Global and site-specific characterization of the SUMO proteome by mass spectrometry

Hendriks, I.A.

Chapter 6

Uncovering Global SUMOylation Signalling Networks in a Site-Specific Manner

Ivo A. Hendriks[1], Rochelle C. D'Souza[2], Bing Yang[1], Matthias Mann[2] and Alfred C.O. Vertegaal[1]

[1]Department of Molecular Cell Biology, Leiden University Medical Centre, Albinusdreef 2, 2333 ZA, Leiden, the Netherlands
[2]Proteomics and Signal Transduction, Max Planck Institute for Biochemistry, 82152 Martinsried, Germany

Chapter 6 has been submitted for publication

ABSTRACT

SUMOylation is a reversible post-translational modification essential for genome stability. Using high-resolution mass spectrometry, we have studied global SUMOylation in a site-specific manner for the first time, identifying over 3,200 SUMOylation sites in over 1,300 proteins. SUMOylation dynamics were studied in response to SUMO protease inhibition, proteasome inhibition and heat shock. SUMOylation regulates many cellular processes including transcription, DNA repair, chromatin remodelling, pre-mRNA splicing and ribosome assembly. Nearly one in four SUMOylated lysines have previously been reported to be ubiquitylated, indicating extensive competition between SUMO and ubiquitin for acceptor lysines. Crosstalk between SUMO and ubiquitin furthermore involves chimeric ubiquitin-SUMO chain formation via SUMOylation of lysines 6, 11, 27, 48 and 63 in ubiquitin. Thus, SUMOylation is a major dynamic post-translational modification regulating all nuclear processes.

INTRODUCTION

Reversible post-translational modification (PTM) of lysine residues in proteins by Small Ubiquitin-like Modifiers (SUMOs) plays a key role in genome stability and transcription [1-3]. Important SUMO target proteins in the DNA damage response include PCNA [4, 5], BRCA1 [6] and 53BP1 [7]. SUMOs are conjugated to target proteins via an enzymatic cascade involving a dimeric E1 enzyme (SAE1/2), a single E2 enzyme, Ubc9, and a limited number of E3 enzymes [8]. Mice deficient for Ubc9 die at the early post-implantation stage as a result of chromosome condensation and segregation defects, underlining the essential role of SUMOylation to maintain genome stability [9].

Frequently, SUMOylation regulates the function of target proteins by enabling or stabilizing non-covalent protein-protein interactions via SUMO Interaction Motifs (SIMs) [8]. Classical examples of this type of interaction include the binding of SUMOylated RanGAP1 to the nucleoporin RanBP2 [10] and the binding of the SRS2 helicase to SUMOylated PCNA [11, 12].

Interestingly, SAE1 and -2 were identified in a screen for Myc-synthetic lethal genes [13]. In cells with hyperactive Myc, inactivation of SAE2 caused mitotic catastrophe and cell death, indicating that Myc-driven tumours are dependent on SUMOylation. These findings provide a therapeutical perspective for SUMOylation inhibitors as a potential therapy for Myc-dependent tumours. In addition, the SUMOylation system plays a more wide-spread role in carcinogenesis [14].

Despite great interest in SUMOylation in the fields of genome stability, transcriptional regulation, nuclear organization, signal transduction and from a clinical point of view, global insight in protein SUMOylation is limited. Mass-spectrometry (MS)-based proteomics has enabled global insight in different PTMs in a site-specific manner [15, 16], including phosphorylation [17, 18], acetylation [19], methylation [20], glycosylation [21] and ubiquitylation [22-26]. Affinity purification strategies include immunoprecipitation and TiO_2 and Fe^{3+} immobilized metal affinity chromatography. Highly potent SUMO proteases [27], inconvenient C-terminal tryptic SUMO tags and low stoichiometry combined with suboptimal purification methods have hampered the identification of SUMO acceptor lysines in target proteins [16, 28-31]. We have developed an efficient purification strategy to study global protein SUMOylation in a site-specific and dynamic manner.

RESULTS

A Strategy for Mapping SUMOylated Lysines in Endogenous Proteins Purified from Human Cells

To facilitate the study of SUMOylated proteins, a common method is the employment of epitope-tagged SUMO, in order to allow efficient purification after highly denaturing lysis of cells to inactivate SUMO proteases. We enriched SUMOylated

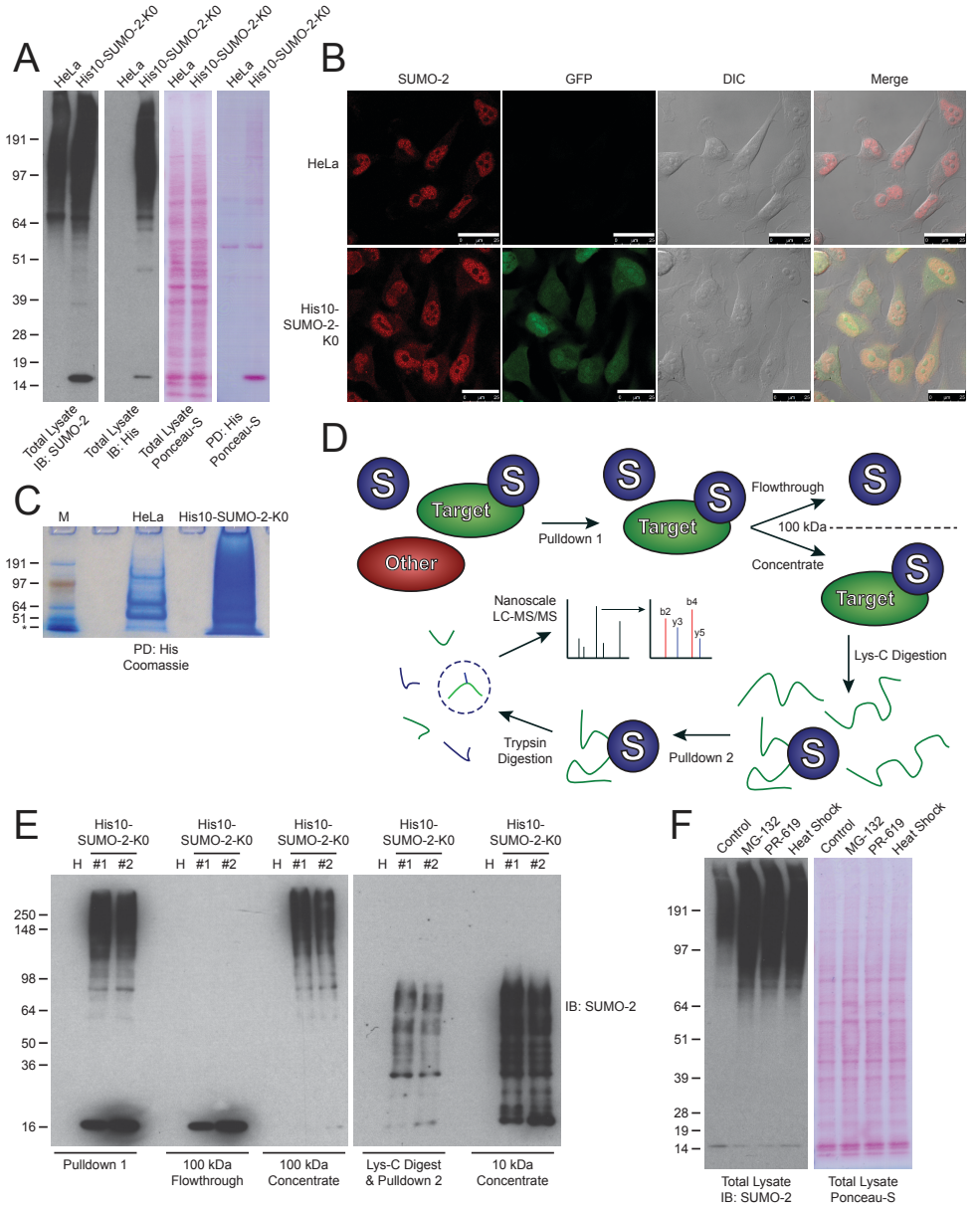


Figure 1. A strategy for mapping SUMO-2 acceptor lysines in endogenous proteins.

A) HeLa cells were infected with a bicistronic lentivirus encoding His10-SUMO-2-K0-Q87R (H10-S2-K0) and GFP separated by an Internal Ribosome Entry Site (IRES). Flow cytometry was used to select cells stably expressing low levels of H10-S2-K0. Total lysates were analysed by immunoblotting to confirm the expression of H10-S2-K0. Ponceau-S staining is shown as a loading control. Additionally, His10-pulldown was performed to enrich SUMOylated proteins, and Ponceau-S was performed to indicate high-specificity enrichment.

B) Stable cell lines were investigated by confocal fluorescence microscopy to confirm the predominantly nuclear localization of H10-S2-K0. Scale bars represent 25 µm.

C) A single-step His10-pulldown was performed on approximately 50 million HeLa and H10-S2-K0

peptides from a HeLa cell line stably expressing His10-tagged SUMO-2. This tag is small and compatible with denaturing buffer conditions. These stable cells were established using a bicistronic lentivirus encoding His10-SUMO-2 and GFP separated by an IRES. After infection, a population expressing this construct at low levels was obtained using FACS-sorting. Low expression levels were confirmed by immunoblotting (**Figure 1A**). As expected, the protein located predominantly in the nucleus (**Figure 1B**) [32, 33]. The His10 tag enabled single round purification with a high yield and purity in contrast to the His6 tag commonly used in the field (**Figure 1C**).

In order to enrich for SUMOylated peptides, we used a SUMO-2 form that is resistant to cleavage by endopeptidase Lys-C. Lysine-deficient (K0) SUMO-2 behaves very similar to wild-type SUMO-2, except for SUMO polymerization [31]. Our purification strategy consisted of a first round of His10-SUMO-2 purification, a filter step to concentrate SUMO-2 conjugates while simultaneously separating SUMO-2 conjugates from free SUMO-2, digestion by Lys-C and a second round of purification followed by a trypsin digest (**Figure 1D and 1E**). The second round of purification enabled enrichment at the site-level, greatly reducing the complexity of the final sample.

The C-terminal tryptic fragment of wild-type human SUMO-2 is 32 amino acids and due to its size, is not compatible with efficient mapping of SUMO-2 acceptor lysines. In contrast, yeast SUMO, Smt3, contains a conveniently located arginine that results in a five amino acid C-terminal tryptic fragment. We have generated a SUMO-2 Q87R mutant mimicking yeast SUMO to enable the identification of SUMO-2 acceptor lysines by mass spectrometry.

SUMOylation is a dynamic process, regulated via crosstalk with the ubiquitin-proteasome system [34, 35], sensitive to heat shock [32] and sensitive to the broad range SUMO – and ubiquitin protease inhibitor, PR-619. Treatments with the proteasome inhibitor MG-132, heat shock and PR-619 resulted in significant accumulation of SUMO-2 conjugates (**Figure 1F**).

cells, and purified SUMOylated proteins were analysed by SDS-PAGE and Coomassie staining.
D) Schematic overview of the H10-S2-K0 SUMOylation site purification strategy. A direct purification step was followed by concentration of SUMOylated proteins, which were subsequently digested with endopeptidase Lys-C. The H10-S2-K0 bearing the SUMOylated peptide was re-purified, concentrated, digested with trypsin, and finally analysed by high-resolution nanoscale LC-MS/MS.
E) Immunoblotting analysis was used to confirm the efficiency of the purification steps described in section D. Concentration of purified SUMOylated proteins under denaturing conditions resulted in highly efficient separation of free SUMO from SUMO conjugates. Lys-C digestion and re-purification resulted in a highly concentrated mixture of SUMOylated peptides.
F) Cells stably expressing H10-S2-K0 were mock treated, or treated with MG-132, PR-619 or heat shock. Prior to SUMO-site enrichment, total lysates were analysed by SDS-PAGE and immunoblotting analysis to ensure efficacy of the treatments. Ponceau-S staining is shown as a loading control.

Site-Specific Identification of SUMOylated Lysines Reveals over 3,200 Sites Located in over 1,300 Proteins

Tryptic digests of in-solution digested re-purified SUMOylated peptides were directly analysed by nanoscale LC-MS/MS, without further fractionation. Using 2-hour LC gradients, we identified in the range of 500 to 1200 SUMOylation sites from single runs depending on treatment conditions. The main experiment was performed in biological quadruplicate, with experiments measured in technical duplicate. Cellular treatment experiments were performed simultaneously, and all different conditions were measured in technical duplicate.

3,246 SUMOylation sites were identified in 1,364 proteins at a false discovery rate (FDR) below 1%. Mass accuracy was within 3 ppm for 97.3% of all identified sites, and within 6 ppm for all sites, with an average absolute mass error of 0.72 ppm. The majority of identified SUMOylation sites had an Andromeda peptide score in the range of 60-120 (**Figure 2A**). The precise SUMO-2 acceptor lysines were pinpointed in over 98% of the SUMOylated peptides. The overall purity of our method was demonstrated by the presence of a SUMO-2 acceptor lysine in 27.5% of the peptides identified in the final purified fractions. This is a major step forward for the field and enabled the analysis of SUMOylation at a proteome-wide level in a site-specific manner for the first time. Our dataset includes an extensive number of well-known SUMO target proteins, e.g. RanGAP1, PML, Topoisomerase-1, -2α, -2β, PCNA, BLM, BRCA1, RanBP2, RNF168, SAFB2, confirming the validity of the approach.

In total 1,104 sites were identified from cells grown at regular cell culture conditions (**Figure 2B and 2C**). The dynamic nature of SUMOylation was underlined by the identification of 2,160 additional SUMOylation sites in response to heat shock, proteasome inhibitor treatment and / or SUMO protease inhibitor treatment, representing two thirds (66.2%) of the total identified sites (**Figure 2B**). The peak intensities of the SUMOylated peptides identified in response to MG-132, PR-619 or heat shock were higher compared to the control sample, indicating that not only the number of acceptor sites increased upon treatment, but also the stoichiometry of SUMOylation (**Figure 2D**). We quantified the changes in overall protein SUMOylation after the cellular treatments, and found significant changes in the SUMOylation state of hundreds of proteins (**Figure S1**).

SUMOylation sites per protein ranged from a single site in nearly half of all SUMO target proteins, with 10 or more sites in 65 proteins, and 20 or more sites in only 5 proteins – MIS18BP1 (20 sites), NKTR (21 sites), FBN1 (24 sites), ZNF451 (28 sites) and GTF2I (29 sites) (**Figure 2E**). Thus, most proteins are only conjugated to one or two SUMO moieties.

We compared all SUMOylated proteins and sites identified in this study to previous studies on SUMOylation. On average, we identified 58% of previously reported SUMO target proteins (**Figure S2A**), and expanded the known amount of SUMOylated proteins by nearly one thousand. We found 62% of all previously

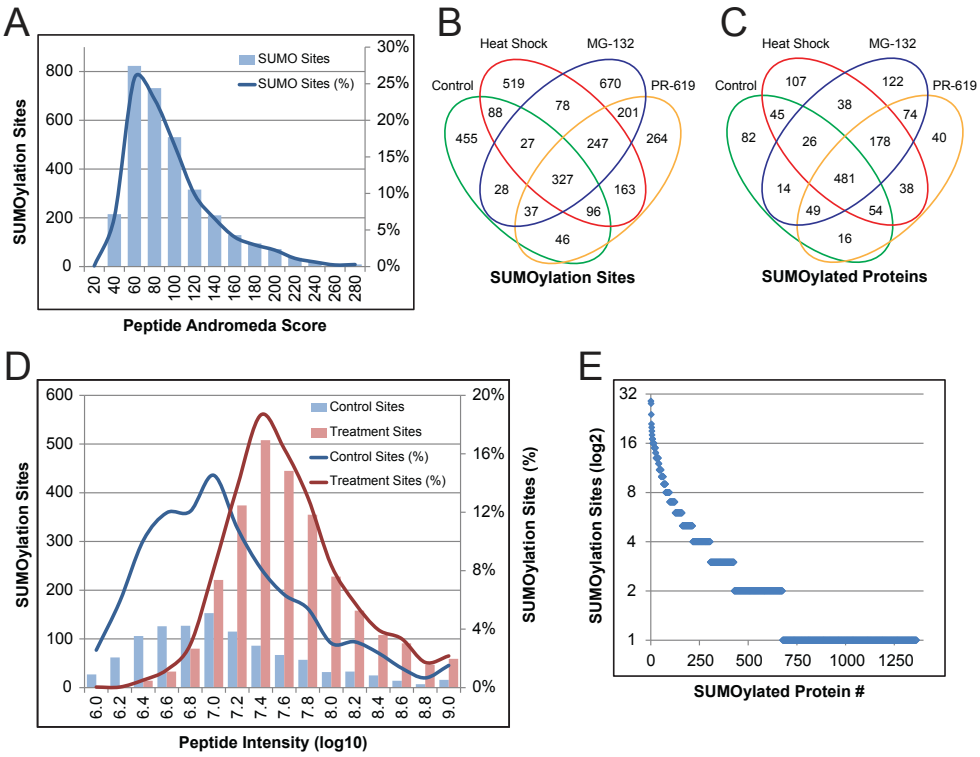


Figure 2. Overview of mass spectrometry results.
A) Overview of the amount of SUMOylated peptides identified with their respective Andromeda peptide scores.
B) Schematic representation of the amount of SUMOylation sites discovered in relation to the cellular treatments used. Over 3,200 SUMOylation sites were identified in total.
C) Schematic representation of the amount of SUMOylated proteins discovered in relation to the cellular treatments used. Over 1,300 SUMOylated proteins were identified in total.
D) Overview of the SUMOylated peptides identified under control conditions, as compared to the SUMOylated peptides exclusively discovered after cellular treatments. The absolute and relative amount of SUMOylation sites are plotted against the peptide intensity.
E) Overview of the amount of SUMOylation sites identified per protein.

reported MS/MS-identified SUMOylation sites (**Figure S2B**), as reported in the PhosphoSitePlus database (PSP; PhosphoSitePlus®, www.phosphosite.org, [36]), and our study expands by more than twentyfold the number of known MS/MS-identified SUMOylation sites.

SUMOylation is a key post-translational modification in all eukaryotes, but is absent in prokaryotes. We studied phylogenetic conservation of SUMOylation with respect to conservation of entire proteomes (**Figure S3A**). SUMOylated proteins are significantly more conserved than total proteomes. Within orthologues, SUMOylation is the most conserved post-translational modification together with ubiquitylation. When considering conservation of proteins with no orthologues,

SUMO is more conserved than phosphorylation, but less than ubiquitylation and acetylation (**Figure S3B**). This difference is indicative of an increased frequency of SUMOylation occurring on proteins that are absent in lower eukaryotes.

We investigated the potential overlap between the identified SUMO-acceptor lysines with other post-translational lysine modifications. For this purpose, we extracted all known human MS/MS-identified ubiquitylation, acetylation and lysine-methylation sites from PSP, and cross-compared modification sites (**Figure 3A and 3B**). SUMOylation is known to compete with ubiquitin for acceptor lysines in target proteins [37]. However, the extent of this crosstalk is currently unclear. We compared the identified SUMO acceptor lysines to acceptor lysines for ubiquitin derived from PSP and found that nearly one in four (24.2%) SUMOylation sites are also known to be ubiquitylated, indicating extensive competition between SUMOylation and ubiquitylation (**Figure 3A and 3B**). Competition between SUMOylation and acetylation and between SUMOylation and lysine-methylation occurs less frequently, although this could be related to the smaller number of acetylation and methylation sites currently identified. From the perspective of all known ubiquitylation, acetylation and methylation sites, SUMOylation occurs on roughly 2% of these sites.

Interestingly, crosstalk between SUMOylation and other post-translational modifications includes regulation of enzymatic components including 37 kinases, 32 proteins with intrinsic phosphatase activity, 24 ubiquitin-protein ligase family members, six ubiquitin proteases, five acetyltransferases, seven deacetylases, 20 methyltransferases and ten demethylases.

We found 21 SUMOylated peptides exclusively in combination with phosphorylation, and an additional 25 SUMOylated peptides together with phosphorylation in a non-unique fashion. Phosphorylation occurred relatively close to the lysine, and was found both upstream and downstream of the SUMOylation lysine. Phosphorylation occurred predominantly at the +5 position (**Figure 3C**), in agreement with the earlier described PDSM [31, 38]. We observed 3 phosphorylation sites at +2 relative to the SUMOylated lysine, which could serve as the acidic charge required for efficient SUMOylation instead of glutamic or aspartic acid.

Two SUMOylated peptides were found exclusively in combination with acetylation, with both of these events occurring on histones, and an additional SUMOylated peptide from PML together with acetylation in a non-unique fashion. These sites could indicate acetylation-dependent SUMOylation, suggesting a novel type of crosstalk between these two major modifications.

Furthermore, we found direct modification of endogenous ubiquitin by SUMO-2 on lysines 11 and 63 under control conditions (**Figure 3D**). After cellular treatments, we additionally observed SUMO-2 modification of ubiquitin on lysines 6, 27 and 48 (**Figure 3D**). Interestingly, we also identified the ubiquitin-like modifier Nedd8 to be modified by SUMO-2 on lysine 48 after cellular treatment with the proteasomal inhibitor MG-132 (**Figure 3D**). Thus, mixed chain formation between

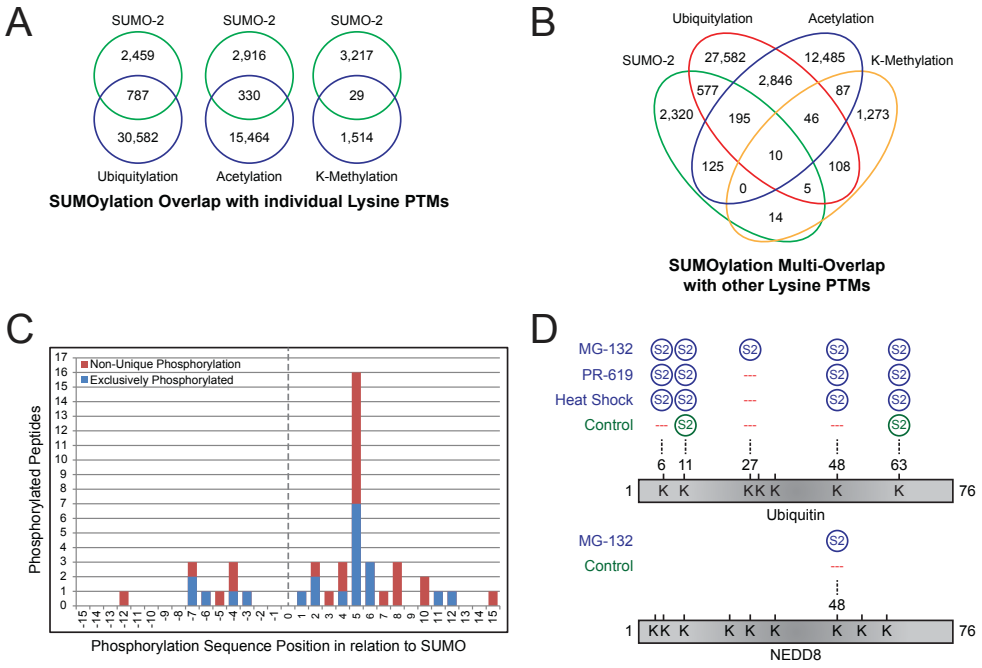


Figure 3. SUMO is extensively involved in PTM crosstalk.
A) Schematic representation of the overlap of the identified SUMOylated lysines as compared to the other lysine post-translational modifications ubiquitylation, acetylation and lysine-methylation.
B) Similar to section A, overlap between SUMOylation, ubiquitylation, acetylation and lysine-methylation.
C) Schematic overview of phosphorylation sites adjacent to SUMOylated lysines, and their amino acid spacing in relation to the central lysine. Some peptides were exclusively found to be SUMOylated in combination with phosphorylation (blue). Non-unique phosphorylation sites on SUMOylated peptides were also discovered (red).
D) Schematic representation of the identified SUMOylation sites on ubiquitin and Nedd8.

ubiquitin and ubiquitin-like family members is more extensive than previously thought. Furthermore, we detected mixed chain formation between all SUMO family members.

Insight in the SUMOylation consensus motif

In contrast to some other major post-translational modifications (**Figure S4**), SUMOylation is known to occur on the classical consensus motif ΨKxE [39] or ΨKx[ED] [40], where Ψ is a large hydrophobic amino acid. Previously, we have found that other residues are also used at the Ψ position [31]. Our dataset provides an important opportunity to obtain more insight in the SUMOylation consensus motif. More than half of the identified sites from untreated cells matched the consensus motif KxE with a very modest frequency of KxD type sites (**Figure 4A**). We studied the KxE type SUMOylation motif in RanGAP1, a highly SUMOylated protein [10, 41].

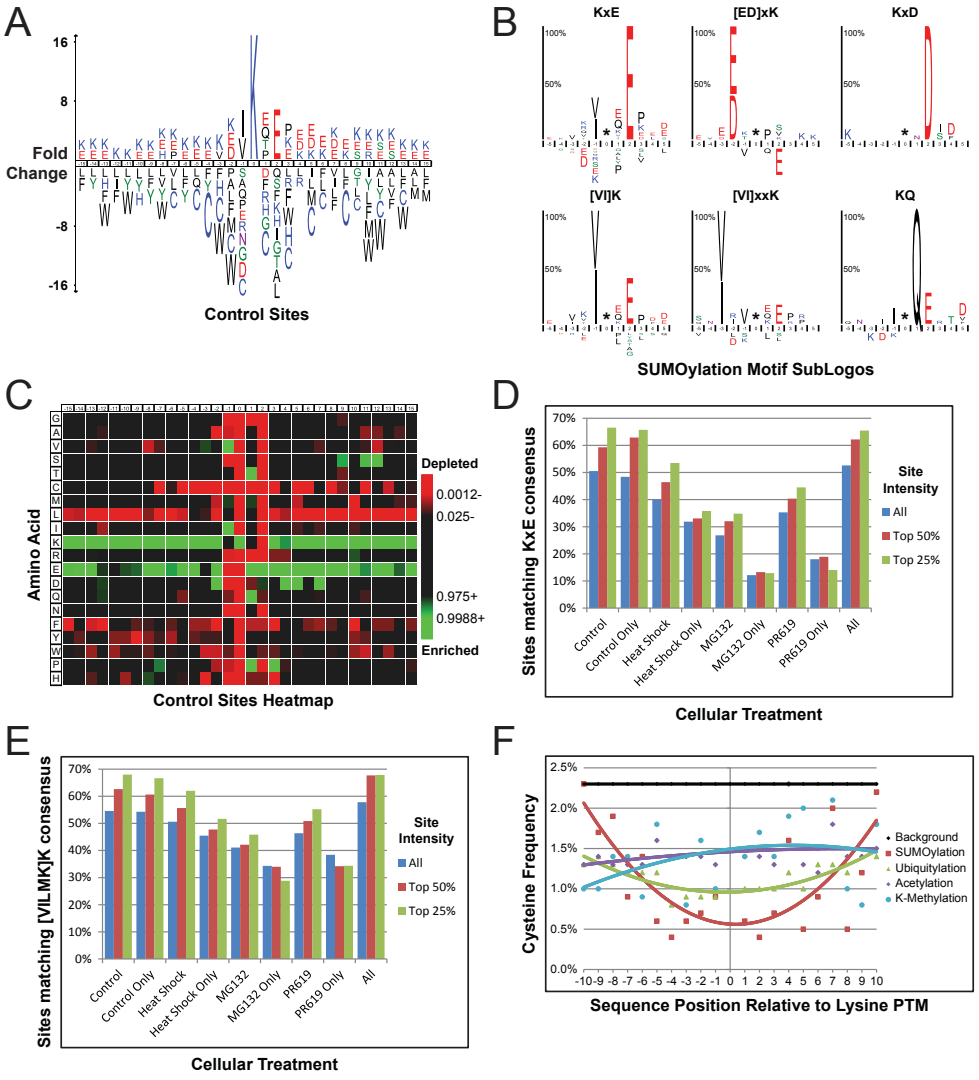


Figure 4. Novel insight in the SUMOylation consensus motif.

A) IceLogo of all SUMOylation sites identified under control conditions. The height of the amino acid letters corresponds to fold-change. All amino acid changes were significant with $p < 0.03$.

B) SubLogos of various consensus motifs. The height of amino acid letters represents the %-change enrichment or depletion of the motif set as compared to the reference set. All amino acid changes were significant with $p < 0.05$.

C) As section A, but in heatmap format. Green is indicative of a statistical enrichment as compared to randomly expected, and red is indicative of a depletion. All amino acid changes were significant with $p < 0.05$.

Replacing E526 for D resulted in a significant drop in SUMOylation (**Figure S5A and S5B**). As a negative control, we included our previously published ΔGL RanGAP1 mutant [31].

The acidic residue is not necessarily located two positions downstream of the SUMOylated lysine but can also be found two positions upstream with a higher frequency for aspartic acid compared to the regular SUMOylation motif (**Figure 4B**), in agreement with the inverted consensus motif [ED]xK that we proposed based on a very small number of identified SUMOylation sites [31]. We confirmed the relevance of the E300 residue in the inverted motif covering ETV6 SUMOylated lysine 302 (**Figure S5C and S5D**). Replacing this residue for alanine significantly reduced SUMOylation, whereas replacing it for aspartic acid had little effect in agreement with the inverted consensus motif [ED]xK.

We superimposed the 1,104 SUMOylation sites identified under control conditions, including a sequence window ranging from -15 to +15 amino acids, and compensated the amino acid frequency against the randomly expected frequency across the human proteome (**Figure 4A and 4C**). The highest degree of enrichment was observed for valine and isoleucine at -1, and glutamic acid at +2, with over half of all control SUMOylation sites adhering to this consensus. When only considering the top 50% and top 25% of most abundant SUMOylation sites for glutamic acid at +2, this frequency increased to 59% and 67% respectively, indicating that the lysines situated in SUMOylation consensus motifs are efficiently SUMOylated (**Figure 4D**). We observed a similar trend for the hydrophobic amino acids at -1 (**Figure 4E**). A further expanded consensus motif, taking statistical local enrichments into account, amounted to [IVL]-K-[QTP]-E-P. Interestingly, the adherence to the consensus motif dropped moderately for sites exclusively mapped after heat shock treatment, and decreased drastically after MG-132 and PR-619 treatment, where sites matching KxE reached as low as 12% (**Figure 4D**).

Considering the 30 amino acid region flanking SUMOylated lysines, it is evident that this region is enriched for lysine and glutamic acid (**Figure 4A and 4C**). Thus, SUMOylated lysines are frequently located in regions enriched for charged residues, indicative of solvent exposure. Interestingly, SUMOylated regions are depleted for phenylalanine, tryptophan, tyrosine, leucine, and most notably for cysteine (**Figure 4A and 4C**). When comparing SUMOylation sites identified after the different treatments (**Figure S6A**) to the sites identified in the sample purified

D) Schematic representation of the cysteine frequency close to all identified control SUMOylation sites, as well as other PTMs, ranging from -10 to +10 amino acids of the modified lysine. For all PTMs, a 2nd order polynomial trend line was calculated. The randomly expected cysteine frequency is indicated in black.

E) Overview of the amount of SUMOylation sites matching the short consensus motif KxE in different subsets of sites corresponding to different cellular treatments. Additionally, per subset, the top 25% intense, the top 50% intense, or all sites are shown.

F) As section E, but for the short consensus motif [VILMK]K.

from non-treated cells, especially in the case of MG-132 and PR-619, we observed a large increase in cysteine, tryptophan, phenylalanine and tyrosine in the direct vicinity (**Figure S6B-D**). This is indicative of regions of proteins normally buried in the hydrophobic core.

Since SUMOs are transferred along an enzymatic cascade via thioester formation, the reduced frequency of cysteines near SUMOylated lysines under standard conditions could help to avoid the formation of thioesters between SUMOs and target proteins. Interestingly, a reduced frequency for cysteine can also be observed for ubiquitylated regions, probably for the same reason as proposed for SUMOylation (**Figure 4F**). Reduced frequencies for cysteine were less pronounced for regions flanking methylated or acetylated lysines (**Figure 4F**).

Insight in SUMOylated protein groups

Protein domains that are frequently SUMOylated include the Krüppel associated box (KRAB) domain which is a repressor domain found in many zinc finger protein-based transcription factors (**Figure 5A**). Other domains include zinc fingers, PHD fingers and RRM1, which are domains that serve important roles in binding of DNA, RNA or other proteins, and are often found in nuclear or chromatin-associated proteins.

We investigated the subcellular localization of SUMOylated proteins by Gene Ontology Cellular Compartments classes, and plotted all identified proteins and sites (**Figure 5B**). SUMOylation was found to be an almost exclusively nuclear modification, with cytoplasmic modification occurring primarily on proteins that are also annotated as nuclear. Enrichment analysis showed the highest ratio for chromatin-associated proteins, closely followed by all nuclear proteins (**Figure 5B**). The first identified SUMOylation site, K524 in the nuclear-pore component RanGAP1 is located in an unstructured region of the protein [42]. SUMOylation is thought to occur predominantly in unstructured regions [40]. To investigate the localized structural properties of proteins around sites of SUMOylation, we *in silico* folded all 3,246 sites including the 30 amino acid sequence window, as well as nearly 5,000 lysines randomly chosen from SUMOylated proteins as a reference set. We performed secondary structure prediction of the modified lysine, and classed the structure as α-helix, β-sheet, or otherwise coiled (**Figure 5C**). Our results indicate that there was a significant reduction in SUMOylation of unstructured regions, a modest reduction in SUMOylation of α-helices and a significant increase in SUMOylation of β-sheets compared to expected frequencies. This trend was most striking for KxE type SUMOylation sites. We additionally observed an increased tendency for SUMOylated regions to be solvent exposed (**Figure 5D**).

SUMO modifies highly interconnected functional networks of proteins

Genome stability, transcription and translation are three important biological processes as evidenced by term enrichment analysis for Gene Ontology Biological

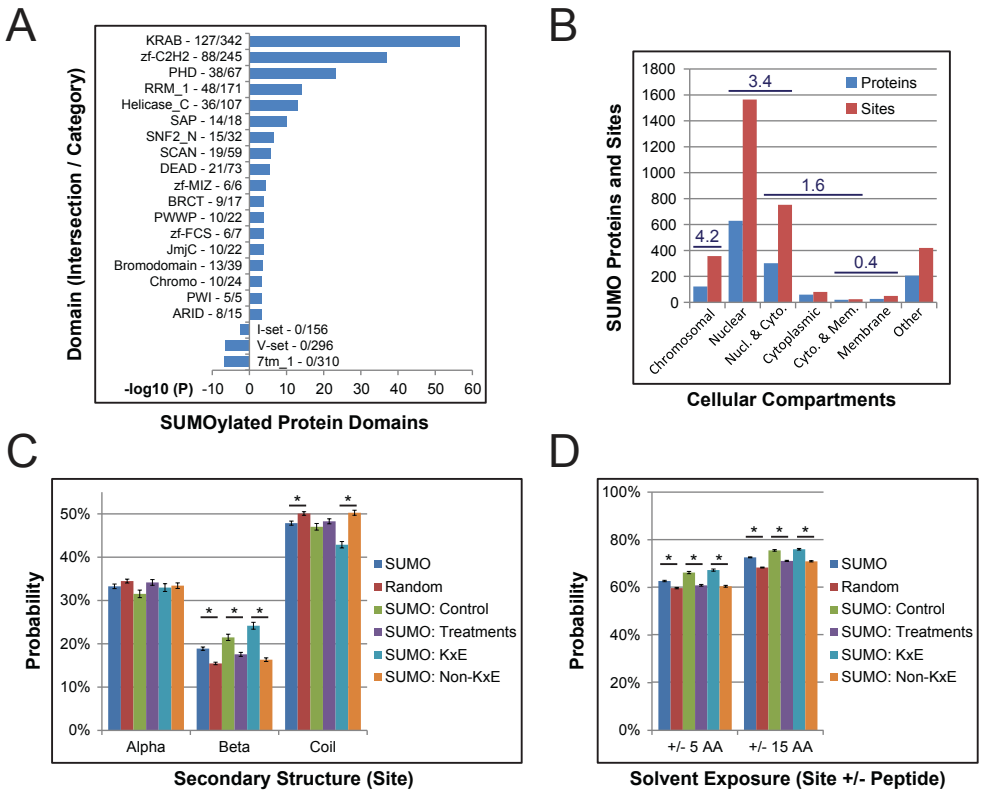


Figure 5. SUMOylation is a predominantly nuclear event, and is often found on proteins bearing RNA or DNA interaction domains.

A) Overview of protein domain families overrepresented or underrepresented in all identified SUMOylated proteins, ranked by -Log10 p-value. SUMOylated proteins bearing these domains as compared to total known proteins containing these domains are indicated.

B) Overview of the amount of SUMOylated proteins and sites in relation to their subcellular localization. Gene Ontology Cellular Compartments (GOCC) enrichment ratios are indicated for the applicable categories. The large majority of all identified SUMOylated proteins were nuclear and/or chromatin-associated.

C) Overview of the predicted secondary structure of different subsets of SUMOylated lysines. The random set corresponds to lysines randomly selected from SUMOylated proteins. Differences indicated with an asterisk (*) were significant with $p < 0.001$.

D) Overview of the predicted solvent exposure of different subsets of SUMOylated lysines, including either -5 to +5 amino acids, or -15 to +15 amino acids. The random set corresponds to lysines randomly selected from SUMOylated proteins. Differences indicated with an asterisk (*) were significant with $p < 0.001$.

Processes involving the identified SUMO targets (**Figure 6A**). Furthermore, nucleic acid metabolism, chromosome organization, DNA repair, cell cycle regulation, RNA splicing, histone modification, and nuclear body organization are amongst the most enriched processes. For Gene Ontology Molecular Functions, in absolute numbers,

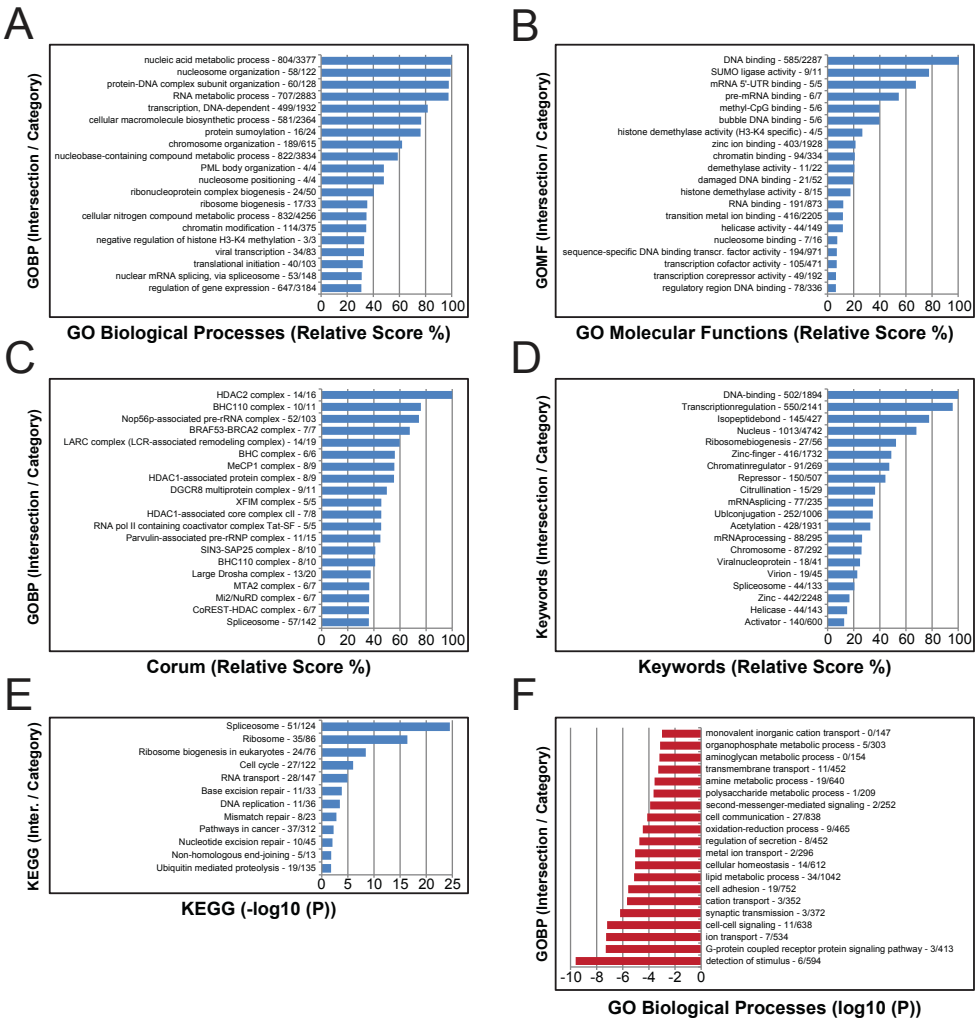


Figure 6. Term enrichment analysis reveals SUMO's involvement in many biological processes and protein complexes.

A) All identified SUMOylated proteins were annotated with Gene Ontology Biological Processes terms, and compared against the annotated human proteome. Categories were scored by a combination of enrichment ratio and p-value. The amount of hits as compared to the category size is indicated.

B) As section A, but for Gene Ontology Molecular Functions.

C) As section A, but for CORUM complexes.

D) As section A, but for Keywords.

E) As section A, but using KEGG annotation. Categories were scored by their p-value.

F) As section A, but showing underrepresented categories. Categories were scored by their p-value.

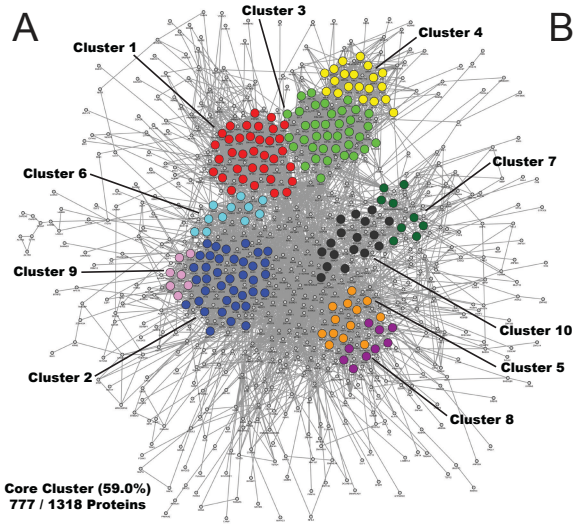
we identified 585 DNA binding proteins and 403 Zinc binding proteins as the largest functional groups of SUMO target proteins (**Figure 6B**). SUMO also modified significant amounts of subunits from known CORUM protein complexes, including Nop56p pre-rRNA, BRAF53-BRCA2, LARC, BHC, MeCP1 and HDAC1 and -2 protein complexes (**Figure 6C**). Additional analyses by keywords (**Figure 6D**) and KEGG terms also highlighted SUMO's regulation of many pivotal cellular processes, including an enrichment for proteins known to be involved in cancer pathways (**Figure 6E**), such as TP53, STAT1, FOS, JUN and SMAD4.

SUMOylated proteins form a very complex, highly organized network of interacting proteins as visualized using a Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) network analysis (**Figure 7A**). Nearly 60% of all identified proteins are part of one main functional cluster, at high STRING confidence. We performed STRING analyses on a per-treatment basis, at medium and high STRING confidences to assess protein-protein interaction enrichment ratios (**Figure 7B**). Overall, a ten-fold increase in interactions was observed as compared to expected, with the SUMO target proteins from untreated cells showing a twenty-fold increase.

Subsequently, MCODE analysis revealed highly interconnected sub-clusters within the core network, including ten sub-clusters with interconnectivity scores ranging from 7 up to 35 (**Figure 7A**). Three dominant clusters involve many functionally related proteins from the spliceosome, the ribosome and cell cycle related factors (**Figure 7C**). Other clusters contain proteins involved in chromatin remodelling, histone deacetylases, histone methyltransferases, regulation of ubiquitin-protein ligases, and proteins involved in ribonucleoprotein complex formation.

We performed STRING analysis on groups of proteins that matched the highest scoring terms from the enrichment analysis (**Figure 6**), and observed a very high connectivity between these sub-clusters of SUMOylated proteins (**Figure S7 and S8**), demonstrating SUMO's ability to regulate functional clusters of proteins involved in many crucial nuclear processes, including the DNA damage response, cell cycle regulation, transcription, replication and splicing.

SUMOylation, ubiquitylation and acetylation were found to compete for lysine residues (**Figure 3A and 3B**). To further investigate this, we performed STRING analysis on the subset of proteins containing these lysines. We found over 80% of these proteins to be situated in a single functional network (**Figure 8A, 8B and 8C**). Beyond the observed overlap between modification sites, we found these clusters of proteins to be highly modified by SUMO, averaging 5 SUMOylation sites per protein (**Figure 8D**). Additionally, we observed a high degree of enrichment for protein-protein interactions as compared to expected (**Figure 8E**), and a higher degree of enrichment as compared to the full SUMO network (**Figure 7B**). Thus, SUMOylation appears to function in concert with other major PTMs, and co-regulates a tight functional cluster of heavily modified and dynamic proteins.



B

STRING Interaction Confidence: 0.4+				
	Enrichment	Proteins	Interactions	P-value
All proteins	8.9	1318	12230	<1.0E-15
Control	11.2	694	3911	<1.0E-15
Control Only	14.7	95	46	<1.0E-15
Heat Shock	9.6	925	8228	<1.0E-15
Heat Shock Only	10.7	115	140	<1.0E-15
MG-132	9.2	940	6731	<1.0E-15
MG-132 Only	7.0	131	47	<1.0E-15
PR-619	9.6	884	6726	<1.0E-15
PR-619 Only	7.2	41	5	7.5E-04
Treatments Only	9.1	632	3144	<1.0E-15

STRING Interaction Confidence: 0.7+				
	Enrichment	Proteins	Interactions	P-value
All proteins	10.1	1318	5272	<1.0E-15
Control	13.4	694	1707	<1.0E-15
Control Only	25.2	95	22	<1.0E-15
Heat Shock	11.1	925	3364	<1.0E-15
Heat Shock Only	11.7	115	53	<1.0E-15
MG-132	10.3	940	3010	<1.0E-15
MG-132 Only	8.7	131	34	<1.0E-15
PR-619	11.0	884	2939	<1.0E-15
PR-619 Only	11.5	41	3	2.5E-03
Treatments Only	10.3	632	1399	<1.0E-15

STRING Interaction Enrichment Analysis

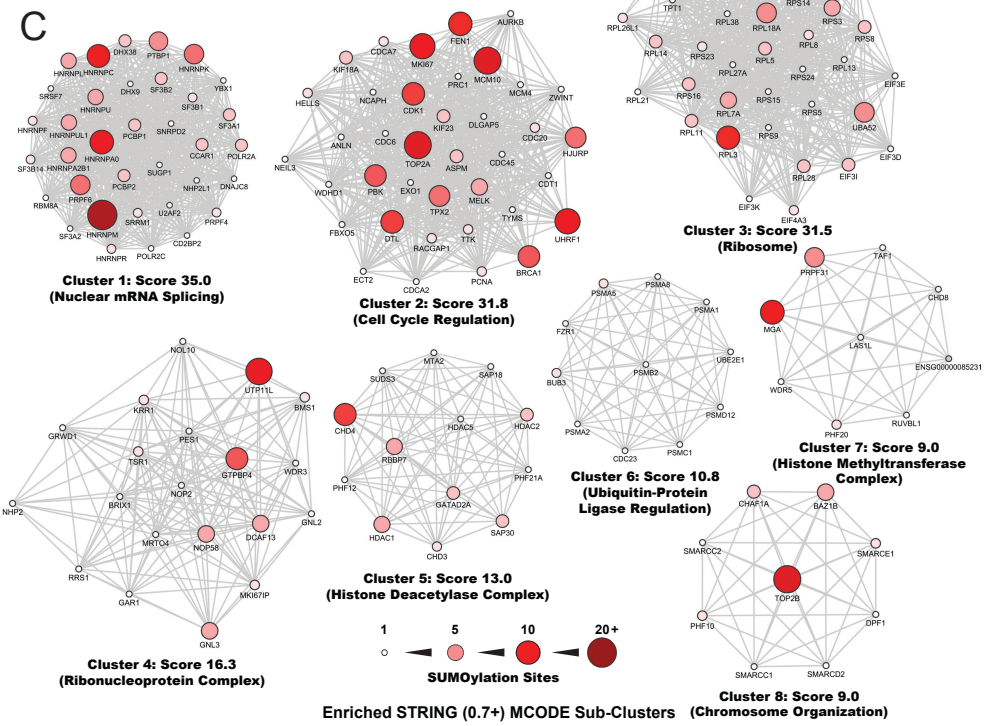


Figure 7. SUMO modifies highly interconnected functional networks of proteins.

A) STRING network analysis of all identified SUMOylated proteins, with a STRING interaction confidence of 0.7 or greater. MCODE was used to extract the most highly interconnected functional clusters from the network, which are indicated in different colours. High-resolution images are available as online supporting information.

B) Overview of STRING analyses using different subsets of SUMOylated proteins. Enrichment is an absolute ratio derived from the observed amount of interactions divided by the expected amount

DISCUSSION

We have developed novel methodology for global identification of SUMOylation sites, enabling us to map over 3,200 SUMO acceptor lysines in over 1,300 proteins. This provides detailed insight in the function of this post-translational modification. All nuclear processes are orchestrated by SUMOylation including transcription, DNA repair, chromatin remodelling, pre-mRNA splicing and ribosome assembly.

This is an important step forward compared to the relatively small number of SUMOylation sites that we found previously [31] and provides a rich dataset for the scientific community to enable follow-up studies. This is particularly relevant for half of the identified sites that are not located in a SUMOylation consensus motif and therefore elude *in silico* prediction. Furthermore, we have obtained novel insight in the consensus motif for SUMOylation and found that SUMOylation sites are frequently located in domains enriched for charged residues. Moreover, protein regions harbouring SUMOylated lysines are depleted for cysteines, possibly to limit thioester formation between SUMOs and target proteins.

The SUMO acceptor lysines could for the first time be properly compared to identified ubiquitin acceptor lysines and acetylation and methylation sites [19, 26]. A striking overlap was found between SUMO acceptor lysines and ubiquitin acceptor lysines, indicating extensive competition between these modifications. Additionally, it is known that these modifications can occur consecutively [43]. Moreover, our findings provide novel insight in crosstalk between SUMOylation and ubiquitylation since we found that five lysines in ubiquitin can be used as SUMO acceptor lysines, indicating complex heterogeneous SUMO-ubiquitin chains that open up exciting new avenues to investigate mechanisms and biological relevance of this novel type of signal transduction. Most likely, many other SUMOylation sites remain to be discovered as a result of cell-type specific SUMOylation or stimulus-dependent SUMOylation that could be investigated with the developed methodology.

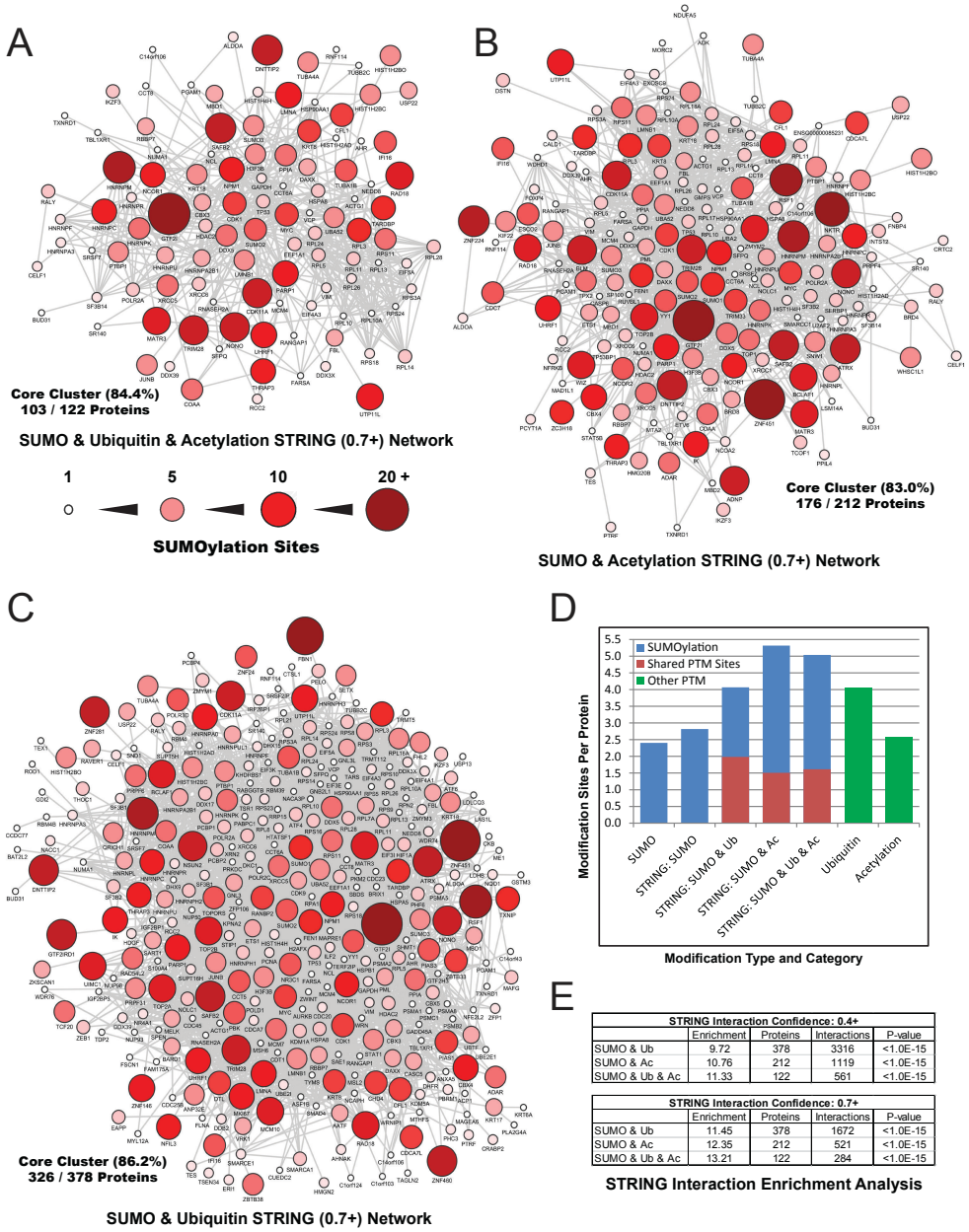
EXPERIMENTAL PROCEDURES

Plasmids

The His10-SUMO-2-K0-Q87R we described and used in this manuscript has the following amino acid sequence: MAHHHHHHHHHHGGSMSEERPREGVRTENDHINLRVAGQDGSVVQFRIRRHTPLSRLMRAYCER-QGLSMRQIRFRFDGQPINETDTPAQLEMEDEDTIDVFRQQTGG. The corresponding nucleotide sequence was cloned in between the PstI and XhoI sites of the plasmid pLV-CMV-IRES-GFP [47].

of interactions. All analyses were performed at STRING confidence of 0.4+ and STRING confidence of 0.7+. The network in section A corresponds to 'All Proteins' at confidence 0.7+. P-values for all individual analyses are indicated.

C) Schematic overview of the eight highest scoring MCODE sub-clusters from section A. The size and colour of the individual proteins corresponds to the amount of SUMOylation sites identified in the protein.



STRING Interaction Confidence: 0.4+

	Enrichment	Proteins	Interactions	P-value
SUMO & Ub	9.72	378	3316	<1.0E-15
SUMO & Ac	10.76	212	1119	<1.0E-15
SUMO & Ub & Ac	11.33	122	561	<1.0E-15

STRING Interaction Confidence: 0.7+

	Enrichment	Proteins	Interactions	P-value
SUMO & Ub	11.45	378	1672	<1.0E-15
SUMO & Ac	12.35	212	521	<1.0E-15
SUMO & Ub & Ac	13.21	122	284	<1.0E-15

Figure 8. SUMO and other PTMs compete for modification of lysines in a set of highly SUMOylated and functionally interconnected proteins.

A) STRING network analysis of all proteins where SUMOylation, ubiquitylation and acetylation occur on at least one of the same lysines. STRING interaction confidence was set at 0.7 or greater, and the size and colour of individual proteins corresponds to the amount of SUMOylation sites per protein.

B) As section A, but with SUMOylation and acetylation occurring on at least one of the same lysines.

C) As section A, but with SUMOylation and ubiquitylation occurring on at least one of the same

Cell culture & cell line generation
HeLa and U2-OS cells were cultured in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS and 100 U/mL penicillin and streptomycin (Invitrogen). HeLa cells stably expressing His10-SUMO-2-AllKR-Q87R were generated through lentiviral infection with a virus encoding His10-SUMO-2-AllKR-Q87R -IRES-GFP. Two weeks after infection, cells were fluorescence-sorted for a low expression level of GFP using a FACSAria II (BD Biosciences). Cells were passed through a 100 µm capillary at a pressure of 138 kPa, and selected for $7.5*10^2$-$3*10^3$ GFP BF 530/30-A intensity, which was 2.5-10 times higher than the background cellular auto-fluorescence of $3*10^2$.

Treatments, transfection & lentiviral infection
In order to accumulate SUMOylated proteins, in addition to mock-treated control cells; other batches of cells were treated with 10 µM MG-132 (Sigma) dissolved in DMSO for 7 hours, treated with 20 µM PR-619 (Millipore) dissolved in DMSO for 7 hours, or incubated at 43°C for 1 hour. For transfection, cells were cultured in DMEM lacking penicillin and streptomycin. Transfections were performed using 2.5 µg of polyethylenimine (PEI) per 1 µg of plasmid DNA, using 1 µg of DNA per 1 million cells. Transfection reagents were mixed in 150 mM NaCl and incubated for 15 minutes before adding it directly to the cells. Cells were split after 24 hours and investigated after 48 hours. Lentiviruses were generated essentially as described previously [48]. Infections were performed with a multiplicity of infection of 2 and using a concentration of 8 µg/mL polybrene in the medium. 24 hours after infection the medium was replaced. Cells were split 72 hours after infection, and passed every further 72 hours afterwards until processing.

Purification of His10-SUMO-2-AllKR-Q87R – Stage 1
Per single MS/MS run to identify SUMO-2 sites, one single fully confluent 15-cm dish of His10-SUMO-2-AllKR-Q87R (~20 million cells) was prepared. Cells were washed three times on the plate with ice-cold PBS, prior to scraping cells and collecting them in a 15 mL tube. Cells were centrifuged at 250 RCF and re-suspended in ice-cold PBS. A small aliquot of cells was kept separately and lysed in 2% SDS, 1% NP-40, 150 mM NaCl, and 50 mM TRIS, buffered at pH 7.5. All PBS was aspirated from the main batch of cells. Subsequently, the cell pellets were lysed in 10 pellet volumes of guanidine Lysis Buffer (6 M guanidine-HCl, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 8.0), by alternating vigorous vortexing for 5 seconds with vigorous shaking for 5 seconds, for 30 seconds total. Immediately following lysis, the samples were snap frozen using liquid nitrogen, and optionally stored at -80°C. Lysates were thawed at room temperature, and subjected to sonication using a microtip sonicator at a power of 30 Watts. Sonication bursts of 5 seconds per 5 mL lysate were used, mixing the sample and allowing it to cool to room temperature in between, up to a total sonication time of 15 seconds. Lysates were optionally equalized using the bicinchoninic acid assay (BCA, Pierce). Subsequently, lysates were supplemented by addition of imidazole to 50 mM and β-mercaptoethanol to 5 mM. 20 µL (dry volume) Ni-NTA agarose beads (Qiagen) were prepared per 1 mL lysate, by washing them four times with guanidine Lysis Buffer supplemented with imidazole to 50 mM and β-mercaptoethanol to 5 mM. The equilibrated beads were added to the lysates and allowed to tumble overnight at 4°C.

lysines.
D) Schematic overview of the amount of SUMOylation sites identified per protein in total, or within the core STRING network (Figure 7A), or within the three networks corresponding to sections A, B and C. Sites shared with the respective other PTMs are displayed in red. Additionally, the total amount of modification sites per protein for ubiquitylation and acetylation are shown.
E) Overview of STRING analyses depicted in sections A, B and C. All analyses were performed at STRING confidence of 0.4+ and STRING confidence of 0.7+. P-values for all individual analyses are indicated.

Following overnight incubation, beads were pelleted by centrifugation at 500 RCF, and washed using at least 5 bead volumes of the following Wash Buffers in order. Wash Buffer 1: 6 M guanidine-HCl, 0.1% Triton X-100, 10 mM imidazole, 5 mM β-mercaptoethanol, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 8.0. During the first wash, beads were transferred to 1.7 mL graduated siliconized microcentrifuge tubes (Sigma), and from this point on siliconized tubes were used exclusively (unless stated otherwise) to prevent loss of beads, proteins or peptides. Wash Buffer 2: 8 M urea, 0.1% Triton X-100, 10 mM imidazole, 5 mM β-mercaptoethanol, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 8.0. Wash Buffer 3: 8 M urea, 10 mM imidazole, 5 mM β-mercaptoethanol, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 6.3. This wash was performed while mixing for 15 minutes, and tubes were changed afterwards. Wash Buffer 4: 8 M urea, 5 mM β-mercaptoethanol, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 6.3. This wash was performed while mixing for 15 minutes, and tubes were changed afterwards. Wash Buffer 5 (same as Wash Buffer 4): 8 M urea, 5 mM β-mercaptoethanol, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 6.3. This wash was performed while mixing for 15 minutes, and tubes were changed afterwards. Subsequently, all Wash Buffer was removed from the beads, and proteins were elution using one bead volume of Elution Buffer (7 M urea, 500 mM imidazole, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 7.0). This elution was performed while shaking at 1100 RPM at room temperature, for 30 minutes. The elution procedure was repeated another two times, and all three elutions were carefully aspirated using a flattened gel loading tip. Subsequently, all elutions were pooled and passed through 0.45 µM filters (Ultrafree, Millipore, twice pre-washed with Elution Buffer) to clear any remaining beads from the samples. Next, samples were concentrated over 100 kDa cut-off spin filters (Vivacon 500, Sartorius Stedim, pre-washed with Elution Buffer), using a temperature-controlled centrifuge set to 20°C and centrifuging at 8,000 RCF. Concentration rate was approximately 50 µL per minute and concentration was performed until <10-50 µL of sample remained. Note that while the filters are rated for 100 kDa, proteins smaller than 25 kDa will not pass through the filter due to their denatured state in 7 M urea, and thus only SUMO attached to its target proteins will be retained on the filter. After concentration, the proteins remaining on the filters were washed once using 250 µL of Elution Buffer minus imidazole, and re-concentrated. Final concentrated SUMOylated proteins were removed from the filters by centrifuging the filters while placed upside down into an open regular 1.5 mL microcentrifuge tube (Eppendorf). Centrifugation steps were performed over four 10-second steps, in-between rotating the filters 180 degrees to allow efficient removal of all SUMOylated proteins. Afterwards, proteins were transferred to siliconized tubes. At this point, samples were snap frozen and stored at -80°C. The single-stage purified SUMOylated proteins are compatible with both in-gel and in-solution digestion protocols and subsequent mass spectrometric analysis, or may be target-specifically investigated by SDS-PAGE and immunoblotting analysis. Alternatively, re-purification of SUMOylated peptides is performed by following Stage 2 of the protocol.

Purification of His10-SUMO-2-AllKR-Q87R – Stage 2

The concentration of purified SUMOylated proteins was determined using the Bradford essay (Quick Start, BioRad). After determination of the amount of SUMOylated proteins, volumes of the samples were increased to 50 µL using Elution Buffer (minus imidazole), and β-mercaptoethanol was added to a concentration of 10 mM. Next, Sequencing Grade Endoproteinase Lys-C (Promega) was added to the samples in a 1:25 enzyme-to-protein ratio, and incubated for 4 hours at room temperature, still, and in the dark. Subsequently, another 10 mM of fresh β-mercaptoethanol was added to the samples, followed by an additional amount of Lys-C equal to the first amount. The second incubation was performed overnight, at room temperature, still and in the dark. Following complete digestion of the SUMOylated proteins by Lys-C, the digests were transferred to 15 mL tubes and diluted with an amount of guanidine Lysis Buffer equal to half the amount used to lyse the initial cell pellet. The samples were then supplemented with addition of imidazole to 50 mM, and β-mercaptoethanol to 5 mM. Next, 40 µL (dry volume) Ni-NTA agarose beads (Qiagen) were prepared per 1 mL sample, by washing them four times with guanidine Lysis Buffer supplemented with imidazole to 50 mM and β-mercaptoethanol to 5 mM. The equilibrated beads were added to the lysates and allowed to

tumble at 4°C for 5 hours. Following incubation, beads were pelleted by centrifugation at 500 RCF, and washed using at least 5 bead volumes of the following Wash Buffers in order. Wash Buffer 1: 6 M guanidine-HCl, 0.1% Triton X-100, 10 mM imidazole, 5 mM β-mercaptoethanol, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 8.0. Wash Buffer 2: 8 M urea, 0.1% Triton X-100, 10 mM imidazole, 5 mM β-mercaptoethanol, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 8.0. Wash Buffer 3: 8 M urea, 10 mM imidazole, 5 mM β-mercaptoethanol, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 6.3. This wash was performed while mixing for 15 minutes, and tubes were changed afterwards. Wash Buffer 4: 8 M urea, 5 mM β-mercaptoethanol, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 6.3. This wash was performed while mixing for 15 minutes, and tubes were changed afterwards. Wash Buffer 5 (same as Wash Buffer 4): 8 M urea, 5 mM β-mercaptoethanol, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 6.3. This wash was performed while mixing for 15 minutes, and tubes were changed afterwards. Subsequently, all Wash Buffer was removed from the beads, and proteins were elution using one bead volume of Elution Buffer (7 M urea, 500 mM imidazole, 100 mM sodium phosphate, 10 mM TRIS, buffered at pH 7.0). This elution was performed while shaking at 1100 RPM at room temperature, for 20 minutes. The elution procedure was repeated another two times, and all three elutions were carefully aspirated using a flattened gel loading tip. Subsequently, all elutions were pooled and passed through 0.45 µM filters (Ultrafree, Millipore, twice pre-washed with Elution Buffer) to clear any remaining beads from the samples. Next, samples were concentrated on 10 kDa cut-off spin filters (Vivacon 500, Sartorius Stedim, pre-washed with Elution Buffer), using a temperature-controlled centrifuge set to 20°C and centrifuging at 14,000 RCF. Concentration rate was approximately 10 µL per minute and concentration was performed until <10-25 µL of sample remained. After concentration, the proteins remaining on the filters were washed twice using 250 µL of Elution Buffer minus imidazole, and re-concentrated. Final concentrated SUMOylated peptides were removed from the filters by centrifuging the filters while placed upside down into an open regular 1.5 mL microcentrifuge tube (Eppendorf). Centrifugation steps were performed over four 10-second steps, in-between rotating the filters 180 degrees to allow efficient removal of all SUMOylated peptides. Afterwards, SUMOylated peptides were transferred to siliconized tubes. At this point, samples were snap frozen and stored at -80°C. The double-purified SUMOylated peptides are compatible with in-solution digestion protocols and subsequent mass spectrometric analysis aimed at determination of specific sites of protein SUMOylation.

In-solution digestion

SUMOylated peptides were supplemented with ammonium bicarbonate (ABC) to 50 mM. Subsequently, dithiothreitol (DTT) was added to a concentration of 1 mM, and samples were left to incubate at room temperature for 30 minutes. Next, chloroacetamide was added to a concentration of 5 mM, and samples were incubated at room temperature for 30 minutes. After alkylation, another 5 mM of DTT was added, and samples were left to incubate at room temperature for 30 minutes. At this point, samples were gently diluted 4-fold using 50 mM ABC. Subsequently, an amount of Sequencing Grade Modified Trypsin (Promega) was added equal to 25% of the Lys-C initially used in a single digestion step. Digestion with trypsin was performed overnight, at room temperature, still and in the dark.

LC-MS/MS analysis

In-solution digested peptides were cleaned, desalted and concentrated on triple-disc C18 reverse phase StageTips [49], before being eluted twice with 25 µL 80% acetonitrile in 0.1% formic acid. Desalted peptides were vacuum centrifuged at room temperature until 10% of the original volume remained, prior to online nanoflow liquid chromatography-tandem mass spectrometry. The analysis of in-solution digested peptides was performed using an EASY-nLC system (Proxeon) connected to a Q-Exactive (Thermo) using Higher-Collisional Dissociation (HCD) fragmentation. Separation of peptides was performed using 20 cm long analytical columns (ID 75 µm, Polymicro Avantes) packed in-house with 1.8 µm C18 beads (Reprospher 100), using a 120 minute gradient from 5% to 75% acetonitrile in 0.1% formic acid and a flow rate of 250 nL per minute. The mass spectrometer was operated in data-dependent acquisition mode using a top 10 method. Full-scan MS spectra were acquired with a

target value of 3E6 and a resolution of 70,000, with a scan range from 300 to 1,750 m/z. HCD tandem MS/MS spectra were acquired using a target value of 1E5, a resolution of 17,500, and a normalized collision energy of 25%. All charges lower than 2 and higher than 6 were rejected, and all unknown charges were rejected. The underfill ratio was set to 0.1%, and a dynamic exclusion of 20 seconds was used. Alternatively, the underfill ratio was set to 1.0% with the dynamic exclusion time set to 10 seconds.

Data processing

MaxQuant version 1.4.1.2 was used to analyse all RAW data [50, 51]. MS/MS spectra were filtered and deisotoped and the 15 most abundant fragments for each 100 m /z were retained. MS/MS spectra were filtered for a mass tolerance of 6 ppm for precursor masses, and a mass tolerance of 20 ppm was used for fragment ions. Peptide and protein identification was performed through matching the identified MS/MS spectra versus a target/decoy version of the complete human Uniprot database, in addition to a database of commonly observed mass spectrometry contaminants. Up to 3 missed tryptic cleavages were allowed. Cysteine carbamidomethylation was set as a fixed peptide modification. Protein N-terminal acetylation, peptide N-terminal carbamylation, methionine oxidation, QQTGG and pyro-QQTGG were set as variable peptide modifications. QQTGG was set as a lysine-specific modification, with a monoisotopic mass of 471.20776, and not allowed to occur at the C-terminal end of peptides. Pyro-QQTGG may spontaneously form out of the tryptic QQTGG remnant as a result of cyclization of the N-terminal glutamine. Pyro-QQTGG (pQQTGG) was set as s lysine-specific modification, with a monoisotopic mass of 454.18121, and not allowed to occur at the C-terminal end of peptides. In order to increase identification confidence, diagnostic peaks were searched within MS/MS spectra corresponding to SUMOylated peptides. For QQTGG; b5-QQTGG, b4-QQTG, b3-QQT and b2-QQ were accepted as diagnostic peaks. For pQQTGG; b5-pQQTGG, b4-pQQTG, b3-QQT and b2-QQ were accepted as diagnostic peaks. In addition to the above variable modifications, lysine ubiquitylation, lysine acetylation, and serine/threonine/tyrosine phosphorylation were individually added as further variable modifications in separate searches. For protein identification, peptides with all above variable modifications were accepted, and protein identification by at least one single SUMO-modified unique peptide was performed. Peptides were accepted with a minimum length of 6 amino acids, a maximum size of 5 kDa, and a maximum charge of 6. The processed data was filtered by posterior error probability (PEP) to achieve a protein false discovery rate (FDR) of below 1%, a peptide-spectrum match FDR of below 1%, and in addition a site decoy fraction of 1% was set. SUMO-site peptides were additionally filtered to have an Andromeda score of at least 20, a localization score of at least 20, and a localization probability of at least 70%.

IceLogo generation

For SUMOylation site analysis of all identified sites, amino acid sequence windows of 15 amino acids downstream as well as 15 amino acids upstream of the modified lysine were extracted from the corresponding proteins, generating 31 amino acid (31AA) sequence windows. IceLogo software version 1.2 [52] was used to overlay 31AA SUMOylation site sequence windows in order to generate a consensus sequence, and compensated against the expected random occurrence frequencies of amino acids across all human proteins (iceLogo). Alternatively, subsets of modification sites were compared directly to other subsets of modification sites, generating consensus sequences showing differential occurrence of amino acids between the subsets (SubLogo). Heatmaps were generated in a similar fashion to IceLogos. For IceLogos, SubLogos and heatmaps, all amino acids shown as enriched or depleted are significant with $p < 0.05$, unless otherwise indicated.

Secondary structure analysis

31AA sequence windows corresponding to SUMOylation sites were analysed using NetSurfP version 1.1 [53] in order to predict localized protein surface accessibility and secondary structure. For each amino acid within the 31AA windows, probabilities for alpha-helix, beta-strand, and coil were calculated. Additionally, amino acids were predicted to be buried or solvent-exposed, with a threshold of

25% exposure. In addition to calculating the properties of the central lysine, the average properties of the central lysine +/- 5 AA were calculated, as well as the average properties of the entire 31AA sequence windows. As a reference set, random lysines were extracted from all proteins identified to be SUMOylated, and 31AA sequence windows were assigned to these random sites (4,839). Duplicate sequence windows were discarded. The reference set was processed identically as compared to the SUMOylation sites set.

SUMOylation and PTM site overlap analysis

For comparative analysis, all 103 SUMOylation sites identified by Matic et al. [31] were assigned to matching Uniprot IDs and 31AA sequence windows were parsed. Furthermore, a selection was made from all PhosphoSitePlus (PSP; PhosphoSitePlus®, www.phosphosite.org, [36]) SUMOylation sites, where 141 SUMOylation sites identified by MS/MS were retained. Additionally, 31,369 MS/MS-identified ubiquitylation sites, 15,794 acetylation sites, and 1,543 lysine-methylation sites were extracted from PSP, and 31AA sequence windows were assigned. From within each data set, duplicate sequence windows were removed. Perseus software was used to generate a matrix where all sequence windows from all PTMs were cross-referenced to each other. Corresponding parental proteins were assigned, and multiple modifications targeting the same lysines within the same proteins were further investigated using STRING network analysis.

Term enrichment analysis

Statistical enrichment analysis for protein and gene properties was performed using Perseus software. The human proteome was annotated with Gene Ontology terms [54], including Biological Processes (GOBP), Molecular Functions (GOMF), and Cellular Compartments (GOCC). Additional annotation was performed with the Kyoto Encyclopaedia of Genes and Genomes (KEGG)[55], Protein families (Pfam)[56], Gene Set Enrichment Analysis (GSEA)[57], Keywords, and Comprehensive Resource of Mammalian Protein Complexes (CORUM)[58] terms for comparative enrichment analysis. SUMOylated proteins or subgroups of SUMOylated proteins were compared by annotation terms to the entire human proteome, using Fisher Exact Testing. Benjamini and Hochberg FDR was applied to p-values to correct for multiple hypotheses testing, and final corrected p-values were filtered to be less than 2%.

STRING network analysis

STRING network analysis was performed using the online STRING database [59], using all SUMOylated proteins or subgroups of SUMOylation proteins as input. Protein interaction enrichment was performed based on the amount of interactions in the networks, as compared to the randomly expected amount of interactions, with both variables directly derived from the STRING database output. Enrichment analysis was performed allowing network interactions at medium or greater confidence (p>0.4), and separately by allowing network interactions at high or greater confidence (p>0.7). Visualization of interaction networks was performed using Cytoscape version 3.0.2 [60], and highly interconnected sub-clusters were localized using the Cytoscape plugin Molecular Complex Detection (MCODE) version 1.4.0-beta2 [61]. For visualization of the core SUMOylation networks, a STRING confidence of 0.7+ was used. For visualization of term enrichment sub-networks, a STRING confidence of 0.4+ was used. For sub-cluster localization the following settings were used; a degree cut-off of 3, a node score cut-off of 0.2, a maximum depth of 3, a K-Core of 5, and haircut.

Phylogenetic conservation analysis

Perseus software was used to annotate the human proteome using phylogenetic conservation scores from the Ensembl Database [62], which contains phylogenetic orthologous information ranging over many eukaryotic species. Only transcripts and genes marked as 'known' were extracted from the Ensembl Database, and only unique results were retrieved. Phylogenetic conservation target scores from 62 eukaryotic species were mapped by their human Ensembl Protein ID to human Uniprot IDs. Subsequently, all proteins identified as SUMOylated in this work were aligned to the annotated human proteome. All MS/MS-identified phosphorylation, acetylation, ubiquitylation and methylation sites

were extracted from PSP, mapped to their respective parental proteins, and aligned to the annotated human proteome. Phylogenetic conservation scores were calculated for the entire human proteome as compared to all individual 62 eukaryotic species as a reference, and scores were separately calculated for SUMOylation in addition to all other PTMs. For phylogenetic conservation within orthologues, all human proteins lacking an orthologue were excluded from the analysis on a per-species basis. For phylogenetic conservation outside of orthologues, all human proteins lacking an orthologue were set to 0% conservation.

SUMO target protein overlap analysis

For SUMO target protein analysis, all proteins identified in this work with at least one SUMO site were selected. Proteins successfully identified by various peptides, but lacking a SUMO-peptide, were ignored. For comparative analysis, identified SUMO target proteins were compared to other studies. SUMO-2 target proteins from Becker et al. [45] were selected by a SUMO-2 / Control ratio of greater than 2, and additionally filtered for a SUMO-2 intensity of greater than 10% as compared to SUMO-1 and control. All SUMO target proteins (SUMO-1 and SUMO-2) from Becker et al. were selected by a SUMO-1 / Control ratio or a SUMO-2 / Control ratio of greater than 2. SUMO-2 target proteins from Golebiowski et al. [28] were selected for a SUMO-2 / Control SILAC ratio of greater than 1.5. Heat-shock inducible SUMO-2 target proteins from Golebiowski et al. were filtered by a SUMO-2-HEAT / SUMO-2 SILAC ratio of greater than 1.5, in addition to a (SUMO-2-HEAT / SUMO-2 SILAC ratio) * (SUMO-2 / Control SILAC ratio) of greater than 1.5. Putative poly-SUMO-modified proteins from Bruderer et al. [44] were selected as all identified proteins. Putative poly-SUMO-modified proteins from Bruderer et al. were additionally filtered for an observed molecular weight shift of 2 times the molecular weight of SUMO or greater, as compared to the expected protein molecular weight. SUMO-2 target proteins from Matic et al. [31] were considered to be all proteins in which at least one site of SUMOylation was identified. Where required, gene IDs were mapped to the corresponding Uniprot IDs. Additionally, where multiple Uniprot IDs were listed for a singular protein identification, a major Uniprot ID was selected by selecting the first Uniprot ID in the list starting with a P, or otherwise the first Uniprot ID starting with a Q, or otherwise the first Uniprot ID in the list. Perseus [63] software was used to generate a complete gene list for all known human proteins, and all identified SUMO target proteins from our study as well as the above-mentioned studies were aligned based on matching Uniprot IDs.

Primary antibodies

Primary antibodies used in this study were Mouse α SUMO-2 (8A2, Abcam), Rabbit α SUMO-2 (raised against the C-terminal part of SUMO-2), Mouse α His (HIS-1, Sigma), Mouse α HA (HA.11, Sanbio), Rabbit α SART-1 (raised against SART-1 peptides).

Electrophoresis and immunoblotting

Protein samples were size-fractionated on Novex 4-12% Bis-Tris gradient gels using MOPS running buffer (Invitrogen). Size-separated proteins were transferred to Hybond-C membranes (Amersham Biosciences) using a submarine system (Invitrogen). Gels were Coomassie stained according to manufacturer's instructions (Invitrogen). Membranes were stained for total protein loading using 0.1% Ponceau-S in 5% acetic acid (Sigma). Membranes were blocked using PBS containing 0.1% Tween-20 (PBST) and 5% milk powder for one hour. Subsequently, membranes were incubated with primary antibodies as indicated, in blocking solution. Incubation with primary antibody was performed overnight at 4°C. Afterwards, membranes were washed three times with PBST and briefly blocked again with blocking solution. Next, membranes were incubated with secondary antibodies (donkey-anti-mouse-HRP or rabbit-anti-goat-HRP) for one hour, before washing three times with PBST and two times with PBS. Membranes were then treated with ECL2 (Pierce) as per manufacturer's instructions, and chemiluminescence was captured using Biomax XAR film (Kodak).

Microscopy

Cells were seeded on glass coverslips, and fixed 24 hours later for 10 minutes in 3.7% paraformalde-

hyde in PHEM buffer (60 mM PIPES, 25 mM HEPES, 10 mM EGTA, 2 mM MgCl2 pH 6.9) at 37°C. After washing with PBS, cells were permeabilized with 0.1% Triton-X100 for 10 minutes, washed with PBST, and blocked using TNB (100 mM TRIS pH 7.5, 150 mM NaCl, 0.5% Blocking Reagent (Roche)) for 30 minutes. Cells were incubated with primary antibody as indicated, in TNB for one hour. Subsequently cells were washed five times with PBST, and incubated with secondary antibodies (Goat α Mouse Alexa 594 (Invitrogen)) in TNB for one hour. Next, cells were washed five times with PBST and dehydrated using alcohol, prior to embedding them in Citifluor (Agar Scientific) containing 400 ng/µL DAPI (Sigma) and sealing the slides with nail varnish. Images were recorded on a Leica SP5 confocal microscope system using 488 nm and 561 nm lasers for excitation, a 63X lens for magnification, and were analysed with Leica confocal software.

ACKNOWLEDGMENTS

This work is supported by the Netherlands Organization for Scientific Research (NWO) (A.C.O.V.), the European Research Council (A.C.O.V.), and the Max Planck Society (M.M). The authors declare no conflict of interest.

Reference List

1. Jackson, S. P., and Durocher, D. (2013) Regulation of DNA damage responses by ubiquitin and SUMO. *Mol. Cell* 49, 795-807
2. Ulrich, H. D., and Walden, H. (2010) Ubiquitin signalling in DNA replication and repair. *Nat. Rev. Mol. Cell Biol.* 11, 479-489
3. Gill, G. (2005) Something about SUMO inhibits transcription. *Curr. Opin. Genet. Dev.* 15, 536-541
4. Hoege, C., Pfander, B., Moldovan, G. L., Pyrowolakis, G., and Jentsch, S. (2002) RAD6-dependent DNA repair is linked to modification of PCNA by ubiquitin and SUMO. *Nature* 419, 135-141
5. Stelter, P., and Ulrich, H. D. (2003) Control of spontaneous and damage-induced mutagenesis by SUMO and ubiquitin conjugation. *Nature* 425, 188-191
6. Morris, J. R. et al. (2009) The SUMO modification pathway is involved in the BRCA1 response to genotoxic stress. *Nature* 462, 886-890
7. Galanty, Y. et al. (2009) Mammalian SUMO E3-ligases PIAS1 and PIAS4 promote responses to DNA double-strand breaks. *Nature* 462, 935-939
8. Flotho, A., and Melchior, F. (2013) Sumoylation: a regulatory protein modification in health and disease. *Annu. Rev. Biochem.* 82, 357-385
9. Nacerddine, K. et al. (2005) The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev. Cell* 9, 769-779
10. Mahajan, R., Delphin, C., Guan, T., Gerace, L., and Melchior, F. (1997) A small ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore complex protein RanBP2. *Cell* 88, 97-107
11. Pfander, B., Moldovan, G. L., Sacher, M., Hoege, C., and Jentsch, S. (2005) SUMO-modified PCNA recruits Srs2 to prevent recombination during S phase. *Nature* 436, 428-433
12. Papouli, E. et al. (2005) Crosstalk between SUMO and ubiquitin on PCNA is mediated by recruitment of the helicase Srs2p. *Mol. Cell* 19, 123-133
13. Kessler, J. D. et al. (2012) A SUMOylation-dependent transcriptional subprogram

is required for Myc-driven tumorigenesis. *Science* 335, 348-353
14. Bettermann, K., Benesch, M., Weis, S., and Haybaeck, J. (2012) SUMOylation in carcinogenesis. *Cancer Lett.* 316, 113-125
15. Olsen, J. V., and Mann, M. (2013) Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol. Cell Proteomics.*
16. Vertegaal, A. C. (2011) Uncovering ubiquitin and ubiquitin-like signaling networks. *Chem. Rev.* 111, 7923-7940
17. Huttlin, E. L. et al. (2010) A tissue-specific atlas of mouse protein phosphorylation and expression. *Cell* 143, 1174-1189
18. Olsen, J. V. et al. (2006) Global, in vivo, and site-specific phosphorylation dynamics in signaling networks. *Cell* 127, 635-648
19. Choudhary, C. et al. (2009) Lysine acetylation targets protein complexes and co-regulates major cellular functions. *Science* 325, 834-840
20. Guo, A. et al. (2013) Immunoaffinity Enrichment and Mass Spectrometry Analysis of Protein Methylation. *Mol. Cell Proteomics.*
21. Zielinska, D. F., Gnad, F., Wisniewski, J. R., and Mann, M. (2010) Precision mapping of an in vivo N-glycoproteome reveals rigid topological and sequence constraints. *Cell* 141, 897-907
22. Kim, D. Y., Scalf, M., Smith, L. M., and Vierstra, R. D. (2013) Advanced proteomic analyses yield a deep catalog of ubiquitylation targets in Arabidopsis. *Plant Cell* 25, 1523-1540
23. Emanuele, M. J. et al. (2011) Global identification of modular cullin-RING ligase substrates. *Cell* 147, 459-474
24. Povlsen, L. K. et al. (2012) Systems-wide analysis of ubiquitylation dynamics reveals a key role for PAF15 ubiquitylation in DNA-damage bypass. *Nat. Cell Biol.* 14, 1089-1098
25. Wagner, S. A. et al. (2011) A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles. *Mol. Cell Proteomics* 10, M111
26. Kim, W. et al. (2011) Systematic and quantitative assessment of the ubiquitin-modified proteome. *Mol. Cell* 44, 325-340
27. Hickey, C. M., Wilson, N. R., and Hochstrasser, M. (2012) Function and regulation of SUMO proteases. *Nat. Rev. Mol. Cell Biol.* 13, 755-766
28. Golebiowski, F. et al. (2009) System-wide changes to SUMO modifications in response to heat shock. *Sci. Signal.* 2, ra24
29. Galisson, F. et al. (2011) A novel proteomics approach to identify SUMOylated proteins and their modification sites in human cells. *Mol. Cell Proteomics* 10, M110
30. Lamoliatte, F. et al. (2013) Targeted Identification of SUMOylation Sites in Human Proteins Using Affinity Enrichment and Paralog-specific Reporter Ions. *Mol. Cell Proteomics* 12, 2536-2550
31. Matic, I. et al. (2010) Site-specific identification of SUMO-2 targets in cells reveals an inverted SUMOylation motif and a hydrophobic cluster SUMOylation motif. *Mol. Cell* 39, 641-652
32. Saitoh, H., and Hinchey, J. (2000) Functional heterogeneity of small ubiquitin-related protein modifiers SUMO-1 versus SUMO-2/3. *J. Biol. Chem.* 275, 6252-6258
33. Vertegaal, A. C. et al. (2004) A proteomic study of SUMO-2 target proteins. *J. Biol. Chem.* 279, 33791-33798
34. Schimmel, J. et al. (2008) The ubiquitin-proteasome system is a key component of the SUMO-2/3 cycle. *Mol. Cell Proteomics* 7, 2107-2122
35. Tatham, M. H., Matic, I., Mann, M., and Hay, R. T. (2011) Comparative proteomic analysis

identifies a role for SUMO in protein quality control. *Sci. Signal.* 4, rs4

36. Hornbeck, P. V. et al. (2012) PhosphoSitePlus: a comprehensive resource for investigating the structure and function of experimentally determined post-translational modifications in man and mouse. *Nucleic Acids Res.* 40, D261-D270
37. Desterro, J. M., Rodriguez, M. S., and Hay, R. T. (1998) SUMO-1 modification of Ikappa-Balpha inhibits NF-kappaB activation. *Mol. Cell* 2, 233-239
38. Hietakangas, V. et al. (2006) PDSM, a motif for phosphorylation-dependent SUMO modification. *Proc. Natl. Acad. Sci. U. S. A* 103, 45-50
39. Rodriguez, M. S., Dargemont, C., and Hay, R. T. (2001) SUMO-1 conjugation in vivo requires both a consensus modification motif and nuclear targeting. *J. Biol. Chem.* 276, 12654-12659
40. Gareau, J. R., and Lima, C. D. (2010) The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition. *Nat. Rev. Mol. Cell Biol.* 11, 861-871
41. Matunis, M. J., Coutavas, E., and Blobel, G. (1996) A novel ubiquitin-like modification modulates the partitioning of the Ran-GTPase-activating protein RanGAP1 between the cytosol and the nuclear pore complex. *J. Cell Biol.* 135, 1457-1470
42. Bernier-Villamor, V., Sampson, D. A., Matunis, M. J., and Lima, C. D. (2002) Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1. *Cell* 108, 345-356
43. Huang, T. T., Wuerzberger-Davis, S. M., Wu, Z. H., and Miyamoto, S. (2003) Sequential modification of NEMO/IKKgamma by SUMO-1 and ubiquitin mediates NF-kappaB activation by genotoxic stress. *Cell* 115, 565-576
44. Bruderer, R. et al. (2011) Purification and identification of endogenous polySUMO conjugates. *EMBO Rep.* 12, 142-148
45. Becker, J. et al. (2013) Detecting endogenous SUMO targets in mammalian cells and tissues. *Nat. Struct. Mol. Biol.* 20, 525-531
46. Roukens, M. G. et al. (2008) Identification of a new site of sumoylation on Tel (ETV6) uncovers a PIAS-dependent mode of regulating Tel function. *Mol. Cell Biol.* 28, 2342-2357
47. Vellinga, J. et al. (2006) A system for efficient generation of adenovirus protein IX-producing helper cell lines. *J. Gene Med.* 8, 147-154
48. Tiscornia, G., Singer, O., and Verma, I. M. (2006) Production and purification of lentiviral vectors. *Nat. Protoc.* 1, 241-245
49. Rappsilber, J., Mann, M., and Ishihama, Y. (2007) Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. *Nat. Protoc.* 2, 1896-1906
50. Cox, J. et al. (2011) Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome. Res.* 10, 1794-1805
51. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367-1372
52. Colaert, N., Helsens, K., Martens, L., Vandekerckhove, J., and Gevaert, K. (2009) Improved visualization of protein consensus sequences by iceLogo. *Nat. Methods* 6, 786-787
53. Petersen, B., Petersen, T. N., Andersen, P., Nielsen, M., and Lundegaard, C. (2009) A generic method for assignment of reliability scores applied to solvent accessibility

predictions. *BMC. Struct. Biol.* 9, 51
54. Ashburner, M. et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.* 25, 25-29
55. Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012) KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, D109-D114
56. Punta, M. et al. (2012) The Pfam protein families database. *Nucleic Acids Res.* 40, D290-D301
57. Subramanian, A. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A* 102, 15545-15550
58. Ruepp, A. et al. (2008) CORUM: the comprehensive resource of mammalian protein complexes. *Nucleic Acids Res.* 36, D646-D650
59. Franceschini, A. et al. (2013) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* 41, D808-D815
60. Shannon, P. et al. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498-2504
61. Bader, G. D., and Hogue, C. W. (2003) An automated method for finding molecular complexes in large protein interaction networks. *BMC. Bioinformatics.* 4, 2
62. Flicek, P. et al. (2013) Ensembl 2013. *Nucleic Acids Res.* 41, D48-D55
63. Cox, J., and Mann, M. (2012) 1D and 2D annotation enrichment: a statistical method integrating quantitative proteomics with complementary high-throughput data. *BMC. Bioinformatics.* 13 Suppl 16, S12

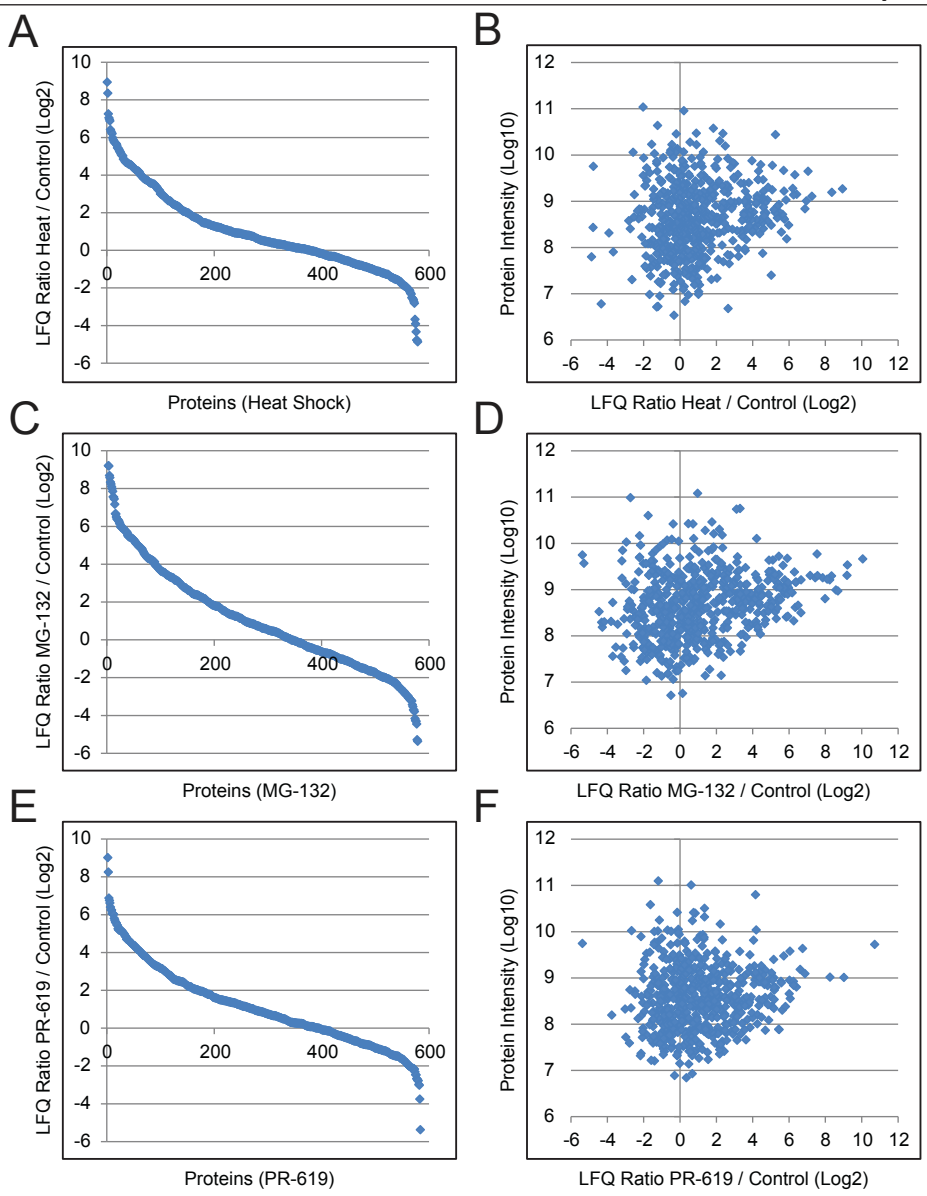


Figure S1. SUMOylation of proteins changes drastically upon heat shock, MG-132 treatment, and PR-619 treatment.

A) Schematic overview of all SUMOylation proteins identified to be differentially regulated after heat shock, by using Label Free Quantification (LFQ).

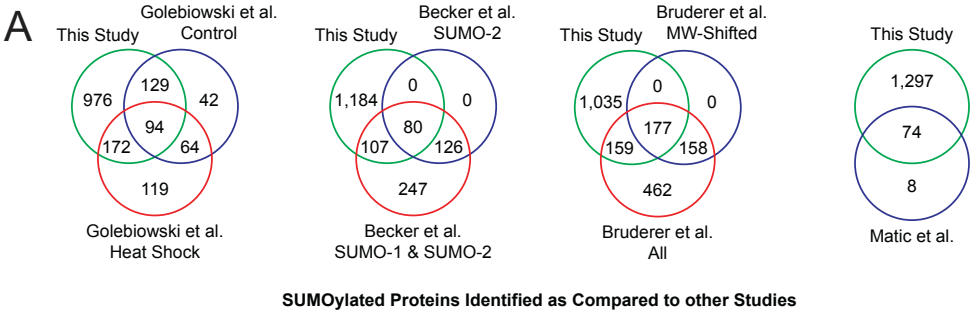
B) Schematic overview of all SUMOylation proteins identified to be differentially regulated after heat shock, with the protein LFQ ratio plotted against the total peptide peak intensity of each protein.

C) As section A, but for MG-132 treatment.

D) As section B, but for MG-132 treatment.

E) As section A, but for PR-619 treatment.

F) As section B, but for PR-619 treatment.



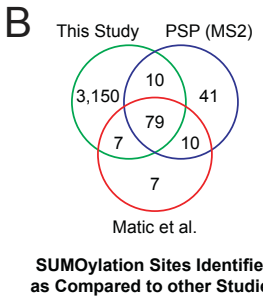


Figure S2. A comparison of SUMOylated proteins and sites identified in this work, to proteins and sites previously identified in other SUMOylation studies.

A) Schematic overview of SUMOylated proteins identified in this study, as compared to previously published studies. From Golebiowski et al. [28], control SUMOylated proteins and heat shock SUMOylated proteins were classed separately. From Bruderer et al. [44], all SUMOylation proteins and molecular-weight corrected SUMOylated proteins were classed separately. From Becker et al. [45], SUMO-2 modified proteins and all SUMO-1 and SUMO-2 modified proteins were classed separately.

B) Schematic overview of SUMOylation sites identified in this study, as compared to Matic et al. [31], as well as all known MS/MS-identified SUMOylation sites from PhosphoSitePlus (PSP).

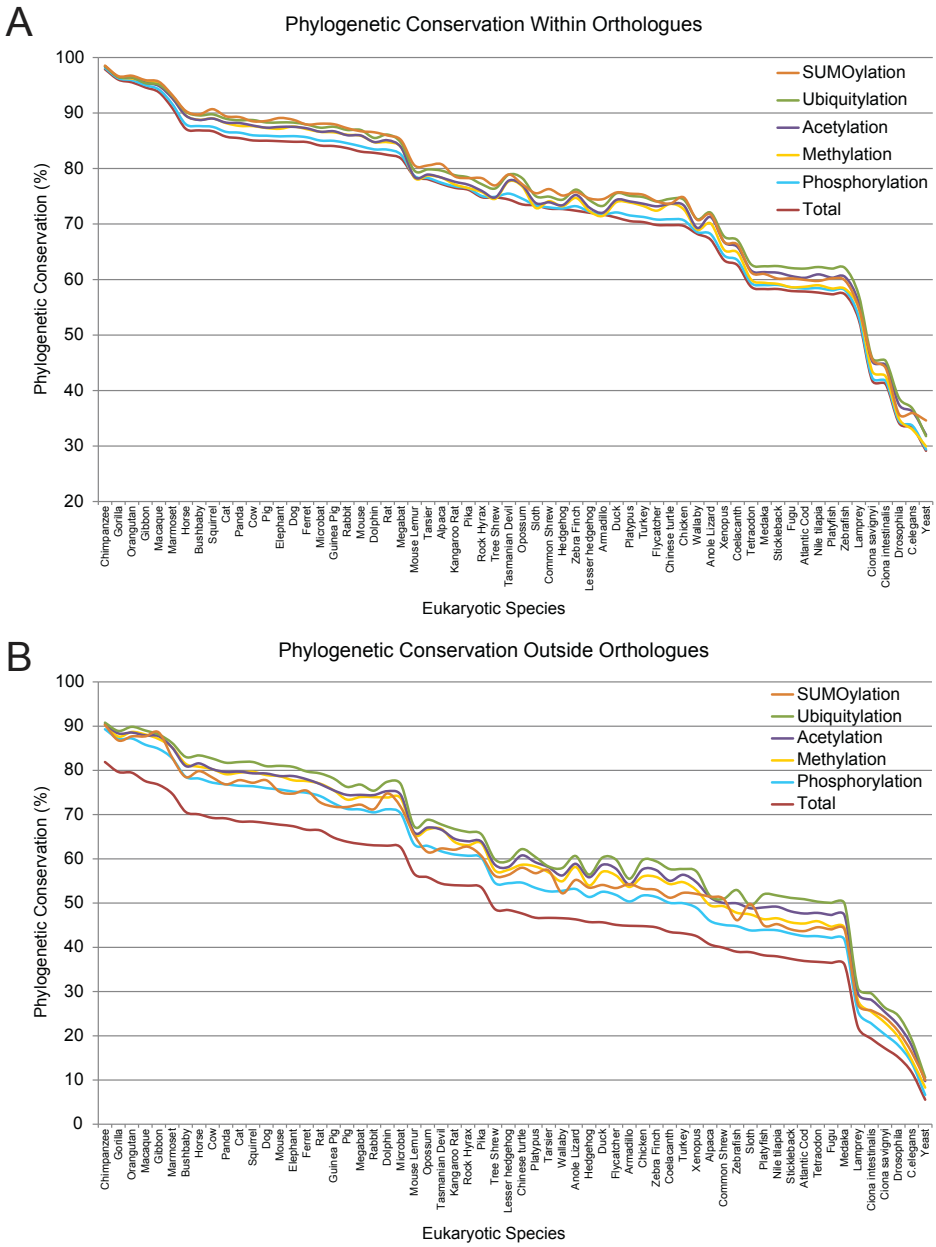


Figure S3. SUMOylation occurs on phylogenetically highly conserved proteins.
A) Phylogenetic conservation analysis (within orthologues) of SUMOylation, as compared to the other major PTMs ubiquitylation, acetylation, methylation and phosphorylation, as well as all human proteins (total). The conservation line descends by average percentage of conservation from human to 62 Ensembl-annotated eukaryotic organisms. The analysis was performed within orthologues, and only genes that have an orthologue were considered on a per-organism basis.
B) Phylogenetic conservation analysis outside of orthologues. Synonymous to section A, with the exception that genes lacking an orthologue were considered to be 0% conserved on a per-organism basis.

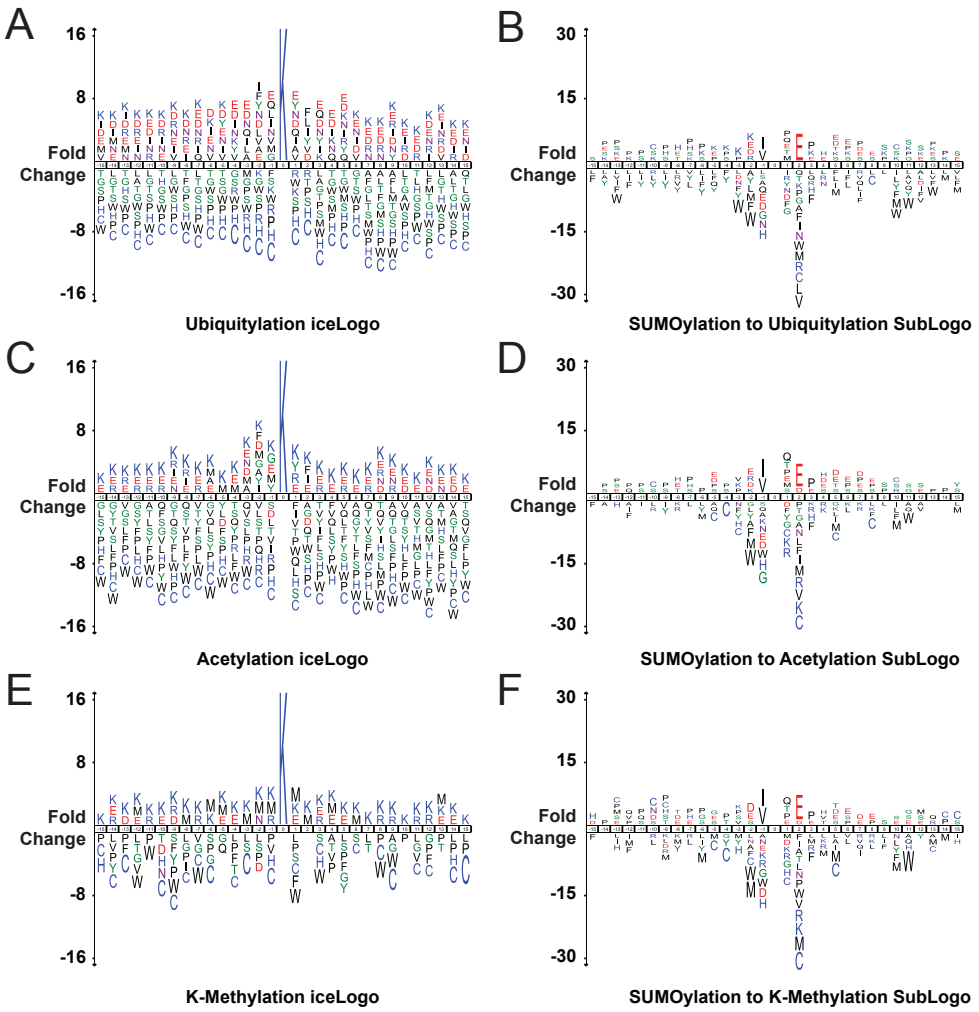


Figure S4. SUMO preferentially occurs on a consensus motif, as opposed to the other major lysine modifications.

A) IceLogo of all known ubiquitylation sites, within a sequence window ranging from -15 to +15 of the modified lysine. Amino acids occurring at an elevated rate as opposed to randomly expected are displayed above the centre, depleted amino acids are displayed below. The height of the amino acid letters correlates to fold-change. All amino acid changes are significant with $p < 0.05$.

B) Sub-IceLogo comparing all SUMOylation control sites versus all ubiquitylation sites.

C) As section A, but for acetylation.

D) As section B, but for acetylation.

E) As section A, but for lysine-methylation.

F) As section B, but for lysine-methylation.

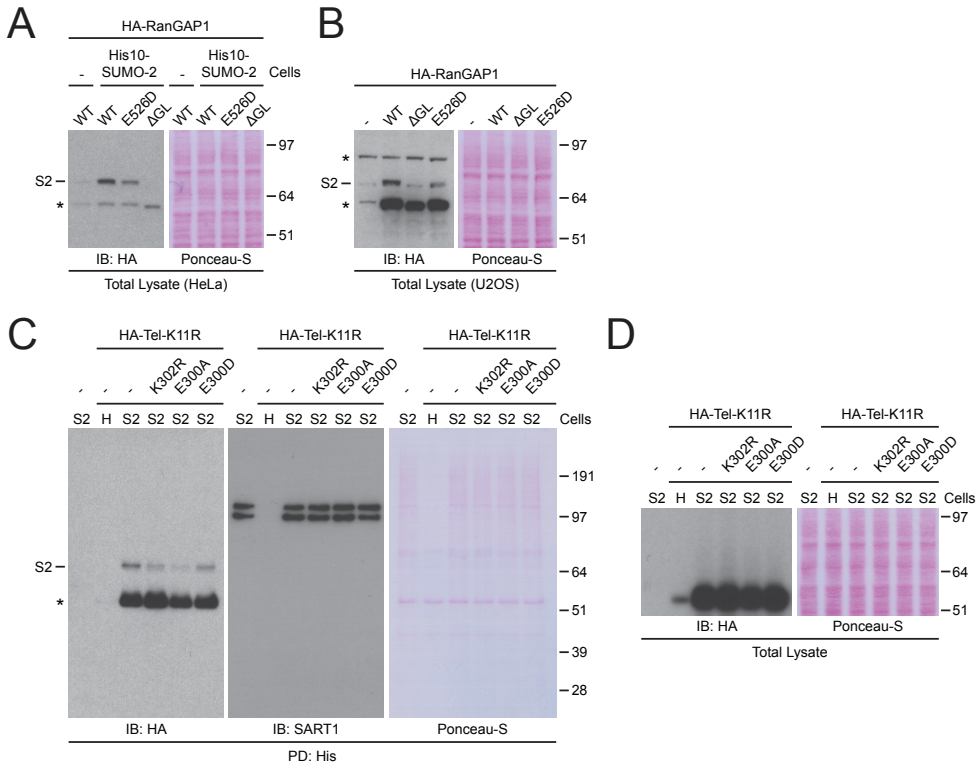


Figure S5. The glutamic acid in RanGAP1's SUMOylation consensus motif is required for efficient SUMOylation. Tel/ETV6 harbours an inverted ExK SUMOylation site at lysine-302.

A) HeLa cells or HeLa cells expressing His10-SUMO-2 were transfected with wild-type (WT) HA-RanGAP1, HA-RanGAP1 E526D, or a SUMOylation-deficient RanGAP1 (ΔGL). Total lysates were analysed by SDS-PAGE and immunoblotting. "S2" indicates SUMO-modified RanGAP1, asterisks indicate non-specific bands. Ponceau-S staining is shown as a loading control.

B) Similar to section A, but using U2-OS cells.

C) HeLa cells or HeLa cells expressing His10-SUMO-2 were transfected with HA-Tel-K11R, HA-Tel-K11R-K302R, HA-Tel-K11R-E300A, or HA-Tel-K11R-E300D. Following lysis, His10-pulldown was performed to enrich SUMOylated proteins. Pulldown fractions were analysed by SDS-PAGE and immunoblotting using the indicated antibodies. "S2" indicates SUMO-modified Tel, asterisks indicate non-specific bands. SART1 is shown as a pulldown control, and Ponceau-S staining is shown as a pulldown and loading control. HA-Tel-K11R was used to monitor SUMOylation of K300, because Tel harbours a dominant major SUMOylation site at lysine-11 [46], which was also identified in our screen.

D) Total lysates corresponding to section C, analysed by SDS-PAGE and immunoblotting. Ponceau-S staining is shown as a loading control.

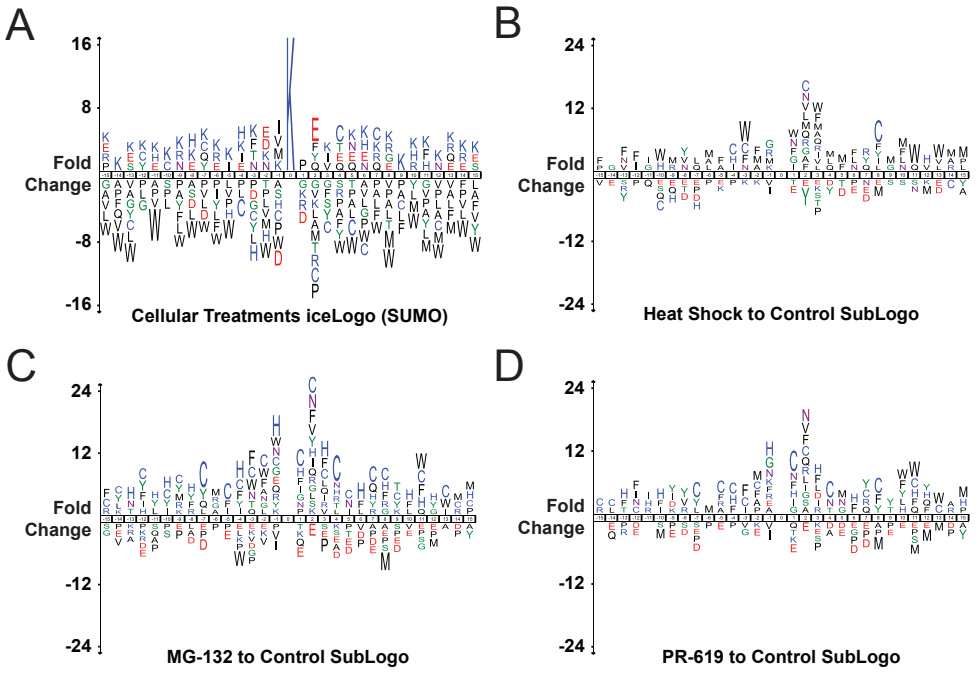


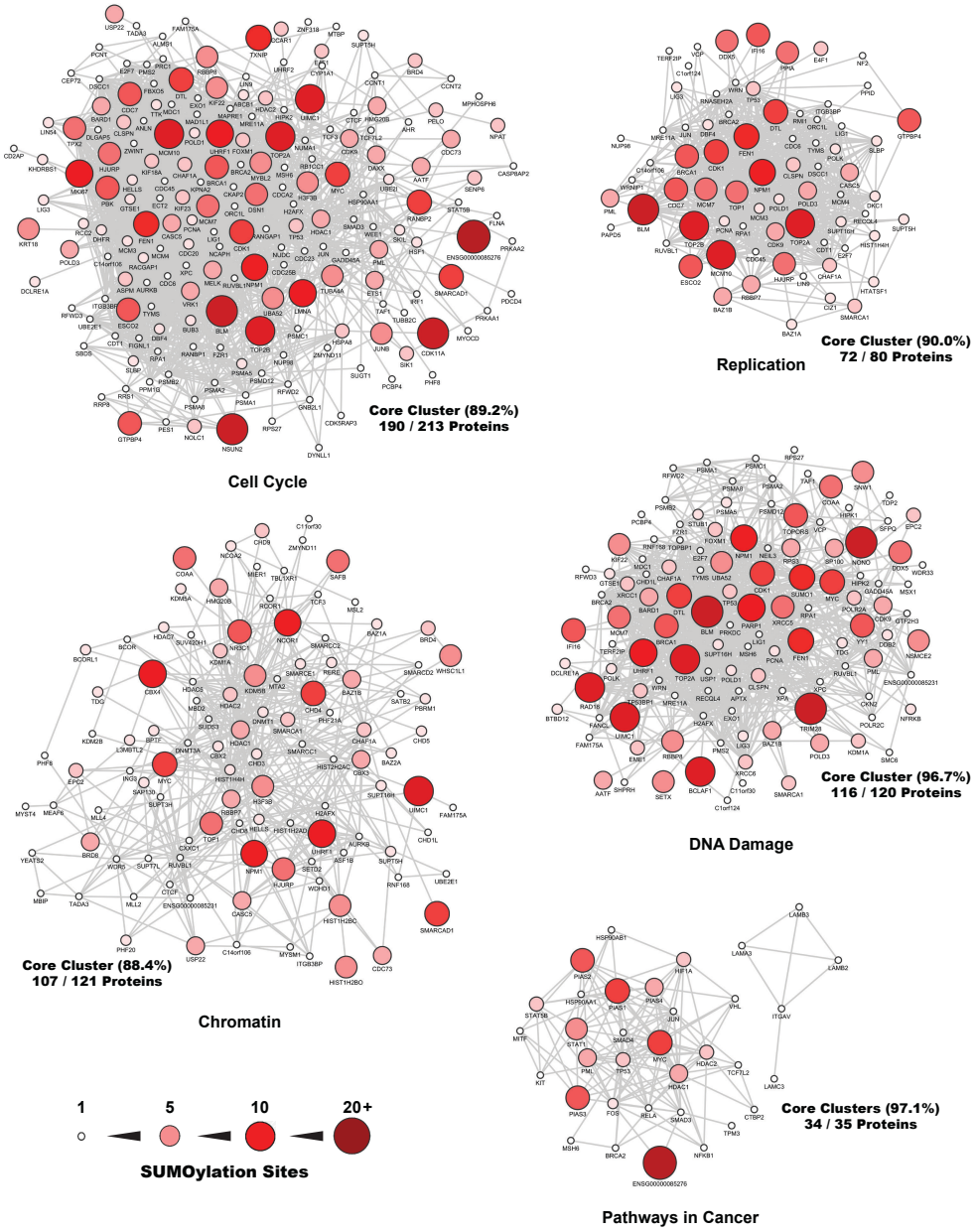
Figure S6. SUMOylation consensus conjugation specificity is dynamic in response to different cellular stresses.

A) IceLogo of all identified SUMOylation sites exclusive to cellular treatments, within a sequence window ranging from -15 to +15 of the modified lysine. All amino acid changes are significant with $p < 0.05$.

B) SubLogo of all heat shock SUMOylation sites as compared to control SUMOylation sites. All amino acid changes are significant with $p < 0.05$.

C) As section B, but for MG-132 treatment.

D) As section B, but for PR-619 treatment.



STRING Interaction (0.4+) Networks of GO-Annotated SUMOylated Proteins

Figure S7. SUMOylation targets cell-cycle, replication, chromatin, DNA damage and cancer-related protein networks.

STRING network analysis of subsets of SUMOylated proteins, filtered by the indicated annotations. STRING interaction confidence was set at 0.4 or greater, and the size and colour of individual proteins corresponds to the amount of SUMOylation sites per protein.

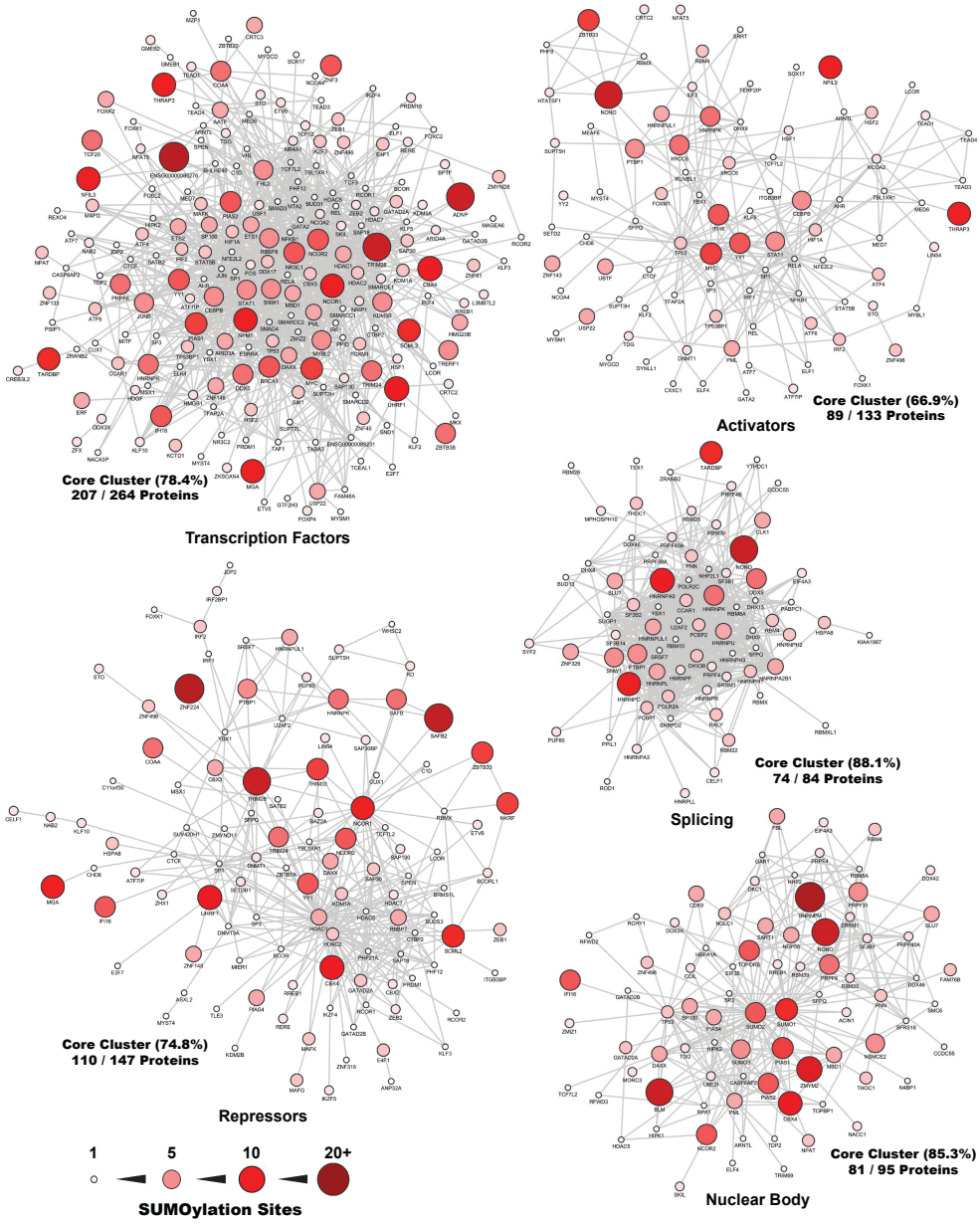


Figure S8. SUMOylation targets transcriptional activators and repressors, splicing proteins, and nuclear body components.

STRING network analysis of subsets of SUMOylated proteins, filtered by the indicated annotations. STRING interaction confidence was set at 0.4 or greater, and the size and colour of individual proteins corresponds to the amount of SUMOylation sites per protein.