Identification of proximal SUMO-dependent interactors using SUMO-ID

Orhi Barroso-Gomila[1], Fredrik Trulsson[2], Veronica Muratore[1], Iñigo Canosa[1], Laura Merino-Cacho[1], Ana Rosa Cortazar[1,5], Coralia Perez[1], Mikel Azkargorta[1,3,4], Ibon Iloro [1,3,4], Arkaitz Carracedo[1,5,6,7], Ana M Aransay[1,3], Felix Elortza[1,3,4], Ugo Mayor[6,7], Alfred C. O. Vertegaal[2], Rosa Barrio[1*], James D. Sutherland[1*].

1. Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Bizkaia Technology Park, Building 801A, 48160 Derio, Spain.

2. Cell and Chemical Biology, Leiden University Medical Center (LUMC), 2333 ZA Leiden, The Netherlands

3. CIBERehd, Instituto de Salud Carlos III, C/ Monforte de Lemos 3-5, Pabellón 11, Planta 0, 28029 Madrid, Spain.

4. ProteoRed-ISCIII, Instituto de Salud Carlos III, C/ Monforte de Lemos 3-5, Pabellón 11, Planta 0, 28029 Madrid, Spain.

5. CIBERONC, Instituto de Salud Carlos III, C/ Monforte de Lemos 3-5, Pabellón 11, Planta 0, 28029 Madrid, Spain.

6. Ikerbasque, Basque Foundation for Science, 48011 Bilbao, Spain

7. Biochemistry and Molecular Biology Department, University of the Basque Country (UPV/EHU), E-48940, Leioa, Spain.

23 (*) Corresponding authors: jsutherland@cicbiogune.es, rbarrio@cicbiogune.es

24 R Barrio ORCID: 0000-0002-9663-0669

25 J D Sutherland ORCID: 0000-0003-3229-793X

26

27 Lead Contact: Rosa Barrio rbarrio@cicbiogune.es

28

ABSTRACT

The fast dynamics and reversibility of posttranslational modifications by the ubiquitin family pose significant challenges for research. Here we present SUMO-ID, a technology that merges proximity biotinylation by TurboID and protein-fragment complementation to find SUMO-dependent interactors of proteins of interest. We developed an optimized split-TurboID version and show SUMO interaction-dependent labelling of proteins proximal to PML and RANGAP1. SUMO-dependent interactors of PML are involved in transcription, DNA damage, stress response and SUMO modification and are highly enriched in SUMO Interacting Motifs, but may only represent a subset of the total PML proximal proteome. Likewise, SUMO-ID also allowed us to identify novel interactors of SUMOylated SALL1, a less characterized SUMO substrate. Thus, SUMO-ID is a powerful tool that allows to study the consequences of SUMO-dependent interactions, and may further unravel the complexity of the ubiquitin code.

INTRODUCTION

Ubiquitin-like (UbL) proteins belong to a superfamily of small proteins that attach covalently to target substrates in a transient and reversible manner. The UbL family includes Small Ubiquitin-like Modifiers (SUMOs). The mammalian SUMO family consists of at least three major *SUMO* paralogues (*SUMO1*,*-2*,*-3*). Human SUMO2 and SUMO3 share 97% sequence identity, whereas they share 47% of sequence identity with SUMO1[1]. Protein SUMOylation is a rigorously regulated cycle involving an enzymatic machinery that acts in a stepwise manner. Briefly, the C-terminal di-glycine motif of mature SUMOs binds lysines in substrates through the sequential action of E1 SUMO-activating enzyme SAE1/SAE2, E2 conjugating enzyme UBC9 and SUMO E3 ligases[2]. If required, SUMO as well as the substrate can be recycled by the action of SENPs that cleave the isopeptide bond. Like Ub, SUMO has internal lysines that can be further modified, extended as SUMO chains, modified by Ub chains to target degradation, or even modified by smaller moieties, like acetyl groups[3-5]. Together, these constitute the concept of the "SUMO code" and the ongoing challenge is to understand how these modifications drive distinct substrate outcomes and cellular fates.

SUMO plays crucial roles in nuclear processes underlying health and disease such as the DNA damage response, cell cycle regulation, transcription and proteostasis[6]. SUMO is known to control vital biological processes including development[7] and cholesterol homeostasis[8]. Improvements in mass spectrometry technology have led to the identification to date of more than 40,700 SUMO sites within 6,700 SUMO substrates[9]. While cell-wide proteomics approaches can help to understand global SUMO signaling[10], better tools are needed that allow the study of the cause and consequences of particular SUMOylation events for individual substrates.

SUMO can also interact non-covalently with SUMO interacting motifs (SIMs) found in some proteins. SIMs are β strands composed of an hydrophobic core motif that

interacts with the hydrophobic residues of the SIM-binding groove of SUMOs to form an intramolecular β-sheet[11]. A well characterized role of the SUMO-SIM interaction concerns the SUMO-targeted Ub ligases (STUbL). The two described human STUbLs, RNF4 and Arkadia/RNF111, recognize poly-SUMOylated substrates through their SIMs and ubiquitylate them, leading to their proteasomal degradation[12,13]. The SUMO-SIM interaction also plays critical roles in assembling protein complexes: interaction of the SIM1 of Ran Binding Protein 2 (RanBP2) with the SUMOylated version of Ran GTPase-activating protein 1 (RanGAP1) is crucial for the RanBP2/RanGAP1*SUMO1/UBC9 E3 ligase complex formation[14].

Another intriguing function of SUMO-SIM interaction is the targeting of proteins to Promyelocytic Leukemia Nuclear Bodies (PML NBs). PML NBs are membrane-less ring-like protein structures found in the nucleus. They are bound to the nuclear matrix, make contacts with chromatin fibers[15] and associate with transcriptionally active genomic regions[16]. They consist of a shell composed of PML proteins that surround an inner core in which client proteins localize. Due to the heterogeneity of client proteins, PML NBs have diverse nuclear functions (reviewed in[17,18]). The *PML* gene contains 9 exons and numerous splicing variants. All PML isoforms contain the N terminal TRIpartite Motif (TRIM) that is responsible for PML polymerization and NB formation[19], binding to Arsenic Trioxide (ATO)[20] and may act as an oxidative stress sensor[18]. PML also contains a phospho-SIM located at its exon 7 and shared by most PML isoforms[21]. Almost all PML isoforms contain three putative SUMO sites: K65, K160 and K490. PML SUMOylation is a well characterized signal for RNF4-mediated ubiquitylation and degradation[22].

Proximity-dependent labelling methods are based on promiscuous labeling enzymes that produce reactive molecules that covalently bind neighbor proteins. Labeled proteins can be then purified and identified using affinity-purification coupled to mass

spectrometry methods[23]. Proximity-dependent biotin identification (BioID)[24] uses a promiscuously active *Escherichia coli* biotin ligase (BirA*) generated by a point mutation (R118G) to biotinylate lysines in nearby proteins within an estimated range of 10 nm[25]. By fusing BirA* to specific proteins, BioID efficiently identifies interactors at physiological levels in living cells[26] . It has been extensively used in the Ub field, for instance, to identify substrates of E3 ligases[27,28]. Recently, a more efficient version of BioID, termed TurboID, has been developed[29], being this more suitable for transient protein-protein interaction (PPI) detection. Several studies have developed split-versions and applied "protein fragment complementation" to BioID and TurboID, where proximal biotinylation is dependent on the proximity of the fusion partners, opening new opportunities for spatial and temporal identification of complex-dependent interactomes[30,31].

To study how SUMOylation and SUMO-SIM interactions can lead to other roles and fates for particular substrates poses particular challenges. SUMOylation occurs transiently and often in a small percentage of a given substrate. Modified proteins can be readily deSUMOylated and SUMO can be recycled and passed to other substrates. SUMO-SIM interactions are also difficult to analyze due to their weak affinity (Kd 1-100 μM). To overcome those technical issues, we developed SUMO-ID, a new strategy based on Split-TurboID to identify interactors of specific substrates dependent on SUMO conjugation or interaction. Using PML as a model, we demonstrate that SUMO-ID can enrich for factors that depend on PML-SUMO interaction. Importantly, the identified proteins are represented among proximal interactors of PML identified using full-length TurboID. We also applied SUMO-ID to a less-characterized SUMO substrate, Spalt Like Transcription Factor 1 (SALL1), and identified both known and novel interactors that depend on intact SUMOylation sites in SALL1. SUMO-ID is thus a powerful tool to study

119 transient and dynamic SUMO-dependent interaction events. The developed methodology
120 is generic and therefore widely applicable in the Ub and UbL field to identify readers of
121 these modifications for individual target proteins to improve our insight in non-covalent
122 signal transduction by Ub and UbL.

123 **RESULTS**

124 **Identification of SUMO-dependent interactions: the SUMO-ID strategy**

125 We posited that Split-TurboID, in which one fragment is fused to SUMO and the
126 complementary fragment to a protein of interest, could identify transient SUMO-
127 dependent interactors (Fig. 1). Upon covalent SUMOylation or non-covalent SUMO-SIM
128 interaction, both fragments are brought together, presumably close enough to allow
129 refolding of the TurboID enzyme. In the presence of biotin, the reconstituted TurboID
130 can then label proximal complexes, which can be purified by streptavidin pull-down and
131 identified by liquid chromatography-mass spectrometry (LC-MS). Due to the high
132 affinity of streptavidin-biotin interaction, harsh cell lysis and stringent washes that
133 significantly reduce unspecific protein binding can be applied. We named this approach
134 "SUMO-ID".

135 **T194/G195 Split-TurboID enables SUMO-ID studies**

136 We applied the previously described E256/G257 BioID split-point[30] to TurboID,
137 but found it unsuitable for SUMO-ID. While SUMO-dependent reconstitution of the
138 E256/G257 was observed in our pilot experiments, the NTurboID^{256}-fusions had a
139 significant background biotinylating activity (Figure S1). This is likely due to residual
140 biotin binding and activation by the intact NTurboID^{256} biotin-binding pocket. We
141 examined the BirA structure to identify a new TurboID split-point that would yield two
142 completely inactive fragments (see File S1). The biotin binding pocket of BirA is
143 composed of three β-strands (strands 5, 8 and 9), the N-terminus of helix E and the 110-

128 loop (Fig. 2A). We split TurboID at T194/G195, so that the resulting NTurboID[194] fragment (NTbID) carries the principal 110-128 biotin binding loop and the β-strands 5 and 8, while the CTurboID[195] fragment (CTbID) carries the β-strand 9 necessary to the formation of the biotin binding β-sheet.

We tested T194/G195 Split-TurboID for SUMO-ID. CTbID-SUMOs were incorporated into substrates in a very efficient manner (Fig. 2B). NTbID and CTbID alone were catalytically inactive and yielded no biotinylation. Combining NTbID-substrate and CTbID-SUMOs resulted in a high-yield biotinylation activity of TurboID after 16 hours of biotin exposure. Modification of NTbID-substrates by CTbID-SUMOs (Fig. 2B, FLAG blot, black arrowheads) and its corresponding biotinylation activity (Fig. 2B, biotin blot, black arrowheads) were efficiently detected, notably in the case of PML protein. Free biotinylated CTbID-SUMOs, that might come from recycling of previously labeled moieties, were observed (Fig. 2B, biotin blot, white squares). We also examined by immunofluorescence, and confirmed that the streptavidin signal recognizing the biotinylated substrates is dependent on fragment-complementation (Fig. 2C). Thus, T194/G195 Split-TurboID biotinylation activity is dependent on fragment complementation, with reduced or no leaky biotinylation of the two fragments, so it could be useful for SUMO-ID and for studying other protein-protein interactions.

We applied the rapamycin-inducible dimerization system, based on FKBP (12-kDa FK506-binding protein) and FRB (FKBP-rapamycin-binding domain)[32], which has been used previously to evaluate the PPI dependency of Split-BioID reconstitution[30]. We fused NTbID and CTbID to FRB and FKBP, respectively, and stably expressed the constructs in HEK293FT cells. We tested short and long rapamycin treatments, together with short biotin labeling times, to evaluate self-biotinylation activity of the reconstituted TurboID. As previously shown, the NTbID, even though it contains the principal biotin

binding 110-128 loop, was catalytically inactive after 24 hours of biotin treatment (Fig. 2D). We observed that biotinylation activity of the reconstituted TurboID correlated well with rapamycin and biotin treatments, showing its dependency on PPI and biotin labeling times. 24 hours of rapamycin treatment led to a 16-fold higher FKBP/FRB PPI dependent biotinylation activity at 2 hours of labeling time. Altogether, these data demonstrate that T194/G195 Split-TurboID fragments have low intrinsic affinity and high biotinylation activity at short biotin labeling times, making them suitable for SUMO-ID.

SUMO-ID detects both covalent and non-covalent SUMO-dependent interactions using short biotin-labeling times

Interaction of a protein with SUMO can be via covalent SUMOylation or non-covalent SUMO-SIM interaction. We used PML, which can both be SUMOylated and has a well-characterized SIM domain, in conjunction with SUMO wild type (WT) or mutants that lack the C-terminal di-glycine (ΔGG) necessary for covalent conjugation. We used a stable HEK293FT cell line expressing NTbID-PML, into which CTbID-SUMOs were transfected, using short biotin-labeling times (0.5-2 hours). We observed that CTbID-SUMO1/2 transfections led to high SUMO-dependent biotinylation activity after only 2 hours of biotin treatment (Fig. 3A, biotin blot, black arrowhead). Additionally, ATO treatment, which induces PML SUMOylation, further enhanced the SUMO-dependent biotinylation. With 30 minutes of biotin treatment, CTbID-SUMO1/2$^{\Delta GG}$ induced biotinylation of unmodified NTbID-PML, likely through SUMO-SIM interactions (Fig. 3A, biotin blot, white arrowhead). With longer biotin labeling (2 hours), biotinylation of endogenous SUMO modified NTbID-PML was also detected, more strongly in the case of ATO treatment (Fig. 3A, biotin blot, black arrowhead). Biotinylated free CTbID-SUMO1/2$^{\Delta GG}$ were detected, while the WT counterparts were not biotynilated at 2 hours (Fig. 3A, biotin blot, white squares), supporting that recycling

of biotinylated SUMOs may be linked to longer labeling times. Indeed, additional experiments showed that appearance of free biotinylated CTbID-SUMOs increased with longer labelling times (Figure S2). Altogether, these results demonstrate that SUMO-dependent biotinylation activity for specific targets, especially at short biotin labeling times, may be a useful strategy for identifying specific SUMO-dependent interactors of those proteins.

Reduced labeling times and lower expression levels reduced SUMO recycling, but still allowed some degree of recycling (and therefore loss-of-specificity) to occur, so we incorporated two further modifications. First, we designed SUMO isopeptidase-resistant versions of CTbID-SUMOs (SUMO non-cleavable, or SUMOnc, Fig. 1B)[33]. This could also reduce the target identification derived from SUMO-SIM interactions involving free unincorporated CTbID-SUMOs, since most CTbID-SUMOncs would be incorporated into substrates. The same strategy was applied to Ub (CTbID-Ubnc). Secondly, we transferred non-cleavable CTbID-SUMOs into pTRIPZ, an all-in-one doxycycline-inducible (Dox) lentiviral vector (Fig. 1B). Regulated expression would offset any deleterious effects stemming from non-cleavable SUMO isoforms, and provide useful experimental control (i.e. non-induced *versus* induced). Inducible TRIPZ-CTbID-SUMOnc showed enhanced SUMOylation compared to the constitutive WT SUMO versions (Fig. 3B). Free non-incorporated versions of SUMO1/2nc and Ubnc were not detectable (Fig. 3B, MYC blot, white squares). Stable cell populations were established (HEK293FT, U2OS and RPE-1 cells) for each (SUMO1nc, SUMO2nc, Ubnc). Validation of TRIPZ-CTbID-SUMO2nc by WB and immunofluorescence is shown (Figure S3). We then introduced constitutively-expressed NTbID-PML into TRIPZ-CTbID-SUMO1nc or -SUMO2nc cells and proved that biotinylation occurs in PML NBs as expected, in a doxycycline dependent manner (Figure S4). These data show that the use of regulated

SUMOnc versions leads to both high activity and specificity needed for the SUMO-ID approach.

To further validate the specificity of SUMO-dependent biotinylation activity with PML, we generated control cells carrying NTbID-PML^{3MAS}, a mutated version of PML lacking the three principal SUMOylation sites (K65, K160 and K490) and the best-characterized SIM domain. While strong SUMO-ID biotinylation activity was observed with the WT version of PML, this biotinylation activity was completely abrogated in the case of PML^{3MAS} (Fig. 3C). This lack of biotinylation activity was specific to SUMO, as ubiquitylation-dependent biotinylation activity was observed in TRIPZ-CTbID-Ubnc / NTbID-PML^{3MAS} double stable cell line (Figure S5). NTbID-PML^{WT} forms true NBs, while NTbID-PML^{3MAS} forms NB-like bodies, as reported previously[34] (Fig. 3D). Biotinylation driven by SUMO-ID was observed in NBs containing NTbID-PML^{WT}, and it was enhanced after 2 hours of ATO treatment, but not in the NB-like structures containing NTbID-PML^{3MAS}. Thus, these results show that SUMO-ID biotinylation activity is dependent on substrate-SUMO interaction.

SUMO-ID identifies SUMO-dependent interactors of PML

Since PML NBs are known hubs of SUMO-dependent signaling[17,18], we wondered which interactions in NBs via PML are SUMO-dependent, so we performed SUMO-ID using NTbID-PML^{WT} compared to NTbID-PML^{3MAS}, each combined with TRIPZ-CTbID-SUMO2nc. Biotinylated proteins were purified by streptavidin pull-down and sequenced by LC-MS (Table S1). 59 high-confidence SUMO-dependent interactors of PML were enriched in PML^{WT} SUMO-ID compared to PML^{3MAS} SUMO-ID (Fig. 4A). Among those, SUMO E3 ligases (PIAS1, PIAS2, PIAS4, TRIM28), transcriptional regulators (TRIM22, TRIM24, TRIM33, GTF2I, IRF2BP2, IFI16, ZNF280B, MED23, MEF2D, SNW1, RPAP3), and DNA repair proteins (RMI1, BLM, SLX4, XAB2) were

identified. Of note, PIAS1 is known to induce PML SUMOylation[35] and SUMO-SIM interaction of BLM is necessary for its targeting to PML bodies[36], which highlights the specificity of the SUMO-ID strategy. Of particular interest, GTF2I and IRF2BP2, identified here by SUMO-ID, form fusion proteins with RARA and cause Acute Promyelocytic Leukaemia (APL, see Discussion)[37,38]. We validated these two proteins, as well as TRIM33 and UBC9, as SUMO-dependent interactors of PML by WB (Fig. 4B).

STRING networking of SUMO-dependent interactors of PML shows a highly interconnected cluster related to protein SUMOylation, DNA damage response and transcriptional regulation (Fig. 4C), while GO enrichment also highlighted protein SUMOylation and transcriptional regulation, as well as DNA repair and stress response pathways (Fig. 4D; Table S2). Collectively, this data show that the SUMO-ID strategy can efficiently identify SUMO-dependent interactors of PML, and that SUMO interaction with PML reinforces essential processes.

PML SUMO-ID hits localize to PML NBs

We checked whether some of the SUMO-dependent interactors of PML localize to NBs. We generated a YFP-PML cell line by inserting YFP into the endogenous PML locus in U2OS cells (Figure S6), and looked for co-localization of selected SUMO-dependent PML interactors by confocal microscopy. Within individual cells, we observed frequent and multiple co-localization events for PIAS4, TRIM24, TRIM33 and UBC9 in PML NBs (Fig. 5), whereas PIAS2, GTF2I and IRF2BP2 colocalizations were less frequent, suggesting heterogeneity in PML NB composition that may depend on different factors (including, but not limited to SUMOylation density, subnuclear localization, cell cycle stage, other PTMs, or contrastingly, technical limitations with antibodies or fixations).

SUMO-dependent interactions are a subset of PML proximal proteome

PML NBs are membraneless structures thought to behave as phase-separated liquids and with high heterogeneity in composition[39]. These characteristics make their purification very challenging, and no proteomic data are nowadays available. Therefore, to compare the obtained PML SUMO-ID specific sub-proteome with the regular PML interactome, we decided to characterize a comprehensive PML and PML^{3MAS} proximity interactome using standard full-length TurboID (FLTbID). We generated stable U2OS cell lines for FLTbID-PML^{WT}, FLTbID-PML^{3MAS} and FLTbID alone, and treated them or not with ATO to induce PML SUMOylation. High confidence PML proximal proteome was composed of 271 proteins that were enriched in FLTbID-PML^{WT} samples compared to FLTbID alone (Fig. 6A, Table S3). STRING networking showed a main core cluster composed of 73.6% of the identified proteins (Fig. 6B). The most representative subclusters were composed of 1) RNA splicing and mRNA processing proteins, 2) transcription, RNA biosynthesis and DNA damage response proteins and 3) replication and SUMOylation related proteins. This largely aligned with Gene Ontology (GO) enrichment analysis, which revealed that PML proximal interactors participate in replication, transcription, RNA splicing, DNA damage response, cell cycle regulation, SUMOylation and ubiquitylation, and telomere maintenance, consistent with fact that PML in U2OS regulates the ALT mechanism[40] (alternative lengthening of telomeres; Fig. 6C, Table S4).

SUMOylation of PML is thought to be a controlling factor for composition and dynamics of NBs. Are all NB interactions linked to PML dependent on SUMO? To answer this question, we subdivided the PML interactome into SUMO-dependent or -independent interactors, by comparing FLTbID-PML^{WT} and FLTbID-PML^{3MAS} samples. We observed some proteins that likely localize to PML NBs[41], such as NCOR-1, STAT3,

294 JUN, BRCA2 and HDAC9, were also enriched in TurboID-PML^{3MAS}, suggesting SUMO-
295 independent targeting to PML NBs (Table S3). Importantly, many of SUMO-dependent
296 interactors identified by SUMO-ID are part of SUMO-dependent PML NBs proteome
297 using standard TurboID, including PIAS2, PIAS4, TRIM24, TRIM33 and IRF2BP2 (Fig.
298 7A; Table S3), supporting the validity of SUMO-ID to identify SUMO-dependent
299 interactors. Interestingly, scores of some PML interactors decreased after ATO treatment
300 (TRIM24, TRIM33, SENP5), suggesting that those proteins may rapidly undergo
301 dissociation or degradation in response to PML SUMOylation. We confirmed such effect
302 for TRIM24 by WB (Fig. 7B). Altogether, these data confirm that SUMO-ID identified
303 hits are a subset of the SUMO-dependent PML proximal proteome.

304 **SUMO-dependent interactors of PML are enriched in SIMs**

305 We expected that many of the SUMO-dependent PML interactors might do so via
306 SUMO-SIM interactions and, therefore, should contain or be enriched in SIMs. To test
307 this, we designed and executed an *in-silico* SIM enrichment analysis. We generated 1000
308 random lists of 59 proteins (the size of the SUMO-ID identified protein list) and evaluated
309 the presence of SIMs (Table S5). The median of single SIM and multiple SIM presence
310 in the random lists were 45.76% and 23.73% respectively (Fig. 7C). SUMO-ID identified
311 proteins showed a much higher content of SIMs, with single SIM and multiple SIM
312 presence values of 96.61% and 89.83% respectively. It is noteworthy that around 83% of
313 the identified SIMs in PML SUMO-ID list were preceded or followed within the first 4
314 amino-acids by acidic residues (D, E) or a Serine. Since longer proteins are expected to
315 have more SIMs, we then normalized the SIM content with the size of proteins on the
316 lists to obtain the value of "SIMs per thousand of amino acids" (STAA) (Table S5). The
317 values obtained with the random lists showed a Gaussian distribution (d'Agostino and
318 Pearson normality test, K2 value 3.836, *p*-value 0.15) (Fig. 7D). The median of the values

obtained with the random lists was 4.85 (Log2 = 2.28) STAA, while for PML SUMO-ID list was 18.42 (Log2 = 4.20) STAA, which translates to a SIM enrichment value of 3.8 times higher than the random lists (empirical *p*-value < 0.001). These results show that SUMO-dependent interactors of PML are highly enriched in SIMs.

SUMO-ID identifies interactors of SUMOylated SALL1

To test the sensitivity and discovery potential of the SUMO-ID, we applied this technique to SALL1, a transcriptional repressor that is SUMOylated[42,43], but of which nothing is known about the causes or consequences of this modification. Using TRIPZ-CTbID–SUMO1nc or SUMO2nc HEK293FT stable cell lines, we introduced NTbID–SALL1^{WT} or SALL1$^{\Delta SUMO}$ (with mutations in 4 major SUMOylation consensus sites) and evaluated SUMO-ID by WB. Efficient SUMO-ID biotinylation activity was observed when using SUMO2nc (Fig. 8A, black arrowhead). NTbID–SALL1^{WT} localizes to the nucleus, forming nuclear bodies with high SUMO-ID activity, and NTbID–SALL1$^{\Delta SUMO}$ also forms aggregates in the cytoplasm (Fig. 8B). Specificity of SALL1 SUMO-ID was confirmed in cells, as biotinylation occurs only in SALL1^{WT} upon doxycycline induction and biotin supplementation. SALL1 SUMO-ID identified potential SUMO-dependent interactors of SALL1 such as the transcription factors TLE3, DACH1/2 and ARID3B, as well as NuRD complex proteins GATAD2A/B, MTA1/2 and RBBP4/7 (Fig. 8C; Table S6), already known as SALL1 interactors[44]. We also identified components of the SUMOylation machinery, such as PIAS1. We confirmed that SUMOylated SALL1 was biotinylated and purified via SUMO-ID (Fig. 8D, black arrowheads) as well as NuRD complex proteins GATAD2B, MTA2 and RBBP4 (Fig. 8D). MCODE subclustering of the STRING interaction network showed a highly interconnected cluster composed of NuRD complex proteins (Fig. 8E) that was also enriched as GO term (*p*-value $2.40 \cdot 10^{-4}$). Thus, SUMO-ID is sensitive and specific, allowing the study of SUMO-dependent

interactors for proteins of interest, opening new avenues of understanding how SUMO can affect their function.

DISCUSSION

The fast dynamics and reversibility of SUMOylation, and the low affinity of SUMO-SIM interactions pose significant challenges not only for SUMO research, but for respective studies of Ub and other UbLs. The use of His-tagged K0-SUMO to isolate substrates and map SUMOylation sites has been instrumental to show the widespread presence of this modification in the human proteome[9,45]. Direct purification of SUMOylated proteins using immunoprecipitation is a gold standard and can be applied to cells and tissues[46], but is also challenging because SUMOylation might affect a small proportion of low abundance proteins, and perhaps only under certain conditions (e.g a discrete cell cycle phase or upon DNA damage). Recently, the NanoBiT-based ubiquitin conjugation assay (NUbiCA) was described that uses a split-luciferase approach to allow a quantitative assessment of Ub-modified proteins[47]. Bimolecular fluorescence complementation (BiFC) approaches employ a split fluorescent protein that enables the localization of UbL-modified proteins in yeast or human cells to be monitored[48-50]. If applied to UbLs and substrates, the BiCAP method[51], which allows purification of reconstituted GFP using GFP nanobodies, could likely enrich modified substrates and perhaps interactors. However, none of these approaches captures the dynamic environment of specific UbL-modified proteins, often characterized by weak and transient interactions.

Here we describe SUMO-ID, a powerful technique that allows the study of the causes and consequences of SUMO-dependent interactions for specific proteins of interest. The fast biotinylation activity of TurboID and the specificity obtained with “protein fragment complementation” permit SUMO-ID to specifically biotinylate

interactors of substrates in a SUMO-dependent manner. Combined with sensitive proteomic methods, SUMO-ID allows the identification of specific interactors, potentially revealing enzymatic machinery responsible for the SUMOylation as well as interactors that may be stabilized or recruited as a consequence of the modification. Like approaches using BiFC, the subcellular localization of SUMO-modified substrates using SUMO-ID can also be inferred, through the use of fluorescent streptavidin. However, caution should be taken with the mentioned factors in order to maintain specificity, such is the use of non-cleavable forms of UbLs or the application of short biotin labelling times. This strategy might compromise the identification of SUMO isopeptidases since their binding to SUMOylated substrate is likely affected.

At the core of SUMO-ID is Split-TurboID, which individual halves should ideally have no activity, as with all split-protein approaches. For SUMO-ID, we initially applied the E256/G257 split point described for Split-BioID[30] to the fast-labelling TurboID derivative, but found that the N-terminal half (1-256) retained substantial biotinylation capacity. We speculate that this is because the "biotin pocket" is still intact and might allow leaky release of biotinoyl-AMP. Leaky biotinylation of TurboID 1-256 was also observed by Cho and colleagues in their recently published report on Split-TurboID[31]. Their final design was based on a L73/G74 split point which showed efficient proximity-dependent reconstitution and biotinylation, but still leaves the biotin pocket intact in C-terminal 74-321 half, opening the possibility of leaky biotinylation during longer labelling times or in stable cell lines. To avoid this, we developed and validated T194/G195 Split-TurboID that separates the β-strands 5 and 8 from the β-strand 9, completely abrogating any residual biotinylation activity of the fragments.

Here we used SUMO-ID to unravel the role of PML SUMOylation in PML NBs function. We identified 59 proteins as SUMO-dependent PML interactors that participate

394 in essential nuclear processes such as protein SUMOylation, transcriptional regulation,
395 DNA repair and stress response. There is growing evidence that PML interaction with
396 SUMO might allow partners to localise into PML NBs through SUMO-SIM
397 interactions[36,52]. We demonstrated that most of the proteins identified by SUMO-ID are
398 indeed part of the proteome of SUMO-dependent PML NBs and that they are enriched in
399 SIMs, suggesting SUMO-SIM interaction dependency. It has been proposed that, after
400 such partner recruitment, proteins might undergo SUMOylation by the PML NB-
401 localized SUMO machinery that reinforces their sequestration[53]. In fact, PML NBs are,
402 together with the nuclear rim, the major targets of active SUMOylation[54]. Our data
403 reinforce this enzyme/substrate co-concentration model as we observed that
404 SUMOylation machinery enzymes (UBC9, PIAS1, PIAS2, PIAS4 and TRIM28) localize
405 to PML NBs in a SUMO-dependent manner and 80% of the SUMO-dependent PML
406 interactors (47 out of 59) are SUMO substrates[9,10].

407 To compare our list of SUMO-dependent *versus* general interactors of PML, we
408 performed a TurboID assay for PML, with cells alone or treated with ATO, and identified
409 271 proteins. ATO induces PML NB formation, subsequent PML SUMOylation, partner
410 recruitment and finally PML degradation[22,55]. It is used to treat APL, a type of Acute
411 Myelocytic Leukaemia (AML), which is mainly caused by the t(15;17) translocation that
412 fuses PML to RARA. Interestingly, two of our SUMO-ID hits, IRF2BP2 and GTF2I, also
413 form fusion proteins with RARA and cause APL, albeit less commonly than PML
414 fusions[37,38]. We validated that both localize to PML NBs. While many of the SUMO-ID
415 candidates show increased peptide intensity in PML NBs after ATO treatment, we
416 observed that some of them decreased. IRF2BP2 and TRIM24, which has also been linked
417 to AML[56,57], showed reduced levels after ATO treatment, suggesting that they might
418 undergo degradation. In line with this idea, the 11S proteasome components are recruited

419 into mature PML NBs and their localization is enhanced with ATO treatment[58],
420 suggesting that mature PML NBs may also act as proteolytic sites. In fact, inhibition of
421 ubiquitylation accumulates SUMOylated proteins within PML NBs[59], suggesting that
422 many clients that are targeted to PML NBs and that are SUMOylated, might undergo
423 ubiquitylation and degradation. Altogether, these data provide further insight into the role
424 of PML SUMOylation in NB biology and open new ways of looking at the mechanisms
425 of ATO in APL treatment.

426 The successful application of SUMO-ID to SALL1, a poorly characterized SUMO
427 substrate, illustrates the sensitivity and utility of SUMO-ID. Although SALL1
428 SUMOylation levels are vanishingly low under physiological conditions, SUMO-ID
429 revealed SUMO-dependent enrichment of the NuRD complex proteins GATAD2A/B,
430 MTA1/2 and RBBP4/7. The association between SALL1, a transcriptional repressor, and
431 the NuRD complex, a repressive histone deacetylase complex, has been previously
432 described[44]. The interaction is mediated by an N-terminal 12 amino acid motif of
433 SALL1[44]. Once recruited, SUMOylation of SALL1 might serve to stabilize the repressor
434 complex via SUMO-SIM interactions, with predicted SIMs present in multiple NuRD
435 complex subunits. As histone SUMOylation is also linked to transcriptional repression[60],
436 SUMO-SIM interactions might further reinforce the SALL1/NuRD complex and drive
437 histone deacetylation at SALL1 targets. In addition, we also found TLE3, DACH1/2 and
438 ARID3B transcription factors as SUMO-dependent interactors of SALL1. TLE3, a
439 transcriptional repressor of the Groucho/TLE family, interacts with HDAC2 (another
440 NuRD complex component) and can regulate acetylation levels[61]. Both TLE3 and the
441 tumor suppressor DACH1 are negative regulators of Wnt signaling[62,63]. Interestingly,
442 SALL1 has been shown to enhance Wnt signaling[64]. Perhaps interaction with
443 SUMOylated SALL1 serves to counteract these negative effects.

In summary, we demonstrate here that SUMO-ID, based on the 194/195 Split-TurboID reconstitution, can facilitate the identification of SUMO-dependent interactions with a protein of interest. It has little or no background, with high biotinylation activity when expressed at low levels and with short biotin incubation time. We believe that this technique improves sensitivity and selectivity when applied to infrequent SUMOylation events and low-affinity of SUMO-SIM interactions. This strategy can be applied to other UbL modifications (e.g Ub-ID shown in Figure S5), and the 194/195 Split-TurboID may be useful for other applications in cell biological studies.

MATERIALS AND METHODS

Cell Culture

U2OS (ATCC HTB-96) and HEK293FT (Invitrogen) were cultured at 37°C and 5% CO_2 in Dulbecco's modified Eagle Medium (DMEM) supplemented with 10% fetal bovine serum (FBS, Gibco) and 1% penicillin/streptomycin (Gibco). Human telomerase reverse transcriptase immortalized retinal pigment epithelial cells (hTERT-RPE1, ATCC CRL-4000) were cultured in DMEM:F12 (Gibco) supplemented with 10% FBS, 2 mM L-Glutamine and 1% penicillin and streptomycin. Cultured cells were maintained through 20 passages maximum and tested negative for mycoplasma.

Cloning

TurboID was a kind gift of A. Ting (Addgene #107171)[29]. PMLIVa^{WT} and PMLIVa^{3MAS} were previously described[21]. SUMO1, SUMO2, Ub, RANGAP1 and UBC9 ORFs were amplified from U2OS cell cDNA by high-fidelity PCR (Platinum SuperFi DNA Polymerase; Invitrogen). All constructs were generated by standard cloning or by Gibson Assembly (NEBuilder HiFi Assembly, NEB) using XL10-Gold bacteria (Agilent). Depending on the construction, plasmid backbones derived from EYFP-N1 (Clontech/Takara), Lenti-Cas9-blast (a kind gift of F. Zhang; Addgene #52962) or TRIPZ

(Open Biosystems/Horizon) were used. After assembly, all vectors were validated by sequencing. Additional details for constructs are described in Table S7. Cloning details are available upon request.

Lentiviral transduction

Lentiviral expression constructs were packaged in HEK293FT cells using calcium phosphate transfection of psPAX2 and pMD2.G (kind gifts of D. Trono; Addgene #12260, 12259) and pTAT (kind gift of P. Fortes; for TRIPZ-based vectors). Transfection media was removed after 12-18 hours and replaced with fresh media. Lentiviral supernatants were collected twice (24 hours each), pooled, filtered (0.45 μm), and supplemented with sterile 8.5% PEG6000, 0.3 M NaCl, and incubated 12-18 hours at 4°C. Lentiviral particles were concentrated by centrifugation (1500 x g, 45 minutes, 4°C). Non-concentrated virus (or dilutions thereof) were used to transduce HEK293FT, and 8x concentrated virus was used for U2OS and hTERT-RPE1 cells. Drug selection was performed as follows: 1 μg/ml puromycin (Santa Cruz) for U2OS and HEK293FT cells, 5 μg/ml for hTERT-RPE1 cells; 5 μg/ml blasticidin (Santa Cruz) for U2OS, HEK293FT and hTERT-RPE1 cells.

CRISPR-Cas9 genome editing

Human PML encodes multiple isoforms, but most differ at the 3' end. To target EYFP into the first coding exon, shared by most PML isoforms, an sgRNA target site was chosen (CTGCACCCGCCCGATCTCCG) using Broad institute GPP sgRNA Designer[65]. Custom oligos were cloned into px459v2.0 (a kind gift of F. Zhang; Addgene #62988). A targeting vector was made by amplifying 5' and 3' homology arms using U2OS genomic DNA, as well as the EYFP ORF (see Table S7 for oligo details). These fragments were assembled by overlap extension using high fidelity PCR and the resulting amplicon was TOPO-cloned and sequence-confirmed. Lipofectamine 2000 (Invitrogen) was used

to transfect U2OS with linear targeting vector and px459 encoding SpCas9, puromycin resistance, and the PML-targeting sgRNA. 24 hours post-transfection, cells were selected for additional 24-36 hours with 2 µg/ml puromycin. Cells were plated at low density and clones were examined by fluorescence microscopy. Clones with clear nuclear body signals were manually picked and expanded. YFP-PML insertions and copy number were validated by PCR, sequencing, and Western blotting.

Transient transfections and drug treatments

HEK293FT cells were transiently transfected using calcium phosphate method. U2OS cells were transiently transfected using Effectene Transfection Reagent (Qiagen). After 24 hours of transfection, cells were treated with biotin (50 µM; Sigma-Aldrich) for indicated exposure times. For stably transduced TRIPZ cell lines, induction with doxycycline (1 µg/ml; 24 hours; Sigma-Aldrich) was performed prior to biotin treatment. ATO (1 µM; 2 hours; Sigma-Aldrich) treatments were performed (with or without biotin, depending on experiment) prior to cell lysis or immunostaining.

Western blot analysis

Cells were washed 2x with PBS to remove excess biotin and lysed in highly stringent washing buffer 5 (WB5; 8 M urea, 1% SDS in 1X PBS) supplemented with 1x protease inhibitor cocktail (Roche) and 50 µM NEM. Samples were then sonicated and cleared by centrifugation (25000 x g, 30 minutes, RT). 10-20 µg of protein was loaded for SDS-PAGE and transferred to nitrocellulose membranes. Blocking was performed in 5% milk in PBT (1x PBS, 0.1% Tween-20). Casein-based blocking solution was used for anti-biotin blots (Sigma). Primary antibodies were incubated over-night at 4°C and secondary antibodies 1 hour at room temperature (RT). Antibodies used: anti-biotin-HRP (1/1000), anti-Myc (1/1000), anti-alpha-Actinin (1/5000) (Cell Signaling Technology); anti-Flag (1/1000), anti-GTF2I (1/1000) (Sigma-Aldrich); anti-BirA (1/1000; SinoBiological);

519 anti-IRF2BP2 (1/1000), anti-UBC9 (1/1000), anti-TRIM24 (1/1000), anti-TRIM33
520 (1/1000), anti-PIAS2 (1/1000), anti-PIAS4 (1/1000), anti-GATAD2B (1/1000), anti-
521 MTA2 (1/1000), anti-RBBP4 (1/1000), anti-PML (1/1000), anti-GAPDH (1/5000), anti-
522 beta-Actin (1/5000), anti-alpha-Tubulin (1/5000) (Proteintech); anti-PML (1/1000)
523 (Bethyl); anti-GFP (1/1000) (SantaCruz); anti-Mouse-HRP, anti-Rabbit-HRP (1:5000)
524 (Jackson ImmunoResearch). Proteins were detected using Clarity ECL (BioRad) or Super
525 Signal West Femto (ThermoFisher). Quantification of bands was performed using ImageJ
526 (v2.0.0-rc-69/1.52n) software and normalized against loading controls (GAPDH, actin,
527 tubulin or alpha-actinin depending on experiments).

528 **Immunostaining and confocal microscopy**

529 U2OS and HEK293FT cells were seeded on 11 mm coverslips (25,000 cells per well; 24
530 well plate). HEK293FT coverslips were pre-treated with poly-L-lysine. After washing 3
531 times with 1x PBS, cells were fixed with 4% PFA supplemented with 0.1% Triton X-100
532 in PBS for 15 minutes at RT. Then, coverslips were washed 3 times with 1x PBS.
533 Blocking was performed for 30 minutes at RT in blocking buffer (2% fetal calf serum,
534 1% BSA in 1x PBS). Primary antibodies were incubated for 2 hours at 37°C and cells
535 were washed with 1x PBS 3 times. Primary antibodies used: anti-BirA (1/500;
536 SinoBiological); anti-Myc (1/200; Cell Signaling Technology); anti-GTF2I (1/100;
537 Sigma-Aldrich); anti-IRF2BP2, anti-UBC9, anti-TRIM24, anti-TRIM33, anti-PIAS2,
538 anti-PIAS4, anti-CBX4 (all 1/100; Proteintech); anti-B23 (1/100) (Santa Cruz); anti-
539 SC35 (1/100) (BD biosciences); anti-SUMO2/3 (1/100) (DSHB). Then secondary
540 antibodies (together with fluorescent streptavidin) were incubated for 1 hour at 37°C,
541 followed by nuclear staining with DAPI (10 minutes, 300 ng/ml in PBS; Sigma Aldrich).
542 Antibodies used: anti-Rabbit Alexa Fluor 488, anti-Mouse Alexa Fluor 568, anti-Rabbit
543 Alexa Fluor 568 (all 1/200; ThermoFisher). Streptavidin Alexa Fluor 594 (1/200, Jackson

ImmunoResearch) was used. Fluorescence imaging was performed using confocal microscopy (Leica SP8 Lightning) with 63x Plan ApoChromat NA1.4. To obtain the signal histograms for co-localization studies in Fig. 5, we used the plot profile tool in ImageJ (v2.0.0-rc-69/1.52n).

Pulldown of biotinylated proteins

Cleared lysates from WB5 lysis buffer were adjusted to the same protein concentration before incubating them with 1/50 (v_{beads}/v_{lysate}) equilibrated NeutrAvidin-agarose beads (ThermoFisher) over-night at RT. Due to the high-affinity interaction between biotin and streptavidin, beads were subjected to stringent series of washes, using the following WBs ($v_{WB}/2v_{lysate}$), all made in 1x PBS: 2x WB1 (8 M urea, 0.25% SDS); 3x WB2 (6 M Guanidine-HCl); 1x WB3 (6.4 M urea, 1 M NaCl, 0.2% SDS); 3x WB4 (4 M urea, 1 M NaCl, 10% isopropanol, 10% ethanol and 0.2% SDS); 1x WB1; 1x WB5; and 3x WB6 (2% SDS). Biotinylated proteins were eluted in 80 µl of Elution Buffer (4x Laemmli buffer, 100 mM DTT) through heating at 99°C for 5 minutes and subsequent vortexing. Beads were separated by centrifugation (25000 x g, 2 minutes).

Liquid Chromatography Mass Spectrometry (LC-MS/MS)

A stable TRIPZ–MYC-CTurboID^{195}–SUMO2nc U2OS cell line was transduced with either EFS–FLAG-NTurboID^{194}–PML^{WT} or EFS–FLAG-NTurboID^{194}–PML^{3MAS} for PML SUMO-ID experiments. Selection in blasticidin (5 µg/ml) and puromycin (1 µg/ml) was performed. Expression was validated by Western blot and immunostaining prior to scale-up for mass spectrometry. The TurboID-PML experiments used U2OS stable cell lines expressing low and equivalent levels of PML^{WT}–TurboID, PML^{3MAS}–TurboID and TurboID alone, selected by blasticidin (5 µg/ml), and treated or not with ATO for 2 hours. For SALL1 SUMO-ID, a HEK293FT stable cell line expressing low levels of TRIPZ–MYC-CTurboID^{195}–SUMO2nc (selected with puromycin, 1 µg/ml) was transiently

569 transfected with EFS-FLAG-NTurboID^{194}–SALL1^{WT} or EFS-FLAG-NTurboID^{194}–
570 SALL1$^{\Delta SUMO}$ (a mutant carrying Lys>Arg mutations at K571, K592, K982, K1086).
571 Three independent pulldown experiments (8 x 10^7 cells per replicate, 8 ml of lysis) were
572 analyzed by LC-MS/MS.

573 Samples eluted from the NeutrAvidin beads were separated in SDS-PAGE (50% loaded)
574 and stained with Sypro Ruby (Invitrogen) according to manufacturer's instructions. Gel
575 lanes were sliced into 3 pieces as accurately as possible to guarantee reproducibility. The
576 slices were subsequently washed in milli-Q water. Reduction and alkylation were
577 performed using dithiothreitol (10 mM DTT in 50 mM ammonium bicarbonate) at 56°C
578 for 20 min, followed by iodoacetamide (50 mM chloroacetamide in 50 mM ammonium
579 bicarbonate) for another 20 minutes in the dark. Gel pieces were dried and incubated with
580 trypsin (12.5 µg/ml in 50 mM ammonium bicarbonate) for 20 minutes on ice. After
581 rehydration, the trypsin supernatant was discarded. Gel pieces were hydrated with 50 mM
582 ammonium bicarbonate, and incubated overnight at 37ºC. After digestion, acidic peptides
583 were cleaned with TFA 0.1% and dried out in a RVC2 25 speedvac concentrator (Christ).
584 Peptides were resuspended in 10 µl 0.1% formic acid (FA) and sonicated for 5 min prior
585 to analysis.

586 PML samples were analyzed in a novel hybrid trapped ion mobility spectrometry –
587 quadrupole time of flight mass spectrometer (timsTOF Pro with PASEF, Bruker
588 Daltonics) coupled online to a nanoElute liquid chromatograph (Bruker). This mass
589 spectrometer takes advantage of a novel scan mode termed parallel accumulation – serial
590 fragmentation (PASEF). Sample (200 ng) was directly loaded in a 15 cm Bruker
591 nanoelute FIFTEEN C18 analytical column (Bruker) and resolved at 400 nl/min with a
592 100 minutes gradient. Column was heated to 50ºC using an oven.

For the analysis of SALL1 samples peptides were eluted from stage-tips in a solvent composed of deionized water/acetonitrile/FA at a ratio of 50/50/0.1 v/v/v. Peptides were lyophilized and dissolved in solvent A composed of deionized water/FA at a ratio of 100/0.1 v/v and subsequently analyzed by on-line C18 nano-HPLC MS/MS with a system consisting of an Ultimate 3000 nano gradient HPLC system (ThermoFisher), and an Exploris 480 mass spectrometer (ThermoFisher). Fractions were loaded onto a cartridge precolumn (5 mm x ID 300 µm, C18 PepMap 100 A, 5 µm particles (ThermoFisher)), using solvent A at a flow of 10 µl/minute for 3 minutes and eluted via a homemade analytical nano-HPLC column (50 cm × ID 75 µm; Reprosil-Pur C18-AQ 1.9 µm, 120 A; Dr. Maisch GmbH). The gradient was run from 2% to 40% solvent B (water/acetonitrile/FA at a ratio of 20/80/0.1 v/v/v) in 40 minutes. The nano-HPLC column was drawn to a tip of ~10 µm and acted as the electrospray needle of the MS source. The temperature of the nano-HPLC column was set to 50°C (Sonation GmbH). The mass spectrometer was operated in data-dependent MS/MS mode for a cycle time of 3 seconds, with a HCD collision energy at 28 V and recording of the MS2 spectrum in the orbitrap, with a quadrupole isolation width of 1.6 Da. In the master scan (MS1) the resolution was set to 60,000, and the scan range was set to 300-1600, at an Automatic Gain Control (AGC) target of 3×10^6 with automated fill time. A lock mass correction on the background ion m/z=445.12 was used. Precursors were dynamically excluded after n=1 with an exclusion duration of 30 seconds, and with a precursor range of 10 ppm. Charge states 2-6 were included. For MS2 the first mass was set to 120 Da, and the MS2 scan resolution was 30,000 at an AGC target of 75,000 with automated fill time.

Mass Spectrometry data analysis

Raw MS files were analyzed using MaxQuant (version 1.6.17)[66] matching to a human proteome (Uniprot filtered reviewed *H. sapiens* proteome, UP000005640) with a

maximum of 4 missed cleavages and with precursor and fragment tolerances of 20 ppm and 0.05 Da. Label-Free Quantification was enabled with default values except for a ratio count set to 1. Slices corresponding to same lanes were considered as fractions. Biotinylation on lysine and on protein N-term was included as a variable modification for SALL1 SUMO-ID data, and biotinylated peptides were set to be included for quantification. Matching between runs and matching unidentified features were enabled. Only proteins identified with at least one peptide at FDR<1% were considered for further analysis. Data were loaded onto the Perseus platform (version 1.6.14)[67] and further processed (Log2 transformation, imputation, median normalization when needed). A t-test was applied in order to determine the statistical significance of the differences detected. Proteins detected with at least 2 peptides (except when otherwise specified) and in at least 2 of the 3 replicates were included.

Network analysis was performed using the STRING app version 1.4.2[68] in Cytoscape version 3.7.2[69], with a high confidence interaction score (0.7). Transparency and width of the edges were continuously mapped to the String score (text mining, databases, coexpression, experiments, fusion, neighborhood and cooccurrence). The Molecular COmplex DEtection (MCODE) plug-in version 1.5.1[70] was used to identify highly connected subclusters of proteins (degree cutoff of 2; Cluster finding: Haircut; Node score cutoff of 0.2; K-Core of 2; Max. Depth of 100). Gene ontology analysis was performed using g:Profiler web server[71].

The mass spectrometry proteomics data corresponding to PML SUMO-ID, PML-TurboID and SALL1 SUMO-ID experiments have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository[72] with the dataset identifiers PXD021770, PXD021809 and PXD021923, respectively.

SIM enrichment analysis

A thousand lists with the same number of proteins as PML SUMO-ID list (59) have been randomly generated from the human proteome (Uniprot filtered reviewed *H. sapiens* proteome, UP000005640). All those lists have been analyzed by adapting the script from[73] and running it on Python version 2.7.5, to obtain the content and number of SIM motifs per protein (ψ-ψ-X-ψ; ψ-X-ψ-ψ; ψ-ψ-ψ; where ψ is either a L, I or V and X is any amino acid) and the number of SIMs per thousand of amino acids (STAA). After removing three outliers (lists 46, 782, 794; ROUT method, Q=1%), STAA values from the random lists were normalized to Log2 and validated for Gaussian distribution (d'Agostino and Pearson normality test). Enrichments were computed using R software v3.6.0 and calculated as the ratio between PML SUMO-ID STAA value and the median of STAA values from the random lists. Empirical p-values have been calculated by counting the number of random lists whose STAA values were as extreme as the PML SUMO-ID STAA value. The raw data from the SIM enrichment analysis and the script can be found in Table S5 and Supplementary Source 1, respectively.

ACKNOWLEDGEMENTS

We are thankful to Iraide Escobes for her work in the Proteomics Platform at CIC bioGUNE and Arnoud de Ru in the Center for Proteomics and Metabolomics at the LUMC. RB and ACOV acknowledge funding by the grant 765445-EU (UbiCODE Program). RB acknowledges funding by grants BFU2017-84653-P (MINECO/FEDER, EU), SEV-2016-0644 (Severo Ochoa Excellence Program), SAF2017-90900-REDT (UBIRed Program) and IT1165-19 (Basque Country Government). Additional support was provided by the Department of Industry, Tourism, and Trade of the Basque Country Government (Elkartek Research Programs) and by the Innovation Technology Department of the Bizkaia County. FE is at Proteomics Platform, member of ProteoRed-

667 ISCIII (PT13/0001/0027). FE and AMA acknowledge CIBERehd. AC acknowledges the
668 Basque Department of education (IKERTALDE IT1106-16), the MCIU (PID2019-
669 108787RB-I00 (FEDER/EU), Severo Ochoa Excellence Accreditation SEV-2016-0644,
670 Excellence Networks RED2018-102769-T), the AECC (GCTRA18006CARR), La Caixa
671 Foundation (ID 100010434), under the agreement LCF/PR/HR17, and the European
672 Research Council (Starting Grant 336343, PoC 754627, Consolidator grant 819242).
673 CIBERONC was co-funded with FEDER funds. UM acknowledges the Basque
674 Government Department of Education (IT1165-19) and the Spanish MCIU (SAF2016-
675 76898-P (FEDER/EU)).

676 AUTHOR CONTRIBUTIONS

677 O.B.-G., J.D.S. and R.B designed experiments, analyzed data and wrote the manuscript.
678 O.B.-G., F.T., V.M., I.C., A.R.C., C.P., M.A., I.I. and J.D.S. developed experimental
679 protocols and performed experiments. A.C., A.M.A., F.E., U.M. and A.C.O.V. provided
680 scientific resources.

681 COMPETING INTERESTS

682 The authors declare no competing interests.

683 REFERENCES

684 1. Flotho, A. & Melchior, F. Sumoylation: a regulatory protein modification in
685 health and disease. *Annu Rev Biochem* **82**, 357-385 (2013).
686 2. Gareau, J.R. & Lima, C.D. The SUMO pathway: emerging mechanisms that
687 shape specificity, conjugation and recognition. *Nat Rev Mol Cell Biol* **11**, 861-
688 871 (2010).

689 3. Ullmann, R., Chien, C.D., Avantaggiati, M.L. & Muller, S. An acetylation
690 switch regulates SUMO-dependent protein interaction networks. *Mol Cell* **46**,
691 759-770 (2012).
692 4. Vertegaal, A.C. SUMO chains: polymeric signals. *Biochem Soc Trans* **38**, 46-49
693 (2010).
694 5. Aillet, F. *et al.* Heterologous SUMO-2/3-ubiquitin chains optimize
695 IkappaBalpha degradation and NF-kappaB activity. *PLoS One* **7**, e51672 (2012).
696 6. Pichler, A., Fatouros, C., Lee, H. & Eisenhardt, N. SUMO conjugation - a
697 mechanistic view. *Biomol Concepts* **8**, 13-36 (2017).
698 7. Talamillo, A. *et al.* The role of SUMOylation during development. *Biochem Soc*
699 *Trans* **48**, 463-478 (2020).
700 8. Talamillo, A. *et al.* SUMOylation in the control of cholesterol homeostasis.
701 *Open Biol* **10**, 200054 (2020).
702 9. Hendriks, I.A. *et al.* Site-specific mapping of the human SUMO proteome
703 reveals co-modification with phosphorylation. *Nat Struct Mol Biol* **24**, 325-336
704 (2017).
705 10. Hendriks, I.A. & Vertegaal, A.C. A comprehensive compilation of SUMO
706 proteomics. *Nat Rev Mol Cell Biol* **17**, 581-595 (2016).
707 11. Kerscher, O. SUMO junction-what's your function? New insights through
708 SUMO-interacting motifs. *EMBO Rep* **8**, 550-555 (2007).
709 12. Kumar, R., Gonzalez-Prieto, R., Xiao, Z., Verlaan-de Vries, M. & Vertegaal,
710 A.C.O. The STUbL RNF4 regulates protein group SUMOylation by targeting
711 the SUMO conjugation machinery. *Nat Commun* **8**, 1809 (2017).

712 13. Sriramachandran, A.M. *et al.* Arkadia/RNF111 is a SUMO-targeted ubiquitin
713 ligase with preference for substrates marked with SUMO1-capped SUMO2/3
714 chain. *Nat Commun* **10**, 3678 (2019).
715 14. Werner, A., Flotho, A. & Melchior, F. The RanBP2/RanGAP1*SUMO1/Ubc9
716 complex is a multisubunit SUMO E3 ligase. *Mol Cell* **46**, 287-298 (2012).
717 15. Eskiw, C.H., Dellaire, G. & Bazett-Jones, D.P. Chromatin contributes to
718 structural integrity of promyelocytic leukemia bodies through a SUMO-1-
719 independent mechanism. *J Biol Chem* **279**, 9577-9585 (2004).
720 16. Wang, J. *et al.* Promyelocytic leukemia nuclear bodies associate with
721 transcriptionally active genomic regions. *J Cell Biol* **164**, 515-526 (2004).
722 17. Bernardi, R. & Pandolfi, P.P. Structure, dynamics and functions of
723 promyelocytic leukaemia nuclear bodies. *Nat Rev Mol Cell Biol* **8**, 1006-1016
724 (2007).
725 18. Lallemand-Breitenbach, V. & de The, H. PML nuclear bodies: from architecture
726 to function. *Curr Opin Cell Biol* **52**, 154-161 (2018).
727 19. Jensen, K., Shiels, C. & Freemont, P.S. PML protein isoforms and the
728 RBCC/TRIM motif. *Oncogene* **20**, 7223-7233 (2001).
729 20. Zhang, X.W. *et al.* Arsenic trioxide controls the fate of the PML-RARalpha
730 oncoprotein by directly binding PML. *Science* **328**, 240-243 (2010).
731 21. Shen, T.H., Lin, H.K., Scaglioni, P.P., Yung, T.M. & Pandolfi, P.P. The
732 Mechanisms of PML-Nuclear Body Formation. *Mol Cell* **24**, 805 (2006).
733 22. Tatham, M.H. *et al.* RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required
734 for arsenic-induced PML degradation. *Nat Cell Biol* **10**, 538-546 (2008).
735 23. Kim, D.I. & Roux, K.J. Filling the Void: Proximity-Based Labeling of Proteins
736 in Living Cells. *Trends Cell Biol* **26**, 804-817 (2016).

737 24. Choi-Rhee, E., Schulman, H. & Cronan, J.E. Promiscuous protein biotinylation
738 by Escherichia coli biotin protein ligase. *Protein Sci* **13**, 3043-3050 (2004).
739 25. Kim, D.I. *et al.* Probing nuclear pore complex architecture with proximity-
740 dependent biotinylation. *Proc Natl Acad Sci U S A* **111**, E2453-2461 (2014).
741 26. Roux, K.J., Kim, D.I., Raida, M. & Burke, B. A promiscuous biotin ligase
742 fusion protein identifies proximal and interacting proteins in mammalian cells. *J*
743 *Cell Biol* **196**, 801-810 (2012).
744 27. Coyaud, E. *et al.* BioID-based Identification of Skp Cullin F-box (SCF)beta-
745 TrCP1/2 E3 Ligase Substrates. *Mol Cell Proteomics* **14**, 1781-1795 (2015).
746 28. Dho, S.E. *et al.* Proximity interactions of the ubiquitin ligase Mind bomb 1
747 reveal a role in regulation of epithelial polarity complex proteins. *Sci Rep* **9**,
748 12471 (2019).
749 29. Branon, T.C. *et al.* Efficient proximity labeling in living cells and organisms
750 with TurboID. *Nat Biotechnol* **36**, 880-887 (2018).
751 30. Schopp, I.M. *et al.* Split-BioID a conditional proteomics approach to monitor the
752 composition of spatiotemporally defined protein complexes. *Nat Commun* **8**,
753 15690 (2017).
754 31. Cho, K.F. *et al.* Split-TurboID enables contact-dependent proximity labeling in
755 cells. *Proc Natl Acad Sci U S A* **117**, 12143-12154 (2020).
756 32. Banaszynski, L.A., Liu, C.W. & Wandless, T.J. Characterization of the
757 FKBP.rapamycin.FRB ternary complex. *J Am Chem Soc* **127**, 4715-4721
758 (2005).
759 33. Bekes, M. *et al.* The dynamics and mechanism of SUMO chain deconjugation
760 by SUMO-specific proteases. *J Biol Chem* **286**, 10238-10247 (2011).

761 34. Shen, T.H., Lin, H.K., Scaglioni, P.P., Yung, T.M. & Pandolfi, P.P. The
762 mechanisms of PML-nuclear body formation. *Mol Cell* **24**, 331-339 (2006).
763 35. Rabellino, A. *et al.* The SUMO E3-ligase PIAS1 regulates the tumor suppressor
764 PML and its oncogenic counterpart PML-RARA. *Cancer Res* **72**, 2275-2284
765 (2012).
766 36. Zhu, J. *et al.* Small ubiquitin-related modifier (SUMO) binding determines
767 substrate recognition and paralog-selective SUMO modification. *J Biol Chem*
768 **283**, 29405-29415 (2008).
769 37. Li, J. *et al.* GTF2I-RARA is a novel fusion transcript in a t(7;17) variant of acute
770 promyelocytic leukaemia with clinical resistance to retinoic acid. *Br J Haematol*
771 **168**, 904-908 (2015).
772 38. Yin, C.C. *et al.* Identification of a novel fusion gene, IRF2BP2-RARA, in acute
773 promyelocytic leukemia. *J Natl Compr Canc Netw* **13**, 19-22 (2015).
774 39. Banani, S.F. *et al.* Compositional Control of Phase-Separated Cellular Bodies.
775 *Cell* **166**, 651-663 (2016).
776 40. Chung, I., Osterwald, S., Deeg, K.I. & Rippe, K. PML body meets telomere: the
777 beginning of an ALTernate ending? *Nucleus* **3**, 263-275 (2012).
778 41. Van Damme, E., Laukens, K., Dang, T.H. & Van Ostade, X. A manually curated
779 network of the PML nuclear body interactome reveals an important role for
780 PML-NBs in SUMOylation dynamics. *Int J Biol Sci* **6**, 51-67 (2010).
781 42. Netzer, C., Bohlander, S.K., Rieger, L., Muller, S. & Kohlhase, J. Interaction of
782 the developmental regulator SALL1 with UBE2I and SUMO-1. *Biochem*
783 *Biophys Res Commun* **296**, 870-876 (2002).
784 43. Pirone, L. *et al.* A comprehensive platform for the analysis of ubiquitin-like
785 protein modifications using in vivo biotinylation. *Sci Rep* **7**, 40756 (2017).

786 44. Lauberth, S.M. & Rauchman, M. A conserved 12-amino Acid motif in Sall1
787 recruits the nucleosome remodeling and deacetylase corepressor complex. *The*
788 *Journal of biological chemistry* **281**, 23922-23931 (2006).
789 45. Hendriks, I.A. & Vertegaal, A.C. A high-yield double-purification proteomics
790 strategy for the identification of SUMO sites. *Nat Protoc* **11**, 1630-1649 (2016).
791 46. Becker, J. *et al.* Detecting endogenous SUMO targets in mammalian cells and
792 tissues. *Nat Struct Mol Biol* **20**, 525-531 (2013).
793 47. Le Boulch, M., Brossard, A., Le Dez, G., Leon, S. & Rabut, G. Sensitive
794 detection of protein ubiquitylation using a protein fragment complementation
795 assay. *J Cell Sci* **133** (2020).
796 48. Sung, M.K. *et al.* Genome-wide bimolecular fluorescence complementation
797 analysis of SUMO interactome in yeast. *Genome Res* **23**, 736-746 (2013).
798 49. Fang, D. & Kerppola, T.K. Ubiquitin-mediated fluorescence complementation
799 reveals that Jun ubiquitinated by Itch/AIP4 is localized to lysosomes. *Proc Natl*
800 *Acad Sci U S A* **101**, 14782-14787 (2004).
801 50. Komiya, M. *et al.* A genetic screen to discover SUMOylated proteins in living
802 mammalian cells. *Sci Rep* **7**, 17443 (2017).
803 51. Croucher, D.R. *et al.* Bimolecular complementation affinity purification
804 (BiCAP) reveals dimer-specific protein interactions for ERBB2 dimers. *Sci*
805 *Signal* **9**, ra69 (2016).
806 52. Lin, D.Y. *et al.* Role of SUMO-interacting motif in Daxx SUMO modification,
807 subnuclear localization, and repression of sumoylated transcription factors. *Mol*
808 *Cell* **24**, 341-354 (2006).
809 53. Sahin, U. *et al.* Oxidative stress-induced assembly of PML nuclear bodies
810 controls sumoylation of partner proteins. *J Cell Biol* **204**, 931-945 (2014).

54. Saitoh, N. *et al.* In situ SUMOylation analysis reveals a modulatory role of RanBP2 in the nuclear rim and PML bodies. *Exp Cell Res* **312**, 1418-1430 (2006).

55. Lallemand-Breitenbach, V. *et al.* Arsenic degrades PML or PML-RARalpha through a SUMO-triggered RNF4/ubiquitin-mediated pathway. *Nat Cell Biol* **10**, 547-555 (2008).

56. Gechijian, L.N. *et al.* Functional TRIM24 degrader via conjugation of ineffectual bromodomain and VHL ligands. *Nat Chem Biol* **14**, 405-412 (2018).

57. Li, C., Xin, H., Shi, Y. & Mu, J. Knockdown of TRIM24 suppresses growth and induces apoptosis in acute myeloid leukemia through downregulation of Wnt/GSK-3beta/beta-catenin signaling. *Hum Exp Toxicol* **39**, 1725-1736 (2020).

58. Lallemand-Breitenbach, V. *et al.* Role of promyelocytic leukemia (PML) sumolation in nuclear body formation, 11S proteasome recruitment, and As2O3-induced PML or PML/retinoic acid receptor alpha degradation. *J Exp Med* **193**, 1361-1371 (2001).

59. Sha, Z., Blyszcz, T., Gonzalez-Prieto, R., Vertegaal, A.C.O. & Goldberg, A.L. Inhibiting ubiquitination causes an accumulation of SUMOylated newly synthesized nuclear proteins at PML bodies. *J Biol Chem* **294**, 15218-15234 (2019).

60. Shiio, Y. & Eisenman, R.N. Histone sumoylation is associated with transcriptional repression. *Proc Natl Acad Sci U S A* **100**, 13225-13230 (2003).

61. Jangal, M. *et al.* The transcriptional co-repressor TLE3 suppresses basal signaling on a subset of estrogen receptor alpha target genes. *Nucleic Acids Res* **42**, 11339-11348 (2014).

62. Flack, J.E., Mieszczanek, J., Novcic, N. & Bienz, M. Wnt-Dependent Inactivation of the Groucho/TLE Co-repressor by the HECT E3 Ubiquitin Ligase Hyd/UBR5. *Mol Cell* **67**, 181-193 e185 (2017).

63. Yan, W. *et al.* Epigenetic regulation of DACH1, a novel Wnt signaling component in colorectal cancer. *Epigenetics* **8**, 1373-1383 (2013).

64. Sato, A. *et al.* Sall1, a causative gene for Townes-Brocks syndrome, enhances the canonical Wnt signaling by localizing to heterochromatin. *Biochem Biophys Res Commun* **319**, 103-113 (2004).

65. Doench, J.G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat Biotechnol* **34**, 184-191 (2016).

66. Cox, J. & Mann, M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* **26**, 1367-1372 (2008).

67. Tyanova, S. *et al.* The Perseus computational platform for comprehensive analysis of (prote)omics data. *Nat Methods* **13**, 731-740 (2016).

68. Snel, B., Lehmann, G., Bork, P. & Huynen, M.A. STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene. *Nucleic Acids Res* **28**, 3442-3444 (2000).

69. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504 (2003).

70. Bader, G.D. & Hogue, C.W. An automated method for finding molecular complexes in large protein interaction networks. *BMC Bioinformatics* **4**, 2 (2003).

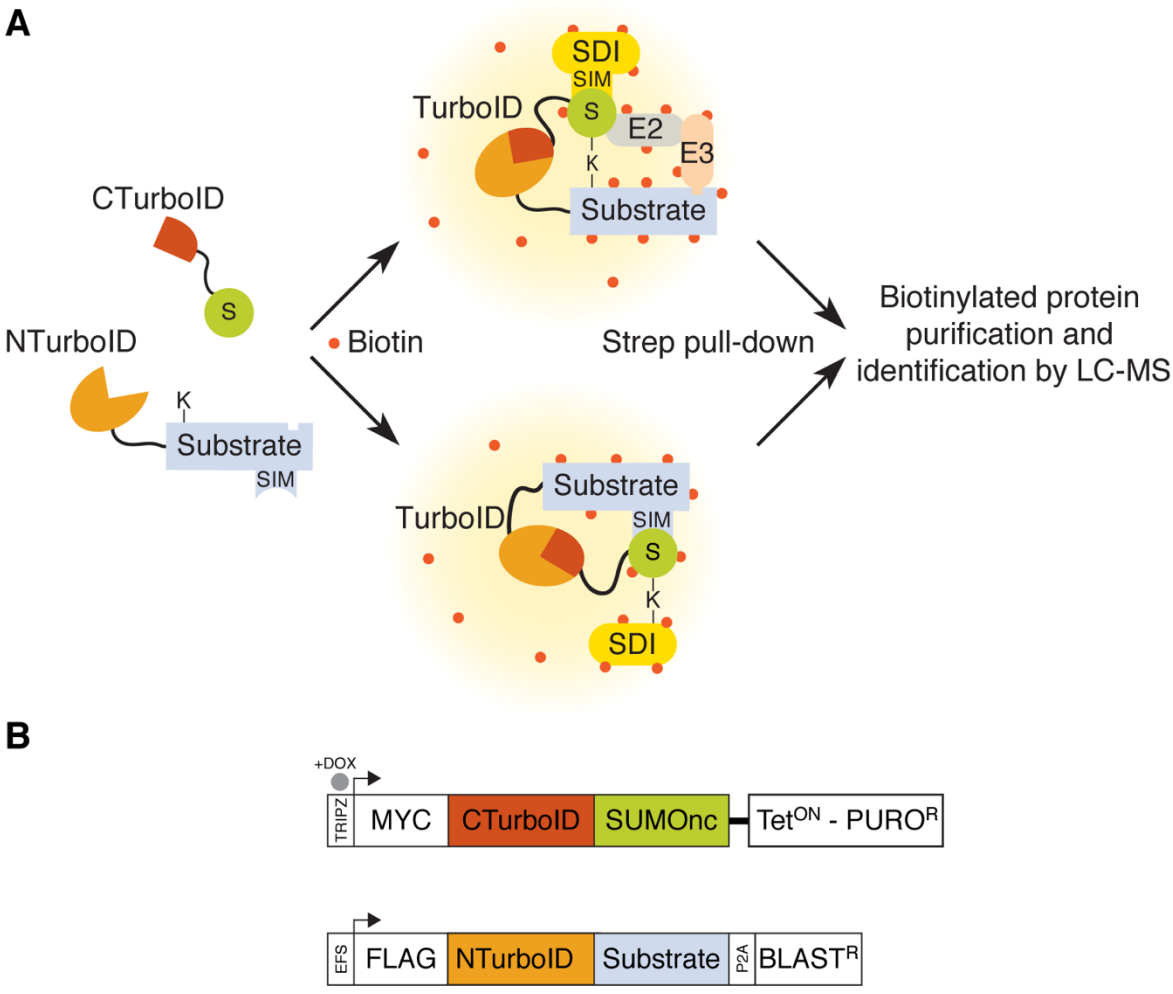
71. Reimand, J. *et al.* g:Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Res* **44**, W83-89 (2016).

72. Vizcaino, J.A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res* **44**, 11033 (2016).

73. Sun, H. & Hunter, T. Poly-small ubiquitin-like modifier (PolySUMO)-binding proteins identified through a string search. *J Biol Chem* **287**, 42071-42083 (2012).

74. Weaver, L.H., Kwon, K., Beckett, D. & Matthews, B.W. Corepressor-induced organization and assembly of the biotin repressor: a model for allosteric activation of a transcriptional regulator. *Proc Natl Acad Sci U S A* **98**, 6045-6050 (2001).

75. Sehnal, D., Rose, A.S., Koča, J., Burley, S.K. & Velankar, S. in Proceedings of the Workshop on Molecular Graphics and Visual Analysis of Molecular Data 29–33 (Eurographics Association, Brno, Czech Republic; 2018).



874

875 **<u>Figure 1:</u> Identification of SUMO-dependent interactions: the SUMO-ID strategy.**

876 Schematic representation of the SUMO-ID strategy **(A)** and the constructs used **(B)**.

877 $BLAST^R$, blasticidin resistant cassette; CTurboID, C-terminal TurboID; DOX,

878 doxycycline; NTurboID, N-terminal TurboID; $PURO^R$, puromycin resistant cassette; S,

879 SUMO; SIM: SUMO Interacting Motif; SDI: SUMO-dependent interactor; SUMOnc,

880 SUMO non-cleavable; Strep, streptavidin.

881

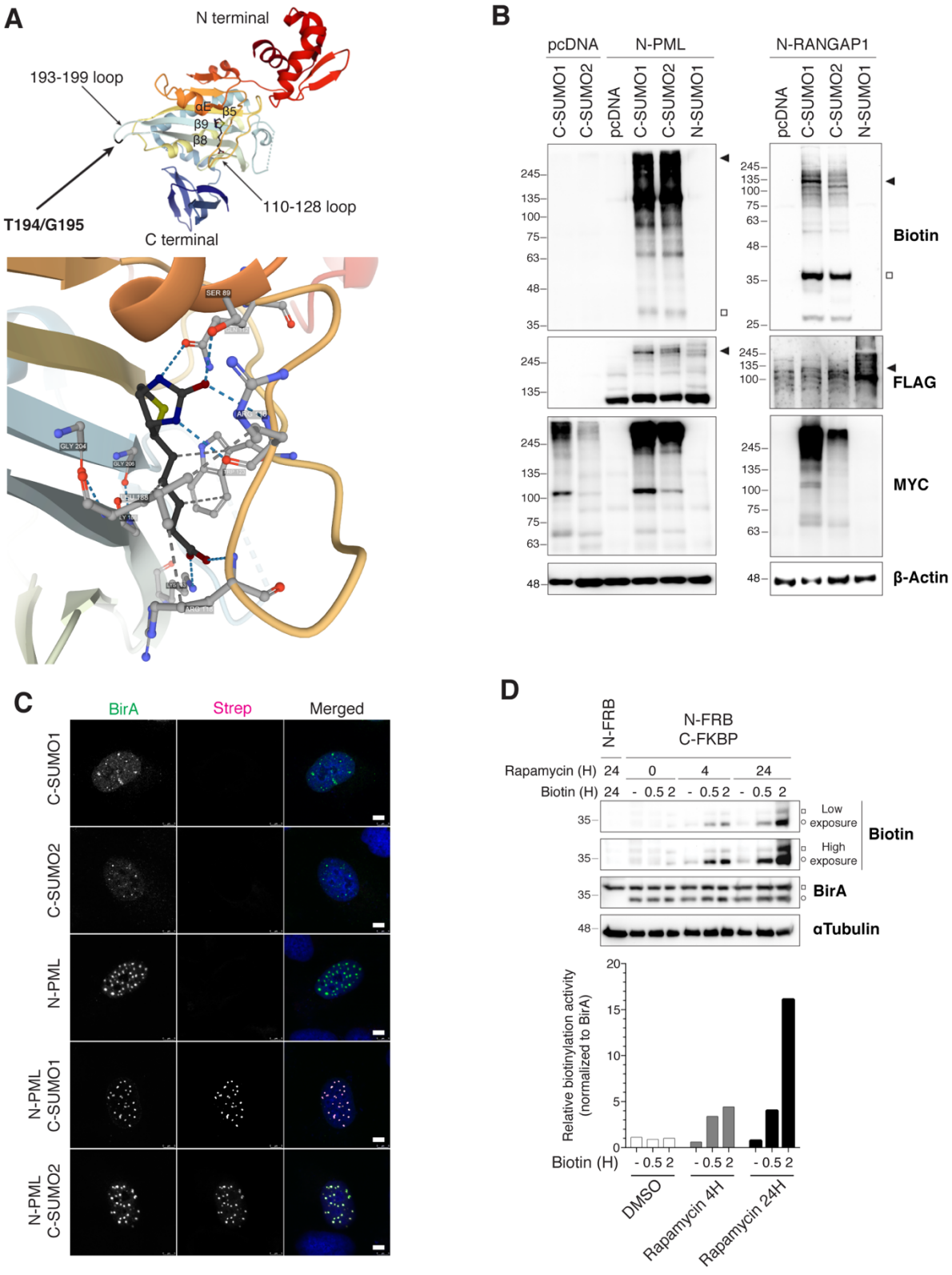


Figure 2: T194/G195 Split-TurboID is suitable for SUMO-ID studies. (A) Structure of the *E. coli* BirA with bound biotin (PDB ID: 1HXD[74,75]), depicting the T194/G195 split point and the BirA-Biotin interaction. T194/G195 split point breaks the 193-199 loop

886 that connects the biotin-interacting β8 and β9 strands (see File S1). **(B)** Western blot of
887 HEK293FT cells that were transiently transfected with combinations of the FLAG-
888 NTurboID[194] (N) or MYC-CTurboID[195] (C) fused to PMLIVa, RANGAP1 or SUMO1/2
889 and treated with 50 μM of biotin for 16 hours. Black arrowheads indicate SUMO-ID
890 activity derived from MYC-CTurboID[195]-SUMOylated FLAG-NTurboID[194]-substrates.
891 White squares indicate biotinylated free MYC-CTurboID[195]-SUMOs. Neither FLAG-
892 NTurboID[194] nor MYC-CTurboID[195] showed any detectable background biotinylating
893 activity. **(C)** Immunostainings of transiently transfected U2OS cells treated with 50 μM
894 of biotin for 16 hours, showing the fragment-complementation dependency of SUMO-ID
895 and its correct localization within the cell, enriched at PML NBs as expected for
896 SUMOylated PML. Nuclei are stained with DAPI (blue) and biotinylated material with
897 fluorescent streptavidin (Strep, magenta). BirA antibody recognizes both NTurboID[194]
898 and CTurboID[195] (green). Black and white panels show the single green and magenta
899 channels. Scale bar: 5 μm. **(D)** Western blot of HEK293FT stable cells for NTurboID[194]-
900 FRB alone or in combination with CTurboID[195]-FKBP, treated or not with 1 μg/mL of
901 rapamycin and 50 μM of biotin at indicated time-points. BirA antibody recognizes both
902 NTurboID[194] and CTurboID[195]. White squares and circles indicate NTurboID[194]-FRB and
903 CTurboID[195]-FKBP, respectively. Self-biotinylating activity of the reconstituted TurboID
904 was measured and normalized to expression levels (BirA blot). Molecular weight markers
905 are shown to the left of the blots in kDa.

906

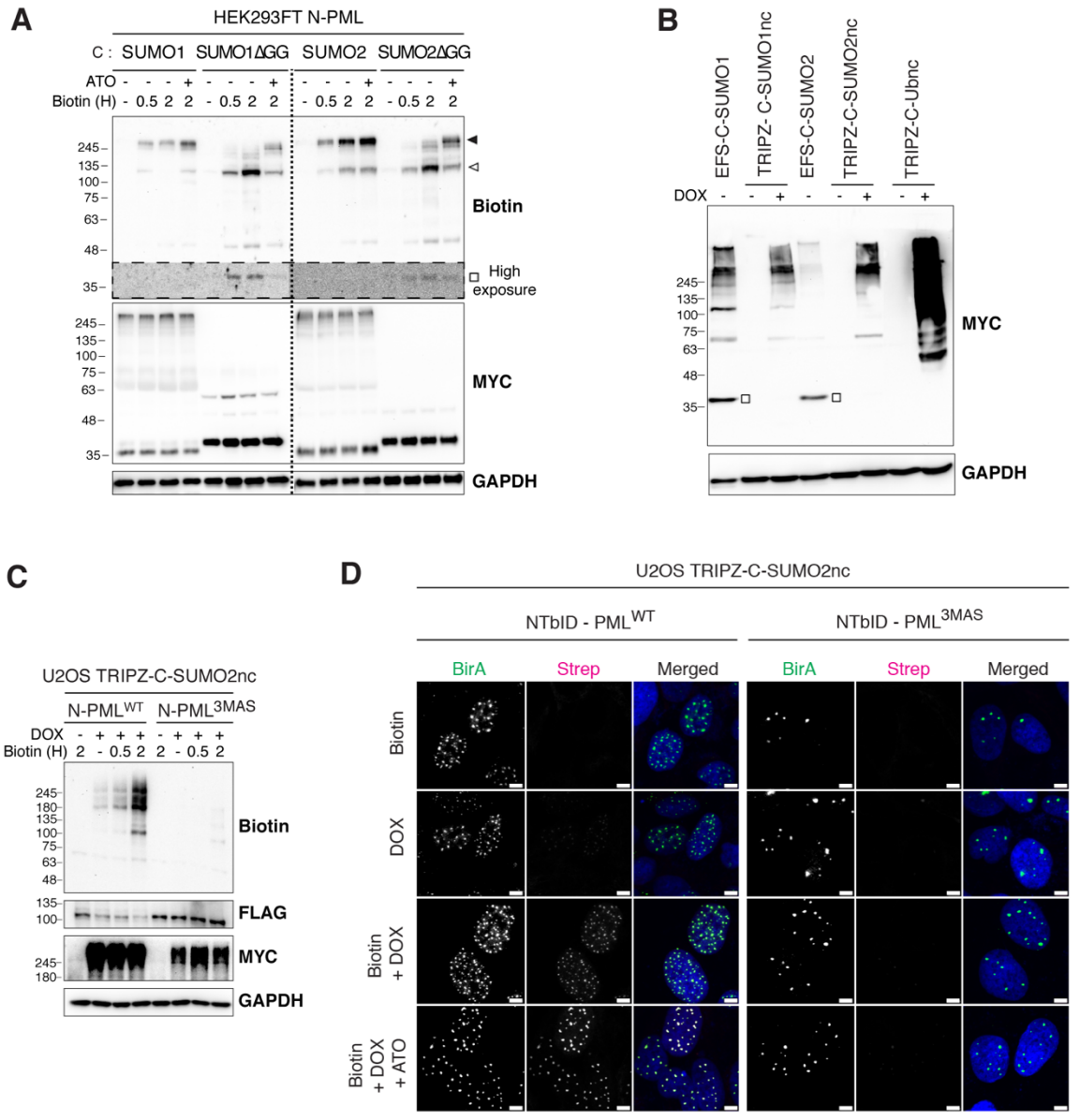
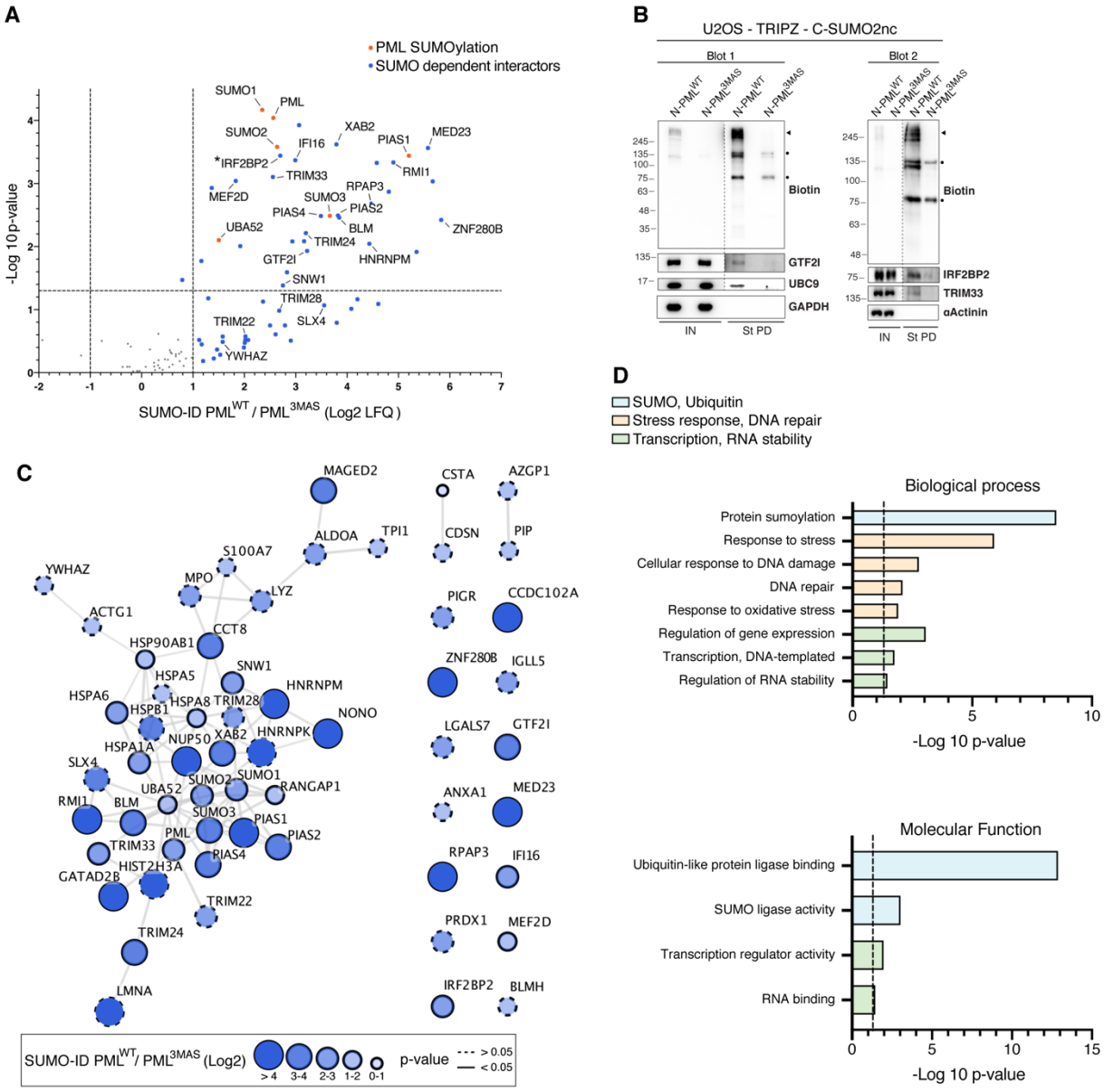


Figure 3: SUMO-ID is specific for SUMO-dependent interactions. (A) Western blot of HEK293FT FLAG-NTurboID[194]-PMLIVa stable cell line transfected with different combinations of MYC-CTurboID[195] (C) and SUMO^{WT} or SUMO$^{\Delta GG}$. Cells were treated or not with 1 µM of ATO for 2 hours and 50 µM of biotin at indicated time points. White square indicates biotinylation of unconjugated MYC-CTurboID[195]- SUMO$^{\Delta GG}$. White arrowhead points to SUMO-SIM interaction mediated SUMO-ID. Black arrowhead shows PML-SUMOylation derived SUMO-ID. **(B)** Western blot of HEK293FT transfected with constitutive MYC-CTurboID[195]-SUMO1/2 or doxycycline-inducible

and isopeptidase-cleavage resistant (nc) MYC-CTurboID^{195}-SUMO1/2nc or MYC-CTurboID^{195}-Ubnc. Doxycycline was added or not at 1 µg/mL for 24 hours. White squares point to free/unconjugated MYC-CTurboID^{195}-SUMOs. **(C)** Western blot of U2OS double stable cell lines for FLAG-NTurboID^{194}-PMLIVa^{WT} or the SUMO/SIM mutant FLAG-NTurboID^{194}-PMLIVa^{3MAS} together with doxycycline-inducible TRIPZ-MYC-CTurboID^{195}-SUMO2nc. Doxycycline was added or not at 1 µg/mL for 24 hours. 50 µM of biotin was added at indicated time-points. PML SUMO-ID showed a high PML/SUMO interaction dependency. **(D)** Confocal microscopy of the same cells as in (C), treated or not with doxycycline (1 µg/mL, 24 hours), biotin (50 µM, 2 hours) and ATO (1 µM, 2 hours). Nuclei are stained with DAPI (blue) and biotinylated material with fluorescent streptavidin (Strep, magenta). BirA antibody shows NTurboID^{194}-PML staining (green). Black and white panels show the single green and magenta channels. Colocalization of the streptavidin and NTurboID^{194}-PMLIVa^{WT} signal is observed within PML NBs, that depends on PML-SUMO interaction. Scale bar: 5 µm. Molecular weight markers are shown to the left of the blots in kDa in (A-C).



<u>Figure 4:</u> SUMO-ID identifies SUMO-dependent interactors of PML. (A) Volcano plot of LC-MS analysis comparing streptavidin pull-downs of U2OS double stable cell lines for TRIPZ-MYC-CTurboID[195]-SUMO2nc together with FLAG-NTurboID[194]-PMLIVa^{WT} or FLAG-NTurboID[194]-PMLIVa^{3MAS}. Cells were treated with 1 µg/mL of doxycycline for 24 hours and 50 µM of biotin for 2 hours. 59 high-confidence SUMO-dependent PML interactors were defined. Asterisk (*) indicates that IRF2BP2 was detected with one peptide but further validated by Western blot and immunofluorescence. **(B)** Western blot validations of PML SUMO-dependent interactors identified by SUMO-

ID in (A). Blots 1/2 represent 2 independent experiments. UBC9 was added as an expected positive control. Dots indicate endogenous carboxylases. Black arrowheads point to specific PML SUMO-ID biotinylating activity. IN: input; St PD: streptavidin pull-down. Molecular weight markers are shown to the left of the blots in kDa. **(C)** STRING network analysis of the 59 SUMO-dependent interactors of PML identified in (A). A highly interconnected cluster related to protein SUMOylation/ubiquitylation, transcriptional regulation, DNA repair and RNA stability proteins is depicted. Color, transparency and size of the nodes were discretely mapped to the Log2 enrichment value as described. The border line type was discretely mapped to the p-value as described. **(D)** Gene ontology analysis of the 59 SUMO-dependent interactors of PML identified in (A). Biological processes and molecular functions related to SUMOylation/ubiquitylation, stress response, DNA repair, transcription and RNA stability were significantly enriched. Dotted line represents the threshold of the p-value (0.05).

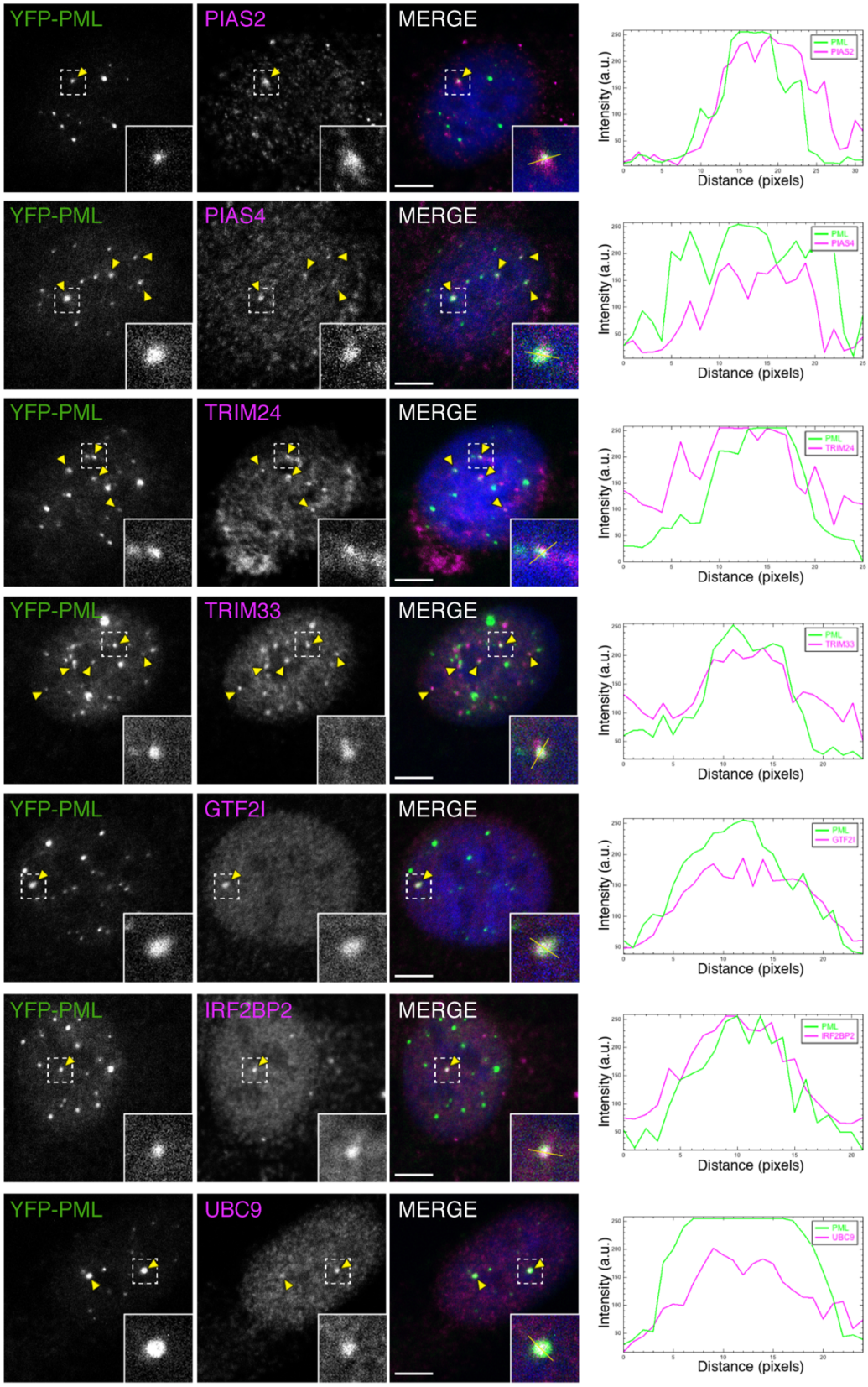


Figure 5: SUMO-dependent interactors of PML localize to PML NBs. Confocal microscopy analysis of PML SUMO-ID identified proteins in U2OS YFP-PML *knock-in*

cell line. UBC9 was added as an expected positive control. Yellow arrowheads indicate colocalization events. Dotted line-squares show the selected colocalization events for digital zooming and the signal profile plotting shown to the right. Nuclei are stained with DAPI (blue), YFP-PML is shown in green and the indicated proteins in magenta. Black and white panels show the single green and magenta channels. Scale-bar: 5 μm.

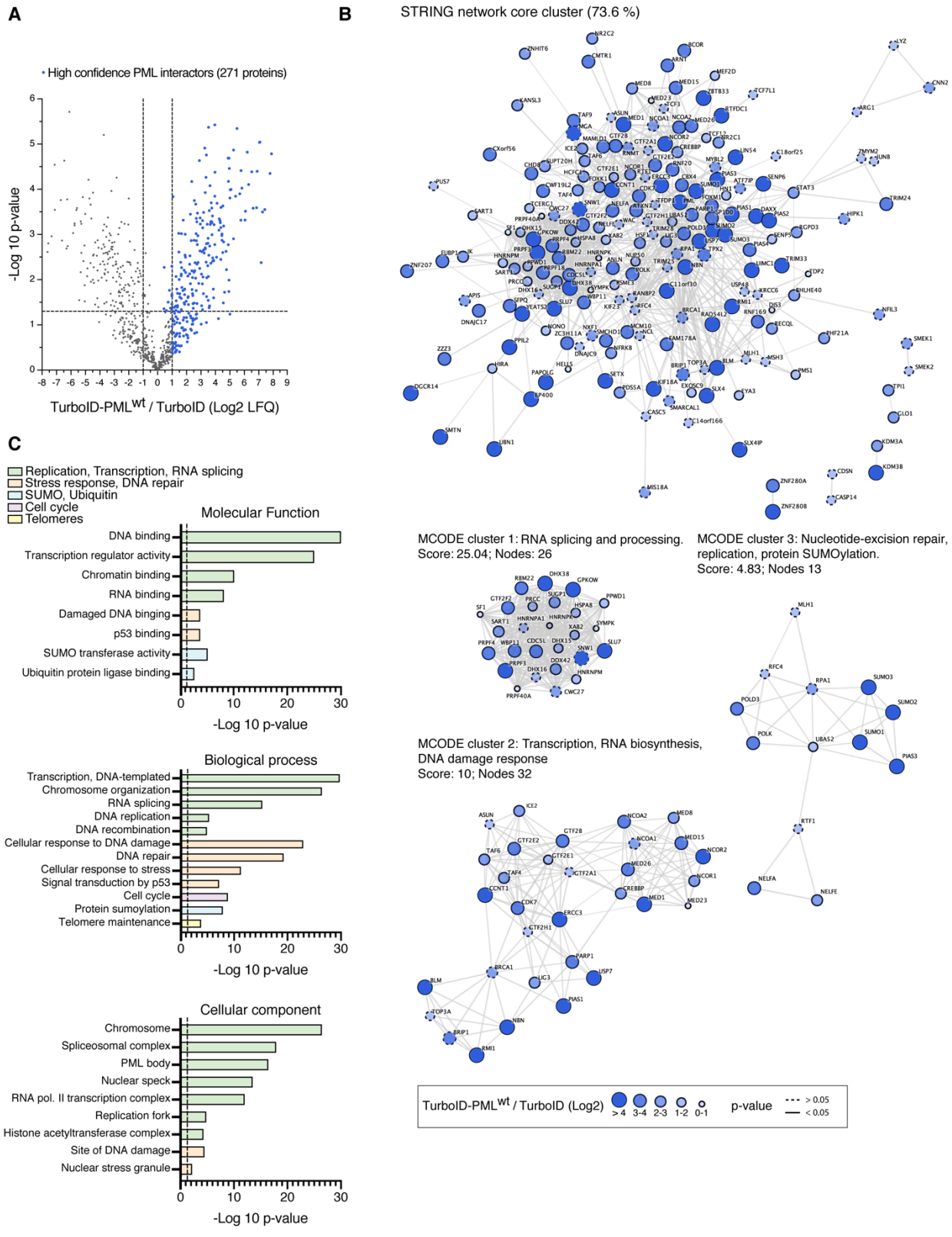
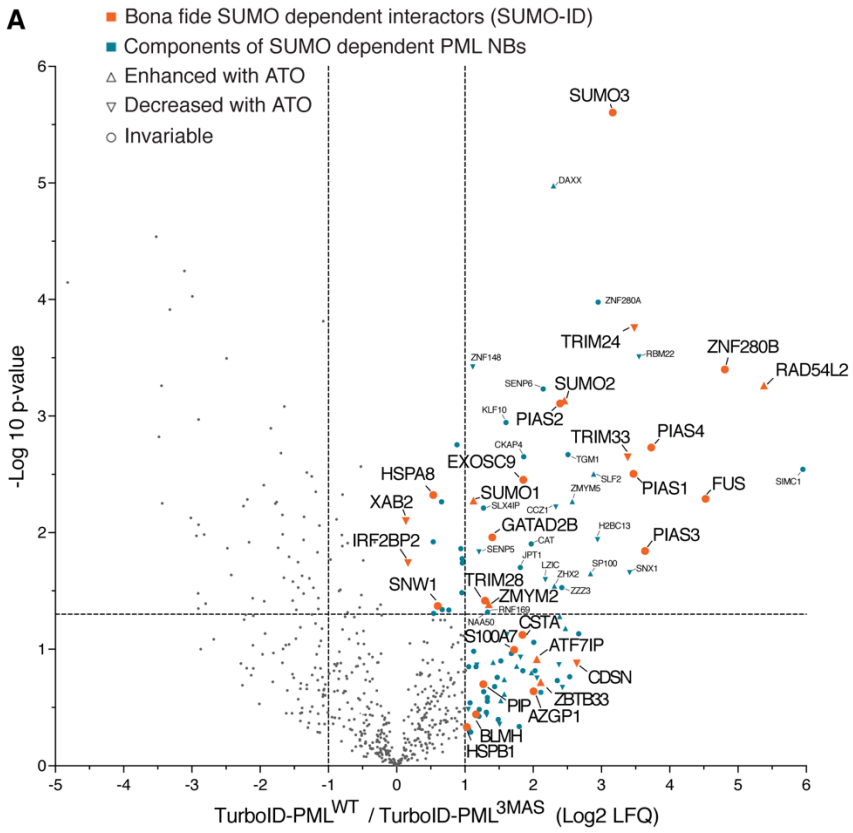
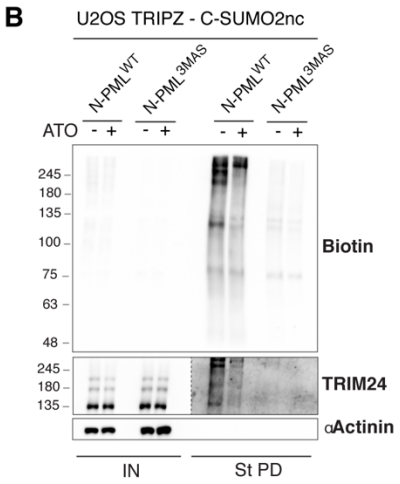


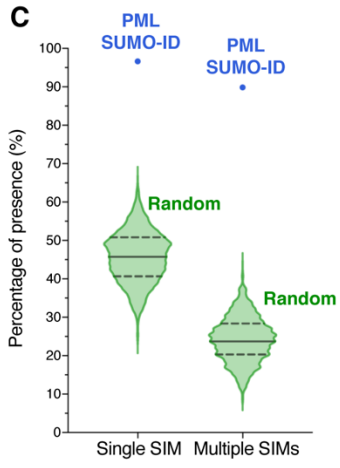
Figure 6: Characterization of the whole PML NBs proteome. (A) Volcano plot of LC-MS analysis comparing streptavidin pull-downs of U2OS stable cell lines for TurboID-PMLIVa^{WT} or TurboID alone. Cells were treated with 50 µM of biotin for 2 hours. High-

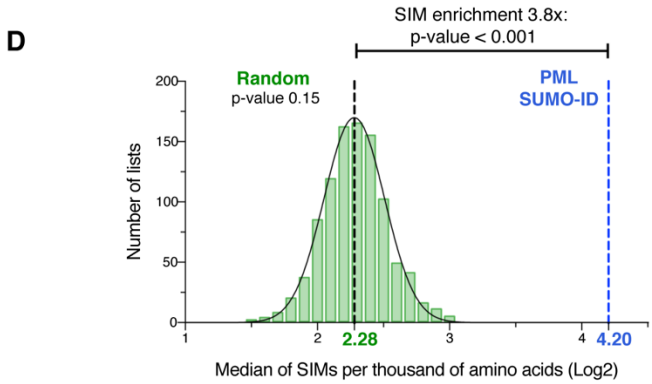
968 confidence PML proteome composed of 267 proteins is shown as blue dots. **(B)** STRING
969 network analysis of the whole PML NBs proteome defined in (A) shows a high
970 interconnected network composed of the 73.6% of the proteins. Highly interconnected
971 sub-clusters were characterized using MCODE. Color, transparency and size of the nodes
972 were discretely mapped to the Log2 enrichment value as described. The border line type
973 was discretely mapped to the p-value as described. **(C)** Gene ontology analysis of the
974 whole PML NBs proteome defined in (A). Depicted biological processes, molecular
975 functions and cellular components were significantly enriched. Dotted line represents the
976 threshold of the p-value (0.05).

977









978

Figure 7: Proteins identified by PML SUMO-ID are a subset of the SUMO-dependent PML NBs proteome and are enriched in SIMs. (A) Volcano plot of LC-MS analysis comparing streptavidin pull-downs of U2OS stable cell lines for TurboID-PMLIVa^{WT} or TurboID-PMLIVa^{3MAS}. Cells were treated with 50 μM of biotin for 2 hours. Proteins enriched in TurboID alone compared to TurboID-PMLIVa^{WT} were previously removed for the comparison. PML SUMO-ID identified proteins (including 1 peptide identified proteins) are highlighted in orange. LC-MS data on the effect of the ATO treatment (1 μM; 2 hours) for TurboID-PMLIVa^{WT} enriched proteins is represented with symbols as described. **(B)** WB validation of the effect of ATO treatment (1 μM; 2 hours) on TRIM24 by PML SUMO-ID. Cells were treated with 1 μg/mL of doxycycline for 24 hours and 50 μM of biotin for 2 hours. After streptavidin pulldown, decreased levels of SUMO-PML interacting TRIM24 upon ATO treatment is observed. **(C)** Violin plots comparing the percentage of single SIM and multiple SIM presence in 1000 random lists and PML SUMO-ID list. The 1000 random lists contain the same number of proteins (59) as the SUMO-ID list. **(D)** SIM presence was normalized by the length of the proteins to obtain the value of SIMs per thousand of amino acids (STAA). Gaussian distribution of STAA median values of the random lists was validated (d'Agostino and Pearson normality test, p-value = 0.15), and PML SUMO-ID SIM enrichment factor was calculated. The dotted black line represents the median STAA value of random lists. The dotted blue line represents the STAA value of the PML SUMO-ID list.

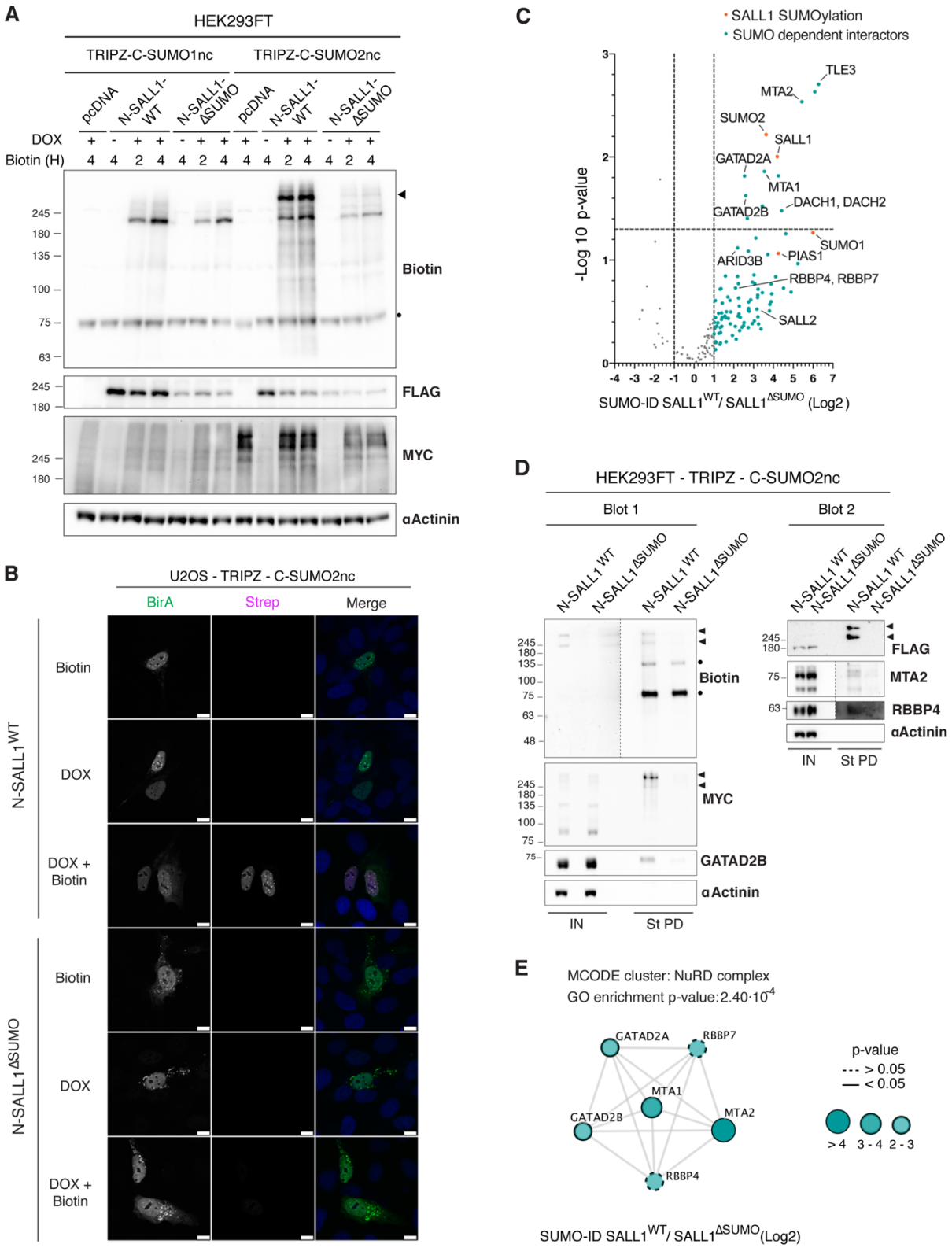


Figure 8: SUMO-ID identifies interactors of SUMOylated SALL1. (A) WB of HEK293FT stable cell lines for TRIPZ-MYC-CTurboID[195]-SUMO1nc/SUMO2nc transfected with FLAG-NTurboID[194]-SALL1[WT] or the SUMO site mutant FLAG-

NTurboID^{194}-SALL1$^{\Delta SUMO}$. Cells were treated or not with 1 µg/mL of doxycycline for 24 hours and 50 µM of biotin at indicated time points. Efficient SALL1 SUMO-ID biotinylating activity was detected for SUMO2nc (black arrowhead). Dot indicates endogenous carboxylase. **(B)** Confocal microscopy of U2OS stable cell line for TRIPZ-MYC-CTurboID^{195}-SUMO2nc transfected with FLAG-NTurboID^{194}-SALL1^{WT} or the SUMO site mutant FLAG-NTurboID^{194}-SALL1$^{\Delta SUMO}$. Cells were treated or not with 1 µg/mL of doxycycline for 24 hours and 50 µM of biotin for 4 hours. Nuclei are stained with DAPI (blue) and biotinylated material with fluorescent streptavidin (Strep, magenta). BirA antibody shows NTurboID^{194}-SALL1 staining (green). Black and white panels show the single green and magenta channels. Nuclear colocalization of FLAG-NTurboID^{194}-SALL1^{WT} and streptavidin signal was observed. **(C)** Volcano plot of LC-MS analysis comparing streptavidin pull-downs of HEK293FT TRIPZ-MYC-CTurboID^{195}-SUMO2nc stable cell line transfected with FLAG-NTurboID^{194}-SALL1^{WT} or the SUMO site mutant FLAG-NTurboID^{194}-SALL1$^{\Delta SUMO}$. Cells were treated with 1 µg/mL of doxycycline for 24 hours and 50 µM of biotin for 4 hours. Potential interactors of SUMOylated SALL1 are depicted. **(D)** Western blot validations of SUMOylated SALL1 interactors found in (C). NuRD complex proteins GATAD2B, MTA2 and RBBP4 were confirmed. Black arrowheads point to SUMOylated SALL1 signal. Dots indicate endogenous carboxylases. **(E)** STRING network analysis of the SALL1 SUMO-ID list and MCODE clustering identifies the NuRD complex as a highly interconnected subcluster. Gene ontology analysis also identified the NuRD complex as an enriched term. Color, transparency and size of the nodes were discretely mapped to the Log2 enrichment value as described. The border line type was discretely mapped to the p-value as described. Molecular weight markers are shown to the left of the blots in kDa in (A, D).