the second exon coding for a larger protein weighing 4.2kDa more than its respective prototype. It is proposed here that exon 2 of the SUMO3var2 transcript likely contains a weak splicing donor that is sometimes ignored, leading to the retention of a small fragment of the intronic sequence between exon2 and exon 3.
Figure 6- SUMO1/2/3 Alternatively Spliced mRNA Transcripts Informatic Predicted Properties
A: SUMO1 variants, denoted here SUMO1var1/2/3. SUMO1var1/2 code for prototype SUMO proteins weighing 11.5kDa, although they have different exon5 splicing. SUMO1var3 lacks exon2, coding for a truncated protein, SUMO1α (8.8kDa). SUMO2 variants, denoted here SUMO2var1/2. SUMO2var1 translates into SUMO2 prototype (10.9 kDa). SUMO2var2 lacks Exon3 and codes for a truncated protein, SUMO2α (8.1 kDa). SUMO3 variants, denoted here SUMO3var1/2. SUMO1var1 codes for prototype weighing 11.6 kDa. SUMO3var2 is the only SUMO variant that contains an intronic inclusion in Exon2 and codes for a larger protein, SUMO3α (15.8 kDa). E: Exon. The thin lines connecting the exons represent introns, the thick black lines represent the SUMO genes, and the numbers underneath indicate base pairs from the transcriptional start of the gene. V= variant. In every case, V1 corresponds to the normally spliced transcript.

B: JASSA SUMOylation motif predictor shows all SUMO1/2/3 amino acid sequence alignments indicating SUMO1 and SUMO1α have an amino acid difference of 25 residues resulting in a 75.2% sequence similarity. The skipping of exon 2 to generate the truncated SUMO1α protein is also lacking 5 Lysine residues at positions: 7, 16, 17, 23, and 25. SUMO2 and SUMO2α differ by 24 residues resulting in a 74.7% sequence similarity. SUMO3 and SUMO3α have the biggest difference of 33 residues with a 73% sequence similarity.

C: Table summarizing the differences between the prototype SUMOs and their respective isoforms.

Informatic alignments using JASSA (Joined Advanced SUMOylation Site and Sim Analyzer) were conducted to further assess the properties of these alternatively spliced isoforms of the SUMO1/2/3 proteins. JASSA is an online predictor that incorporates a scoring system based on a specific position frequency matrix generated from the positions of experimental SUMOylation sites to predict if these isoforms contain functional properties at the amino acid level (145). Corresponding with different web tools, JASSA exhibits superior executions owed to applied innovative qualities concerning improved evaluations of the extrapolation. These improvements consist of recognizing database hits that are equivalent to the query sequences and constitute rival sites not only to secondary structural elements but also to the 3D fold of the protein of interest retrieved from deposited PDB files (145). The JASSA prediction alignments revealed that since SUMO1α shows a complete deficit of the second exon that translates to a lack of the highly conserved Lysine7, important for SUMO1/2/3 conjugation, as well as 5 other lysine residues within the second exon with an amino acid difference of 25 residues and 75.2% sequence similarity. The JASSA predictions further show that SUMO2 and SUMO2α’s difference in exon 2 indicates a 24 amino acid difference with a 74.7% sequence similarity, however, there are no missing lysine residues. Lastly, SUMO3 and SUMO3α differ from each other by 33 amino acid
differences with the lowest sequence similarity of 73%, the lowest of the respective SUMO isoforms. However, these differences are not predicted to translate to any missing or extra lysine residues thus raising the question of whether both SUMO2α and SUMO3α are capable of not only conjugating but forming chains.

Although evidence of the existing SUMO1/2/3 variants is strongly supported on the NCBI database, according to the literature there have been no reports describing the existence of these variants. Given how common alternative splicing produces isoforms and the RNA-seq data available, it is a logical summation that alternative splicing may be a potential candidate for the regulation of the SUMO1/2/3 genes, *ultimately regulating itself* through a negative feedback mechanism (based on the existence of different SUMO1/2/3 variants) resulting in *de novo synthesis* of multiple proteins being coded from the same gene. While the speed of the changes observed in our lab’s previous studies of global SUMOylation levels suggests post-transcriptional regulatory mechanisms at play, it is, therefore, crucial to measure the abundance of SUMO1/2/3 transcripts during infection and other stress conditions to determine whether the changes in global SUMOylation observed entail changes at the mRNA level as well. In these studies, we have characterized the alternative splicing of the three SUMO paralogs and determined their functional properties.

To facilitate our study of the SUMO system at the transcript level, a primer design approach was created to make use of the differences in the SUMO1/2/3 transcripts by inserting the primers at the exon/exon junctions to ensure strict amplification of the SUMO1/2/3 messenger RNA transcripts (*Fig. 7*). To distinguish SUMO1var1 (345 bp) and SUMO1var2, we designed a primer to target the area of the 5’ UTR that is spliced out in SUMO1var2 (258 bp). SUMO1var3 (198 bp) is distinguished by targeting the exon junction formed by the splicing of exon2 out.
SUMO2var1 (242 bp) and SUMO2var2 (157 bp) are distinguished from each other using this approach by targeting the exon junction formed through the skipping of exon 3 in SUMO2var2.

**Figure 7- Strategy for Independent Targeting of the SUMO1/2/3 Variants**

A) Primers were designed to target exon/exon junctions to ensure proper amplification of mRNA transcripts only. B) Primer design validation through PCR using RNA extracted from Hek293A cells and confirmed via agarose gel electrophoresis. C) Final confirmation of primer design via sequencing.

Due to the intronic inclusion generated in SUMO3var2 (239 bp), our primer design approach targets this area while the primer designed for SUMO3var1 (179 bp) targets the exon junction formed by exon2 and exon3. These data confirm not only does the cell produce each of the SUMO1/2/3 variant transcripts, but our method for amplifying each variant independently in a single qRT-PCR reaction using precisely 100ng of purified RNA is successful (Fig. 7). This methodology serves as the basis for the further characterization of the SUMO1/2/3 prototype and alternatively spliced mRNA transcripts for further downstream applications.
However, to have a true understanding of the precise levels of each transcript within the cell, it was necessary to then develop a true means of quantification. Calibration curves are an essential tool for providing these copy number estimates, here denoted, $CNest$, for each SUMO1/2/3 variant transcript to establish true overall SUMO1/2/3 combined copy numbers as well as the respective proportions between each of the paralogs. As outlined in Methods and Materials, beginning with calculating the predicted weights for each variant, dilutions were carried out through the quantification of T7 transcribed RNA that indicated the amount of each dilution required for subsequent qRT-PCR reactions run in triplicates. The quantification of the SUMO1/2/3 variants reveals an R-value close to 1, giving a consistently greater range of change within the orders of magnitude present within the serial dilutions. Gel electrophoresis was performed as a last step quality check (data not shown). The generated calibration curves then made it feasible to detect the exact copy number of each variant found within A549, HEK293, and Calu-3 cells, all of which are immortalized cell lines. However, to avoid the detection being based on this characteristic alone, total RNA was purified from PBM cell lines (Fig. 8).
Figure 8- SUMO1/2/3 mRNA Normalcy Levels Across Four Mammalian Cell Lines

Cloning of the SUMO1/2/3 variants to establish calibration curves reveal the alternatively spliced SUMO1/2/3 variants are all produced within the eukaryotic cell in different amounts compared to their respective prototypes. Quantification of all SUMO1/2/3 standard levels in four cell lines using triplicate purifications was obtained from A549, HEK293A, Calu-3 and PBMC cell lines using three separate RNA purifications all ran in triplicate qRT-PCR reactions using MyGo machinery. A549 cell SUMO1/2/3 variant levels indicate this cell line with the highest presence of SUMO1var2 as well as SUMO2var2. Calu-3 SUMO1/2/3 levels exhibit the highest presence of SUMO3var2. To avoid transcript readouts being based solely in immortalized cells, Peripheral Blood Mononucleated Cells obtained from three separate live donors reveal the least amount of SUMO3var2 but the highest levels of SUMO1var3.

The purified RNA from the four mammalian cell lines was used to carry out final one-step qRT-PCR reactions ran on MyGo machinery in triplicates using precisely 100ng of mRNA per reaction. The calibration curve readouts indicate the Hek293A cell contains 19x10^6 total SUMO transcripts, the highest amount amongst the compared cell lines. The A549 cell line readouts indicate 11x10^6
total SUMO transcripts, followed by Calu-3 with $9 \times 10^6$ total SUMO transcripts. Lastly, the PBM cell, the only model derived from live donors exhibits the lowest of total SUMO transcripts with $2 \times 10^6$ CNest (copy number estimates). Across all four cell lines, SUMO2 represents the most SUMO paralog transcripts, with the HEK293A cell line exhibiting the highest SUMO2 levels constituting 90% of SUMO transcripts. Of the alternatively spliced transcripts encoding the SUMO1/2/3 protein isoforms, SUMO1var3 levels, the transcript coding for the isoform, SUMO1$\alpha$, is present at almost a 10-fold lower than that of SUMO1var1 and SUMO1var2. While SUMO2var1 and SUMO2var2 follow a similar trend, the same is not true for SUMO3$\alpha$’s transcript which represents 15-38% of total SUMO3 in the cell, the only alternatively spliced variant predicted to be conjugatable. It is also noteworthy that SUMO2/3 variants levels in PBM cells are present in the lowest amounts except for SUMO1var3, the only cell line that isn’t exclusively immortalized. Another observation is regardless of cell sample origin is the transcripts encoding for each of the alternative isoforms represent the lowest levels. Therefore, by looking at the total pool of SUMO transcripts as well as the respective ratios of each SUMO1/2/3 paralog gives a sense of total SUMO contribution as a system and by individual paralog giving further proof that these alternatively spliced variants found on the NCBI database could contribute differential functional effects to not only cellular responses but towards the regulation of the SUMO system itself. It is on this basis that further characterization was needed on the behavior of these SUMO1/2/3 transcripts.

The cloning of the SUMO1/2/3 variants to establish calibration curves most importantly revealed the alternatively spliced SUMO1/2/3 variants are all produced within the eukaryotic cell in different amounts compared to their respective prototypes. In pursuit of the characterization of the alternatively spliced transcripts in accordance with previous studies, our validated quantitative...
approach for detecting levels of all SUMO1/2/3 variants was then used for testing the response to stressors all known to trigger global increases in SUMOylation to assess any differences in stress response amongst the alternatively spliced transcripts (Fig. 9). We hypothesize to mitigate cellular stress, that SUMO gene expression profiles are altered. More specifically, the transcripts coding for the prototypical SUMO1/2/3 proteins would be upregulated, while the transcripts encoding for the SUMOα isoforms would be downregulated in response to the various stressors. Three separate stress treatments: heat shock at 43°C for 1-hour, cold shock at 27°C for 24 hours, and H1N1 infections for 12 hours were repeated three times to obtain purified RNA used in triplicate qRT-PCR reactions using 100ng of purified mRNA. It is important to note that all purified RNA species were verified by gel electrophoresis to ensure the quality of purification. Control transcripts were chosen based on their known documented responses to each respective treatment. The influenza Matrix gene was chosen due to late expression upon infection demonstrating an average of 256-fold increase indicative of successful uptake and production of viral particles in both A549 and HEK293A cells (Fig. 9A/B). RBM3 was the gene used to ensure the cold shock stimulated the appropriate cellular response at 27°C for 24 hours which showed a 4-fold upregulation compared to the control treatment (Fig. 9A/B). Lastly, the HSP70 transcript was used for heat shock which showed an average 7-fold increase upon treatment in both A549 and Hek293A cell lines (Fig. 9A/B).

Changes in the relative abundance of the SUMO variants were triggered by the three aforementioned types of stress. Total RNA was purified and probed for the CNeat of each variant. Untreated cell cultures were plated at equal cell densities and maintained for equal time frames as forms of controls. Changes in the relative abundance of each SUMO variant were calculated by subtracting the control from the stressed CNeat. All data are shown on the log2 scale which
represents the average values obtained from the triplicate measurements in three independent experiments.

**Figure 9- SUMO1/2/3 Variants Transcriptomic Profile Upon Cellular Stress**
Transcripts coding for alternative SUMO1/2/3 isoforms upon exposure to various stressors using triplicate purifications obtained from A-C: Control Transcripts known to be upregulated for each treatment show each treatment did induce the correct cellular response. D: A549 and E: HEK293A cell lines were stressed using three separate RNA purifications all run in triplicate qRT-PCR reactions using MyGo machinery. F: SUMO1/2/3 transcript responses to SARS-CoV-2 infection in Calu-3 cells. Y-axes indicate the log2 fold change of each SUMO1/2/3 transcript using generated calibration curves. X-axes are indicative of each stress treatment. G-I: Heat maps illustrating the extent of differential SUMO1/2/3 gene expression in response to stress. Representation of averaged control and stress conditions to illustrate the SUMO1/2/3 proportions in a cell. SUMO3var2 represents the transcript with the least abundance, however, this transcript experiences the most variation upon each stress treatment. Analyses were done based on 100ng of RNA.

Upon 12-hour Influenza A viral infection, SUMO1var1 exhibited a 2-fold upregulation in A549 cells (Fig. 9C) and a significant 3.5-fold upregulation in the Hek293A cell lines (Fig. 9D). Upon cold shock, the A549 cell line showed a 2-fold significant upregulation and a 3-fold upregulation in HEK293A cells. Heat shock induced a 1.2-fold downregulation in both cell lines. SUMO1var2 showed 1.5 downregulation upon viral infection in both cell lines, while cold shock induced an average of 2.1 upregulation. Heat shock, however, induced a 1.2-fold downregulation of the SUMO1var1 transcripts. SUMO1var3 showed a significant 2.7-fold decrease upon viral infection in A549 cells and a 1.8-fold decrease in the HEK293A cell line. Cold shock induced a 2.1-fold upregulation in A549 cells and a significant 3.4-fold upregulation in the HEK293A cell line. Lastly, heat shock treatments showed a 1-fold upregulation in both cell lines. SUMO2 is the first SUMO gene whose two variants show different profiles in the two cell lines. SUMO2var1 showed a significant 2.4-fold upregulation upon H1N1 infection in A549 cells and a 1.3-fold downregulation in the HEK293A cell lines. Cold shock induced a 1-fold upregulation in A549 and a 1-fold downregulation in HEK292A cells. However, heat shock treatments caused a 1-fold downregulation in both cell lines. SUMO2var2 was upregulated 1.3-fold upon infection in A549 cells but caused a 1-fold decrease in HEK293A cell lines. Cold shock again caused the opposite regulation effects in each cell line; A549 cells experienced a 1.1-fold downregulation whereas HEK293A cells showed a 1.1-fold upregulation. Heat-shock treatments were the only ones that caused a similar trend in both cell lines exhibiting a 1.2 upregulation in SUMO2var2 transcripts.
SUMO3var1 transcripts were upregulated 1.8-fold upon H1N1 infection in A549 cells but exhibited a 1.2-fold decrease in the HEK293A cell line. Cold shock induced the opposite 1.4-fold effect in each cell line, A549 cells showed upregulation and HEK293A showed a downregulation. Heat shock showed a 1-fold upregulation in A549 cells and a 1-fold downregulation in HEK293A cells. SUMO3var2 transcripts exhibited a 2-fold increase upon viral infection in A549 cells and a 2-fold decrease in HEK293A cells. Cold Shock induced a 1-fold decrease in A549 cells and a 3-fold decrease in the HEK293a cell line. Lastly, heat shock-induced the same 1-fold downregulation in both cell lines.

Although the SUMO1 transcripts appear to have similar responses to the respective treatments across both cell lines, the same does not hold true for the SUMO2 and SUMO3 transcripts who in fact exhibit opposite profiles in response to the stressors. Therefore, these trends show differences in SUMO gene expression profiles in cell lines with different identities. The previous data indicates the individual transcripts’ response to different stressors, further data analyses were carried out by considering the overall SUMO1/2/3 system response to the respective stress treatments. An important first observation is the total pool of SUMO transcripts in each cell goes from $1.5 \times 10^6$ to $1 \times 10^6$ copy number estimates upon Influenza A infection. However, the SUMO system increases at the transcript level by 5 million copies upon a 24 cold shock treatment, and an average increase of 1 million copies upon a 1-hour heat shock treatment (Fig (E/F)). Upon heat shock, there appear to be no significant changes within the SUMO1/2/3 groups further adding to our postulation of SUMO acting as a transcriptional modifier as opposed to at the protein level in response to cellular insults given the respective pathways mechanisms. In conclusion, these data are not consistent with our initial hypothesis as the stresses tested didn’t appear to consistently upregulate the transcripts encoding for the prototypical proteins nor downregulate the transcripts encoding for
the SUMOα isoforms. Instead, the SUMO gene expression profiles are altered in a stress-dependent and cell-specific manner. Therefore, the various SUMOα isoforms might exist at very different concentrations in the cell, and alternative splicing might contribute differently toward regulating the cellular levels of each SUMO protein perhaps with different proteins and isoforms communicating through crosstalk for recognition and management of specific cellular insults.

Given the impact of the current pandemic, it was important to characterize the SUMO1/2/3 stress responses upon two distinct viral infections to observe if the stress responses upon Influenza A infection are conserved. The literature now shows SARS-CoV-2 has been reported to block certain aspects of the cellular SUMOylation response such as the activating enzyme, SAE1, as well as a “dysregulation” of other genes connected to the SUMOylation pathway (146). The SUMOylation system was also implicated in aiding the formation of viral ribonucleoprotein complexes and nucleocapsids (147). Furthermore, it is now proven that SARS-CoV-2 impairs host protein expression to suppress interferon response (148). The interferon gene was used as a control transcript upon SARS-CoV-2 infection based on immune response levels showing a 45-fold change compared to the control transcript (Fig. 10A).
**A. Stress Control Transcript in Calu3 Cells**

**B. SUMO1/2/3 Response to CoV2 Infection in Calu3 Cells**

**C.**

<table>
<thead>
<tr>
<th>Calu-3</th>
<th>CoV-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1V1</td>
<td></td>
</tr>
<tr>
<td>S1V2</td>
<td></td>
</tr>
<tr>
<td>S1V3</td>
<td></td>
</tr>
<tr>
<td>S2V1</td>
<td></td>
</tr>
<tr>
<td>S2V2</td>
<td></td>
</tr>
<tr>
<td>S3V1</td>
<td></td>
</tr>
<tr>
<td>S3V2</td>
<td></td>
</tr>
</tbody>
</table>

**D.**

<table>
<thead>
<tr>
<th>CELLS</th>
<th>mRNA</th>
<th>IAV</th>
<th>27C⁺</th>
<th>43C⁺</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S1V3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S2V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3V1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S3V2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 10- SUMO1/2/3 Variants Transcriptomic Profile Upon SARS-CoV-2 Infection
Transcripts coding for alternative SUMO1/2/3 isoforms upon exposure to SARS-CoV-2 infection using triplicate purifications ran in triplicate qRT-PCR reactions using precisely 100ng total RNA in both infected and non-infected. A: Control Transcript known to be upregulated upon viral infection indicates a proper cellular response. B: SUMO1/2/3 transcript responses to SARS-CoV-2 infection in Calu-3 cells. Y-axes indicate the log2 fold change of each SUMO1/2/3 transcript using generated calibration curves. X-axes are indicative of each alternatively spliced SUMO1/2/3 transcript response. C: Heat map illustrating the extent of the differential SUMO1/2/3 transcripts response to SARS-CoV-2 infection in comparison to D: Heat maps illustrating the extent of differential SUMO1/2/3 gene expression in response to stressors known to induce global levels of cellular SUMOylation.

Upon 12-hour Influenza A viral infection, SUMO1var1 exhibited an average of 2.3-fold upregulation in A549 cells and Hek293A cell lines, however, upon SARS-CoV-2 infection in Calu-3 cells this transcript was downregulated 1.4-fold. SUMO1var2 showed a 1.5-fold downregulation upon H1N1 infection however this downregulation was less upon SARS-CoV-2 with a 1-fold change (Fig. 10B). SUMO1var3 showed on average a 2.2-fold downregulation upon H1N1 infection, however, upon SARS-CoV-2 infection, this changed to a 1.1-fold upregulation in Calu-3 cells. SUMO2var1 transcripts showed an upregulation upon H1N1 in A549 and downregulation in HEK293A cells, however, upon SARS infection in Calu-3 cells, there was a 1.2-fold downregulation, like that of the HEK293A infection profile. SUMO2var2 was upregulated in A549 but downregulated in HEKs which was the same profile observed upon SARS-CoV-2 infection. SUMO3var1 had a strong upregulation in A549 and a slight downregulation in HEK293A, this downregulation was further upon SARS-CoV-2 in Calu3 cells. Lastly, while A549 exhibited a 2-fold upregulation upon H1N1 infection, there was a 2-fold downregulation in HEK293A. This response was not conserved upon SARS-CoV-2 which exhibits a 1-fold upregulation. Interestingly these data indicate the transcripts encoding for the SUMO1α and SUMO3α isoforms are the only ones that exhibit some degree of increase, therefore these data ultimately reveal a type of perturbation in the presence of SARS-CoV-2 in the presence compared to the normal SUMO1/2/3 cellular stress response corroborating with the current studies of this
virus’ ability to directly engage host RNAs to dysregulate essential steps including splicing and
dysregulate a highly conserved cellular stress response (148).

Given the nuances observed in the changes in transcript levels in response to stress, it was
necessary to observe nuclear retention to further characterize the relevance of the SUMO1/2/3
variants. The translocation of proteins in and out of the nucleus induces abundant biochemical
changes such as the entry of nuclear proteins for transcription induction, while their export to the
cytoplasm quells the transcription of target genes (149). Hence, these procedures endure as
fundamental methods to assist in the analysis of explicit intracellular occurrences and portrayal of
protein tasks (150). Therefore, subcellular fractionation was performed to determine whether the
variants are potentially translated in both A549 and HEK293A cell lines using Norgen Biotek’s
Cytoplasmic and Nuclear Fractionation Kit as outlined in Methods in Materials (Fig. 11). qRT-
PCR analyses using strictly 100ng of purified RNA confirmed the existence of all the variants in
both isolated nuclear and cytoplasmic extracts in both cell lines (Fig. 11). The control transcripts
used in these studies show a significant inverse relationship characteristic of transcripts known to
be abundantly expressed in each respective fraction indicative of successful fractionation
procedures (Fig. 11A/B). The U2 gene was chosen as the nuclear transcript control due to its
transcriptomic role in the U2 component of the spliceosome machinery indicating an 80% nuclear
presence in both A549 and HEK293A cell lines (Fig. 11A/B). As a cytoplasmic control, the S14
gene was selected due to comprising the transcriptomic portion of the 40S ribosomal protein
known to be abundantly expressed in the cytosol which exhibits a 90% cytosolic presence using
our fractionation technique. As a last measure control to verify sufficient fractionation, samples
were lysed for western blot and probed for GAPDH which shows a strong profile only within the
cytosol and not the nuclear fraction (Fig. 11C).
A. Fractionation Control Transcripts in A549 Cells

<table>
<thead>
<tr>
<th>Fraction</th>
<th>%Nuclear</th>
<th>%Cytosolic</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1V1</td>
<td>68.5%</td>
<td>31.5%</td>
</tr>
<tr>
<td>S1V2</td>
<td>22.8%</td>
<td>77.2%</td>
</tr>
<tr>
<td>S1V3</td>
<td>57.1%</td>
<td>42.9%</td>
</tr>
<tr>
<td>S2V1</td>
<td>39.7%</td>
<td>60.3%</td>
</tr>
<tr>
<td>S2V2</td>
<td>16.9%</td>
<td>83.1%</td>
</tr>
<tr>
<td>S3V1</td>
<td>21.5%</td>
<td>78.5%</td>
</tr>
<tr>
<td>S3V2</td>
<td>63.5%</td>
<td>36.5%</td>
</tr>
</tbody>
</table>

B. SUMO1/2/3 Transcript Levels in Nuclear/Cytosolic Fractions in A549 Cells

C. Table of SUMO1/2/3 Transcript Levels in Nuclear/Cytosolic Fractions in HEK293A Cells

<table>
<thead>
<tr>
<th>Fraction</th>
<th>%Nuclear</th>
<th>%Cytosolic</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1V1</td>
<td>64.3%</td>
<td>35.7%</td>
</tr>
<tr>
<td>S1V2</td>
<td>25.6%</td>
<td>74.4%</td>
</tr>
<tr>
<td>S1V3</td>
<td>55.0%</td>
<td>45%</td>
</tr>
<tr>
<td>S2V1</td>
<td>33.3%</td>
<td>66.7%</td>
</tr>
<tr>
<td>S2V2</td>
<td>19.2%</td>
<td>80.8%</td>
</tr>
<tr>
<td>S3V1</td>
<td>19.5%</td>
<td>80.5%</td>
</tr>
<tr>
<td>S3V2</td>
<td>91.8%</td>
<td>8.2%</td>
</tr>
</tbody>
</table>
Figure 11- Detection of SUMO1/2/3 Variants in Cytoplasmic Fractions Provides Hints of Translational Roles

A/B: Fractionated RNA from three respective purifications of A549 and Hek293A cells show all SUMO1/2/3 variants are exported out of the nucleus in different amounts. The alternatively spliced SUMO1/2/3 transcripts that encode SUMO1/2/3 isoform proteins are underlined in red. Statistics were performed using a Two-Way Anova with p-values = .05 and above being considered significant and graphed.

C: Table summarizing percentages of each SUMO1/2/3 variant present in respective nuclear/cytosolic fraction.

D/E: Control Transcripts for genes known to have a strong presence in respective nuclear/cytosolic fractions in both A549 and Hek293A cell lines.

F: Agarose gel electrophoresis confirmation of RNA from three respective purifications of A549 and Hek293A cells are sufficiently fractionated.

While SUMO1var1 and SUMO1var2 code for the same protein, their nuclear/cytosolic RNA level profiles are contrasting displaying significant disparities in each fraction. This is a peculiar observation given that at the transcript level their only differences are in the 5’ UTR in which the fifth exon is spliced into two for SUMO1var2. However, SUMO1var the transcript encoding for the SUMO1α isoform is present in relatively equal amounts in the respective fractions. When comparing SUMO2 and SUMO3 fractionated RNA levels, their profiles are also contrasted, however, SUMO2var2 is the transcript that is present within all four cell lines at the lowest levels ironically is the transcript with the highest cytoplasmic levels which could be indicative of a potential strong amplificatory reach. SUMO3 on the other hand has a strong nuclear presence, the highest of all transcripts in Hek293A cells.

Given the clear distinctions of the alternative SUMO isoforms at both the transcription level in response to various stressors, a final subcellular fractionation was used to assess whether the SUMO1/2/3 transcripts nuclear export profiles shift upon 24-hour cold shock in A549 and HEK293A cell lines, the stress treatment that caused the most significant change (Fig. 12).
Figure 12 - The SUMO1/2/3 alternative transcripts nucleo-cytosolic profiles shift at 27°C

Sub-cellular fractionation was carried out at 27°C and 37°C as a control. U2, S14, and RBM3 gene transcripts were also used as additional controls. Red circles are indicative of nuclear fractions, green circles are indicative of cytosolic transcripts. Larger areas indicate an upregulation upon cold shock incubation at 27°C. Overall SUMO transcripts increase by two million upon cold shock in both A549 and HEK293A cell lines.
SUMO1var1 which under standard conditions is 68% nuclear and 32% cytosolic, however, upon cold shock, an extra 14% of total SUMO1var1 transcripts were exported in HEK293A cells (Fig. 12A). SUMO1var2 transcript experienced an increased export of 7% in A549 cells, however, SUMO1var3, the transcript encoding for the SUMO1α protein isoform, experienced no nuclear export fluctuations upon cold shock treatment (Fig. 12A). SUMO2var1 exhibited a 7% increase in export in the HEK293A cell line, while SUMO2var2 showed a 9% upregulation in nuclear export (Fig. 12A). SUMO1var1 exhibits a 5% increase in export upon cold shock in the A549 cell line while SUMO3var2 showed the most significant 20% increase in export in both cell lines. While these results examine how the individual SUMO1/2/3 transcripts respond to cold shock treatment, it is important to note how the SUMO system reacts as a whole (Fig. 12B). The reddish hues represent the nuclear portion, and the greenish hues are all cytosolic transcripts with the legends indicative of which transcript contributes the most. When comparing the 37°C and 27°C treatments, there are significant changes. As the areas increase in both cases, the total number of transcripts is increased upon cold shock, indicating that there is some degree of regulation at the transcriptional level. While this upregulation in transcripts is not significant, it does provide a basis for the total SUMO1/2/3 contribution to the overall increases observed. It also makes the point of increased cytosolic pools and decreased nuclear pools overall, which is an additional contributor to the increase in global cellular SUMOylation observed. These data ultimately reveal the control mechanisms: 1) changes in the nucleocytoplasmic distribution of transcripts, 2) changes in total transcript numbers, and 3) changes at the level of specific SUMO transcript variants.

The establishing of alternatively spliced SUMO1/2/3 cytoplasmic transcripts and these indications that those pools appear to be under some level of control indicate they may be playing
a regulatory role, as well as potential other roles at the level of translation potentially co-existing with the prototypical SUMO1/2/3 proteins.

<table>
<thead>
<tr>
<th>Variant</th>
<th>Pattern searched</th>
<th>Sequence ID</th>
<th>#matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMO1V3</td>
<td>5' ATCATGTCTGACCAGGATAGCAGTGAGATT 3'</td>
<td>SRR14200578</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200579</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR13200580</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200582</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5' TCATGTCTGACCAGGATAGCAGTGAGAT 3'</td>
<td>SRR14200578</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200579</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200580</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200582</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5' CATGTCTGACCAGGATAGCAGTGAGA 3'</td>
<td>SRR14200578</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200579</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200580</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200582</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td><strong>Total hits &amp; Likely independent hits</strong></td>
<td>18 &amp; 6</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variant</th>
<th>Pattern searched</th>
<th>Sequence ID</th>
<th>#matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMO2V2</td>
<td>5' TATGTGAACGACAGTTGAAATGGAGGAT 3'</td>
<td>SRR14200581</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5' ATGTGAACGACAGTTGAAATGGAGGA 3'</td>
<td>SRR14200581</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5' TCATGAGAGGCAGTTGCGGACCTTGGCT 3'</td>
<td>SRR14200580</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200582</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td><strong>Total hits &amp; Likely independent hits</strong></td>
<td>5 &amp; 3</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variant</th>
<th>Pattern searched</th>
<th>Sequence ID</th>
<th>#matches</th>
</tr>
</thead>
<tbody>
<tr>
<td>SUMO3V2</td>
<td>5' TACTGCAGAGGCGAGGTCGGCACTTGGCT 3'</td>
<td>SRR14200580</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200582</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5' ACTGCAGAGGCGAGGTCGGCACTTGC 3'</td>
<td>SRR14200580</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200582</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5' CTGCAGAGGCGAGGTCGGCACTTGG 3'</td>
<td>SRR14200580</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200582</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5' ATGCAGCGCCCGAGGCGCTTGCAATGAGG 3'</td>
<td>SRR14200581</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200582</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5' TGCAGCGCCCGAGGCGCTTGCAATGAG 3'</td>
<td>SRR14200581</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200582</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>5' GCAGCGCGCCCGAGGCGCTTGCAATGA 3'</td>
<td>SRR14200581</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td></td>
<td>SRR14200582</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td><strong>Total hits &amp; Likely independent hits</strong></td>
<td>15 &amp; 5</td>
<td></td>
</tr>
</tbody>
</table>

Studies giving no hits for the sequences shown above: SRP122522, SRP362491, and SRP286677.

Figure 13- Ribosome-Sequencing Confirmation of SUMO1/2/3α’s Translation
The SUMOα-specific sequences were used to search publicly available Ribo-seq datasets. Sequences underlined are present in the downstream exon, whereas the sequences not underlined are in the upstream exons.

Ribosome profiling is now considered a powerful tool for the global monitoring of in vivo translation through deep sequencing of ribosome protected mRNA fragments (151). This methodology is a proven technique in the sense given ribosome associated RNAs are cross-linked with the ribosomes and fractionated away from free mRNA. The sample is treated with RNase to chew away all mRNA not directly protected by a ribosome. Once the RNase is inactivated, the cross-links are broken, the RNA leftover is purified away from the ribosomes, and the mRNA sequences left are corresponding to ribosome-protected sequences. The SUMOα-specific sequences were used to search publicly available Ribo-seq datasets (Fig. 13). Sequences underlined are present in the downstream exon, whereas the sequence not underlined is in the upstream exon. Provides evidence that sequence-specific mRNAs coding for the SUMOα’s was protected by a ribosome hence it was being translated and therefore produced in the cell.

Given the above confirmation of the in vivo translation of the SUMO1/2/3α isoforms, investigations of the SUMO1/2/3α isoforms then began at the informatic level to assess differences in tertiary structures that are essential for SUMO conjugation (Fig. 14). Structurally, because the second exon of SUMO1var3 is spliced out, the mature mRNA is predicted to code for a shorter version of the protein missing a substantial segment of the n-terminal region, spanning the first beta-strand and part of the second beta-strand. Therefore, this novel SUMO1 isoform could translate into a distorted protein that is lacking a major conjugation site, thus altering its function. Alpha-fold predicts the very heart of the SUMO2α core is severely disrupted in one of the alphahelical structures of the beta-strands are completely gone while the second is partially gone due to the splicing out of the third exon.
Figure 14 - Artificial Intelligence Predicted Functional Properties
Artificial intelligence predicts the deletions generated by the alternative splicing of the SUMO1/2α transcripts to contain disruptions in their respective tertiary structures affecting the central globular structure and therefore their functionality. While the intronic inclusion that occurs in the SUMO3α transcript is predicted to not affect the structure, therefore, predicted to be conjugatable.

Lastly, the server predicted SUMO3α to be the only one that is conjugatable given it is the only SUMO gene whose alternatively spliced variant involves one intronic inclusion. It is important to note that while the alternative isoforms still contain the alpha helix, the changes in beta strands do affect the central globular structure and in theory, functionality.

Given the Alpha-fold functionality predictions in conjunction with SUMO1/2/3 detection within the cytoplasm, functional characteristics were then investigated via the cloning of all SUMO1/2/3 protein isoforms into a His-Tagged as well as a YFP-Tagged vector. The clones were then overexpressed in their mature form (diglycine motif at C-end) by transfection in a mammalian cell line to determine if conjugation to endogenous SUMO targets by immunoblot analyses as outlined in Methods and Materials occurs (Fig. 15). These data confirmed the prediction of SUMO3α being the only alternatively spliced SUMO1/2/3 isoform that is capable of conjugating to cellular substrates (Fig. 15D Lane 7). SUMO1/2α are detected in free form but in low amounts indicative of instability with potential regulatory roles at the proteasomal degradation level. This postulation is further confirmed in Panel E Lanes 3 and 5 by YFP-Transfection resulting in a significantly increased signal that appears to stabilize the alternatively truncated SUMO1/2 proteins. One interesting and peculiar observation to be made is the informatic predictions (Fig. 6B) show the normal prototype SUMO1 vs SUMO1α amino acid sequences are lacking five lysine residues resulting in a complete lack of conjugation abilities for SUMO1α. However, the same is not true for SUMO2 and SUMO2α whose amino acid alignments show no missing lysine residues, yet SUMO2α is un-conjugatable indicative of SUMO2 conjugation occurring through a non-consensus motif.
Figure 15- Characterized Functional Properties of the SUMO1/2/3α Isoforms
A: To determine whether the SUMO12/3α can conjugate to cellular substrates, plasmids coding for SUMO1/2/3His-tags.
B: SUMO1/2/3 YFP-tags were transiently transfected into Hek293A cells and 48h post-transfection the cells were collected and SUMOylation assessed by western blot on a 10% SDS-Page Gel using anti-S Tag antibodies and GAPDH as a control for loading normalization.
Given the confirmed differences in transcript levels in conjunction with their export out of the nucleus, it was then necessary to investigate another major facet of the SUMOylation system, localization. We hypothesized that given the SUMO3α-isoform exerts similar functional effects to that of the SUMO3 prototype, then the SUMO3α-isoform should exhibit the same cellular localization. We began by transiently transfecting Hek293A cells expressing SUMO1/2/3 and SUMO1/2/3α respectively and then subjected to fluorescence confocal microscopy to detect YFP-fusions of SUMO (green channel) and Phalloidin (actin filaments red channel), and DAPI (DNA) staining from three representative fields of cells post resulted in SUMO1/2/3α localization levels appearing to be significantly altered in contrast to their respective prototypes (Fig. 16). The profile for SUMO1 indicates typical localization to the nuclear membrane as corroborated by Su et.al 2002 (63) showing a dotted pattern in the nucleus, however, the signal for SUMO1α exhibits a more diffused fluorescence in the nucleus as well as cytoplasmic dots (Fig. 16A). The signals for SUMO2 are also strikingly different, the SUMO2 prototype shows a dotted pattern in the nucleus while SUMO2α transfected cells exhibit a dominantly diffused nuclear and some diffused cytosolic signal (Fig. 16C). Lastly, while SUMO3 proteins have been typically documented to localize to the cytoplasm, in our artificial system, there is a strong dotted and diffused nuclear signal with no signal in the cytoplasm (Fig. 16E). In contrast to SUMO3α transfected cells exhibit dotted signal in the nucleus but no diffused fluorescence in the nucleus and show some diffused fluorescence in the cytoplasm (Fig. 16F). Further co-localization assays will be needed to determine the precise localization of the SUMO1/2/3 proteins and their respective α-isoforms.
Figure 16- SUMO1/2/3α His-Tagged Plasmids Affect SUMO1/2/3 Localization
Transiently transfected Hek293A cells expressing SUMO1/2/3 and SUMO1/2/3α respectively and then subjected to fluorescence confocal microscopy to detect the YFP-fusions of SUMO (green channel) and Phalloidin (actin filaments red channel), and DAPI (DNA) staining from three representative fields of cells. A/B) SUMO1 and SUMO1α localization levels. C-D) SUMO2 and SUMO2α localization levels. E-F) SUMO3 and SUMO2α localization levels appear to be strictly nuclear showing nuclear diffusion.

Collectively, the proven viability of the SUMO isoform transcripts, their higher gene expression upon cold shock, and preferential protein localization within the nuclear compartment adds to our characterization of these SUMO isoforms as potentially indispensable components for nuclear machinery.
5. Conclusions

The differences in transcript proportions in standard and stressed cells, altered levels of transcripts in the cytoplasm vs nucleus, the changes in functionality effects exhibited by SUMO3α’s ability to conjugate, and overall differences in localization have not been previously reported. Our primer design provides the specific targeting of the distinct regions to quantify the SUMO1/2/3 variants independently indicative of a successfully developed experimental approach. Moreover, the processing of the SUMO1/2/3 transcripts with 5’ capping and 3’ polyadenylation of the transcripts to achieve nuclear export implies that the SUMO1/2/3 variant mRNA species are not merely aberrant products of splicing. We have determined in a given cell, the transcript encoding for the prototype of each SUMO1/2/3 protein is the transcript with the largest presence. Amongst the respective SUMO paralog groups, the SUMO2 prototype transcript (SUMO2var1) represents on average 83% of all SUMO transcripts and 99.4% of all SUMO2 transcripts. The remaining 0.6% of SUMO2var2 transcripts could explain why Matunis et al. 2021 still observed substantial SUMO2 levels observed in their knock-out cell lines (152).

Our approach offers a novel examination at SUMO1/2/3 expression levels using copy number variance estimates indicating these transcripts are extremely sensitive to multiple stressors across multiple cell lines triggering significant changes. Of the transcripts encoding the alternative protein isoforms, our data indicates SUMO3var2 behaves more contrarily than SUMO2var2 and SUMO1var3 in the sense that there are more SUMO3var2 transcripts signifying a larger presence of the SUMO3 isoform, SUMO3α. However, only SUMO3var2 transcripts undergo significant changes upon viral infection. Our lab has previously shown SUMOylation regulates the activity of various Influenza viral proteins by altering the normal proportion of SUMOylated vs. non-SUMOylated viral proteins and is largely upregulated during influenza infections. More
importantly, **large increases in global cellular SUMOylation (~3 fold and higher) have been proven to block influenza viral replication.** Given our data’s overall response to stress, we now hypothesize the main mechanism regulating the increase in SUMOylation during IAV infection must be translational or post-translational. Due to these critical roles played by the SUMO1/2/3 proteins in cellular homeostasis and our confirmed differences in localization, it is advantageous for viral proteins to preferentially become SUMOylated to circumvent proteasomal degradation compared to being occupied by ubiquitin considering they share consensus sites. In contrast to heat shock conditions which result in serious misfolding of proteins, therefore upregulating the SUMO1/2/3 levels at the protein level could then be the band-aid that prevents further degradation upon stress through the increase of solubility.

According to the literature, the cellular heat and cold shock responses target respectively different molecular pathways. Cellular cold shock responses typically engage in enhancing transcription, whereas heat shock cellular responses typically target protein folding. It has been shown that the SUMO1/2/3 proteins exert a stronger effect on transcriptional machinery by ensuring a fast and effective assembly of the transcriptional complex. Therefore, a more pronounced effect in SUMO1/2/3 gene transcription would be evident in the stress response pathway that focuses on transcript fidelity. Based on our data trends, the increase in fold gene expression occurs upon cold shock treatment, however, heat shock treatment exhibits a decrease in SUMO1/2/3 transcription. Despite the stark differences in gene expression, our internal controls indicate transcription activity in both conditions, thus the transcriptomic profile of each treatment is specific to the activation or inactivation of genes necessary to mitigate a proper cellular response. This is indicative of SUMO’s vital role in transcriptional machinery assembly and regulation, which further corroborates with viral infection where SUMO has been documented to conjugate
to the RDRPs thus facilitating further viral replication. Although SUMO is considered a post-translational modifier, perhaps, its role in transcription is more vital to the cell as a modulator as opposed to the typically thought protein stabilizer.

A more peculiar observation is the nuclear-cytosolic export profile of SUMO1var2 given this transcript is the only one that doesn’t code for a distinct protein isoform, yet this variant has the highest nuclear export levels as well as significant responses to each respective treatment in both cell lines indicating a conserved transcriptomic cellular stress response. Therefore, this production of this transcript could be to contribute to regulatory functions as shown by the widespread analyses indicating as many as half of all fluxes in mRNA quantities in response to stress are caused by fluctuating rates of decay (28). Moreover, microarray experiments have revealed that 40–50% of gene expression fluctuations in response to cellular signals occur at the level of the mRNA stability (29). Nevertheless, the detection of all SUMO1/2/3 variant transcripts in different amounts in both the cytoplasmic and nuclear RNA fractions also indicates the SUMO system could be regulated at more than just the post-transcriptional level as shown by the increase in SUMO3var2 nuclear export at 27°C. Therefore, export of this transcript as well as the other isoform transcripts then change according to the needs of the cells to ultimately generate isoform versions of the respective SUMO1/2/3 proteins. It is proposed here the altered states of hydrophobicity given the differences in their tertiary structures could be further facilitating the cellular stress response. Altered functional proteins do occur within nature such as Ubiquitin B, in which an interpolated frameshift mutation ramifies into a truncated peptide that is lacking the C-terminal glycine (153). The abnormal aggregation of this peptide, known as UBB+1, has further been linked to Alzheimer's disease (153).
Until now, tools allowing the distinctions of SUMO2 and SUMO3 both functionally and biochemically in vivo have been limited due to the high SUMO2/3 sequence conservation (65). One study found through tryptic digests and MALDI-TOF mass spectrometry analysis of SUMO2 conjugation assays that self-conjugation occurs via Lys11 within a SUMO modification motif and further mutation of lysine 11 in SUMO2 and SUMO3 blocked the assembly of polymeric chains (78). These studies state SUMO2 appears to be a better substrate for chain formation than does SUMO3 indicating the 3-amo acid difference in the N-terminal region appears to be partially inhibitory to the conjugation of SUMO3. Not only have we shown the differences between SUMO2 and SUMO3 at the transcript level but also at the protein level given the fact that we have shown SUMO3α is the only alternatively spliced protein product capable of conjugation which is peculiar given that the extra amino acids within the intronic inclusion contain no lysine residues but are completely lacking in both SUMO1 and SUMO2 proteins. Since our data shows the altered version of SUMO3 proteins are conjugatable, this then raises the question of whether the intronic inclusion codes for a non-consensus SUMO motif not previously characterized and how this SUMO3α isoform contributes essential distinct functional roles to the cellular stress response. These observations also raise the possibility of the α-isoforms forming a competitive substrate affinity as is the case for other post-translational modifications. A simple direct way to test this would be to conduct an in vitro binding competition assay using all the components necessary for SUMO1/2/3 conjugation and test which variant preferably binds to targets.

While the informatic predictions show the normal prototype SUMO1 vs SUMO1α amino acid sequences are lacking five lysine residues resulting in a complete lack of conjugation abilities for SUMO1α. However, the same is not true for SUMO2 and SUMO2α whose amino acid alignments show no missing lysine residues yet SUMO2α is un-conjugatable indicative of SUMO2
conjugation occurring through a non-consensus motif. Additional regulatory mechanisms can be ascertained by the differences in conjugation ability as well as the localization which is indicative of differences in target substrates. The detection of the un-conjugatable SUMO1/2α in free form in low amounts can be potentially explained by the instability of these translated proteins contributing regulatory roles at the proteasomal degradation level. This postulation is further confirmed through the YFP-fusion constructs resulting in a significantly increased signal that appears to stabilize the alternatively truncated SUMO1/2α proteins. This is a peculiar observation is that 70-80% of these transcripts encoding for the SUMO1/2α isoforms are exported, the highest of the seven variants aside from SUMO1var2. However, as stated earlier, given it is considered more evolutionarily efficient for the cell to dictate mRNA localization because mRNA transcripts are capable of being translated tens to hundreds of times in response to local stimuli rather than transporting individual proteins back and forth different cellular compartments (34). These data apply to this innovative notion branded as the “RNA Signature” hypothesis, in which each mRNA transcript contains an individualized set of regulatory features that regulate its transport, localization, and translational control (35).

Considering the cytoprotective effects associated with global increases in cellular SUMOylation, molecules capable of increasing global cellular SUMOylation could be used to induce prophylactic and therapeutic broad-spectrum resistance to influenza infection, trigger pre-adaptive resistance to extreme environmental temperatures, and diminish the tissue damage triggered by acute ischemic events, including heart attacks, strokes, and perioperative ischemic events. Our hypothesis of the alternative SUMO1/2/3 isoforms’ functional capabilities may still play a role in cell survival. These observations point to the mechanism used by alternative SUMO1/2/3 isoforms to inhibit SUMO conjugation by interacting with these SUMO1/2/3
consensus sites on the target SUMO substrate and termination of conjugation chains serving as a dominant-negative inhibitor by establishing competitive reversible interactions with the SUMO1/2/3 interacting motifs of targeted proteins. Once normal SUMO1/2/3 upregulation has occurred upon stress, essentially serving as a “break” on the SUMOylation system could be useful from a therapeutic point of view as well in instances when the SUMOylation system is on overdrive and can no longer be naturally shut down such is the case of uncontrollable growth (cancer) (5). Therefore, the ability to manipulate SUMOylation conjugation in conjunction with chemotherapies might offer a new aspect to cancer treatments as indicated by our SUMO2 data.

We now hypothesize if the SUMO2var1 transcript is expressed 50X times more than SUMO1 and SUMO3 in a given cell type, then artificially altering the proportion of SUMO2var1 vs SUMO2var2 in that cell type could inactivate the SUMOylation system for that cell type, while the effects might be substantially lesser for other cell types. This provides for some degree of specificity that is ultimately not achievable when using global inhibitors of the SUMOylation cascade, such as the current Ubc9- or SAE2/1-targeted chemical inhibitors. This is further justified by Anufrieva et al. 2018 who showed decreased splicing efficiency and global intron retention is a novel stress response mechanism that may promote the survival of malignant cells following therapy.

While SUMOylation regulation presently assumes transcript levels for SUMO1/2/3 do not fluctuate to any significant degree and therefore increases in cellular SUMOylation observed upon stress are merely products of changes in SUMO conjugation/deconjugation driven by fluctuations of specific ligases. We now present two changes to this: 1) it does not acknowledge the existence of alternatively spliced variants for SUMO1/2/3, as this has not been formally reported; and 2) ignores the substantial increases in global SUMOylation triggered by stress require a concurrent
increase in the total levels of the SUMO modifiers in the cell. Clearly, alternative splicing, translational regulation, and feedback from the proteins encoded by the alternatively spliced transcripts play important regulatory roles in the cellular SUMOylation response. Thus, our transcriptomic approach challenges the current dogma of how global cellular SUMOylation levels are regulated, introducing a new theory in which alternative splicing along with the different protein isoforms produced have key functions in the regulation of the cellular levels and activity of the SUMO proteins. It is proposed here that small molecules capable of triggering large (>2 fold) global increases in SUMOylation may exert protective effects stimulating potential crosstalk between the alternative variants and isoforms beginning at the alternative splice level through an indirect negative feedback mechanism providing further insight to the mechanisms regulating global SUMOylation levels in the cell.
6. Limitations of the Study

PBM cells were initially used to simply assess the alternative SUMO variants in a different cell type to ensure the numbers we were seeing in the Hek293A and A549 cell lines weren’t exclusive to that of immortalized cell lines alone, because immortalized cell lines have gene profiles that are not physiologically similar to normal cell lines (74). PBM cell lines are a suitable option because they are nucleated blood cells and were derived from clinical samples taken from healthy patients ideally providing a more selective response to the immune system (75). However, a draw-back to this approach is they are difficult to work with and not sustainable in tissue culture, thus could no stress conditions on this cell type could be conducted to observe any fluctuations in the SUMO1/2/3 variants. A pitfall in the cell types that were used to study stress responses, Hek293A and A549, is they may not be as physiologically relevant due to these cells having disrupted regulatory and senescence pathways (76), ultimately affecting SUMO expression and function. An alternative strategy would be to begin experiments in animal models to achieve a more thorough understanding of our experimental designs. Finally, the collection of samples from different tissues would provide a more thorough assessment of any tissue-specific effects under the different stress conditions posited to increase global amounts of SUMOylation within the cell as seen on the NCBI database.

Based on our Alpha-Fold analyses, we expect SUMO2α to retain its Ubc9- binding ability and therefore inhibit SUMO2/3 chain formation by a competitive mechanism. Thus, a need to assess the half-life of SUMO2α’s using a pulse-chase approach and perform binding competition assays. Include a wider set of cellular stress conditions including time courses as well as tissue-specific samples for both stress treatments and nuclear vs cytosol assay. Further analyses involving the detection and measurement of endogenous SUMO isoform protein expression would also be
needed to complement RNA levels through the use of recombinant protein techniques and generation of antibodies to show the SUMO1/2/3α isoforms are synthesized and analyze stress responses on the protein level.
7. Future Directions

During physiological tension such as oxidative stress, SENP1 upregulation has been documented, consequently decreasing, and disrupting the global SUMO conjugation (154). This disruption often leads to the subsequent ubiquitination of previously SUMOylated proteins for proteasomal degradation; a crucial process for protein-turnover rate to assist in the mitigation of stress (154). Increased expression of these SUMO protease enzymes would thus result in lowered levels of SUMO1/2/3 conjugation including SUMOylation levels within the nucleus. Because the SUMOylation system serves as a histone modifier, most often but not always involved in transcriptional repression (155), SENP1 upregulation during oxidative stress, would therefore also target these SUMO1/2/3 regulated histone modifications, perhaps increasing SUMO1/2/3 expression. This increase in SUMO expression could lead to the increased half-life of cytoprotective proteins through the direct interaction of nascent SUMO1/2/3 proteins with the lysine residues normally found in ambiguous ubiquitin/SUMO consensus sites of substrates, thus decreasing their protein turnover rate during cellular insult. We hypothesize here, in response to stress, increased SENP1 expression likely enhances pro-SUMO1/2/3 maturation making more active SUMO1/2/3 available for conjugation thereby facilitating global SUMOylation and, at the same time on the background of increased SUMOylation, selectively deconjugates, via SENP1 pathways, SUMO1/2/3 from specific substrates proteins causing an increased presence of SUMO1/2/3 localized within the nucleus. Transient overexpression of SUMO1/2/3 would terminate upon the SUMOylation, perhaps at the level of HDACs of histone sites in a stress insult-dependent manner. To mitigate the cytoprotective strategy of the SUMOylation system, SUMO1/2/3 alternative splice variants thus provide the enzymatic diversity from single transcripts. Ultimately, SUMOylation could result in the remodeling of the spliceosome.
arrangement at multiple steps particularly affecting 5’ donor splice sites which in turn would alter the splicing of RNA transcripts through ADAR1 editing, providing a SUMO-dependent enhancement to the output of singular SUMO transcripts. We hypothesize the SUMO system self-regulates the expression of the primary SUMO1/2/3 paralog transcripts through the SUMOylation of specific components, SRSF1 targets, within the spliceosome complex during the alleviation of physiological stress, engaging in a negative feedback mechanism that downregulates the expression of prototypical SUMO1/2/3 paralogs while upregulating the expression of the alternative SUMO variants and their translated isoforms. The putative molecular mechanism responsible for the shift of SUMO expression from the prototypical paralogs to SUMO variants could result from the SUMOylation of spliceosome components in particular the SR proteins who ensure the coupling of with U1 snRNP by RNA polymerase II at the 5’ splice donor site, resulting in the alternative splicing of SUMO transcripts. It would also be necessary to investigate the effects associated with SENP1 transcripts/activity as well as the E3 ligases given their implications in SUMO’s ability to modify and affect RNA editing. Therefore, transcriptomic studies involving the analysis of the SUMO conjugating ligases in conjunction with the SUMO proteases are crucial to assess potential contributing regulation mechanisms at this level.

Evaluations on SUMOα overexpression effects on global cellular SUMOylation triggered by stress would be needed to assess effects exerted by the SUMOα’s on the global increase in cellular SUMOylation triggered by stress. Following the same methodology of subjecting the two mammalian cell lines to different stressors at 8 hours post-transfection to probe for global SUMOylation levels and cellular stress survival. It is also necessary to further investigate the translation of SUMO 1/2/3 variants isoforms given translational regulation could be a major contributor to the regulation of SUMO levels in the cell. A preliminary assessment will involve a
standard sucrose gradient approach to fractionate cell extracts derived from cells grown under normal and stress conditions into monosomal and polysomal (2 or more ribosomes per transcript) fractions (77). RNA will be purified from each fraction to evaluate the distribution of the variants by qRT-PCR. Due to the short length of the coding sequence present in the mature SUMO variants, we expect that there will not be more than 3 ribosomes per transcript at any given point.
8. Supplementary Data

<table>
<thead>
<tr>
<th></th>
<th>V1</th>
<th>V2</th>
<th>V3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sumo1</td>
<td>NM_003352.4</td>
<td>NM_001005781.1</td>
<td>NM_001005782.1</td>
</tr>
<tr>
<td>Sumo2</td>
<td>NM_006937.3</td>
<td>NM_001005849.1</td>
<td></td>
</tr>
<tr>
<td>Sumo3</td>
<td>NM_001286416.1</td>
<td>NM_006936.2</td>
<td></td>
</tr>
</tbody>
</table>

Figure S1- Accession Numbers of each SUMO1/2/3 Variant
<table>
<thead>
<tr>
<th>FRW</th>
<th>Name</th>
<th>Size</th>
<th>Sequence</th>
<th>Location</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1V1</td>
<td>S1V1V2.FW</td>
<td>345bp</td>
<td>5' - ACTCAAAGTCATGGCAAGGATAGCTGAG-3'</td>
<td>Exon1/Exon2 Junction-invariant 1/2, not var3</td>
<td>SUMO1var1/2</td>
</tr>
<tr>
<td>S1V1</td>
<td>S1V1V3.RV</td>
<td></td>
<td>5' - AGAGATGGGGTGCCAGATAGCGATGAG-3'</td>
<td>Spliced out Exon5 intron in variant 2</td>
<td>SUMO1var1</td>
</tr>
<tr>
<td>S1V2</td>
<td>S1V2V2.FW</td>
<td>258bp</td>
<td>5' - ACTCAAAGTCATGGCAAGGATAGCTGAG-3'</td>
<td>Exon1/Exon2 Junction-invariant 1/2, not var3</td>
<td>SUMO1var1/2</td>
</tr>
<tr>
<td>S1V2</td>
<td>S1V2.RV</td>
<td></td>
<td>5' - CCAGACCCTCAAATTTTAAAAACTAAATGGGTAAG-3'</td>
<td>Exon5/Exon6 Junction-Exclusive of variant 2</td>
<td>SUMO1var2</td>
</tr>
<tr>
<td>S1V3</td>
<td>S1V3.FW</td>
<td>188bp</td>
<td>5' - GTCACTCAGTGGACAGATAGCAGTG-3'</td>
<td>Exon1/Exon3 Junction-Exclusive of variant 3</td>
<td>SUMO1var3</td>
</tr>
<tr>
<td>S1V3</td>
<td>S1V1V3.RV</td>
<td></td>
<td>5' - AGAGATGGGTCCGTCATTTCAATGCC-3'</td>
<td>Spliced out Exon5 intron in variant 2</td>
<td>SUMO1var3</td>
</tr>
<tr>
<td>S2V1</td>
<td>S2V1V2.FW</td>
<td>242bp</td>
<td>5' - ACAGAAAGCAGCAGAGGAGTGCAAG-3'</td>
<td>Present in SUMO2 variants</td>
<td>SUMO2var1/2</td>
</tr>
<tr>
<td>S2V1</td>
<td>S2V1.RV</td>
<td></td>
<td>5' - TGTATCTCTCATCTCCATTTGCCACTGTCG-3'</td>
<td>Present in SUMO2 variants</td>
<td>SUMO2var1</td>
</tr>
<tr>
<td>S2V2</td>
<td>S2V2V2.FW</td>
<td>157bp</td>
<td>5' - ACAGAAAGCAGCAGAGGAGTGCAAG-3'</td>
<td>Exon1/Exon2 Junction-invariant 1/2, not var3</td>
<td>SUMO2var1/2</td>
</tr>
<tr>
<td>S2V2</td>
<td>S2V2.RV</td>
<td></td>
<td>5' - TGTATCTCTCATCTCCATTTGCCACTGTCG-3'</td>
<td>Exon2/Exon4 Junction-Present in var2, not var1</td>
<td>SUMO2var2</td>
</tr>
<tr>
<td>S3V1</td>
<td>S3V1.FW</td>
<td>179bp</td>
<td>5' - TGGCGAGAGCGGAGCTGTCAAGAGCCG-3'</td>
<td>Exon2/Exon3 Junction-Exon 2 is smaller in var1</td>
<td>SUMO3var1</td>
</tr>
<tr>
<td>S3V1</td>
<td>S3V1V2.RV</td>
<td></td>
<td>5' - GGCCTCTCTAGAAGACTGTCGGCGCCAG-3'</td>
<td>Within Exon4- Present in SUMO3 variants</td>
<td>SUMO3var1/2</td>
</tr>
<tr>
<td>S3V2</td>
<td>S3V2.FW</td>
<td>239bp</td>
<td>5' - CTGGCTGCTGTGGCAGCGCCATCCC-3'</td>
<td>Within Exon2 intrinsic inclusion event</td>
<td>SUMO3var2</td>
</tr>
<tr>
<td>S3V2</td>
<td>S3V1V2.RV</td>
<td></td>
<td>5' - GGCCTCTCTAGAAGACTGTCGGCGCCAG-3'</td>
<td>Within Exon4- Present in SUMO3 variants</td>
<td>SUMO3var1/2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>FRW</th>
<th>Name</th>
<th>Size</th>
<th>Sequence</th>
<th>Use</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1N1</td>
<td>Matrix.FWD</td>
<td>233 bp</td>
<td>5' - CGGCAAAACGAGTTGCGAGACC-3'</td>
<td>H1N1Response</td>
</tr>
<tr>
<td>H1N1</td>
<td>Matrix.RV</td>
<td></td>
<td>5' - CTTCCTGACCTAATTGTTCCCGCC-3'</td>
<td>H1N1 Response</td>
</tr>
<tr>
<td>RBM3</td>
<td>RBM3.FW</td>
<td>221bp</td>
<td>5' - TGGTGCCAGCTACTCTAGAGGTTGG-3'</td>
<td>Cold Shock Response</td>
</tr>
<tr>
<td>RBM3</td>
<td>RBM3.RV</td>
<td></td>
<td>5' - GAGTCAGAAATTTTCTTCAGATCTGAGTTC-3'</td>
<td>Cold Shock Response</td>
</tr>
<tr>
<td>HSP70</td>
<td>HSP70.FW</td>
<td>151bp</td>
<td>5' GCGGTGATGACTGCCCCTGAT-3’</td>
<td>Heat Shock Response</td>
</tr>
<tr>
<td>HSP70</td>
<td>HSP70.RV</td>
<td></td>
<td>5' CGGCGGGATGTCAGGTGAG-3’</td>
<td>Heat Shock Response</td>
</tr>
<tr>
<td>U2</td>
<td>U2.FW</td>
<td>169bp</td>
<td>5' - CATCGCTTCTCGGCTTTTGT-3’</td>
<td>Nuclear Control</td>
</tr>
<tr>
<td>U2</td>
<td>U2.RV</td>
<td></td>
<td>5’ - TGGAGGTAAGTGCAATAGACGGGC-3’</td>
<td>Nuclear Control</td>
</tr>
<tr>
<td>S14</td>
<td>S14.FW</td>
<td>166bp</td>
<td>5' - GCCAGACCAGAGATGAAATCCTC-3’</td>
<td>Cytosolic Control</td>
</tr>
<tr>
<td>S14</td>
<td>S14.RV</td>
<td></td>
<td>5’ - CAGGTCAGGGGTCTTGTGCCG-3’</td>
<td>Cytosolic Control</td>
</tr>
</tbody>
</table>

**Figure S2- Primer Sequences for Amplification of all seven SUMO1/2/3 Variants Independently**

Table of primer sequences used for independent detection and amplification of each SUMO1/2/3 variant by specific targeting of exon/exon distinctions found within each transcript sequence. As well as control transcripts used in respective stress and fractionation assays.
SUMO2 remains the abundant paralog upon all stress treatments ranging from an average of 88.20% transcripts, experiences a decrease in transcription upon heat shock comprising 83% of SUMO transcripts and a slightly less decrease upon Influenza A infection (86% of total SUMO transcripts). The largest shift in SUMO proportions occurs within the SUMO3 transcripts, in which SUMO3 var2 which represents on average 25% of the total SUMO3 mRNA was significantly decreased by 19% of transcripts upon Influenza A viral infection.
**Fig. S4** SUMO1/2/3 Stress-Induced Response Levels in the HEK293A Cell Line

SUMO2 remains the abundant paralog upon all stress treatments ranging from an average of 88.20% transcripts, experiences a decrease in transcription upon heat shock comprising 83% of SUMO transcripts and a slightly less decrease upon Influenza A infection (86% of total SUMO transcripts). The largest shift in SUMO proportions occurs within the SUMO3 transcripts, in which SUMO3var2 which represents on average 25% of the total SUMO3 mRNA was significantly decreased by 19% of transcripts upon Influenza A viral infection.
A. Fractionation Control Transcripts in A549 Cells

![Graphical Representation of Subcellular Fractionation at 27°C](image)

B. SUMO1/2/3 Nuclear Export Levels in A549 Cells at 27°C

![Graphical Representation of Subcellular Fractionation at 27°C](image)

C. Table summarizing the percent of each transcript’s export profile.

<table>
<thead>
<tr>
<th></th>
<th>A549</th>
<th></th>
<th>HEK293A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>%CYTO</td>
<td>%NUC</td>
<td>%CYTO</td>
</tr>
<tr>
<td>S1V1</td>
<td>33%</td>
<td>67%</td>
<td>36%</td>
</tr>
<tr>
<td>S1V2</td>
<td>65%</td>
<td>35%</td>
<td>72%</td>
</tr>
<tr>
<td>S1V3</td>
<td>41%</td>
<td>59%</td>
<td>43%</td>
</tr>
<tr>
<td>S2V1</td>
<td>61%</td>
<td>39%</td>
<td>60%</td>
</tr>
<tr>
<td>S2V2</td>
<td>73%</td>
<td>27%</td>
<td>64%</td>
</tr>
<tr>
<td>S3V1</td>
<td>77%</td>
<td>23%</td>
<td>82%</td>
</tr>
<tr>
<td>S3V2</td>
<td>9%</td>
<td>91%</td>
<td>29%</td>
</tr>
</tbody>
</table>

S5- Graphical Representation of Subcellular Fractionation at 27°C

A) Control Transcripts for each respective fraction and cold shock B) Fractionated Response of SUMO1/2/3 transcripts in response to cold shock showing standard deviation. C) Table summarizing the percent of each transcript’s export profile.
9. References

29. C. Cheadle et al., Control of gene expression during T cell activation: alternate regulation of mRNA transcription and mRNA stability. BMC Genomics 6, 75 (2005).


74. C. Hoege, B. Pfander, G. L. Moldovan, G. Pyrowolakis, S. Jentsch, RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. Nature 419, 135-141 (2002).
86. K. Nacerddine et al., The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. Dev Cell 9, 769-779 (2005).


117. Z. Dassouki et al., ATL response to arsenic/interferon therapy is triggered by SUMO/PML/RNF4-dependent Tax degradation. Blood 125, 474-482 (2015).
119. J. J. Driscoll et al., The sumoylation pathway is dysregulated in multiple myeloma and is associated with adverse patient outcome. Blood 125, 2827-2834 (2010).


10. Vita

Myriah Acuña is a native El Pasoan, who earned her Associate of Science in Chemistry from El Paso Community College in 2013. She enrolled in the University of Texas at El Paso and earned her Bachelor of Science in Forensic Biology in 2014. That same year, she entered graduate school at UTEP in the Master of Biological Science program. Upon the completion of her M.Sc., Myriah transferred into the UTEP Pathobiology PH. D program under her same mentor, Dr. German Rosas-Acosta. She proposed her dissertation and entered doctoral candidacy in Fall 2020 while teaching many courses such as anatomy and physiology, phage hunters, microbiology, and molecular cell biology.

This dissertation was typed by:

Myriah L. Acuña
Myriahlorraineacuna@gmail.com