Site-specific characterization of endogenous SUMOylation across species and organs

Ivo A. Hendriks, David Lyon, Dan Su, Niels H. Skotte, Jeremy A. Daniel, Lars J. Jensen & Michael L. Nielsen

Small ubiquitin-like modifiers (SUMOs) are post-translational modifications that play crucial roles in most cellular processes. While methods exist to study exogenous SUMOylation, large-scale characterization of endogenous SUMO2/3 has remained technically daunting. Here, we describe a proteomics approach facilitating system-wide and in vivo identification of lysines modified by endogenous and native SUMO2. Using a peptide-level immunoprecipitation enrichment strategy, we identify 14,869 endogenous SUMO2/3 sites in human cells during heat stress and proteasomal inhibition, and quantitatively map 1963 SUMO sites across eight mouse tissues. Characterization of the SUMO equilibrium highlights striking differences in SUMO metabolism between cultured cancer cells and normal tissues. Targeting preferences of SUMO2/3 vary across different organ types, coinciding with markedly differential SUMOylation states of all enzymes involved in the SUMO conjugation cascade. Collectively, our systemic investigation details the SUMOylation architecture across species and organs and provides a resource of endogenous SUMOylation sites on factors important in organ-specific functions.

1 Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark. 2 Disease Systems Biology Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark. 3 Protein Signaling Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Blegdamsvej 3B, 2200 Copenhagen, Denmark. Correspondence and requests for materials should be addressed to M.L.N. (email: michael.lund.nielsen@cpr.ku.dk)
Small ubiquitin-like modifiers (SUMOs) are post-translational modifications (PTMs) that regulate many cellular processes, including virtually all nuclear functions. SUMOylation is indispensable for eukaryotic life, and most vertebrates express three SUMO family members, SUMO1, SUMO2, and SUMO3. All SUMOs are conjugated by the same set of enzymatic machinery comprising one heterodimeric E1, a single E2, and a handful of E3 enzymes. SUMOylation is a reversible process, and several SUMO-specific proteases exist to facilitate removal of SUMO from lysines. SUMO2/3 are often referred to interchangeably owing to their high degree of homology, and overall the SUMO machinery is conserved across eukaryotes. In mouse, knock-out of SUMO2, which is the most abundantly expressed out of SUMO2/3, is embryonic lethal.

A central goal in biology is to understand the cellular mechanisms governing biological processes, and SUMO has emerged as an important factor in nuclear protein assemblies and a distinguished mark for functionally engaged processes. Another mechanistic feature of SUMO is to spatially target functionally related proteins, emphasizing the importance of understanding the substrate specificity and global regulation pattern of SUMOylation. Moreover, SUMO is essential for maintenance of genome integrity and regulation of intracellular signaling. Hence, SUMO has become an increasingly important subject of study in biological sciences and biomedicine, with the modification linked to various diseases including cancer, where SUMO or the SUMO enzymatic machinery are frequently misregulated, while SUMO-mediated transcriptional regulation is required to facilitate certain types of tumorigenesis. Similarly, alterations in SUMO regulation have been associated with ischemia, Huntington’s disease, diabetes, and heart failure. As a result, inhibitors of SUMOylation are actively being developed, with clinical applications in mind.

To understand the functional behavior of SUMOylation in health and disease, it is pivotal to study SUMO at the endogenous level and within the confines of the relevant model systems. However, despite great biological and clinical interest, current knowledge of endogenous and systemic regulation of SUMOylation remains scarce. Although recent developments within mass spectrometry (MS)-based proteomics have facilitated insights into the extent of SUMOylation in cells, these studies were based upon introduction of exogenously expressed, epitope-tagged, and often mutated SUMO variants. As a result, these methods are unable to investigate endogenous SUMOylation, and infeasible for in vivo tissue analyses without relying on genetic engineering.

Although SUMO was discovered 20 years ago, the interest in studying SUMOylation in a biomedical context using proteomics approaches continues to grow. However, it has remained technically daunting to establish analytical strategies allowing site-specific characterization of endogenous SUMO, partly owing to the overall low stoichiometry of SUMO, along with highly active SUMO-specific proteases only deactivated under harsh buffer conditions. Additionally, when using conventional proteomics approaches, digestion of SUMO2/3 with trypsin leaves a 32-residue mass tag on modified lysines, which complicates MS/MS analysis. Still, contemporary approaches for studying endogenous SUMO have allowed detection of a limited number of natively SUMOylated targets at the protein level, but have employed genetic engineering to purify proteins modified by epitope-tagged SUMO from mice. However, all current methods fail to facilitate systems-wide analysis of specific, endogenous, native, and in vivo SUMO2/3 sites.

Given the biomedicinal importance of understanding the cellular architecture of endogenous SUMO2/3, we set out to map SUMOylation to great depth across cells and tissues. To this end, we developed an MS-based proteomics method that facilitates purification of peptides modified by endogenous, wild-type SUMO2/3, and therein pinpoint modified lysines. The method relies on a well-characterized antibody, and analysis of samples with standard MS approaches and freely available software. From triplicate analyses of Human embryonic kidney 293 (HEK) cells, we identified 14,869 unique SUMO2/3 sites mapping to 3,870 endogenously SUMOylated proteins. However, immortalized cancer cell lines are selected for high proliferation rates and commonly do not represent the complex biological conditions in tissues. Thus, in order to delineate the physiological differences in SUMO architecture between tissues we mapped nearly 2000 SUMO2/3 sites from single-organ analyses of eight different mouse tissues.

Collectively, we describe the largest endogenous SUMO2/3 proteomics resource to date, provide insight into the baseline SUMOylation across cells and various organs, and highlight differences in SUMO architecture between these highly distinct model systems. Because our method facilitates the study of endogenous and in vivo SUMOylation patterns in most vertebrates, we present a final step in equalizing the proteomics playing field for the system-wide study of SUMO2/3.

Results

Enrichment of endogenously SUMOylated peptides. To facilitate enrichment of SUMOylated peptides, we utilized the commercially available SUMO2/3 8A2 antibody, which recognizes the C-terminus of SUMO2/3 conserved across vertebrates, including human, mouse, chicken, Drosophila, Xenopus, and zebrafish. The antibody can also be acquired from the Developmental Studies Hybridoma Bank for in-house production, and has previously been used to enrich endogenously SUMOylated proteins. However, despite identifying a few hundred SUMO substrates, no modification sites were reported.

To facilitate efficient proteomics analysis of SUMO2/3-modified lysines, we devised a peptide-level enrichment strategy utilizing the epitope of the 8A2 antibody, to map within the C-terminal part of SUMO2 (57-IRFRFDGQPI-66). We reasoned that protein digestion using trypsin would cleave within the epitope of SUMO2/3, thus prohibiting recognition by the 8A2 antibody, whereas digestion with endoproteinase Lys-C would leave the epitope intact. We confirmed this using dot blot analysis of tryptic and Lys-C digests of HeLa total lysate (Supplementary Fig. 1A–B). Moreover, as the endoproteinase Lys-C remains active at high molar concentrations of chaotropic buffers, this allows samples to be rapidly lysed under denaturing conditions, ensuring rapid inactivation of SUMO proteases and thereby preventing loss of SUMOylation during sample preparation.

Following Lys-C digestion, all samples were desalted on C8 resin to allow efficient capture of large hydrophobic peptides, bearing in mind that the SUMO2/3 mass remnant generated upon Lys-C digestion entails a total mass of 5.6 kDa (Supplementary Fig. 1C–D). Following desalting, samples were lyophilized and SUMOylated peptides were efficiently enriched by SUMO-IP using the 8A2 antibody (Supplementary Fig. 2A–B). Notably, the amount of antibody we used per milligram of starting protein material was 20 times lower than reported previously (Supplementary Note 1).

Although efficient enrichment of peptides modified by endogenous SUMO is challenging, the single-largest complication of proteomics analysis of SUMO2/3 relates to the large mass remnant that remains after standard protein digestion. To alleviate this, we performed a series of pilot experiments to establish an effective strategy entailing a second digestion step to
complement the Lys-C digestion required for the SUMO-IP step. Comparison of several different proteinases revealed that Asp-N facilitated both consistent and reliable identification of endogenously SUMOylated peptides, yielding four-fold higher identification numbers compared to the other enzymes and generating a mass remnant that uniquely identified peptides as modified by SUMO-2/3 (Supplementary Fig. 2C–E and Supplementary Note 1). Although our strategy entails a digestion, purification, followed by another digestion; serial digestion methods are widely used in the proteomics field16, and yield highly pure samples when executed properly. Conclusively, we adopted a serial digestion strategy using Lys-C and Asp-N as the main workflow for the rest of our proteomics experiments (Fig. 1a).

**Mapping the endogenous SUMOylome in human cells.** Having established an optimized SUMO2/3 purification strategy, we benchmarked our method in HEK cells by profiling endogenous SUMOylation in triplicate under standard cell culture conditions, and in response to either heat shock or the proteasomal inhibitor.

---

**Diagram a:**
- SUMO peptide purification
- Denatured
- Native
- Lys-C digestion
- Peptide pre-purification
- SUMO immunoprecipitation

**Diagram b:**
- HEK
- Control
- Heat
- MG132
- Asp-N digestion
- Fractionation
- LC-MS/MS

**Diagram c:**
- Exogenous, control
- Exogenous, stress
- Endogenous, control
- Endogenous, stress
- SUMO sites (number)

**Diagram d:**
- SUMO sites (fraction)
  - Control
  - Heat
  - MG132
  - This study (HEK)
  - Hendriks, 2017 (HeLa & U2OS)
  - Lamoliatte, 2017 (HEK)
  - Both studies

**Diagram e:**
- Protein copy-number
  - This study
  - Becker, 2013
  - Lamoliatte, 2017
  - Bockel, 2017
  - Ubiquitin
  - Total
MG132 (Fig. 1b)\textsuperscript{28}. All experiments were analyzed using a streamlined MS proteomics approach (Supplementary Note 2).

Collectively, we mapped 14,869 unique endogenous SUMO2/3 sites across the investigated cellular conditions in human cells, with 4476 sites identified under standard growth conditions (Fig. 1c, Supplementary Data 1 & 2). We observed an excellent degree of reproducibility with high Pearson correlations between same-condition replicates (Supplementary Fig. 3A), unsupervised hierarchical clustering of same-condition replicates (Supplementary Fig. 3B), and close grouping of same-condition replicates in principle component analysis (PCA) (Supplementary Fig. 3C). Although not reaching the same number of identified SUMOylation sites compared to our recent study using the exogenous K0-SUMO method\textsuperscript{2}, this study nonetheless represents the largest compendium of endogenous SUMO2/3 sites to date (Fig. 1c).

Still, we identified similar numbers of SUMO2/3 sites as exogenous methods when performed at similar scale and cellular background (Fig. 1d, Supplementary Note 3).

To gain insight into the overall sensitivity of our endogenous approach, we compared the depth of sequencing achieved in our data against other SUMO2/3 proteomics studies. To this end, we compared the copy-numbers of identified SUMO2/3 target proteins, by utilizing protein-per-cell copy-numbers derived from IBAQ analysis of a deep HeLa proteome\textsuperscript{29}. We found that endogenous SUMO2/3-targeted proteins entailed a median copy-number of 60,000, which was reassuringly similar to the largest exogenous SUMO studies\textsuperscript{2,30}, demonstrating that our strategy can reach great depth of sequencing despite the technical challenges associated with the study of endogenous SUMO2/3 (Fig. 1e).

Compared to previous proteomics studies focused on endogenous SUMO\textsuperscript{20,24}, our approach achieved at least 3-fold greater sequencing depth. To emphasize that our method is not limited in sensitivity, we additionally compared our data to a similarly-sized group of 4391 ubiquitylated proteins identified by at least three lysine residues\textsuperscript{31}. We observed a median protein copy-number of 100,000 for these ubiquitylated proteins, which further supported that our method achieves good sensitivity, and suggests that SUMOylation may systemically occur on less abundant proteins compared to ubiquitylation.

To demonstrate that our method does not entail a sequence-specific bias, we analyzed the adherence of identified SUMO2/3 sites to the KxE consensus motif, as this commonly is considered an important benchmark in SUMO proteomics data\textsuperscript{32}. We found 24.9% of all SUMOylated lysines to reside in the KxE-type consensus motif, with 60.1% of total SUMOylation occurring on KxE motifs (Supplementary Data 1). When only considering sites identified under standard growth conditions, 36.4% of sites matched KxE, with 63.0% of total SUMO2/3 residing on KxE motifs. Overall, we validated our method by reconfirming canonical SUMO phenomena including specific consensus sub-motifs and distribution of sites-per-protein (Fig. 2, Supplementary Fig. 4 and Supplementary Note 4).

Next we performed a full comparison of SUMO2/3 sites identified in this study, to those previously reported in proteomics screens\textsuperscript{2,19,21,23,30,33–39}, in addition to SUMO sites reported in the literature via low-throughput approaches (Supplementary Data 3). Out of all SUMO2/3 sites we identified, 67.0% were previously reported in screens primarily using exogenous SUMO, rising to 88.2% when looking at the top 10% highest-scoring sites. Differences in overlap could largely be attributed to variable protein digestion patterns, as most proteomics SUMO screens rely on trypptic digestion, whereas we employed a Lys-C/Asp-N approach.

We compared SUMO2/3 target proteins identified in our study to those identified by other SUMO proteomics studies\textsuperscript{2,19,20,23,30,34–40}. Here, we observed an overlap of 87.8%, visualizing a higher overlap at the protein-level as compared to the site-level, suggesting that mainly sites within the same SUMO2/3 target proteins were identified by exogenous approaches (Supplementary Data 4). When considering the top 10% highest-scoring SUMOylated proteins, the overlap increased to 99.0%. Differences in the cellular SUMOylation response to stress were quantified in HEK cells, and overall found to be comparable to the global stress response observed in a previous study\textsuperscript{4}, with Pearson correlations of $R = 0.56$ and $R = 0.59$ between co-identified proteins and sites in response to MG132, respectively (Supplementary Fig. 5).

A catalogue of in vivo SUMO in eight types of mouse organs.

As our method exhibited good sensitivity, we next sought to understand the physiological differences of SUMOylation across tissues. To this end, we performed SUMO-IP on eight types of mouse organs—brain, heart, kidneys, lungs, liver, skeletal muscle, spleen, and testes (Fig. 3a). Using immunoblot analysis, we validated that our SUMO-IP worked efficiently in mouse organs (Supplementary Fig. 6). To properly assess reproducibility of our method and gain quantitative insight, we performed single-organ analyses across five wild-type animals with all organs used in their entirety, excepting liver of which only half was analyzed.

Across all organs, we identified 1963 SUMOylation sites (Fig. 3b and Supplementary Data 5), mapping to 955 protein-coding genes (Supplementary Data 6). The number of SUMOylation sites per organ type appeared unrelated to the size or protein content of the organs. Overall, we observed the highest number of SUMOylation sites in liver, testis, kidney, and spleen, identifying >700 SUMO2/3 sites by direct MS/MS in these organs, with up to 1131 sites in liver (Fig. 3b). The lowest numbers of SUMOylation...
sites were identified in heart and brain, both falling short of 200 MS/MS-identified sites. Using matching of MS1-level evidence between experiments and organs, we were able to increase identification density, hereby mapping >600 SUMO2/3 sites in all organs, >1000 sites in five organs, and up to 1309 unique sites in liver. Accordingly, the number of sites that could be identified across multiple organs increased considerably with the matching data (Fig. 3c).

Globally, we observed 38.1% of SUMOylated lysine residues in mouse organs to reside in the KxE motif, with 50.3% of all SUMOylation occurring on these motifs. The top 100 sites reached 67% KxE adherence, and the top 1000 sites matched 49.4% KxE (Fig. 3d), a number on par with proteomics studies reporting equal numbers of identified SUMO2/3 sites in cell lines under standard conditions\(^\text{28,39}\). Similar to the human SUMO2/3 data, we could confirm the existence of most known SUMOylation sequence motifs within the mouse data (Fig. 3e), although we did not observe enrichment for aspartic acid residues at −2 when using iceLogo analysis, while threonine residues appeared enriched from −5 to −1. To verify significantly enriched SUMO motifs, we additionally performed Motif-X analysis (Supplementary Data 5 and Supplementary Fig. 7), which confirmed a significant presence of KxE-type motifs. In addition, we observed a [ED]xKP motif, suggesting that the inverted SUMOylation consensus motif is significantly SUMOylated in mouse tissue when the modified lysine is followed by a proline residue. Overall, we found that the KxE adherence is variable between organ types, and did not correlate with site abundance or total number of sites (Fig. 3f). Collectively, our findings support that basic SUMO2/3 targeting preferences are comparable between cultured cells and tissues, with some variation in number of sites and KxE adherence between different organ types.

**Functions of SUMO in organs.** SUMOylation in mouse globally adhered to functions that are canonically associated with SUMO (Fig. 3g and Supplementary Data 7), as determined using term annotation enrichment analysis with multiple-hypothesis corrected two-tailed Fisher’s P-values to ensure an FDR of <2%. The most enriched terms highlighted SUMO2/3 modification of transcription factors, chromatin regulators, numerous macromolecular protein complexes, and a predominantly nuclear localization. To gain insight into SUMO target proteins across organs, we performed label-free relative quantification of SUMO sites and proteins (Tables S5 and S6). Out of 1963 identified SUMO2/3 sites and 955 SUMO2/3 target proteins, we could quantify 46.3% and 77.8%, respectively. Reproducibility of the five replicates was assessed using scatter plot analysis, and we observed a high average Pearson correlation \((R = 0.69)\) between all same-organ replicates, with the highest reproducibility \((R = 0.82)\) in liver (Supplementary Fig. 8A). We moreover observed good reproducibility between the replicate animals with organs clustering together hierarchically (Fig. 4a), and in PCA (Fig. 4b). At the SUMO site level, hierarchical clustering and PCA analysis additionally supported data reproducibility (Supplementary Fig. 8B–C), and similarly outlined correlations between organ types, with the largest deviations in the PCA observed for liver and brain.

In order to assess whether any biological functions were enriched for SUMOylation across organ types, we considered SUMO2/3 target proteins inferred from MS/MS-identified SUMO sites only, and compared the SUMOylated proteins to the respective background proteomes for each organ type as documented in the TISSUES database\(^\text{11}\). Subsequently, we performed annotation enrichment analysis to extract qualitative differences between the SUMOylated proteins and the

---

**Fig. 2** The endogenous SUMO2/3 consensus motif and SUMO dynamics. **a** IceLogo representation of amino acid residues enriched around the top 10% SUMOylated lysines identified in human cells, at position 0, indicated by the asterisk. Residues displayed above the line are enriched, and those displayed below the line are depleted, compared to a reference set of human nuclear proteins. **b** The adherence of all identified SUMO2/3 sites to the KxE motif, in relation to the average of top-scoring sites. **c** Venn diagram of SUMO2/3 sites identified in response to different treatments, drawn to scale. **d** Venn diagram of SUMO2/3 target proteins identified in response to different treatments, drawn to scale.
background proteomes, and moreover checked whether any enriched terms were also unique to the organ when compared to the other tissues we analyzed (Fig. 4c and Supplementary Data 8). All organs were enriched for many of the canonical SUMO functions when compared to the background proteomes (Supplementary Data 8), although we chose to focus on organ-unique terms that were significantly regulated by SUMOylation. Kidney was the only other organ where we did not observe organ-unique pathways to be enriched, likely owing to the considerable overlap we observed with liver in terms of SUMO2/3 sites. In liver, a range of metabolic functions were enriched, including amino acid metabolic pathways and the urea cycle. In lung, we observed enrichment for canonical SUMO functions, including pathways implying a more predominant chromatin-centric localization. We also observed SUMO enrichment in lung for terms associated with myeloid leukocyte and erythrocyte differentiation, suggesting that SUMOylation may play a role in immune system regulation. In muscle, SUMOylation was mainly enriched on muscle-specific proteins, and involved in glycolysis. In spleen, we observed an overall similarity to the SUMOylome in lung, and an enrichment for pathways

Fig. 3 Mapping the mouse SUMOylome. a Experimental design of SUMO proteomics experiments carried out in mice. Eight types of organs were studied, in biological quintuplicate (five animals), and analyzed as six fractions. b Overview of the number of SUMO2/3 sites identified per organ, comparing direct MS/MS identifications, to identifications made when matching MS1-level data between organs. c Overview of the number of organs in which the same SUMO2/3 sites were detected. d The adherence of all identified SUMO2/3 sites to the KxE motif, per organ, relative to the average of top-scoring sites. e IceLogo representation of amino acid residues enriched around all SUMOylated lysines identified in mouse, at position 0, indicated by the asterisk. Residues displayed above the line are enriched, and those displayed below the line are depleted, compared to a reference set of mouse nuclear proteins. f The adherence of MS/MS-identified SUMO2/3 sites to the KxE motif, in relation to the average of top-scoring sites. g Term enrichment analysis comparing identified mouse SUMO2/3 target proteins to mouse proteins known to be expressed in the eight organs analyzed. Numbers in parentheses indicate overlap between the numbers of proteins identified in this study, compared to those present in the reference. Score was derived from the logarithms of enrichment ratios and Benjamini-Hochberg FDR-corrected two-tailed Fisher’s P-values of < 0.02. A full list of all enriched terms, and all relevant scores, is available in Supplementary Data 7.
associated with chromatin remodelers. Finally, SUMOylation in testis was among the most distinct from all other types, with specific SUMO modification of proteins that are not necessarily expressed in a tissue-specific manner. For example, we observed strong SUMOylation enrichment for proteins associated with the DNA damage response, mitotic recombination, helicase activity, telomeres, and the global cellular stress response.

Although we corrected our term enrichment analysis for the expression in progenitors of pancreatic cells (5 / 7) SUMO target proteins

>345

Testis was among the most distinct from all other types, with proteins associated with chromatin remodelers. Finally, SUMOylation in skeletal muscle and heart showed the largest correlate elevated protein expression levels in each organ to the distinct tissues. To assess the extent of this effect, we correlated elevated protein expression levels in each organ to increased levels of SUMOylation (Fig. 4d). Comparatively, SUMOylation in skeletal muscle and heart showed the largest dependence on protein abundance, with modification occurring on proteins predominantly expressed in these organs. Conversely, SUMOylation in brain, lung, spleen, and testis, only displayed a modest correlation with protein abundance, further suggesting
differences in the way SUMOylation is targeted to proteins in these organs.

**SUMOylation in human and mouse.** Although considerable differences exist between cultured human cells and mouse organs, we sought to elucidate whether the SUMOylome would similarly reflect this distinction. To this end, we performed a series of comparisons on modified proteins (Supplementary Data 9 and 10), including subcellular localization of SUMO target proteins (Fig. 5a, b), SUMO-phosphorylation co-modification (Fig. 5c, d, Supplementary Data 11 and 12), and structural targeting preference (Fig. 6a–c). All these analyses are further detailed in Supplementary Note 5. Overall, we observed globally constant properties of SUMOylation, including the predominant nuclear localization, preference for modifying disordered protein regions, and involvement of SUMO-phosphorylation co-modification.

Since we identified a considerable number of endogenous SUMO2/3 sites in human cells and mouse organs, we were interested in evaluating evolutionary conservation of SUMOylation across species. Generally, the SUMO enzymatic machinery is highly conserved across eukaryotes; however, conservation of the individual SUMOylation sites catalyzed by the SUMO enzymes has never been investigated on a larger scale.

We calculated Residue Conservation Scores (RCS) for all lysines in SUMOylated proteins identified in human and mouse. When considering RCS for all lysines, we found an average RCS of 74.3% for lysine residues residing in globular and buried regions, 70.9% for globular and exposed regions, and 65.4% for disordered regions (Fig. 6d). Thus, lysines in disordered regions are globally less conserved than lysines in ordered regions, in agreement with what has been described previously. Strikingly, we found that SUMOylated lysines in ordered and exposed protein regions were significantly less evolutionarily conserved compared to non-modified lysines in the same structural context (Fig. 6e). Overall, SUMOylation preferentially targeted disordered regions, which are intrinsically poorly conserved, and otherwise SUMO2/3 had a tendency to modify less conserved lysines when targeted to globular regions. Thus, we highlight a high degree of SUMO evolution across species, in spite of the underlying structural and sequence-specific targeting preferences being similar between human cells and mouse tissue.

Finally, we assessed the density of SUMOylation in the investigated samples, by quantifying the total amount of SUMO2/3-modified peptides and proteins relative to the background signal (Supplementary Fig. 9 and Supplementary Note 6). We observed a SUMO-IP purity of over 93% from HEK cells, and 42–62% from mouse organs. Moreover, we found that the amount of SUMO2/3 peptides per milligram of starting material was up to 70 times larger when purifying SUMO2/3 from HEK cells as compared to liver, indicative of a large accumulation of SUMO2/3 in actively cycling HEK cells.

**Quantifying the native SUMO2/3 equilibrium.** Within eukaryotic cells, SUMO proteins are expressed as immature precursors which are processed into mature counterparts by SUMO proteases. During the SUMO conjugation cycle, the E1-E2-E3 enzymatic machinery is responsible for the reversible conjugation of mature free SUMO to protein substrates. However, despite the dynamic nature of SUMO conjugation, the exact equilibrium of SUMO pools remains relatively poorly understood, due to the lack of analytical tools to reliably quantify distinct pools of cellular SUMO. Because our method enriches the C-terminal part of SUMO2/3 regardless of whether it is conjugated to another protein, we were able to quantify the cellular pools of conjugated SUMO2/3, unconjugated SUMO2/3, and immature SUMO2 and SUMO3, in cultured cells and tissue (Fig. 7a), hereby providing insight into the regulation of these distinct pools of SUMO. Moreover, analysis of the conjugated pool of SUMO2/3 allows concomitant discrimination between SUMO-chain formation, modification of SUMO E1 (SAE1/UBA2), E2 (UBC9), and E3 enzymes, and conjugation to any other protein.

In HEK cells, we found 93% of SUMO2/3 conjugated to target substrates, while only 6% of SUMO2/3 existed as free SUMO (Fig. 7b). The fraction of conjugated SUMO2/3 increased further in response to cellular stress, up to 96% after MG132, and up to 98% after heat shock. However, considering the global pool of conjugated SUMO2/3 increased by ~50% in response to heat shock (Supplementary Fig. 3D), our data suggest that ongoing synthesis of new SUMO2/3 is likely essential for the SUMO stress response. Under control conditions, ~6% of SUMO2/3 was part of SUMO-chains, ~3% of SUMOylation was found modifying E3 SUMO ligases, ~1% modified the E2, and only a small fraction modified the E1. Interestingly, nearly 1% of the total SUMO pool corresponded to immature SUMO3, which is generally considered to be expressed at a lower level than SUMO2. In response to stress, we observed more E3 ligase modification, indicative of increased SUMOylation activity. Additionally, in support of an increased demand for SUMO maturation, the small pool of immature SUMO2/3 that was detectable under control condition was abolished in response to stress treatments.

Intriguingly, when quantifying the distribution of SUMOylation within mouse organs, we observed a much lower percentage of conjugated SUMO2/3 (Fig. 7c). On average, only 52% of SUMO2/3 was conjugated across all organs, with 47% free SUMO2/3, ~1% immature SUMO2, and only a minor fraction of immature SUMO3. SUMO conjugation ranged from above-average rates of 72% and 70% in liver and lung, to below-average rates of 21% and 35% in brain and heart, respectively.
Correspondingly, in brain, 78% of SUMO2/3 was observed to be unconjugated. Strikingly, the small amount of conjugated SUMO2/3 in brain was found to largely modify the SUMO E1 and E2 enzymes, or integrated into SUMO-chains, with less than half of conjugated SUMO2/3 targeting other proteins. SUMOylation in testis displayed similar patterns, with 35% of SUMO2/3 residing on the E2 enzyme, 22% within SUMO-chains, and only 40% modifying other target proteins. The amount of SUMO-chain formation in testis was significantly higher than in all other organs (Fig. 7c). Outside of brain and testis, SUMOylation predominantly targeted a wider variety of substrates, similar to other organs (Supplementary Fig. 10), and on par with HEK cells. Immature SUMO2 was more abundant in all mouse organs, forming 1.2% of the total pool of SUMO2/3, contrasting HEK cells where immature SUMO3 was more abundant. In heart, significantly more immature SUMO2 was measured in comparison to other organs.

Taken together, we observed considerable differences in the SUMO2/3 equilibrium, highlighting dynamic global requirements for SUMOylation that varied significantly between cultured cell lines and tissues. We validated these findings using immunoblot analyses and a distinct experimental approach (Supplementary Fig. 11 and Supplementary Note 7). Intriguingly, considering our tissue analyses were performed in wild-type mice, our findings support the growing number of observations that SUMO dynamics may be substantially enhanced in many cancers and neurodegenerative diseases. Moreover, our data indicate a requirement for enhanced SUMO2/3 equilibrium in rapidly proliferating cultured cells, which suggests that a high availability and degree of conjugation of SUMO2/3 may occur in, and in turn facilitate, proliferating cancer cells.

Endogenous SUMO chain topology. In contrast to the well-described functions of poly-ubiquitylation, comparatively little is known about the functional role of poly-SUMOylation in vertebrates. Moreover, SUMO chains have typically been investigated in the context of mutated and overexpressed SUMO. Thus, we decided to investigate SUMO chain-topology under endogenous and in vivo conditions, by extracting abundance profiles for all individual SUMO linkages. In HEK cells, we observed ~35% of SUMOylation to reside on K11 in SUMO2 (Fig. 7d), which was previously reported as the major site for poly-SUMOylation and is the only KxE-type consensus motif in SUMO2. Upon heat or proteotoxic stress, modification of K11 increased further to ~50%. However, in contrast to previous chain-topology studies, we observed a notable degree of modification on K7, K21, and K33 in untreated cells. In response to stress, SUMOylation on K21 and K33 was reduced. Modification of SUMO3 generally followed the same trend as SUMO2 (Supplementary Fig. 12A), while for SUMO1, most poly-SUMOylation by endogenous SUMO2/3 occurred via K7, which resides in an inverted ExK-type consensus motif (Supplementary Fig. 12B). Modification on K7 decreased in response to cellular stress, whereas K17, K23, K37, and K45, displayed increased SUMOylation in response to stress. We also observed SUMO-2/3 modification of K78 in SUMO1, with ~10% of all SUMO1 chains occurring via this residue. K78 is the most C-terminal lysine residue in SUMO1, and with no further lysines or arginines towards the C-terminus it is not readily detectable using tryptic approaches. Although our data do not provide direct clues to the functional role of the observed poly-SUMOylation, the chain-topology analysis demonstrates that a wide range of lysine residues within SUMO proteins are involved in endogenous poly-SUMO formation. Moreover, the overall abundance of SUMO residing in poly-chains compared to the overall conjugation of SUMO-phospho co-modified peptides identified by MS/MS in human cell lines. SUMOylated lysines are at position 0, with spacing indicating the residue at which phosphorylation occurred. The black line corresponds to the average length profile of all detected SUMO2/3 site peptides. Asterisks denote significantly different values, as determined by two-tailed Fisher’s exact testing. * P < 0.05, ** P < 0.001. d As (c), but for SUMO-phospho co-modified peptides identified by MS/MS in mouse organs.

**Fig. 5 SUMO2/3 subcellular localization and SUMO-phospho co-modification across human and mouse. a Subcellular localization analysis, based on Gene Ontology Cellular Compartments (GOCC). All SUMO2/3 target proteins (P) and SUMO2/3 sites (S) were assigned a localization based on GOCC terms, ranging from chromatin-associated to extracellular. Subsets of SUMO2/3 target proteins and sites were compared to each other, and to a full human background proteome (right-most value), which represents all proteins present in cells. CM consensus motif (KxE). b As (a), but for SUMO2/3 target proteins and sites identified in mouse organs. The mouse background proteome (right-most value) was derived from the TISSUES database, covers eight organs, and represents all proteins present in the organs. c Schematic overview of SUMO-phospho co-modified peptides identified by MS/MS in human cell lines. SUMO-phospho peptides (number).

| Fraction (cumulative) | Chromatin | Membrane | Cytoplasm | Nucleus | Nucl / Cyto | Cyto / Mem | Membrane | Extracellular |
|----------------------|-----------|----------|-----------|---------|------------|-----------|----------|--------------|
| 0%                   | 100%      | 0%       | 0%        | 0%      | 0%         | 0%        | 0%       | 0%           |
| 20%                  | 80%       | 0%       | 0%        | 0%      | 0%         | 0%        | 0%       | 0%           |
| 40%                  | 60%       | 0%       | 0%        | 0%      | 0%         | 0%        | 0%       | 0%           |
| 60%                  | 40%       | 0%       | 0%        | 0%      | 0%         | 0%        | 0%       | 0%           |
| 80%                  | 20%       | 0%       | 0%        | 0%      | 0%         | 0%        | 0%       | 0%           |
| 100%                 | 0%        | 0%       | 0%        | 0%      | 0%         | 0%        | 0%       | 0%           |

**Fig. 6 Endogenous SUMO chain-topology.** a Subcellular localization analysis, based on Gene Ontology Cellular Compartments (GOCC). All SUMO2/3 target proteins (P) and SUMO2/3 sites (S) were assigned a localization based on GOCC terms, ranging from chromatin-associated to extracellular. Subsets of SUMO2/3 target proteins and sites were compared to each other, and to a full human background proteome (right-most value), which represents all proteins present in cells. CM consensus motif (KxE). b As (a), but for SUMO2/3 target proteins and sites identified in mouse organs. The mouse background proteome (right-most value) was derived from the TISSUES database, covers eight organs, and represents all proteins present in the organs. c Schematic overview of SUMO-phospho co-modified peptides identified by MS/MS in human cell lines. SUMO-phospho peptides (number).
SUMO in cells (Fig. 7b), supported that mono-SUMOylation is the major SUMOylation event in cultured cells.

In mouse organs, we mainly observed SUMO chain formation on SUMO2. Strikingly, we observed a notably different SUMO2 chain topology in mouse organs, with much less modification on the KxE-type K11, while K21 and K33 were observed the most highly modified. With the sequence of SUMO2 fully conserved between human and mouse, this hints at differential regulation of SUMO chain topology in HEK cells compared to organs. Still, the overall relative abundance of poly-chain formation was comparable between human and mouse, this hints at differential regulation of SUMO chain topology in HEK cells compared to organs. Still, the relative presence of 100% is consistent role across cells and organs. SUMO1 was only found to be modified on K7 and K78, with insufficient data for statistical significance, whereas we could not quantify SUMO3 chain formation on SUMO2. Moreover, N-terminal quantification of SUMO3 was obfuscated by two isoforms, Q9Z172-2 and G3UZA7, which we found expressed and SUMO-modified in mouse tissues. Interestingly, both of these SUMO3 isoforms differ remarkably more in their sequence from SUMO3 than SUMO2 does, and could provide direct evidence for the existence and expression of distinct SUMO family members in mouse organs. Overall, we observed a marked difference in SUMO chain topology between cell culture and mouse organs, with SUMO2 in mouse being predominantly modified on lysines other than the canonical K11, and three distinct variants of SUMO3 being expressed and SUMO-modified in mouse organs.

**Discussion**

Here, we describe a proteomics strategy facilitating the unbiased and site-specific study of endogenous SUMO2/3, which is compatible with most vertebrate model systems. As a proof of principle, we applied our method in cell culture, identifying in excess of 14,000 SUMOylated lysines, achieving cellular depth similar to previous studies, but without the requirement for exogenously expressed and mutated SUMO2/3 or other genetic engineering. Because of this, we could utilize our strategy to profile in vivo SUMOylation across eight types of mouse organs, identifying 1963 unique SUMO2/3 sites.

Technically, our developed approach exhibits a notably increased sequencing sensitivity compared to previous endogenous SUMO methodologies, and is the only method that can specifically identify lysine residues modified by SUMO-2/3. To achieve this, we utilized the 8A2 antibody to purify SUMOylated peptides, contrary to another method that purified endogenously SUMOylated proteins. We reason that the increased depth achieved by our method is largely owing to efficient purification of peptides instead of proteins, and through purification in very mild conditions after dissolution of lyophilized peptides, allowing the antibody to exert its full specificity.

Recently, a study was published that facilitates identification of endogenously SUMOylated lysines, through digestion of total lysate with the wild-type α-lytic protease (WALP) enzyme followed by di-glycine enrichment. The WALP-only strategy has
Fig. 7 Insight into the SUMO equilibrium. a Overview of the SUMOylation cycle, highlighting different pools of SUMO2/3 that can be quantified using our purification strategy. Notably, conjugated SUMO2/3 and free SUMO2/3 can be quantified, along with immature free SUMO. Conjugated SUMO2/3 was further split into SUMO covalently conjugated to the E1, E2, or E3 enzymes, to other SUMO (chain), or to any other target protein. Note that the usual transfer of SUMO from E1 to E2 to target requires the cysteine and does not end up on lysines within the same enzymes, but transfer to lysines in SUMO enzymes can occur when the enzymes are active in close proximity to each other. b Quantification of the SUMO2/3 equilibrium in human cell lines in response to various treatments, visualizing the fraction of total SUMO existing as conjugated to certain target proteins, or as free SUMO. Error bars represent SD, n = 3 cell culture replicates. c As (b), but quantifying the SUMO2/3 equilibrium in different mouse organs. Error bars represent SEM, n = 5 animals. Asterisks indicate significant differences (blue, higher; red, lower) between the indicated organ and the six median organs within the same category, by two-tailed Student’s t-test. **P < 0.001, *P < 0.05. d Quantification of endogenous SUMO2/3 chain architecture in HEK cells in response to various treatments, corresponding to endogenous SUMO-2 modified by SUMO-2/3. Error bars represent SD, n = 3 cell culture replicates. e As (d), but quantifying the SUMO-2 chain architecture in different mouse organs. Error bars represent SEM, n = 5 animals.
various shortcomings, including the inability to distinguish SUMO family members and a high potential for false-positive identification of ubiquitination sites as SUMO54 (Supplementary Note 8). In contrast to the WALP approach, our methodology validates the presence of the SUMO2/3 mass remnant by its unique fragmentation pattern (Supplementary Data 13 and Supplementary Note 9). Furthermore, our strategy is at least fourfold more sensitive compared to all data reported in other endogenous studies24,30, while using comparable equipment, a single fraction of a single sample, and only 60 min of MS time (Supplementary Fig. 13).

Biologically, we observed a striking difference between the SUMO2/3 equilibrium in cell culture and organs. In cell culture we found the large majority of SUMO2/3 to be conjugated, whereas considerable pools of free SUMO2/3 were observed in mouse organs. Similar observations have previously been made for ubiquitin, with the majority of ubiquitin conjugated in cell culture, but a large pool of free ubiquitin existing within the brain51. Additionally, we observed a much greater amount of total SUMO2/3 in cell culture, relative to the total protein content. Considering SUMO2/3 is known to be a dynamic modifier, and canonically linked to cell-cycle regulation, replication, and the DNA damage response52, our observations suggest that rapidly and ever-dividing cancer cells rely on the SUMOylation pathway to maintain their genomic integrity. By contrast, cell division is much less prominent in organs, hence SUMO2/3 is not expected to be present or conjugated to a similarly high extent, but instead exists as a free pool to facilitate the response to cellular stress. The high density of SUMO2/3 in cell lines could also be a result of continuous selection pressure, with increased amounts of endogenous and exogenous stresses creating a dependence on SUMOylation to survive. Indeed, overexpression of SUMO-2/3 has previously been shown to increase the speed at which cells divide in culture35. Ironically, whereas SUMO2/3 canonically serves to protect faithful multiplication of the cell’s genomic content while protecting the cell against stress, it may also ultimately potentiate the rapid and uncontrolled cellular division observed in cancer, and facilitate the dramatic changes observed within the genomic architecture of cancer cells.

Although exhibiting differences in SUMO dynamics and density, the basic targeting preferences of SUMO did not differ between human cells and mouse organs, with SUMOylation predominantly occurring on KxE-type motifs, a preferential occurrence in disordered regions of proteins, and enrichment in the nucleus and similar spatial assemblies (Fig. 8). Still, SUMOylation across organs targeted more specific, often organ-unique functions, although many canonical SUMO2/3 targets were modified across all organs. Similarly, SUMO2/3 predominantly exists as unconjugated in organs, revealing a notable difference in SUMO2/3 equilibrium and dynamics between cultured cells and organs.

Compared to other organs, SUMOylation was more prevalent and more likely to be conjugated instead of free, in liver, lung, spleen, and testis (Fig. 8). In liver, we identified the largest number of sites, although partly because liver was the largest organ analyzed. Nonetheless, SUMOylation appeared to actively play a role in liver, modifying many liver-specific proteins and likely influencing important metabolic pathways. In testis, we observed the highest amount of SUMO2/3 relative to total protein, and a predominant modification of proteins involved in the DNA damage response and replication, which would be supported by the ongoing cell division in the form of spermatogenesis. A high degree of E2 auto-modification and SUMO-chain formation was also found in testis, suggesting that SUMO modification of Ubc9 and SUMO-chain formation may play a critical role in regulation of proteins involved in spermatogenesis.

Indeed, SUMO modification of Ubc9 itself has previously been shown to be important during meiosis53. Moreover, we found ZNF451, a SUMO E3 ligase that can efficiently extend SUMO chains54, to be predominantly SUMOylated in testis, suggesting ZNF451 could be responsible for the higher degree of SUMO chains observed in testis. In lung, we observed the highest degree of SUMO E3 auto-modification, often indicative of their activity.

In brain, we detected relatively few SUMOylation events, especially in relation to the size of the organ. Moreover, nearly 80% of SUMO2/3 in brain appeared to exist in free mature state, and more than half of conjugated SUMO2/3 was found in SUMO-chains and on the E1 and E2 enzymes, leaving only 10% of the SUMO2/3 pool conjugated to various target proteins. As SUMOylation commonly is studied in the context of neurological disease, we found this exceptionally striking. A recent publication supports our findings and suggests that SUMO conjugation of synaptic proteins is not only rare, but often false-positively identified55. Since we investigated the brain SUMOylome in normal young adult mice, our findings could imply that the SUMOylation system potentially exists in a primed state. This is further supported by the high degree of SUMOylation on the enzymatic machinery, which could be indicative of high levels of SUMO-specific protease activity able to
Methods
understanding of the functional role of SUMOylation in health

dynamics emerging as a general hallmark of several cancer types,

Moreover, the unbiased compilation of
strating marked differences in SUMO2/3 architecture compared
to baseline SUMOylation levels across multiple organs, demon-

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.

Collectively, we present a practical and efficient strategy for
mapping the attachment sites of endogenous SUMO2/3 that will
help to pave the way for future endogenous and in vivo SUMO2/3
proteomics endeavors, with SUMO2 being the predominantly
expressed and most dynamic of all SUMO family members.
OMylated peptide purification. Lysophilized peptides were dissolved in ice-cold SUMO-IP Buffer (50 mM MOPS, 10 mM Na2HPO4, 50 mM NaCl, pH 7.2), using 1 mL SUMO-IP Buffer per mg of protein initially in the sample. Samples were cleared of contamination at 3000 g for 30 min at 4 °C, and transferred to 15 mL tubes, with no more than 10 mL of samples per tube. A volume of 50 μL of SUMO-IP beads were added per 10 mL of sample, after which the sample was mixed for 3 h at 4 °C. Beads were briefly pelleted, transferred to 1.5 mL tubes using ice-cold SUMO-IP buffer, and washed once more with SUMO-IP buffer. Next, beads were washed with ice-cold PBS, transferred to clean 1.5 mL LoBind tubes, and washed once more with PBS. Subsequently, beads were washed with ice-cold MQ water, transferred to clean 1.5 mL LoBind tubes, and washed once more with MQ water. Peptides were eluted from the beads using two volumes of ice-cold 0.15% TFA in MQ water, by gently mixing the beads and allowing them to stand on ice for 30 min. The elution was repeated once, and both elutions were pooled. Elutions were cleared by centrifugation through 0.45 μm filters, and pH-neutralized by addition of 1/10th volume of 1 M Na2HPO4. Eluted SUMOylated peptides were frozen at −80 °C until further processing.

Second-stage digestion of SUMOylated peptides. The purified lys-C-digested SUMO2/3 peptides were further digested with various enzymes. All mouse samples and the home-made peptide samples were further digested with Asp-N (Roche), and some initial cell line samples were digested with either trypsin (Thermo), Glu-C (Promega), or wild-type alpha-lytic protease (WALP; Sigma). Asn-N, Glu-C, and WALP digestions were performed in the neutralized SUMO-IP elutions, overnight, and at 30 °C, using 1 μg of enzyme per 50 mg of protein initially in the sample. For trypsin digests, a 10-fold concentration of 2 M. One microgram of trypsin per 50 μg of protein initially in the sample was used, digesting overnight and at room temperature.

Desalting and on-StageTip high-pH fractionation. C18 StageTips were prepared in-house26. For all experiments, four plugs of C18 material (Sigma-Aldrich, Empore® SPE Disks, C18, 47 mm) were layered per StageTip. For single-shot samples desalted at low pH, StageTips were activated with 100 μL 100% methanol, equilibrated with 100 μL 0.1% formic acid, and equilibrated twice with 50 μL 0.5% acetic acid, prior to loading samples. After loading, samples were washed twice with 50 μL 0.5% acetic acid prior to elution. Samples from initial experiments with trypsin, Glu-C, and WALP were desalted at low pH (0.5% acetic acid) and eluted as a single sample with 50% ACN in 0.5% acetic acid, and equilibrated twice with 75 μL 50 mM ammonium hydroxide. The pH of the digested samples was raised by addition of ammonium hydroxide to a final concentration of 20 mM. Samples were prepared for analysis on C18 StageTips, and were washed twice with 75 μL 50 mM ammonium hydroxide. Subsequently, six fractions (F1-6) were eluted by sequential elution with increasing amounts of ACN in 20 mM ammonium hydroxide. For Asp-N digests, elutions were performed with 4, 7, 10, 13, and 25% ACN. For trypsin digests, elutions were performed with 5, 8, 12, 18, 27, and 50% ACN. For WALP digests, elutions were performed with 2, 4, 7, 10, 13, and 40% ACN. Optimization was performed on earlier experiments, and these fractions represent the finalized profiles used for the majority of the samples. All fractions were completely dried in a SpeedVac at 60 °C. Dried peptides were dissolved in 10 μL 0.1% formic acid, and frozen at −20 °C until analysis.

Mass spectrometric analysis. Samples were analyzed on 15 cm long 75 μm internal diameter columns, packed in-house with ReproSil-Pur 120 C18-AQ 1.9 μm beads (Dr. Maisch), connected to an EASY-nLC 1200 system (Thermo). Column elution was performed using a multitude of different gradients, owing to considerable differences in peptide properties depending on high-pH fraction and the enzymes used for digestion of the peptide mixtures. Settings were optimized during earlier experiments, and the following values represent the finalized settings used for the large majority of all samples, and for all samples that were quantified. Single-shot samples were analyzed with 140 min effective gradient time, and fractions with 70 min effective gradient time, with Buffer A (0.1% FA) as the initial buffer, and an increasing amount of Buffer B (80% ACN in 0.1% FA) over time. The following values represent the increasing value of Buffer B across the time-frame of the analytical gradients, and do not include initial ramp-up and washing blocks. Single-shot analyses; Glu-C: 5–30%, trypsin: 25–45%, WALP: 5–20%. Fraction analyses; Asp-N; F1: 13–24%, F2: 14–27%, F3: 15–30%, F6: 17–32%. Trypsin; F1: 26–42%. WALP, F1: 2–5%, F3: 6–5–30%. A column heater was used to warm the column to 60 °C. Ion source was heated using a Nanospray Flex Ion Source (Thermo). Analysis of the ion stream was performed using a Q-Exactive HF mass spectrometer (Thermo). Spray voltage was set to 2 KV, with a capillary temperature of 275 °C, and a S-Lens RF level of 50%. Full scans were made at a resolution of 60,000, with an AGC target of 3,000,000 and a maximum injection time of 60 ms. The following scan ranges were used; Asp-N: 400–1600 m/z, Glu-C: 300–1750 m/z, trypsin: 600–1900 m/z, WALP: 300–1200 m/z. Precursor fragmentation was achieved through higher-energy collision dissociation (HCD) with a normalized collision energy of 25, an AGC target of 100, and an isolation width of 1.3 m/z. The following precursor charge states were considered for MS/MS; Asp-N: 2–6, Glu-C: 2–4, and trypsin: 2–5. A dynamic exclusion of 60 and 45 s was used for single-shot samples and fractions, respectively. For collection and analysis of precursors by MS/MS, loss count was set to 7, MS2 resolution to 60,000, maximum injection time to 120 ms, and intensity threshold to 5000.

Raw data analysis. MS proteomics RAW data are available at the ProteomeXchange Consortium database via the PRIDE partner repository65, under dataset ID PXD008803. All RAW files were analyzed using MaxQuant software version 1.5.3.0.63. For the human cell line data, we downloaded from UniProt66. The human dataset was downloaded on 22 February 2017, and the mouse dataset on 10 May 2017. Laboratory data was processed with MaxQuant software version 1.5.3.0.63. All processed data were additionally manually filtered for peptide-spectrum-match and protein assignment by filtering of MaxQuant output tables. The MaxQuant output tables were filtered for peptide-spectrum-match and protein assignment by filtering of MaxQuant output tables. The MaxQuant output tables were filtered for peptide-spectrum-match and protein assignment by filtering of MaxQuant output tables. The MaxQuant output tables were filtered for peptide-spectrum-match and protein assignment by filtering of MaxQuant output tables.
residue was an aspartic acid or glutamic acid. Duplicate assignments of MS/MS scans to multiple SUMO sites, and duplicate ±25 amino acid sequence windows flanking SUMO-modified lysines, were manually discarded. Stringent manual filtering resulted in discarding 13% of sites from the human Asp-N data, and 18% of sites from the mouse Asp-N data. SUMOylated proteins were reconstructed based on the "proteinGroups.txt" file, and only sites that remained after manual filtering were mapped back to the proteins. SUMO target proteins were defined as those with at least one unmodified peptide, and other peptides were discarded from further analysis. Data pertaining to tissue-specific protein expression levels in mouse were extracted from the TISSUES database.

Identification of SUMO-phospho co-modified peptides. The "evidence.txt" output file from MaxQuant was used to filter for SUMO-phospho co-modified peptides, and only direct MS/MS evidence was considered. A localization delta score of 0 was deemed sufficient for both SUMO and phosphoylation. To combine a scored peptide was calculated from the excess score over Andromeda, delta, and localization scores, primarily weighing the delta score, and co-modified peptides were filtered for a combined score of >20 (i.e. a delta score of >40), applying an additional layer of FDR over the MaxQuant default. The highest-scoring unique modified peptides were retained in the list, and 51 amino acid sequence windows were assigned to both SUMOylation and phosphorylation. Quantitative information was extracted from modificationSpecificPeptides.txt and aligned to the corresponding co-modified peptides.

Quantification and scoring of SUMO sites and proteins. The n = 3 HEK data (cell culture replicates) and n = 5 mouse data (five animals) was quantified, using MaxQuant Q values as thresholds for quantification of peptides. Mouse data were median-normalized to compensate for variations in overall protein abundance between different organs. For the HEK data, proteins were only quantified if detected in 3 out of 3 replicates at least one condition. For the mouse data, proteins were only quantified if detected in at least 4 out of 5 replicates at least one condition. After filtering, unique phospho-SUMO values were globally imputed using Perseus software.

Sites were processed and quantified separately. For MOTIF-X analyses, the online tool (http://motif-x.med.harvard.edu) was used, and all peptides either modified or not were used for the comparison. SUMO target proteins identified in this study were aligned to the scaffold, along with SUMO target proteins identified in other SUMO proteomics studies.

Residues Conservation Scores (RCS) were calculated, using MSA derived from eggNOG and the following parameters: method was set to "trident", matrix to "blossum62", diversity to 1, chemistry to 0.5, and a gap penalty of 2 was used. This resulted in RCS values for all amino acid residues within the MSA. In order to evaluate differences between evolutionary conservation of SUMOylated and non-SUMOylated lysine residues, we performed a two-tailed paired Student’s t-test on the aforementioned RCS values, while correcting for disordered, globular-exposed, and globular-buried regions. P-values <0.001 were deemed significant.

Comparison of proteomic SUMO studies. In order to align human SUMO target proteins, a scaffold was created based on all unique protein-coding genes as downloaded from UniProt on 16 July 2017. Protein identifier and protein names were imported, along with basic protein size and mass information. SUMO target proteins identified in this study were aligned to the scaffold, along with SUMO target proteins identified in other SUMO proteomics studies. Comparison of human SUMO sites, a scaffold was generated from all 51 amino acid sequence windows derived from this study, as well as 51 amino acid sequence windows identified in other SUMO proteomics studies. Duplicate 51 amino acid sequences were excluded, and afterwards all studies were individually mapped to the scaffold. Qualitative and quantitative information from this study was highly consistent with one previous study. We performed whole-genome alignment to the database. Information on low-throughput-identified SUMO sites, ubiquitylation sites, acetylation sites, and methylation sites, was extracted from PhosphoSitePlus on 19 July 2017, and aligned to the database.
separation of peptides and small proteins. Proteins were separated on the same gradient gels, but using MOPS buffer. Electrophoretic separation was achieved using XCell SureLock™ Mini-Cell Electrophoresis systems (Thermo) set to 80 V for 10 min and then 200 V until the loading dye front reached the bottom of the gel (~45 min). To assist in capture and visualization of peptides and small proteins, supported nitrocellulose membranes with 0.2 μm pore-size were used (Protran, Amersham) for immunoblot transfer. Subsequently, XCell SureLock™ Mini-Cell Electrophoresis systems were used to transfer peptides or proteins from the gels to the nitrocellulose membranes. For peptides, transfer buffer contained 25 mM TRIS, 192 mM glycine, and 20% methanol. For proteins, transfer buffer contained 50 mM TRIS, 384 mM glycine, and 10% methanol. For peptides, transfer was performed ice-cold for 1 h at 20 V and 2 h at 25 V. For proteins, transfer was performed ice-cold for 1 h at 25 V, 1 h at 30 V, and 1 h at 35 V. At this stage, membranes acquired either through dot blot (vacuum) transfer or electrophoretic transfer, were handled similarly. Membranes were rinsed with MQ water, and subsequently incubated with Ponceau-S solution (Sigma) for 1 min. After incubation, membranes were rinsed three times with MQ water and kept in 0.1% acetic acid to stabilize Ponceau-S stain during scanning of the membranes. Subsequently, Ponceau-S was rinsed off the membranes by incubation in PBS supplemented with 0.1% Tween-20 (Sigma); PBST. Membranes were blocked for 1 h at room temperature using 8% (w/v) skin milk powder in PBST (8% milk), after which incubation with primary antibody occurred overnight in 8% milk at 4 °C. Subsequently, membranes were washed three times 10 min with ice-cold PBST, re-blocked for 30 min with 10% milk at 4 °C, incubated with secondary antibody in 10% milk at 4 °C, washed three times 20 min with ice-cold PBST, and washed three times 10 min with ice-cold PBS. Visualization was achieved using enhanced chemiluminescence kits (Novex) according to the manufacturer’s instructions. Chemiluminescence was captured using Amersham Hyperfilm ECL (Sigma). Anti-SUMO2/3 R2 antibody (Ambcan, ab81371) was used as the primary antibody at a 1:2500 concentration in all cases. Goat-anti-mouse HRP conjugated secondary antibody (Jackson Immunoresearch, 115–035–003) was used at a concentration of 1:2500. All scans of immunoblots and Ponceau-S-stained membranes displayed in this manuscript are uncropped and display the full molecular weight range.

Data availability. Mass spectrometry RAW data are available at the ProteomeXchange Consortium database via the Proteomics Identifications (PRIDE) partner repository, under dataset ID PXD008003. Other data can be obtained from the corresponding author upon reasonable request.

References
1. 1. Schorova, L. & Martin, S. Sumoylation in Synaptic Function and Dysfunction. Nat. Rev. Mol. Cell Biol. 19, 641–652 (2018).
2. 2. Hendriks, J. A. & Melchior, F. A small protein modifier with global impact. EMBO J. 35, 6–23 (2016).
3. 3. Fero, A. & Melchior, F. Sumoylation: a regulatory protein modification in health and disease. Annu. Rev. Biochem. 82, 357–385 (2013).
4. 4. Kessler, J. D. et al. SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. Science 335, 348–353 (2012).
5. 5. Krumova, P. & Weishaupt, J. H. Sumoylation: a regulatory protein modification in health and disease. Annu. Rev. Biochem. 82, 357–385 (2013).
6. 6. Hendriks, J. A. & Melchior, F. A small protein modifier with global impact. EMBO J. 35, 6–23 (2016).
7. 7. Fero, A. & Melchior, F. Sumoylation: a regulatory protein modification in health and disease. Annu. Rev. Biochem. 82, 357–385 (2013).
8. 8. Kessler, J. D. et al. SUMOylation-dependent transcriptional subprogram is required for Myc-driven tumorigenesis. Science 335, 348–353 (2012).
9. 9. Hendriks, J. A. & Melchior, F. A small protein modifier with global impact. EMBO J. 35, 6–23 (2016).
10. 10. Fero, A. & Melchior, F. Sumoylation: a regulatory protein modification in health and disease. Annu. Rev. Biochem. 82, 357–385 (2013).

Received: 18 January 2018 Accepted: 5 June 2018
Published online: 25 June 2018
