Identification of SUMO targets required to maintain human stem cells in the pluripotent state.

Barbara Mojsa[1], Michael H. Tatham[1], Lindsay Davidson[2], Magda Liczmanska[1], Emma Branigan[1], and Ronald T. Hay[1*]

[1]Division of Gene Regulation and Expression, [2]Division of Cell and Developmental Biology, School of Life Sciences, University of Dundee, Dundee, UK

[*]Corresponding author r.t.hay@dundee.ac.uk

Running title: SUMO targets in human induced pluripotent stem cells

18 **Abbreviations**

19

20 bFGF - basic fibroblast growth factor

21 ESCs - embryonic stem cells

22 FACS - fluorescence activated cell sorting

23 GGK – GlyGly-Lys branched peptide

24 hESCs – human embryonic stem cells

25 hiPSCs – human induced pluripotent stem cells

26 LIF - leukemia inhibitory factor

27 mESCs – mouse embryonic stem cells

28

29

Abstract

To determine the role of SUMO modification in the maintenance of pluripotent stem cells we used ML792, a potent and selective inhibitor of SUMO Activating Enzyme. Treatment of human induced pluripotent stem cells with ML792 initiated changes associated with loss of pluripotency such as reduced expression of key pluripotency markers. To identify putative effector proteins and establish sites of SUMO modification, cells were engineered to stably express either SUMO1 or SUMO2 with TGG to KGG mutations that facilitate GlyGly-K peptide immunoprecipitation and identification. A total of 976 SUMO sites were identified in 427 proteins. STRING enrichment created 3 networks of proteins with functions in regulation of gene expression, ribosome biogenesis and RNA splicing, although the latter two categories represented only 5% of the total GGK peptide intensity. The remainder have roles in transcription and chromatin structure regulation. Many of the most heavily SUMOylated proteins form a network of zinc-finger transcription factors centred on TRIM28 and associated with silencing of retroviral elements. At the level of whole proteins there was only limited evidence for SUMO paralogue-specific modification, although at the site level there appears to be a preference for SUMO2 modification over SUMO1 in acidic domains. We show that SUMO is involved in the maintenance of the pluripotent state in hiPSCs and identify many chromatin-associated proteins as bona fide SUMO substrates in human induced pluripotent stem cells.

Introduction

Pluripotent cells display the property of self-renewal and have the capacity to generate all of the different cells required for the development of the adult organism. The pluripotent state is defined by a specific gene expression programme driven by expression

54 of the core transcription factors OCT4, SOX2 and NANOG that sustain their own expression
55 by virtue of a positive, linked autoregulatory loop, while activating genes required to
56 maintain the pluripotent state and repressing expression of the transcription factors for
57 lineage specific differentiation (1). Once terminally differentiated, somatic cell states are
58 remarkably stable, however, forced expression of key pluripotency transcription factors that
59 are highly expressed in embryonic stem cells (ESCs), including OCT4, SOX2 and NANOG,
60 leads to reprogramming back to the pluripotent state (2-4). Under normal circumstances
61 the efficiency of reprogramming is very low and it is clear that there are roadblocks to
62 reprogramming designed to safeguard cell fates (5,6). The Small Ubiquitin like Modifier
63 (SUMO) has emerged as one such roadblock and reduced SUMO expression decreases the
64 time taken and increases the efficiency of reprogramming in mouse cells (7-9). Three SUMO
65 paralogues, known as SUMO1, SUMO2 and SUMO3 are expressed in vertebrates. Based on
66 almost indistinguishable functional and structural features SUMO2 and SUMO3 are
67 collectively termed SUMO2/3 and share about 50% amino acid sequence identity with
68 SUMO1. SUMO proteomics studies have revealed very high numbers of cellular SUMO
69 substrates (see (10) for a review) and as a consequence SUMO can influence a wide range of
70 biological processes. However, the proportion of the total the cellular pool of a protein that
71 is SUMOylated varies greatly. There are rare examples of almost constitutively SUMOylated
72 proteins such as the Ran GTPase RanGAP1 (11), although the majority of substrates have
73 such low modification stoichiometry that identification from purified SUMO conjugates
74 using Immuno-Blotting is close to the detection limit (for review see (12)). Conjugation of
75 SUMO to protein substrates is mechanistically similar to that of ubiquitin conjugation but is
76 carried out by a completely separate enzymatic pathway. SUMOs are initially translated as
77 inactive precursors that require a precise proteolytic cleavage, carried out by a set of SUMO

78 specific proteases (SENPs), to expose the terminal carboxyl group of a Gly-Gly sequence that
79 ultimately forms an isopeptide bond with the ε–amino group of a lysine residue in the
80 modified protein. The hetrodimeric E1 SUMO Activating Enzyme (SAE1/SAE2) uses ATP to
81 adenylate the C-terminus of SUMO, before forming a thioester with a cysteine residue in a
82 second active site of the enzyme and releasing AMP. SUMO then undergoes a transthioation
83 reaction on to a cysteine residue in the only E2 SUMO conjugating enzyme Ubc9. Assisted by
84 a small group of E3 SUMO ligases, including the PIAS proteins, RanBP2 and ZNF451 the
85 SUMO is transferred directly from Ubc9 onto target proteins (13). Modification of target
86 proteins may be short-lived, with SUMO being removed by a group of SENPs. Together this
87 creates a highly dynamic SUMO cycle where the net SUMO modification status of proteins is
88 determined by the rates of SUMO conjugation and deconjugation (12). Preferred sites of
89 SUMO modification conform to the consensus ψKxE, where ψ represents a large
90 hydrophobic residue (14,15). A conjugation consensus is present in the N-terminal sequence
91 of SUMO2 and SUMO3 and thus permits self-modification and the formation of SUMO2/3
92 chains (16). As a strict consensus is absent from SUMO1, it does not form chains as readily
93 as SUMO2/3 (12). Once linked to target proteins SUMO allows the formation of new
94 protein-protein interactions as the modification can be recognised by proteins containing a
95 short stretch of hydrophobic amino acids termed a SUMO interaction motif (17).

96 Stem cell lines are an excellent model to study the mechanisms that control self-
97 renewal and pluripotency. Mouse ESCs have been widely used for these studies as the cells
98 can also be used in mice for *in vivo* studies. However, they display different characteristics
99 from human ESCs. Mouse ESCs require leukemia inhibitory factor (LIF) and bone
100 morphogenetic protein (BMP) signalling to maintain their self-renewal and pluripotency
101 (18,19). In contrast LIF does not support self-renewal and BMPs induce differentiation in

102 human ESCs (20-22). The maintenance of the pluripotent state of hESCs requires basic
103 fibroblast growth factor (bFGF, FGF2) and activin/nodal/TGF-b signalling along with
104 inhibition of BMP signalling (23,24). These differences may reflect the particular
105 developmental stages at which ESC lines are established *in vitro* from mouse and human
106 blastocysts, or may be due to differences in early embryonic development (25). As hESCs are
107 derived from embryos their use is limited, but human Induced Pluripotent Stem Cells (hiPSC)
108 can be derived by reprogramming normal somatic cells and display most of the
109 characteristics of hESCs (2) and are now widely used to study self-renewal and pluripotency
110 in humans.

111 To determine the role of SUMO modification in hiPSCs we made use of ML792, a
112 highly potent and selective inhibitor of the SUMO Activating Enzyme (26). Treatment of
113 hiPSCs with this inhibitor rapidly blocks SUMO modification allowing endogenous SUMO
114 proteases to strip SUMO from targets. When used over the course of 48 hours hiPSCs
115 treated with ML792 lose the majority of SUMO conjugation but show no large-scale changes
116 to the cellular proteome nor loss of viability, although markers of pluripotency are reduced.
117 Decreased expression of selected pluripotency markers seems to be a consequence of
118 SUMO removal from key targets rather than large scale changes to the proteome. SUMO
119 site proteomic analysis of hiPSCs reveals extensive SUMO modification of proteins involved
120 in transcriptional repression, RNA splicing and ribosome biogenesis. At the protein level
121 most proteins do not appear to display SUMO paralogue-specific modification, while at the
122 site level there is clear evidence of SUMO paralogue specificity. This site-specific selectivity
123 appears, at least in part, to be influenced by proximal amino acids, with generally acidic
124 domains being preferentially modified by SUMO2.
125

126 **Experimental procedures**

127 **Antibodies and inhibitors**. Rabbit antibodies TRIM28 (4124S, 4123S), CTCF (3418S), OCT4A
128 (2890S), SOX2 (23064S), NANOG (3580S), KLF4 (12173S) and mouse antibodies against TRA-
129 1-60 (4746T), TRA-1-81(4745T), SSEA-4 (4755T) and SMA (D4K9N) were from Cell Signalling
130 Technology. The anti-SALL4 (ab29112), anti-TRIM24 (ab70560), anti-NOP58 (ab155556),
131 anti-TRIM24 (ab70560), anti-NESTIN (ab196908) were from Abcam. Mouse antibody against
132 α-Tubulin was from Bethyl Laboratories, mouse anti-LaminA/C antibody was from Sigma
133 (SAB4200236), rabbit anti-mCherry (PA5-34974), rabbit anti-TRIM33 (PA5-82152) and
134 mouse anti-HIS (34650) were from Invitrogen and Qiagen respectively. Anti-CYTOKERATIN17
135 was a gift from R. Hickerson (University of Dundee). Sheep antibodies against SUMO1,
136 SUMO2 (27) and chicken antibodies against PML (28) were generated in-house. Secondary
137 antibodies conjugated with HRP and Alexa fluorophores were from Sigma and Invitrogen,
138 respectively. ML792 (1644342-14-2), MG132 (474787) and N-ethylmaleimide (E3876) were
139 from Sigma Aldrich. Protease Inhibitor cocktail (11836170001) was from Roche. Propidium
140 iodide and DAPI were from Life Technologies.

141

142 **Cloning**. SUMO1-KGG-mCherry and SUMO2-KGG-mCherry PiggyBac expression vectors were
143 generated by GATEWAY cloning. Briefly, SUMO1, SUMO2 and mCherry fragments were PCR
144 amplified using the following resources: 6His SUMO1 T95K (300nt) from pSCAI88 and 6His
145 SUMO2 T90K (300nt) from pSCAI89 with a common forwards primer (5'-
146 CACCatgcatcatcatcatcatcatgct-3') and set of specific mCherry fusing primers (5'-
147 TCACCATACCCCCCTTTTGTTCCTG-3' and 5'-TCACCATACCTCCCTTCTGCTGCT-3'); mCherry
148 from pRHAI4 CMV-OsTIR1-mCherry2-PURO (700 nt) with a set of common overlapping
149 oligos (5'-GGTATGGTGAGCAAGGGCG-3' and 5'-TTATTACTTGTACAGCTCGTCCATG-3').

150 Subsequently, PCR fragments were fused together using overlap extension PCR and TOPO
151 cloned into pENTR™/D-TOPO™ (Invitrogen) and verified by DNA sequencing. The assembled
152 SUMO1-KGG-mCherry and SUMO2-KGG-mCherry sequences were then sub-cloned from the
153 pENTR vector into the destination PiggyBac GATEWAY expression vector paPX1 (gift from A.
154 Dady UUniversity of Dundee)) using LR clonase (ThermoFisher Scientific).

155 **Human Induced pluripotent stem cells (hiPSCs) culture and transfection protocols.** Human
156 ESC lines (SA121 and SA181) were obtained from Cellartis / Takara Bio Europe. All work with
157 hESCs was approved by the UK Stem cell bank steering committee (Approval reference:
158 SCSC17-14). Human iPSC lines were obtained from Cellartis / Takara Bio Europe (ChiPS4) or
159 the HipSci consortium (bubh3, oaqd3, ueah1 and wibj2). Cell lines were maintained in TESR
160 medium (29) containing FGF2 (Peprotech, 30 ng/ml) and noggin (Peprotech, 10 ng/ml) on
161 growth factor reduced geltrex basement membrane extract (Life Technologies, 10 µg/cm^2)
162 coated dishes at 37°C in a humidified atmosphere of 5% CO_2 in air. Cells were routinely
163 passaged twice a week as single cells using TrypLE select (Life Technologies) and replated in
164 TESR medium that was further supplemented with the Rho kinase inhibitor Y27632 (Tocris,
165 10 µM). Twenty four hours after replating Y27632 was removed from the culture medium.
166 To make SUMO1-KGG-mCherry and SUMO2-KGG-mCherry expressing stable cell lines,
167 ChiPS4 cells were transfected using a Neon electroporation system (Thermo Fisher
168 Scientific) with 10 µl tips. Briefly, ChiPS4 cells were dispersed to single cells as described
169 above then 1x10^6 cells were collected by centrifugation at 300xg for 2 minutes and
170 resuspended in 11 µl of electroporation buffer R containing 1 µg of either paPX1-SUMO1-
171 KGG-mCherry or paPX1-SUMO2-KGG-mCherry PiggyBac expression vectors along with 0.2
172 µg of Super PiggyBac transposase (System Biosciences). Electroporation was performed at
173 1150 V, 1 pulse, 30 mSec and cells plated in mTESR containing Y27632. 5 days after

174 electroporation, mCherry positive cells were positively selected by fluorescence activated
175 cell sorting (FACS) using an SH800 cell sorter (Sony). Monoclonal cell lines were prepared
176 from the bulk sorted population by plating at low density on geltrex coated dishes and
177 individual clones picked using 3.2 mm cloning discs (Sigma Aldrich) soaked in TrypLE select.
178 Cell lines were then expanded and analysed to check for expression of mCherry and His-
179 SUMO1/2.
180
181 **Flow cytometry for cell cycle assessment and pluripotency markers.** For cell cycle analysis
182 and staining for pluripotency markers ChiPS4 cells were harvested using standard
183 procedures, washed and fixed with ice cold 70% ethanol or 4% PFA for the analysis of cell
184 cycle or NANOG staining respectively. Next cells were stained with propidium iodide or anti-
185 NANOG primary antibody, followed by Alexa 488 conjugated secondary antibody and
186 analysed by flow cytometry using Canto analyser (Becton Dickson). Data was then analysed
187 using FlowJo 10.
188
189 **Immunofluorescence, Cell Painting assay and high content microscopy**. For IF assays
190 ChiPS4 cells were seeded on µ-Slide 8 Well (Ibidi) or 96 well plates suitable for High Content
191 Microscopy (Nunc). Standard IF procedure was used where appropriate. Briefly, following
192 treatments cells were washed with PBS, fixed with 4% formaldehyde, blocked in 5% BSA in
193 PBS-T and incubated with primary and Alexa conjugated secondary antibodies. Cell Painting
194 was performed as descibed by Bray et al. (30). Imaging and subsequent analysis was
195 performed using IN Cell Analyzer systems (GE Healthcare) and Spotfire (Tibco).
196

197 **Protein sample preparation and Western blotting (WB).** ChiPS4 were maintained in a
198 stable culture as described before and treated with inhibitors for a stated time and dose,
199 usually 400nM ML792 was used for 24h or 48h. For WB cells were washed with PBS +/+ and
200 directly lysed in an appropriate volume of 2x Laemmli buffer (approximately 200 μl of buffer
201 was used per 0.5x10^6 cells) (LD; [4% SDS; 20% Glycerol; 120mM 1 M Tris-Cl (pH 6.8); 0.02%
202 w/v bromophenol blue]) and subsequently sonicated using Bioruptor Twin (Diagenode).
203 Protein content was assessed using BCA Protein Assay (ThermoFisher Scientific) and for
204 most purposes 15 μg of total protein was loaded per lane on SDS-Page gel (NuPage 4-12%
205 polyacrylamide, Bis-Tris with MOPS buffer). Proteins were transferred to PVDF membrane
206 using iBlot™ 2 Gel Transfer Device (Invitrogen). Membranes were blocked for 1h in 5% milk
207 in TBS-T and incubated overnight with primary antibodies and 1h with secondary HRP
208 conjugated antibodies before being developed using enhanced chemiluminescence
209 (ThermoFisher Scientific) and exposed to film.
210
211 **NiNTA purification.** Cells were washed with PBS and scraped in PBS containing 1mM N-
212 ethylmaleimide. The cells were then collected by centrifugation at 300 xg for 5 minutes and
213 the pellets weighed. An aliquot of the cells was lysed in 1.2x NuPage sample buffer
214 (ThermoFisher Scientific) for analysis by Western blotting. The remaining cell pellets
215 (approximately 2 g) were lysed with 5x the pellet weight of lysis buffer (6 M guanidine-HCl,
216 100 mM sodium phosphate buffer (pH 8.0), 10 mM Tris-HCl (pH 8.0), 10 mM imidazole and
217 5 mM 2-mercaptoethanol). DNA was sheared by sonication using a probe sonicator (3min,
218 35% amplitude, 20sec pulses, 20sec intervals on ice and the samples centrifuged at 4000
219 rpm for 15min at 4°C to remove insoluble material). The protein concentration of the lysate
220 was determined using BCA assay and 6.5mg of total protein from each sample was then

221 incubated overnight at 4˚C with 50 µl of packed pre-equlibrated Ni-NTA agarose beads.
222 After the overnight incubation the supernatant was removed and the beads were washed
223 once with 10 resin volumes of lysis buffer, followed by 1 wash with 10 resin volumes of 8 M
224 urea, 100 mM sodium phosphate buffer (pH 8.0), 10 mM Tris-HCl (pH 8.0), 10 mM imidazole
225 and 5 mM 2-mercaptoethanol, and then 6 washes with 10 resin volumes of 8 M urea, 100
226 mM sodium phosphate buffer (pH 6.3), 10 mM Tris-HCl (pH 8.0), 10 mM imidazole and 5
227 mM 2-mercaptoethanol. Proteins were eluted from Ni-NTA agarose beads with 125 µl 1.2x
228 NuPAGE sample buffer for SDS-PAGE.
229
230 **Mass Spectrometry based proteomics and quantitative data analysis**
231 Three proteomic experiments are described in this study;
232 (1) Changes in total proteome of ChiPS4 cells during ML792 treatment
233 ChiPS4 cells were either DMSO treated (0 hours condition), or treated with 400nM ML792
234 for 24 hours or 48 hours. Four replicates of each condition were prepared. Crude cell
235 extracts were made to a protein concentration of between 1 and 2 mg/ml by addition of
236 1.2x NuPAGE sample buffer to PBS washed cells followed by sonication. For each replicate
237 25µg protein was fractionated by SDS-PAGE (NuPage 10% polyacrylamide, Bis-Tris with
238 MOPS buffer— Invitrogen) and stained with Coomassie blue. Each lane was excised into
239 four roughly equally sized slices and peptides were extracted by tryptic digestion (31)
240 including alkylation with chloroacetamide. Peptides were resuspended in 35 µL 0.1% TFA
241 0.5% acetic acid and 10uL of each analysed by LC-MS/MS. This was performed using a Q
242 Exactive mass spectrometer (Thermo Scientific) coupled to an EASY-nLC 1000 liquid
243 chromatography system (Thermo Scientific), using an EASY-Spray ion source (Thermo
244 Scientific) running a 75 µm x 500 mm EASY-Spray column at 45ºC. A 240 minute elution

245 gradient with a top 10 data-dependent method was applied. Full scan spectra (m/z 300–
246 1800) were acquired with resolution R = 70,000 at m/z 200 (after accumulation to a target
247 value of 1,000,000 ions with maximum injection time of 20 ms). The 10 most intense ions
248 were fragmented by HCD and measured with a resolution of R = 17,500 at m/z 200 (target
249 value of 500,000 ions and maximum injection time of 60 ms) and intensity threshold of
250 $2.1x10^4$. Peptide match was set to 'preferred', a 40 second dynamic exclusion list was
251 applied and ions were ignored if they had unassigned charge state 1, 8 or >8. Data analysis
252 used MaxQuant version 1.6.1.0 (32). Default setting were used except the match between
253 runs option was enabled, which matched identified peaks among slices from the same
254 position in the gel as well as one slice higher or lower. The uniport human proteome
255 database (downloaded 24/02/2015 - 73920 entries) digested with Trypsin/P was used as
256 search space. LFQ intensities were required for each slice but LFQ normalization was
257 switched off. Manual LFQ normalization was done by calculating the relative LFQ intensity
258 compared to average LFQ intensity for each protein found in the same slice across all 12
259 samples. For each peptide sample this gave a list of sample LFQ/average LFQ values from
260 which the median was used to normalize all protein LFQ intensities in that sample. The final
261 protein LFQ intensity per lane (and therefore protein sample) was calculated by the sum of
262 normalized LFQ values for that protein intensity in all four slices. Downstream data
263 processing used Perseus v1.6.1.1 (33). Proteins were only carried forward if an LFQ intensity
264 was reported in all four replicates of at least one condition. Zero intensity values were
265 replaced from log2 transformed data (default settings) and outliers were defined by 5% FDR
266 from Student's t-test using an S0 value of 0.1. A summary of these data can be found in
267 Supp. Data. 1. The mass spectrometry proteomics data have been deposited to the

268 ProteomeXchange Consortium via the PRIDE (34) partner repository with the dataset
269 identifier PXD025867.
270
271 (2) Characterisation of ChiPS4 cells stably expressing 6His-SUMO1-KGG-mCherry and
272 6His-SUMO2-KGG-mCherry.
273 Crude cell extracts were prepared in triplicate from parental ChiPS4 cells, ChiPS4-6His-
274 SUMO1-KGG-mCherry and ChiPS4-SUMO2-KGG-mCherry types and fractionated by SDS-
275 PAGE as described above. Samples were prepared and analysed almost identically to
276 experiment 1; gels were sectioned into four slices per lane, tryptic peptides prepared,
277 peptides analysed by LC-MS/MS, and the resultant raw data processed by MaxQuant. The
278 only exceptions were the inclusion of a second sequence database containing the two 6His-
279 SUMO-KGG-mCherry constructs, and the use of MaxQuant LFQ normalization. Two
280 MaxQuant runs were performed; the first aggregating all slices per lane into a single output
281 ("by lane"), and the second considering each slice separately ("by slice"). The former was
282 used to determine cell-specific changes in protein abundance from the proteinGroups.txt
283 file, and the latter used the peptides.txt file to monitor differences in abundance of SUMO-
284 specific peptides between samples, to infer overexpression levels. For the whole cell protein
285 abundance analysis only proteins with data in all three replicates of at least one condition
286 were carried forward. In Perseus zero intensity values were replaced from log2 transformed
287 data (default settings) and outliers were defined by 5% FDR from Student's t-test using an
288 S0 value of 0.1. A summary of these data can be found in Supp. Data. 2.
289 The mass spectrometry proteomics data have been deposited to the ProteomeXchange
290 Consortium via the PRIDE (34) partner repository with the dataset identifier PXD023241.
291

292 (3) Identification of SUMO1 and SUMO2 modified proteins from ChiPS4 cells.

293 Two repeats of this experiment were performed using approximately $0.5x10^8$ cells of
294 ChiPS4-6HisSUMO1-KGG-mCherry and ChiPS4-6HisSUMO2-KGG-mCherry per replicate.
295 Samples were taken at different steps of the protocol to assess different fractions. These
296 were; crude cell extracts, NiNTA column elutions and GlyGly-K immunoprecipitations. The
297 last being the source of SUMO-substrate branched peptides. The whole procedure was
298 carried out as described previously (35). In brief, crude cell lysates were prepared of which
299 approximately 100ug was retained for whole proteome analysis as described for
300 experiments 1 and 2 above. The remaining lysate (~20 mg protein) was used for NiNTA
301 chromatographic enrichment of 6His-SUMO conjugates. Elutions from the NiNTA columns
302 were digested consecutively with LysC then GluC, of which 7% of each was retained for
303 proteomic analysis and the remainder for GlyGly-K immunoprecipitation. The final enriched
304 fractions of LysC and LysC/GluC GG-K peptides were resuspended in a volume of 20□L for
305 proteomic analysis. Peptides from whole cell extracts were analysed once by LC-MS/MS
306 using the same system and settings as described for experiments 1 and 2 above except a
307 180 minute gradient was used with a top 12 data dependent method. NiNTA elution
308 peptides were analysed identically except a top 10 data dependent method was employed
309 and maximum MS/MS fill time was increased to 120ms. GG-K immunoprecipitated peptides
310 were analysed twice. Firstly, 4uL was fractionated over a 90 minute gradient and analysed
311 using a top 5 data dependent method with a maximum MS/MS fill time of 200ms. Secondly,
312 11□L of sample was fractionated over a 150 minute gradient and analysed using a top 3
313 method with a maximum MS/MS injection time of 500ms. Data from WCE and NiNTA
314 elutions were processed together in MaxQuant using Trypsin/P enzyme specificity (2 missed
315 cleavages) for WCE samples and LysC (3 missed cleavages), or LysC+GluC_D/E (considering

316 cleavage after D or E and 8 missed cleavages) for NiNTA elutions. GlyGly (K) and phospho
317 (STY) modifications were selected. The human database and sequences of the two
318 exogenous 6His-SUMO-KGG-mCherry constructs described above were used as search
319 space. In all cases every raw file was treated as a separate 'experiment' in the design
320 template such that protein or peptide intensities in each peptide sample were reported,
321 allowing for manual normalization. Matching between runs was allowed but only for
322 peptide samples from the same cellular fraction (WCE, NiNTA elution or GG-K IP), the same
323 or adjacent gel slice, the same protease and the same LC elution gradient. For example,
324 spectra from adjacent gel slices in the WCE fraction across all lanes were matched, and
325 spectra from all GG-K IPs that were digested by the same enzymes were matched.
326 Normalization followed a similar method as described above where 'equivalent' peptide
327 samples (ie. those from the same gel slice or 'equivalent' peptide samples) from different
328 replicates were compared with one another. For each protein or peptide common to all
329 equivalent peptide samples the intensity in that sample relative to the average across all
330 equivalent samples was calculated. The median of that relative intensity in each peptide
331 sample was used to normalize all protein or peptide intensities from that sample. The final
332 total protein or peptide intensity per replicate was calculated by the sum of all normalized
333 intensities in samples derived from that replicate. It is important to note that peptide
334 samples derived from SUMO1 and SUMO2 cells were considered equivalent for
335 normalization purposes, which assumes largely similar abundances of proteins or peptides
336 between cell types. Zero intensity values were replaced from log2 transformed data
337 (Perseus default settings) and outliers were defined by 5% FDR from Student's t-test using
338 an S0 value of 0.1. A summary of these data can be found in Supp. Data. 3.

339 The mass spectrometry proteomics data have been deposited to the ProteomeXchange
340 Consortium via the PRIDE (34) partner repository with the dataset identifier PXD023257.
341
342 **Bioinformatic analysis of the SUMO site proteomics**. 429 proteins identified with at least
343 one SUMO1 or SUMO2 modification site were uploaded to STRING (36) for network
344 analysis. Only proteins associated by a minimum STRING interaction score of 0.7 (high
345 confidence) were included in the final network. Disconnected nodes were removed.
346 Selected groups of functionally related proteins were resubmitted to STRING create smaller
347 sub-networks. These were visualised in Cytoscape v 3.7.2 (37) allowing the graphical display
348 of numbers of sites identified and total GG-K peptide intensity into the protein networks.
349
350 **RNA preparation and real-time quantitative PCR (RT-qPCR)**. Total RNA was extracted using
351 RNeasy Mini Kit (Qiagen) and treated with the on-column RNase-Free DNase Set (Qiagen)
352 according to the manufacturer's instructions. RNA concentration was then measured using
353 NanoDrop and 1µg of total RNA per sample was subsequently used to perform a two-step
354 reverse transcription polymerase chain reaction (RT-PCR) using random hexamers and First
355 Strand cDNA Synthesis Kit (ThermoFisher Scientific). Each qPCR reaction contained PerfeCTa
356 SYBR Green FastMix ROX (Quantabio), forward and revers primer mix (200 nM final
357 concentration) and 6 ng of analysed cDNA and was set up in triplicates in MicroAmp™ Fast
358 Optical 96-Well or 384-Well Reaction Plates with Barcodes (Applied Biosystems™). The
359 sequences of primers used were as follows: NANOG (hNANOG_FOR624
360 ACAGGTGAAGACCTGGTTCC; hNANOG_REV722 GAGGCCTTCTGCGTCACA), SOX2 (hSOX2
361 _FOR907 TGGACAGTTACGCGCACAT; hSOX2_REV1121 CGAGTAGGACATGCTGTAGGT),
362 OCT4A (hOCT4A_FOR825 CCCACACTGCAGCAGATCA and hOCT4A_REV1064

363 ACCACACTCGGACCACATCC), KLF4 (hKLF4_FOR1630 GGGCCCAATTACCCATCCTT and
364 hKLF4_REV1706 GGCATGAGCTCTTGGTAATGG), TBP (hTBP_FOR896
365 TGTGCTCACCCACCAACAAT; hTBP_REV1013 TGCTCTGACTTTAGCACCTGTT). Data were
366 collected using QuantStudio™ 6 Flex Real-Time PCR Instrument and analysed using a
367 corresponding software (Applied Biosystems™). Relative amounts of specifically amplified
368 cDNA were calculated using TBP amplicons as normalizers.
369
370 **Experimental Design and Statistical Rationale** All proteomics analyses used a label-free
371 quantitation method employing either triplicate or quadruplicate cell cultures (biological
372 replicates). Biological replicates were individual cell cultures grown and treated separately
373 from one-another but originally derived from a single cell line culture just prior to the
374 experiment. For example, comparisons between different cell lines took individual large-
375 scale cultures for each cell line and split the cells equally among replicate culture dishes.
376 Once these had grown to appropriate levels of confluence they were treated as described.
377 Intensity data from multiple mass spectrometry runs of the same peptide sample (technical
378 replicates) were normalized against other samples from the same mass spectrometry run
379 and then quantitative data from identical peptide samples aggregated together by
380 summation prior to statistical analyses. Protein or peptide outliers were determined by two-
381 samples student's t-tests in pairwise comparisons. Specific FDR and S0 values for filtering
382 are described for each experiment above.
383
384 **Results**
385

386 **Inhibition of SUMO modification leads to decreased expression of pluripotency markers in**
387 **hiPSCs.** The pluripotent state in hESCs and hiPSCs is controlled by a network of transcription
388 factors and other chromatin-associated proteins that determine the chromatin environment
389 of key genes (1). To understand the role of SUMO modification in the maintenance of
390 pluripotency in hiPSCs we used ML792 (26) a potent and selective inhibitor of SUMO
391 Activating Enzyme (SAE). This inhibitor has been reported to block proliferation of cancer
392 cells, particularly those overexpressing Myc (26), but has not been evaluated in hiPSCs. To
393 address the role of SUMO modification in hiPSCs, ChiPS4 cells were treated with ML792 in a
394 series of time-course and dose-response experiments. We determined that 400nM ML792
395 was the lowest concentration that effectively reduced SUMO modification after 4 hrs
396 treatment with minimal effects on cell viability. We restricted our analyses to ML792
397 treatment times that did not exceed 48 hrs, such that the early effects of SUMO
398 modification inhibition could be evaluated. Microscopic examination revealed that ML792
399 treatment caused morphological changes with the ChiPS4 cells becoming larger and flatter
400 (Fig. 1a). The rate of proliferation was unchanged after 24 hrs but was slightly reduced after
401 48 hrs (Fig. 1b). DNA staining of the cells and analysis by flow cytometry indicated that the
402 cell cycle distribution after 24hrs was unaltered by ML792 treatment but after 48 hrs
403 displayed an increased proportion of cells in G2 phase and cells with increased DNA content
404 suggesting endoreplication (Fig. 1c). Western blot analysis revealed a loss of high molecular
405 weight SUMO1 and SUMO2 conjugates and concomitant increase in free SUMOs in ChiPS4
406 cells (Fig. 1d). This is likely to be a consequence of the rapid removal of SUMO from
407 modified proteins by SUMO specific proteases (SENPs). Analysis of the protein levels of key
408 pluripotency markers indicated that while the levels of OCT4 and SOX2 were unchanged,
409 inhibition of SUMOylation resulted in a decrease in the protein level of NANOG (Fig. 1d).

410 This appeared to be a consequence of reduced transcription as determination of mRNA
411 levels by reverse transcriptase quantitative polymerase chain reaction (RT-qPCR) after
412 ML792 treatment demonstrated a reduction in *NANOG* mRNA. This was also apparent for
413 *KLF4*, but consistent with the Western blotting results, the levels of *OCT4* and *SOX2* mRNA
414 were unchanged (Fig. 1e).
415
416 **Inhibition of SUMO modification induces phenotypic changes but not large scale**
417 **proteomic changes in hiPSC.** To further investigate the nature and causes of the observed
418 morphological changes induced by the inhibition of SUMO modification, ChiPS4 cells were
419 treated with ML792 for 48h and analysed by phenotypic screening using a cell painting assay
420 (30) (Fig. 2a). Principle component analysis (PCA) indicated that there are clear differences
421 between cells treated with ML792 for 48h and untreated or vehicle (DMSO) treated cells.
422 The main differences were found to occur in the nuclear compartment (Fig. 2a). Feature
423 extraction identified changes in the global size (Area of nuclei) and shape (Nuclei form
424 factor) of the nucleus and the structure of the nucleolus (Nuclei: FITC texture correlation)
425 (Fig. 2a). These findings were validated using traditional immunofluorescence (IF)
426 approaches (Fig. 2b). Consistent with previously presented data, NANOG expression as well
427 as the size and shape of the nucleus are both affected by ML792 treatment. NOP58 was
428 used as a marker of the nucleolus, which undergoes a dramatic increase in size and shape
429 (Fig. 2b). As expected, the classic punctate nuclear localisation pattern of SUMO1 and
430 SUMO2 is altered by their deconjugation from substrates, becoming more diffuse and less
431 tightly associated with the nucleus (Fig. 2b).
432 To evaluate the effect of inhibition of SUMO modification on the global proteome in
433 hiPSCs, total protein extracts were prepared from ChiPS4 cells either untreated or treated

434 with ML792 for 24 or 48 hours and analysed by label-free quantitative proteomics. Data for
435 4741 proteins was obtained (see Supp. Data File 1). Consistent with SUMO conjugation
436 being critical for pluripotency, known pluripotency markers were modestly but significantly
437 reduced during 24 h and 48 h ML792 exposure (Fig. 2c). However, overall protein
438 abundance changes compared at both time-points showed little evidence for large-scale
439 shifts (Fig. 2d), with few functionally related proteins undergoing coordinated regulation
440 according to STRING: Linker histones appeared to be reduced in abundance after 48 hours,
441 but not 24 hours (Fig. 2d) and proteins from the gene ontology group 'Collagen-associated
442 extracellular matrix' had members both up and down-regulated (Supplementary Data file 1).
443 Alone these data do not appear to provide an obvious link between the morphological
444 changes induced by SUMO inhibition and underlying protein abundance changes. This
445 implies that it is an accumulation of multiple small protein abundance changes effected by
446 or in addition to deSUMOylation of key regulators, that explains the consequences of
447 ML792 treatment on hiPSCs.
448
449 **Generation of hiPSC lines for SUMO proteomic analysis.** Our data suggest an important
450 role for SUMO in maintaining the pluripotent state of hiPSCs. Given that previous studies of
451 SUMOylation have typically focused on somatic cells, it was important to establish the
452 SUMOylome of hiPSCs to identify candidate proteins that either directly or indirectly
453 contribute to the observed phenotype. We adapted a proteomic approach that allows sites
454 in proteins modified by SUMO1 and SUMO2/3 to be identified (38). To enable this analysis
455 in hiPSCs the ChIPS4 cell line was engineered to stably express 6His-SUMO-mCherry
456 constructs for either SUMO1 or SUMO2 (Fig. 3a, b) that incorporated the TGG to KGG
457 mutations to facilitate GlyGly-K peptide immunoprecipitation and identification as described

458 previously (38). As mCherry is linked to the C-terminus of SUMO the expressed fusion
459 protein will be processed by endogenous SUMO proteases to expose the C-terminal GlyGly
460 sequence for conjugation and release free mCherry. Single cell clones were selected based
461 on the expression of free mCherry and the level of SUMO paralogue expression (Fig. 3c).
462 Clones selected for SUMO site proteomic experiments were extensively characterised to
463 ensure that the exogenous SUMO was functional and that the cells retained their
464 pluripotency. Western blotting indicated that His-tagged SUMO-KGG paralogues were
465 conjugated to substrates in response to heat shock (Supplementary Fig. 1). Cells expressing
466 SUMO1-KGG and SUMO2-KGG had normal cell cycle profiles (Supplementary Fig. 2a),
467 expressed levels of pluripotency markers comparable to wild type ChIPS4 cells
468 (Supplementary Fig. 2b, c) and retained the ability to differentiate into endoderm, ectoderm
469 and mesoderm (Supplementary Fig. 2d). Analysis of the whole cell proteomes (Fig. 3d) of
470 ChIPS4 cells expressing SUMO1-KGG and SUMO2-KGG identified the expected exogenous
471 mCherry, SUMO1 and SUMO2 peptides (Fig. 3e) while peptides common to both
472 endogenous and exogenous SUMOs showed both types were conjugated to substrates at
473 roughly similar levels (Fig. 3f). Importantly, the engineered cell lines did not show large-scale
474 differences from parental cells in their expressed proteomes (Fig. 3g, h, see also
475 Supplementary Data File 2). Together these data indicate that expression of SUMO mutants
476 did not disrupt the normal pluripotent state or differentiation potential of ChiPS4 cells.
477
478 **Identification of SUMO1 and SUMO2 targets in hiPSCs.** The workflow for the identification
479 of SUMO1 and SUMO2 targets incorporates proteomic analysis at three levels
480 (Supplementary Fig. 3a). This involved analysis of whole cell extracts to monitor total
481 protein levels, analysis of nickel affinity purified proteins to identify SUMO modified

482 proteins and analysis of GG-K immunoprecipitations to define sites of SUMO modification.
483 SUMO1/SUMO2 comparisons can be made at the site level because after LysC digestion
484 both mutants leave the same GG-K adduct on substrates (Supplementary. Fig. 3b), and so
485 intensity comparisons can be used to infer preference for SUMO type at the site level. The
486 experiment was conducted twice in triplicate and PCA indicated that replicates performed
487 at the same time displayed a high degree of clustering (Supplementary Fig. 3c). PCA also
488 indicated clear differences between SUMO1 and SUMO2 as well as between experiments
489 carried out at different times (Supplementary Fig. 3c). This is likely to be a reflection of the
490 precise growth state of the cells when the experiment was conducted. Very few total
491 protein abundance differences between SUMO1 and SUMO2 cell were consistent to both
492 experimental runs, but there were substantial and consistent differences between SUMO1
493 and SUMO2 at both nickel affinity purifications and at the site level (Supplementary Fig. 3d).
494 Across the two experimental runs and aggregating the data for SUMO1 and SUMO2 a total
495 of 976 SUMO sites were identified in 427 proteins. Approximately 84% of these had already
496 been described in at least one of four large-scale SUMO2 site proteomics studies totalling
497 49768 unique sites of non-stem cell origin (see Supplementary Data file 3) (Fig. 4a). DNA
498 methyl transferase 3B (DNMT3B) and the key embryonic stem cell transcription factor SALL4
499 were among a small group of proteins with at least three novel sites in this study (Fig. 4a).
500 Although peptide intensity is a relatively imprecise proxy for abundance, it has been
501 successfully used for SUMO-substrate branched peptides to separate high occupancy from
502 low occupancy sites (39). For our data total GG-K peptide intensity suggests SALL4 is among
503 the most modified SUMO substrate in hiPSCs and contains 17 sites of modification (Fig. 4b).
504 DNMT3B contains 12 sites and is also among the most abundantly modified substrates,
505 while the methyl DNA binding protein MBD1 contains 8 sites and is also in the top cohort of

506 substrates by modified peptide intensity (Fig. 4b). Transcription intermediary factors 1 α
507 and 1 β (TRIM24 and TRIM28) are also both highly modified substrates. The boundary and
508 chromatin isolation factor CTCF is heavily modified with SUMO, consistent with a role for
509 SUMO in chromatin architecture (Fig. 4b). There is also evidence for extensive SUMO chain
510 formation as the branch points from SUMO2/3 chains are amongst the most abundant GG-K
511 peptides (Fig. 4b). Strikingly, the SUMOylated forms of a number of these nuclear substrates
512 could be detected directly in total whole cell lysates from control ChiPS4 cells, but not
513 ML792 treated hiPSCs (Fig. 4c), implicating SUMO as a major regulator of the total pool of
514 these proteins. STRING enrichment analysis of the 427 modified proteins created a network
515 consisting of 3 broad clusters that could be categorised as having functions in ribosome
516 biogenesis, RNA splicing, and regulation of gene expression (Fig. 4d). Despite forming
517 extensive protein networks (Fig. 5a, b), proteins involved in ribosome biogenesis and RNA
518 splicing represented only approximately 5% of the total GGK peptide intensity (Fig. 4b
519 insert). Most of the remainder have roles in transcription and chromatin structure or are
520 closely linked to these functions (Fig. 5c-h). There is a prominent network of zinc-finger
521 transcription factors, closely associated with TRIM28 (Fig. 5c) which contains many of the
522 most heavily SUMOylated proteins identified. The TRIM-ZNF-SUMO axis may play a key role
523 in silencing retroviral elements and this may be particularly important in hiPSCs (40).
524 Histone proteins themselves form a small cluster of SUMO substrates (Fig. 5d) which STRING
525 positioned in the centre of the gene regulation region of the whole network (Fig. 4c). The
526 transcriptional regulators themselves form a bipolar network (Fig. 5e) with the smaller sub-
527 cluster consisting mainly of apparently weakly modified ribosomal proteins and the larger
528 sub-cluster containing many heavily modified chromatin associated proteins. Strikingly,
529 many members of chromatin remodelling complexes such as PRC2 (Fig. 5f), BAF (Fig. 5g) and

530 NURD (Fig. 5h) are among this group. Together these networks and clusters of proteins
531 provide multiple direct and indirect links between SUMO and chromatin structure
532 regulation.
533
534 **Differences between SUMO1 and SUMO2 at the protein and acceptor site level.** The
535 proteomic experimental design allowed quantitative comparisons between SUMO1 and
536 SUMO2 at multiple stages of the purification process (Supplementary Fig. 3a):
537 SUMO1/SUMO2 ratios from crude IPS cell extracts shows there to be few differences (0.03%
538 significant) at the whole proteome level (Fig. 6a). There are also surprisingly few differences
539 (7.8% significant) between NiNTA purifications from the two cell types (Fig. 6a). Exceptions
540 include the well-documented SUMO1 substrate RanGAP1, along with TRIM24 and TRIM33
541 which all appear to show similar levels of SUMO1 preference (Fig. 6a). In contrast, over half
542 of the GGK-containing peptides quantified in both experimental runs and both cell lines
543 showed large and significant difference between SUMO1 and SUMO2 cells (Fig. 6a). Extreme
544 examples of SUMO1 preferential sites include not only RanGAP1 K524, but also TRIM33
545 K776, NFRKB K359, PNN K157, TFPT K216 and two sites in TOP1 (K117 & K134) (Fig. 6a).
546 Conversely, TRIM28 contains two of the most SUMO2-preferntial sites at K507 and K779,
547 and lysines 48 and 63 from ubiquitin also seems to be among the extreme SUMO2 acceptors
548 (Fig. 6a). In the context of whole proteins, three key proteins in our dataset, SALL4, TRIM24
549 and TRIM33 show largely SUMO1 preferential modification, while TRIM28 and CTCF show
550 apparent preference for SUMO2 (Fig. 6b, c). This was broadly confirmed by Western-blot
551 analysis from NiNTA purifications (Fig. 6d), although in some cases the degree of preference
552 was not as striking as expected. Significantly, Western-blot analysis of NiNTA purifications
553 confirmed the SUMO2-modified form of CTCF to be more abundant than SUMO1, but both

554 the modified and unmodified forms of CTCF were purified (Fig. 6d). Thus, it seems likely that
555 large amounts of unmodified CTCF obscured the SUMO2 preference in the NiNTA
556 purification proteomic data. This effect may be rare, but highlights potential weaknesses in
557 proteomics studies when using relatively low-stringency purifications such as 6His/Ni-NTA
558 pull downs.
559
560 **Preference for SUMO2 modification in regions of higher negative charge.** The proteomic
561 data reveal many multiply modified proteins to have a strong overall SUMO paralog
562 selection. TOP1, TRIM33, ZNF462 and ZNF532 all contain multiple SUMO1-specific sites, and
563 conversely, DNMT3a, CHD4, ubiquitin and SUMOs 2 and 3 themselves have a majority of
564 SUMO2-specific sites (Supplementary Fig. 4). However, a number of substrates such as
565 GTF2I, SAFB2 and MGA (Supplementary Fig. 4) have a mixture of sites showing varied SUMO
566 type preference. This broad range of site-specific SUMO paralogue preference raises the
567 question of how specificity is determined. By averaging the SUMO1/SUMO2 peptide
568 intensity ratio data over the two experimental runs we ranked 739 sites by paralogue
569 preference from SUMO1 to SUMO2 (Fig. 7a). Not all sites could be included for sequence
570 analysis as sites close to protein N or C termini lacked amino acids in their sequence
571 windows. Sequence logos of the top and bottom 123 sites (one sixth of total) shows broadly
572 similar motifs for SUMO1 and SUMO2 sites (Fig. 7b), with both showing the characteristic
573 ψKxE motif. However, outside the consensus SUMO2 sites are more prone to D or E in the -2
574 position than SUMO1 sites, and a significant number of SUMO1 sites have P in +3 (Fig. 7b).
575 More generally, D/E residues are more enriched within the 21 residue sequence windows of
576 SUMO2 than SUMO1. This is confirmed by sequence window charge analysis (Fig. 7c) which
577 shows acceptor lysine residues preferentially modified with SUMO2 over SUMO1 are more

578 likely to be embedded in a sequence that is net negatively charged at pH7.4 than those
579 showing SUMO1 preference. This trend is also generally progressive across the spectrum of
580 site-specific SUMO paralogue selection (Supplementary. Fig. 5a-e). For some proteins where
581 SUMO paralogue preference varies in different regions of the sequence, this coincides with
582 local charge variations consistent with SUMO2 preference in acidic domains. SALL4, CHD4
583 and SAFB (Fig. 7d) show that sites preferentially modified by SUMO2 were generally
584 negatively charged while SUMO1-preferential modification took place in regions of the
585 protein of that were more positively charged.

586

587 **Discussion**

588 Our studies highlight an important role for SUMO in maintaining the pluripotent
589 state of ChIPS4 hiPSCs. Using a potent and highly specific inhibitor of the E1 SUMO
590 Activating Enzyme ML792 (26) we blocked de novo SUMO modification which allowed
591 endogenous SUMO specific proteases to remove SUMO from previously modified proteins.
592 In this way ML792 treatment effectively deSUMOylated cellular proteins. In response to 48h
593 of 400nM ML792 exposure ChIPS4 cells showed reduced pluripotency and underwent
594 dramatic morphological changes, particularly to nuclear structure and size. Over the same
595 period of time there appeared to be few large-scale changes to the cellular proteome,
596 implying that the consequences of loss of SUMOylation in hiPSCs primarily involves
597 functional changes to factors critical for nuclear structure and function, probably followed
598 by multiple modest protein abundance changes, which together lead to the observed
599 phenotype. Candidates for these SUMOylated factors were identified by SUMO site
600 proteomic analysis. The major network that accounted for the bulk of SUMO modification
601 sites was associated with negative regulation of gene expression. At the heart of this

602 network is TRIM28, the histone methyl transferase SETDB1 and the Chromobox silencer
603 CBX3. ChIP-seq derived locations of TRIM28, SETDB1 and CBX3 indicate that they are
604 associated with retroviral elements (41). These proteins along with SUMO have previously
605 been shown to function in HERV silencing in mouse ES cells (40,42) and adult human cells
606 (43,44) and this is consistent with our proteomic analysis that indicates that all three of
607 these proteins are heavily SUMO modified (Fig. 4). The TRIM28 co-repressor functions by
608 interacting with DNA bound Kruppel type Zinc finger proteins and these proteins are also
609 identified as being SUMO modified in our proteomic studies and are part of the large
610 TRIM28 centric network of SUMO modified proteins. It is also suggested that a number of
611 developmental genes are repressed by TRIM28/KRAB-ZNFs through H3K9me3 and de novo
612 DNA methylation of their promoter regions (45), thus making TRIM28/ZNFs a crucial link in
613 maintenance of pluripotency in human stem cells. These data are consistent with the wide
614 distribution of SUMO on chromatin (46-49) and with a very different distribution in murine
615 fibroblasts and embryonic stem cells (9). Recently Theurillat et al.(50) reported a SUMO site
616 proteomic analysis of mouse embryonic stem cells. In this case the method used allowed
617 analysis of endogenous SUMO2 but not SUMO1 modification sites. The number of sites
618 identified in this study was rather similar (608 SUMO2 sites in 350 proteins) to that reported
619 here (976 SUMO1 plus SUMO2 sites in 427 proteins), as were the major functional
620 networks. However, the precise targets identified were very different, with only SALL4 of
621 the top 20 mouse ES cell specific proteins being present in the top 50 (based on peptide
622 intensity) SUMO modified proteins in human stem cells reported here (Fig. 4b). Such
623 differences presumably reflect the differences between mESCs and hiPSCs and the
624 developmental stages which these cells represent. While the mouse cells were established
625 from blastocysts the human cells are derived by reprogramming normal somatic cells (25).

626 Analysis of the SUMO proteome of the ChIPS4 hiPSCs shows that well-defined
627 groups of protein are modified. Aside from the TRIM28/ZNF network mentioned above,
628 proteins involved in “ribosome biogenesis” and “splicing” are SUMO modified and this likely
629 impacts on the normal growth and self-renewal of the hiPSCs. However, the largest network
630 of proteins falls into the category of “negative regulation of transcription”, with many
631 chromatin remodellers, chromatin modification and DNA modification enzymes identified as
632 SUMO substrates. Thus, SUMO modification is likely to play an important role in maintaining
633 pluripotency of hiPSCs by repressing genes that either disrupt pluripotency or drive
634 differentiation.

635 During our analysis we also found that a large proportion of modification sites
636 displayed a clear preference for either SUMO1 or SUMO2, and that this seems to be
637 influenced by the charge distribution around the acceptor lysine. Specifically, lysines within
638 acidic domains are more likely to be modified by SUMO2 than SUMO1. While this is by no
639 means a strict rule, these data suggest site-level SUMO paralog selection is at least in part
640 defined by the electrostatic environment of the acceptor lysine. Intriguingly, structures of
641 Ubc9 in complex with either SUMO1 or SUMO2 (51) show significant charge deviation
642 between the two paralogues close to the active site of the E2 enzyme (Supplementary Fig.
643 6), potentially implicating the SUMO molecule itself in site-selection. It is important to note
644 however, that in spite of the clear variety of site-level SUMO paralogue preference, most
645 proteins display little overall SUMO preference when considering the total cellular pool
646 together, meaning examples of entirely SUMO1 or SUMO2/3 modified proteins are probably
647 rare. Understanding the differences between SUMO1 and SUMO2/3 in terms of site-
648 selectivity and downstream functional outcomes will likely be critical in gaining a full
649 understanding of the role of SUMOylation in all higher eukaryotic cell types.

650

651

652 **Acknowledgements**

653 We would like to acknowledge the invaluable help from various facilities at the University of
654 Dundee: Flow Cytometry and Cell Sorting, National Phenotypic Screening Centre, Dundee
655 Imaging Facility, Human Pluripotent Stem Cell Facility. This work was supported by an
656 Investigator Award from Wellcome (217196/Z/19/Z) and a Programme grant from Cancer
657 Research UK (C434/A21747) to RTH. ML is part of the Ubi-Code European Training Network
658 that received funding from the European Union Horizon 2020 research and innovation
659 programme under the Marie Curie grant agreement No. 765445. BM was supported by the
660 European Union's Horizon 2020 research and innovation programme under the Marie
661 Skłodowska-Curie grant agreement No. 704989.

662

663 **Author Contributions**

664 B.M. cloned expression vectors, generated cell lines, designed and performed most
665 experiments using hiPSCs and interpreted data. ML performed NiNTA purifications and
666 Western blotting analyses. L.D. was in charge of cell culture, cell line derivation and quality
667 control and was consulted over experimental design. M.H.T consulted over proteomic
668 experimental design, acquired and processed MS data and conducted bioinformatic and
669 statistical analyses. E.B. performed bioinformatic analysis of the structural data. B.M.,
670 M.H.T., L.D., and R.T.H. contributed to data analysis. B.M., R.T.H., M.H.T. and L.D. wrote the
671 paper. R.T.H. conceived the project.

672

673 **Data availability**

674 Data underlying all Figures and Supplementary Figures are available in the source data file.
675 Source data for proteomics experiments are deposited in PRIDE as described in the Methods
676 section. All other data are available from the corresponding authors on reasonable request.
677

678 **Supplementary data files**

679 This article contains supplemental data including following supplementary data files:

680 ***Supplementary data file 1.*** Summary of the quantitative data from the proteomics
681 experiment to study changes to the cellular proteome during ML792 treatment of ChiPS4
682 cells.

683 ***Supplementary data file 2.*** Summary of the quantitative data from the proteomics
684 experiment to study differences in the cellular proteome among wild type ChiPS4 cells and
685 cells expressing 6His-SUMO1-KGG-mCherry or 6His-SUMO2-KGG-mCherry.

686 ***Supplementary data file 3.*** Summary of the quantitative data from the proteomics
687 experiment to identify SUMO1 and SUMO2 targets from ChiPS4 cells.
688

689 **Author Information**

690 Correspondence and requests for materials should be addressed to R.T.H
691 (r.t.hay@dundee.ac.uk)
692

693 **Figure Legends**

694

695 **Figure 1. Inhibition of SUMO modification leads to loss of hiPSCs pluripotency.**

696 **A.** ChiPS4 cells were treated with SUMO E1 activating enzyme inhibitor ML792 (400 nM) or
697 DMSO vehicle for the indicated time and analysed for morphology, using phase contrast
698 microscopy (scale bar = 100 μm). **B.** ChIPS4 cells were seeded at a standard density of $3x10^5$
699 cells/cm^2 in triplicate and either DMSO or 400nM ML792 treated for the indicated
700 durations. Cells were harvested using TrypleSelect and counted. Data are plotted as a mean
701 (line) ± SEM of individual replicates (dots) n=3. Statistical significance was calculated with t-
702 tests corrected for multiple comparisons using Holm-Sidak's method (* p<0.05 significantly
703 different from the corresponding DMSO control) **C.** Crude cell extracts taken from cells
704 incubated at the indicated time-points with 400nM ML792 were analysed by Western
705 blotting to determine conjugation levels of SUMO1, SUMO2/3 and abundance of key
706 pluripotency markers NANOG, SOX2 and OCT4. Anti-Actin western blot was used as a
707 loading control. Asterisk (*) indicates free SUMO1 or SUMO2/3. **D**. Cells were collected as in
708 **B**, fixed, stained with propidium iodide (PI) and analysed by flow cytometry. Plots are a
709 representative of three independent experiments. **E.** Cells were treated with ML792 and
710 after the indicated time total RNA was extracted. mRNA levels of NANOG, OCT4, SOX2 and
711 KLF4 were determined by quantitative PCR. Relative mRNA expression levels normalized to
712 TBP were plotted as means ± SEM of four independent experiments. **p <0.01; *** p
713 <0.001; **** p <0.0001 significantly different from the corresponding value for untreated
714 control (two-way ANOVA followed by Sidak's multiple comparison test).

715

716 **Figure 2. Inhibition of SUMO modification causes morphological changes but does not**
717 **trigger large-scale proteome changes in hiPSCs.**
718 **A.** Cell painting analysis. ChIPS4 cells were treated with PBS, DMSO vehicle or 400 nM
719 ML792 for 48 h. Cells were then stained, fixed and analysed using high content microscopy.
720 The experiment was performed three times in 8 replicates per condition. Information
721 extracted from Cell Painting analysis was focused on subcellular compartments most
722 affected by ML792 treatment. Most variation between treatments and controls in principal
723 component analysis was captured by PC1. Selected graphs represent quantitation of
724 individual measures contributing to the difference observed in principal component
725 analysis: area of nuclei; nuclei form factor (size and shape of nucleus); nuclei FITC texture
726 correlation (size of nucleolar structures). **B.** ChIPS4 cells were treated for 48 h with DMSO
727 vehicle or 400 nM ML792, fixed and stained with DAPI (blue), anti-SUMO1 or anti-SUMO2
728 (red) and anti-NANOG or anti-NOP58 (green) antibodies. Immunofluorescence (IF) images
729 were obtained using Leica SP8 confocal microscope and a 60x water lens. All images contain
730 25 μm scale bar. **C.** Crude extracts from hiPSCs treated in triplicate with ML792 for 24h or
731 48h or not (0h) were analysed by label-free proteomics and the change in protein
732 abundance of detected pluripotency markers (n=14) was calculated as % of 0h intensity.
733 Paired t-test p values are indicated. Average reduction is 13.2% (24h) and 25.7% (48h). **D.**
734 Scatter plot of Log_2 24h/0h and log_2 48h/0h abundance change for the entire 4741 protein
735 whole cell proteomic dataset. Extreme outliers are indicated. All identified core and linker
736 histones are represented by coloured markers others are in grey. Linker histones were
737 identified by STRING analysis as a functionally related group of proteins that are significantly
738 reduced in abundance at 48h compared with 0h. Core histone proteins are indicated for
739 reference.

740

741 **Figure 3. ChiPS4 cell lines engineered for SUMO1 and SUMO2 site proteomic analysis.**

742 **A.** Overview of the principles behind the SUMO-mCherry construct design and the utility of
743 the expressed proteins for SUMO site proteomic analysis in CHiPS4 cells. The C-terminal
744 mCherry protein used for cell selection is cleaved away from 6His-SUMO by endogenous
745 SUMO proteases allowing conjugation to substrate proteins. **B.** Scheme representing the
746 piggyBac construct design used for generation of ChiPS4 cell lines expressing 6His-SUMO1-
747 KGG-mCherry or 6His-SUMO2-KGG-mCherry. **C.** Flow cytometry analysis of single cell clones
748 using mCherry expression levels to infer SUMO expression levels. **D.** Coomassie stained SDS-
749 PAGE gel fractionating whole cell protein extracts from parental CHiPS4 cells (WT) and the
750 selected 6His-SUMO-KGG-mCherry clones prepared in triplicate. This was used to monitor
751 SUMO overexpression levels using proteomic analysis and each lane was excised into 4
752 sections allowing differentiation between conjugated (slices A-C) and unconjugated (Slice D)
753 SUMO forms. **E.** Peptide intensity data from slices A-D for two peptides each from 6His-
754 SUMO1-KGG (left) and 6His-SUMO2-KGG (centre) that are unique to the exogenous
755 constructs. Data for 20 mCherry peptides are also shown (right). **F.** Peptide intensity data
756 from slices A-C for four peptides from SUMO1 (left) and three from SUMO2 (right) that are
757 common to both the endogenous and exogenous forms of the proteins. **G.** Quantitative
758 data from 3863 proteins identified from the gel shown in D were compared by principal
759 component analysis. **H.** Numerical ratio and unpaired student's t-test results comparing WT
760 parental cells with 6His-SUMO1-KGG-mCherry cells (left), and WT with 6His-SUMO2-KGG-
761 mCherry cells (right). Outliers (red markers) were defined in Perseus by 5% FDR with an S0
762 value of 0.1 (78 outliers from WT vs SUMO1 cells and 64 from WT vs SUMO2 cells). Gene
763 names from extreme outliers are indicated.

764

765 **Figure 4. Functions of SUMO1 and SUMO2 targets in hiPSCs.**

766 **A.** 976 SUMO sites were identified from 6His-SUMO1-KGG and 6His-SUMO2-KGG IPS cells,
767 of which 155 were novel compared with previous high-throughput SUMO site proteomics
768 studies (See Supplementary Data File 2). Proteins with three or more novel sites are
769 highlighted. **B.** Summary of the top 50 SUMO substrates by total SUMO1+SUMO2 GGK-
770 peptide intensity for all identified sites. Gene names are shown with numbers of sites in
771 brackets. Bars are colour coded by category shown in panel C. Insert shows contribution to
772 total GGK peptide intensity of proteins from the categories shown in C (note categories are
773 not mutually exclusive). **C.** STRING interaction network of the 427 IPS SUMO substrates.
774 Only high confidence interactions were considered from 'Text mining', 'Experiments' and
775 'Databases' sources. Network PPI enrichment p-value $<1.0 \times 10^{-16}$. Nodes are coloured by
776 functional or structural group as indicated. **D.** Immunoblot analysis of ChIPS4 cells treated
777 with the SUMO E1 conjugating enzyme inhibitor ML792 or DMSO for 48h. Total protein
778 extracts were probed with anti-TRIM28, anti-TRIM24, anti-CTCF, anti-DNMT3b, anti-SALL4,
779 anti-TRIM33 and anti-tubulin or anti-Lamin A/C antibodies (loading controls). SUMO-
780 modified species above the band for unmodified proteins (*), reduce with ML792
781 treatment.

782

783 **Figure 5. Networks of functionally related SUMO substrates in hiPSCs.**

784 **A-E.** Detailed protein interaction networks derived from the broad categories shown in Fig.
785 3C. Node shade is proportional to $\log_{10}$ total GGK peptide intensity and border thickness
786 indicates numbers of sites found. **F-H** shows data for selected chromatin remodelling
787 complexes. Grey nodes were not identified in the present study but included to complete

788 complexes. TRIM proteins were included in C to link more zinc-finger proteins in the
789 network.
790
791 **Figure 6. SUMO paralogue preference at the site level in hiPSCs is influenced by proximal**
792 **acidic residues.**
793 **A.** Scatter plots of $\log_2$ SUMO1/SUMO2 ratio and -$\log_{10}$ student's t-test p-value for proteins
794 or peptides detected in the different cellular fractions. Red markers were found to be
795 significantly different in both experimental runs. Selected outliers are indicated. Numbers of
796 significantly differing proteins or peptides compared to the entire set of proteins or
797 peptides quantified are shown. **B.** Schematic presentations of selected proteins found to
798 have high total GG-K peptide intensities in the present study. Site positions are indicated by
799 number. SUMO preference is represented by colour and GGK peptide intensity by the size of
800 the site node (see key). Site node border line thickness represents number of experiments in
801 which it was found to show a significant SUMO preference. Protein nodes are labelled with
802 the gene name and are colour is based on SUMO preference calculated by total GGK
803 peptide intensity from SUMO1 cells/total from SUMO2 cells. Edges linking sites to proteins
804 are angled relative to their position in the linear protein sequence with first and last
805 residues at the top of the protein node. **C.** Summary of log2 SUMO1/SUMO2 intensity
806 measured in each of the three cellular fractions analysed. § - No data acquired. **D.**
807 Immunoblot analysis of whole cell extracts (input) and NiNTA purifications from SUMO1 and
808 SUMO2 cells either treated or not with 400nM ML792. * - Position of unmodified protein in
809 NiNTA purifications.
810
811 **Figure 7. Validation of SUMOylation status of abundant SUMO substrates in ChIPS4 cells.**

812 **A.** Log_2 SUMO1/SUMO2 ratio versus rank of ratio for 738 GG-K containing peptides that
813 could be numerically compared between SUMO1 and SUMO2 samples. Top and bottom
814 1/6th (123 sites) by ratio are selected to represent SUMO paralogue-preferential sites
815 (coloured) with the remainder marked in grey. **B.** pLogos (52) for the 123 SUMO1 or SUMO2
816 specific sites as shown in part C using the human proteome as background. Two analyses
817 were performed using the entire 21 residue window (left), and the 17 residue window of
818 the same sequences without the 4 residue SUMO consensus motif (right). The position of
819 the consensus sequence region is boxed with a dashed line (left) or shown by a dotted line
820 (right). **C**. Comparison between SUMO1 and SUMO2-preferential sites for predicted net
821 charge at pH7.4 of the amino-acid region encompassing -10 to +10 residues around the
822 acceptor lysine. Charge predicted using Isoelectric Point Calculator tool (53)**. D.** Schematic
823 presentations summarising site-specific SUMOylation data, and predicted sequence charge
824 distribution for SALL4, CHD4 and SAFB2. Net charge at each amino acid position in the
825 sequence (left) was calculated for sequence windows of sizes from 5 to 100 residues as
826 indicated in the key (see M&M for details). SUMO conjugation sites and 21 residue
827 sequence windows are shown (right).
828

829 **Bibliography**

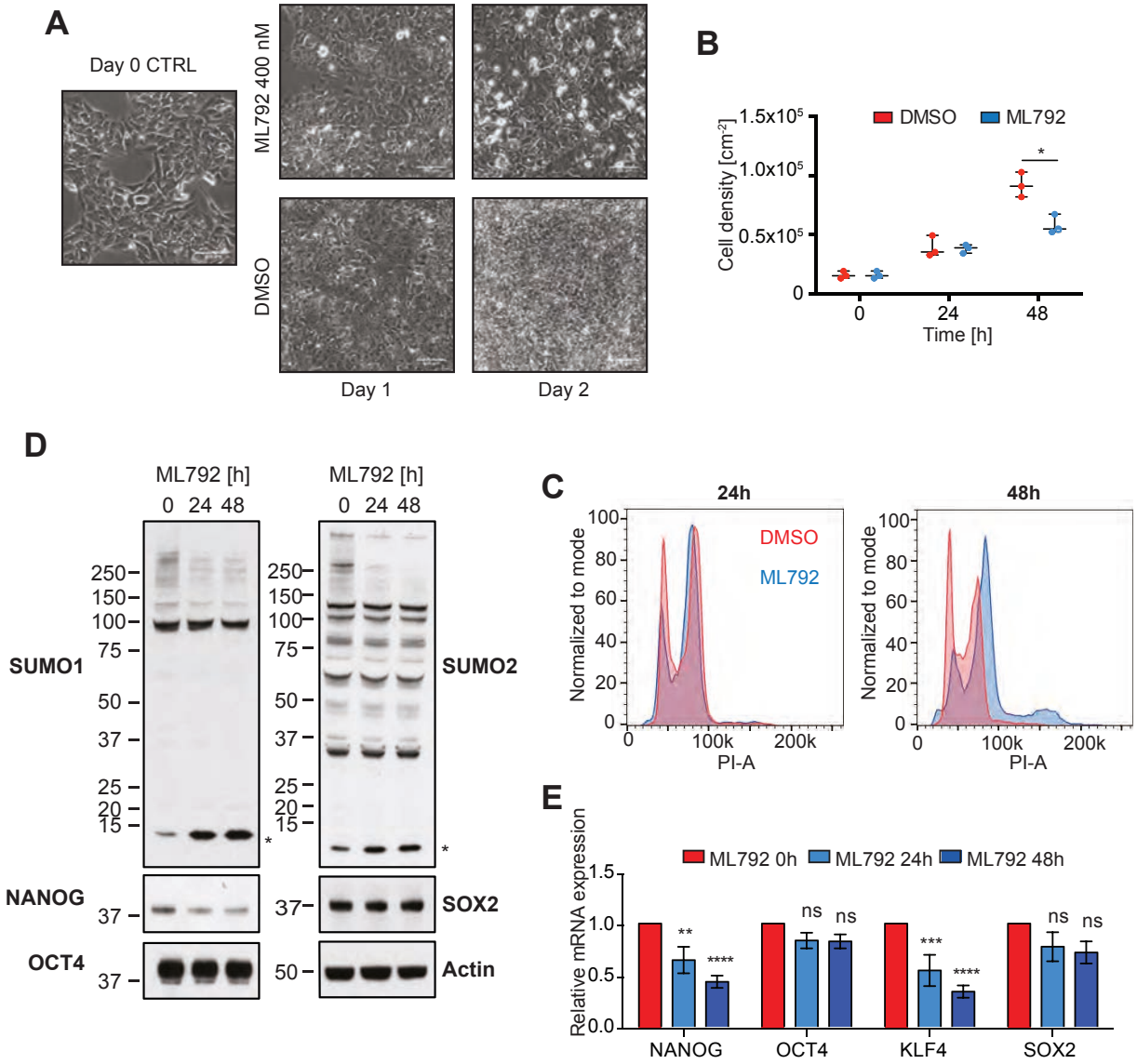
830

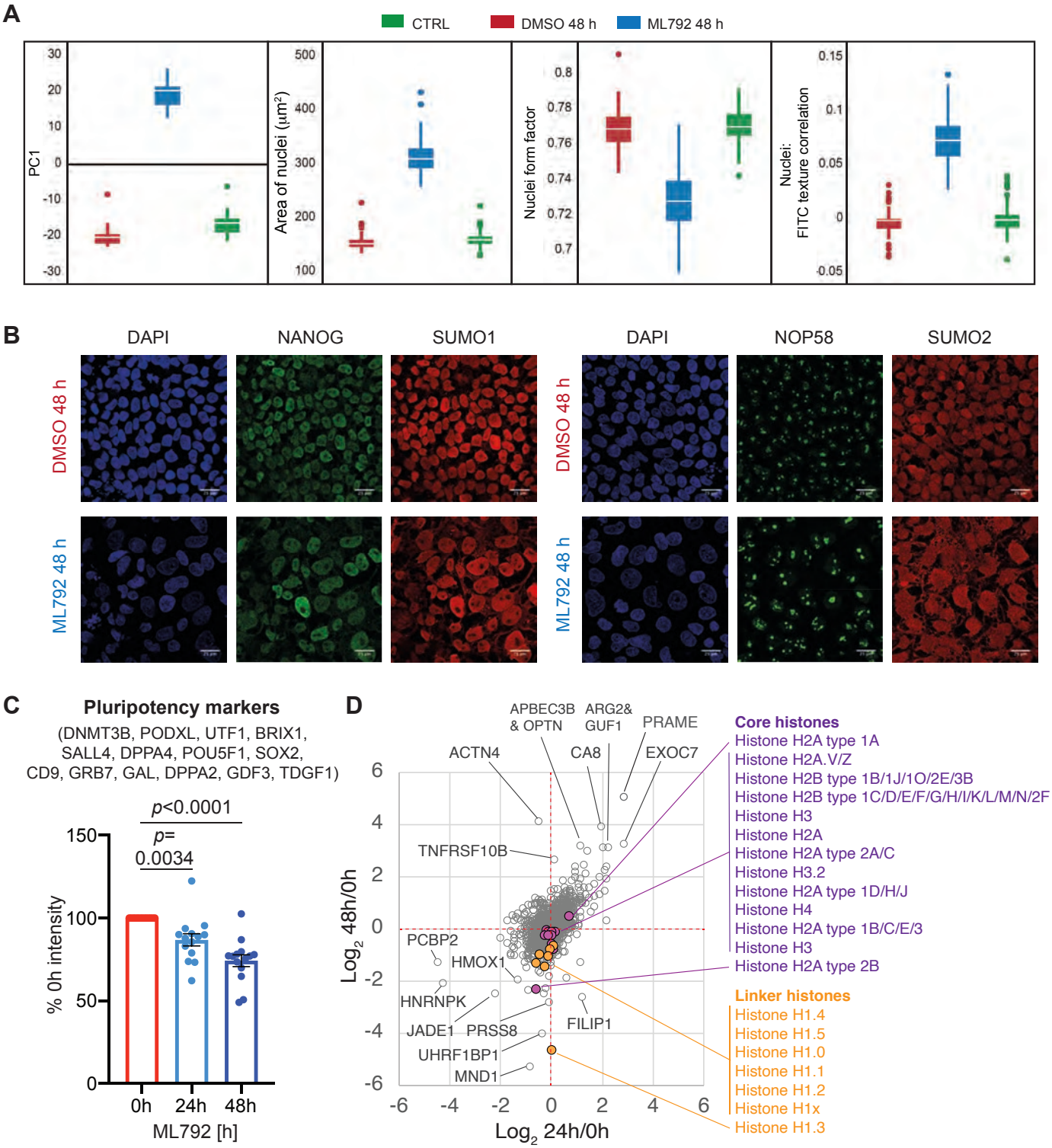
831 1. Young, R. A. (2011) Control of the embryonic stem cell state. *Cell* **144**, 940-954
832 2. Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and
833 Yamanaka, S. (2007) Induction of pluripotent stem cells from adult human fibroblasts
834 by defined factors. *Cell* **131**, 861-872
835 3. Takahashi, K., and Yamanaka, S. (2006) Induction of pluripotent stem cells from
836 mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676
837 4. Yu, J., Vodyanik, M. A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J. L., Tian, S.,
838 Nie, J., Jonsdottir, G. A., Ruotti, V., Stewart, R., Slukvin, II, and Thomson, J. A. (2007)
839 Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**,
840 1917-1920
841 5. Apostolou, E., and Hochedlinger, K. (2013) Chromatin dynamics during cellular
842 reprogramming. *Nature* **502**, 462-471
843 6. Vierbuchen, T., and Wernig, M. (2012) Molecular roadblocks for cellular
844 reprogramming. *Mol Cell* **47**, 827-838
845 7. Borkent, M., Bennett, B. D., Lackford, B., Bar-Nur, O., Brumbaugh, J., Wang, L., Du, Y.,
846 Fargo, D. C., Apostolou, E., Cheloufi, S., Maherali, N., Elledge, S. J., Hu, G., and
847 Hochedlinger, K. (2016) A Serial shRNA Screen for Roadblocks to Reprogramming
848 Identifies the Protein Modifier SUMO2. *Stem Cell Reports* **6**, 704-716
849 8. Cheloufi, S., Elling, U., Hopfgartner, B., Jung, Y. L., Murn, J., Ninova, M., Hubmann,
850 M., Badeaux, A. I., Euong Ang, C., Tenen, D., Wesche, D. J., Abazova, N., Hogue, M.,
851 Tasdemir, N., Brumbaugh, J., Rathert, P., Jude, J., Ferrari, F., Blanco, A., Fellner, M.,
852 Wenzel, D., Zinner, M., Vidal, S. E., Bell, O., Stadtfeld, M., Chang, H. Y., Almouzni, G.,
853 Lowe, S. W., Rinn, J., Wernig, M., Aravin, A., Shi, Y., Park, P. J., Penninger, J. M.,
854 Zuber, J., and Hochedlinger, K. (2015) The histone chaperone CAF-1 safeguards
855 somatic cell identity. *Nature* **528**, 218-224
856 9. Cossec, J. C., Theurillat, I., Chica, C., Bua Aguin, S., Gaume, X., Andrieux, A., Iturbide,
857 A., Jouvion, G., Li, H., Bossis, G., Seeler, J. S., Torres-Padilla, M. E., and Dejean, A.
858 (2018) SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing
859 Distinct Chromatin States. *Cell Stem Cell* **23**, 742-757 e748
860 10. Hendriks, I. A., and Vertegaal, A. C. (2016) A comprehensive compilation of SUMO
861 proteomics. *Nat Rev Mol Cell Biol* **17**, 581-595
862 11. Mahajan, R., Delphin, C., Guan, T., Gerace, L., and Melchior, F. (1997) A small
863 ubiquitin-related polypeptide involved in targeting RanGAP1 to nuclear pore
864 complex protein RanBP2. *Cell* **88**, 97-107
865 12. Hay, R. T. (2005) SUMO: a history of modification. *Mol Cell* **18**, 1-12
866 13. Flotho, A., and Melchior, F. (2013) Sumoylation: a regulatory protein modification in
867 health and disease. *Annu Rev Biochem* **82**, 357-385
868 14. Rodriguez, M. S., Dargemont, C., and Hay, R. T. (2001) SUMO-1 conjugation in vivo
869 requires both a consensus modification motif and nuclear targeting. *J Biol Chem* **276**,
870 12654-12659
871 15. Sampson, D. A., Wang, M., and Matunis, M. J. (2001) The small ubiquitin-like
872 modifier-1 (SUMO-1) consensus sequence mediates Ubc9 binding and is essential for
873 SUMO-1 modification. *J Biol Chem* **276**, 21664-21669

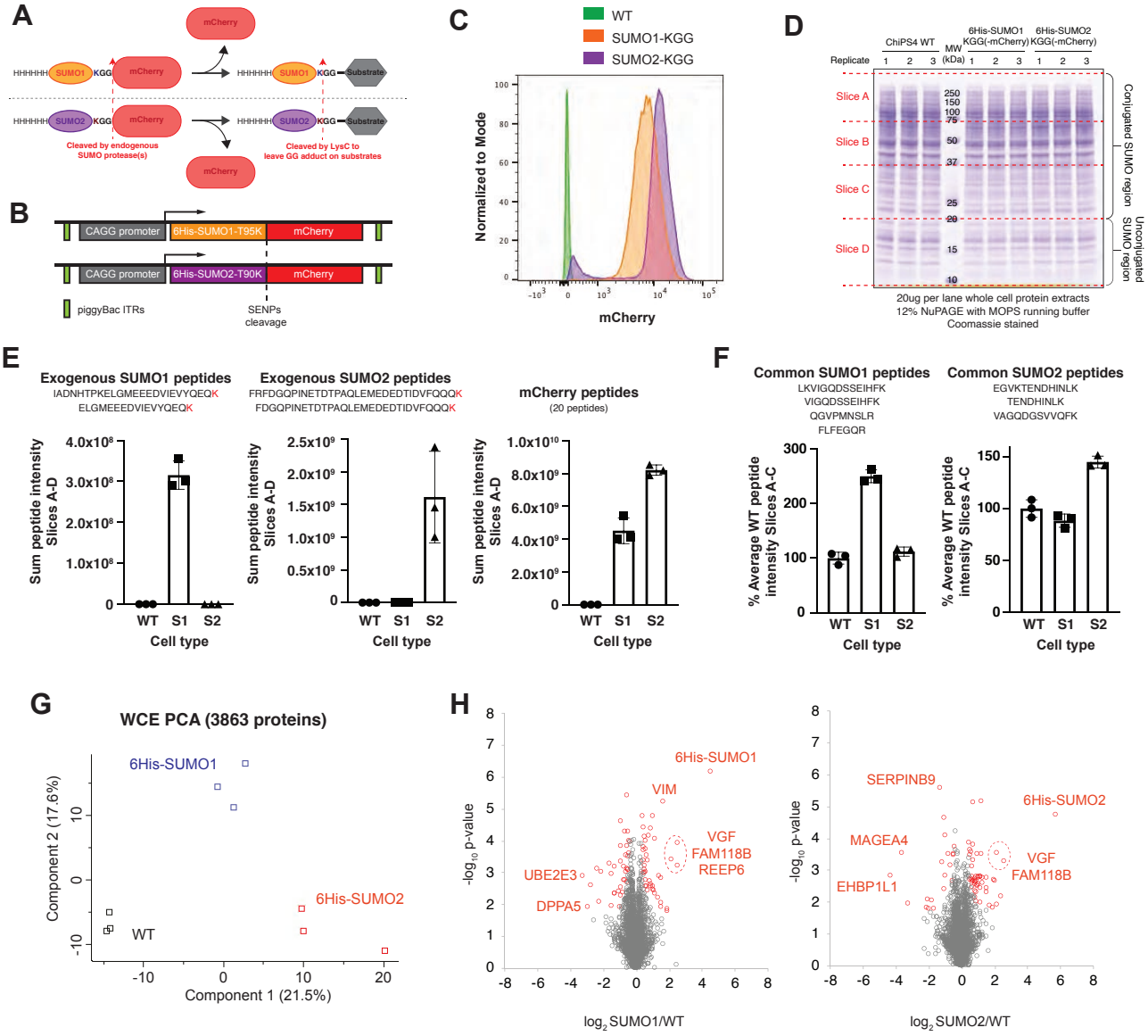
874 16. Tatham, M. H., Jaffray, E., Vaughan, O. A., Desterro, J. M., Botting, C. H., Naismith, J.
875 H., and Hay, R. T. (2001) Polymeric chains of SUMO-2 and SUMO-3 are conjugated to
876 protein substrates by SAE1/SAE2 and Ubc9. *J Biol Chem* **276**, 35368-35374
877 17. Song, J., Durrin, L. K., Wilkinson, T. A., Krontiris, T. G., and Chen, Y. (2004)
878 Identification of a SUMO-binding motif that recognizes SUMO-modified proteins.
879 *Proc Natl Acad Sci U S A* **101**, 14373-14378
880 18. Williams, R. L., Hilton, D. J., Pease, S., Willson, T. A., Stewart, C. L., Gearing, D. P.,
881 Wagner, E. F., Metcalf, D., Nicola, N. A., and Gough, N. M. (1988) Myeloid leukaemia
882 inhibitory factor maintains the developmental potential of embryonic stem cells.
883 *Nature* **336**, 684-687
884 19. Ying, Q. L., Nichols, J., Chambers, I., and Smith, A. (2003) BMP induction of Id
885 proteins suppresses differentiation and sustains embryonic stem cell self-renewal in
886 collaboration with STAT3. *Cell* **115**, 281-292
887 20. Humphrey, R. K., Beattie, G. M., Lopez, A. D., Bucay, N., King, C. C., Firpo, M. T., Rose-
888 John, S., and Hayek, A. (2004) Maintenance of pluripotency in human embryonic
889 stem cells is STAT3 independent. *Stem Cells* **22**, 522-530
890 21. Xu, C., Rosler, E., Jiang, J., Lebkowski, J. S., Gold, J. D., O'Sullivan, C., Delavan-
891 Boorsma, K., Mok, M., Bronstein, A., and Carpenter, M. K. (2005) Basic fibroblast
892 growth factor supports undifferentiated human embryonic stem cell growth without
893 conditioned medium. *Stem Cells* **23**, 315-323
894 22. Xu, R. H., Chen, X., Li, D. S., Li, R., Addicks, G. C., Glennon, C., Zwaka, T. P., and
895 Thomson, J. A. (2002) BMP4 initiates human embryonic stem cell differentiation to
896 trophoblast. *Nat Biotechnol* **20**, 1261-1264
897 23. James, D., Levine, A. J., Besser, D., and Hemmati-Brivanlou, A. (2005)
898 TGFbeta/activin/nodal signaling is necessary for the maintenance of pluripotency in
899 human embryonic stem cells. *Development* **132**, 1273-1282
900 24. Xu, R. H., Peck, R. M., Li, D. S., Feng, X., Ludwig, T., and Thomson, J. A. (2005) Basic
901 FGF and suppression of BMP signaling sustain undifferentiated proliferation of
902 human ES cells. *Nat Methods* **2**, 185-190
903 25. Rossant, J. (2015) Mouse and human blastocyst-derived stem cells: vive les
904 differences. *Development* **142**, 9-12
905 26. He, X., Riceberg, J., Soucy, T., Koenig, E., Minissale, J., Gallery, M., Bernard, H., Yang,
906 X., Liao, H., Rabino, C., Shah, P., Xega, K., Yan, Z.-h., Sintchak, M., Bradley, J., Xu, H.,
907 Duffey, M., England, D., Mizutani, H., Hu, Z., Guo, J., Chau, R., Dick, L. R., Brownell, J.
908 E., Newcomb, J., Langston, S., Lightcap, E. S., Bence, N., and Pulukuri, S. M. (2017)
909 Probing the roles of SUMOylation in cancer cell biology by using a selective SAE
910 inhibitor. *Nature Chemical Biology* **13**, 1164-1171
911 27. Tatham, M. H., Geoffroy, M. C., Shen, L., Plechanovova, A., Hattersley, N., Jaffray, E.
912 G., Palvimo, J. J., and Hay, R. T. (2008) RNF4 is a poly-SUMO-specific E3 ubiquitin
913 ligase required for arsenic-induced PML degradation. *Nat Cell Biol* **10**, 538-546
914 28. Hands, K. J., Cuchet-Lourenco, D., Everett, R. D., and Hay, R. T. (2014) PML isoforms
915 in response to arsenic: high-resolution analysis of PML body structure and
916 degradation. *J Cell Sci* **127**, 365-375
917 29. Ludwig, T. E., Levenstein, M. E., Jones, J. M., Berggren, W. T., Mitchen, E. R., Frane, J.
918 L., Crandall, L. J., Daigh, C. A., Conard, K. R., Piekarczyk, M. S., Llanas, R. A., and
919 Thomson, J. A. (2006) Derivation of human embryonic stem cells in defined
920 conditions. *Nat Biotechnol* **24**, 185-187

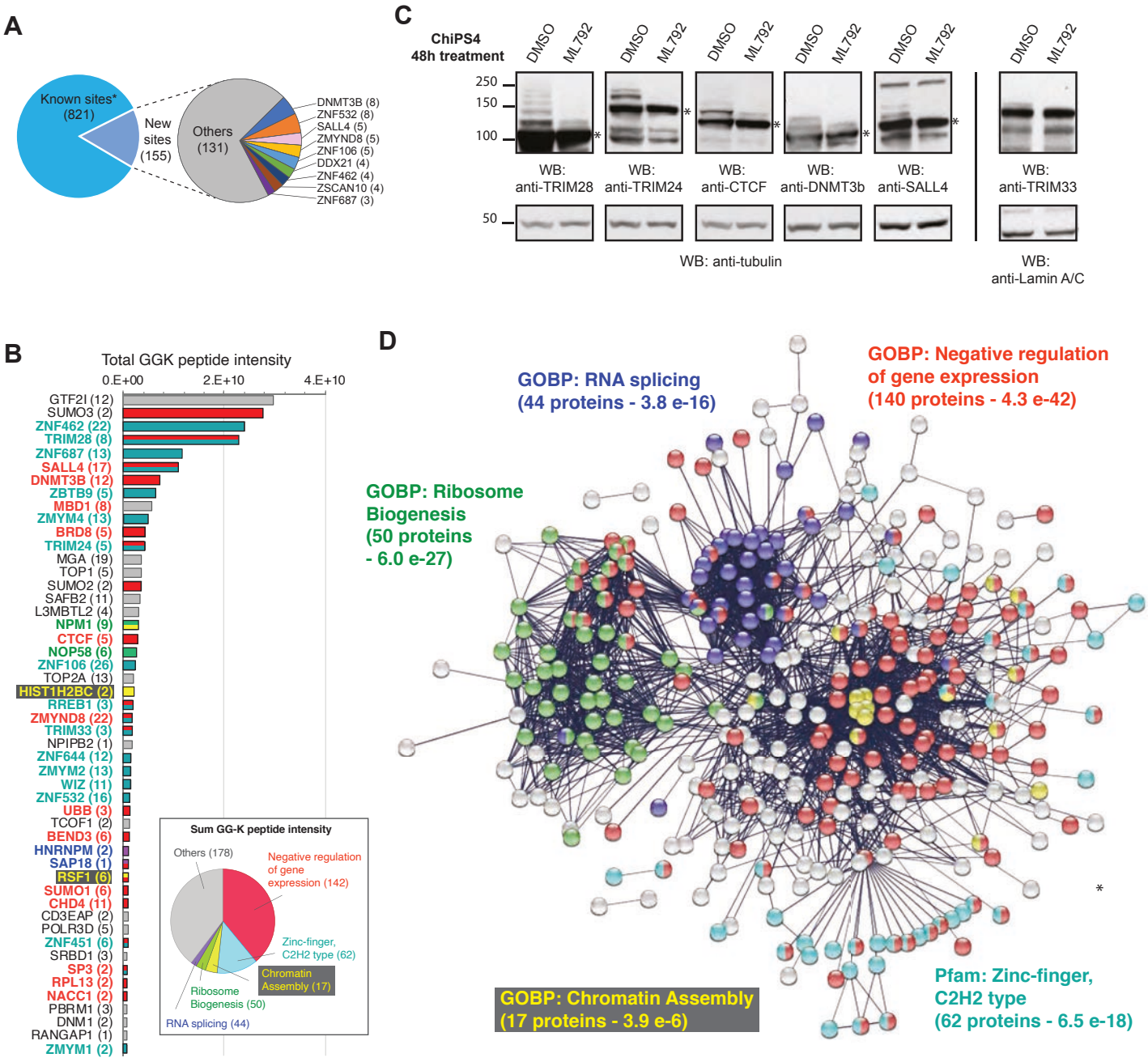
921 30. Bray, M. A., Singh, S., Han, H., Davis, C. T., Borgeson, B., Hartland, C., Kost-Alimova,
922 M., Gustafsdottir, S. M., Gibson, C. C., and Carpenter, A. E. (2016) Cell Painting, a
923 high-content image-based assay for morphological profiling using multiplexed
924 fluorescent dyes. *Nat Protoc* **11**, 1757-1774
925 31. Shevchenko, A., Tomas, H., Havlis, J., Olsen, J. V., and Mann, M. (2006) In-gel
926 digestion for mass spectrometric characterization of proteins and proteomes. *Nat*
927 *Protoc* **1**, 2856-2860
928 32. Cox, J., and Mann, M. (2008) MaxQuant enables high peptide identification rates,
929 individualized p.p.b.-range mass accuracies and proteome-wide protein
930 quantification. *Nat Biotechnol* **26**, 1367-1372
931 33. Tyanova, S., Temu, T., Sinitcyn, P., Carlson, A., Hein, M. Y., Geiger, T., Mann, M., and
932 Cox, J. (2016) The Perseus computational platform for comprehensive analysis of
933 (prote)omics data. *Nat Methods* **13**, 731-740
934 34. Perez-Riverol, Y., Csordas, A., Bai, J., Bernal-Llinares, M., Hewapathirana, S., Kundu,
935 D. J., Inuganti, A., Griss, J., Mayer, G., Eisenacher, M., Perez, E., Uszkoreit, J., Pfeuffer,
936 J., Sachsenberg, T., Yilmaz, S., Tiwary, S., Cox, J., Audain, E., Walzer, M., Jarnuczak, A.
937 F., Ternent, T., Brazma, A., and Vizcaino, J. A. (2019) The PRIDE database and related
938 tools and resources in 2019: improving support for quantification data. *Nucleic Acids*
939 *Res* **47**, D442-D450
940 35. Tammsalu, T., Matic, I., Jaffray, E. G., Ibrahim, A. F., Tatham, M. H., and Hay, R. T.
941 (2015) Proteome-wide identification of SUMO modification sites by mass
942 spectrometry. *Nat Protoc* **10**, 1374-1388
943 36. Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J.,
944 Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., and Mering, C. V.
945 (2019) STRING v11: protein-protein association networks with increased coverage,
946 supporting functional discovery in genome-wide experimental datasets. *Nucleic*
947 *Acids Res* **47**, D607-D613
948 37. Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N.,
949 Schwikowski, B., and Ideker, T. (2003) Cytoscape: a software environment for
950 integrated models of biomolecular interaction networks. *Genome Res* **13**, 2498-2504
951 38. Tammsalu, T., Matic, I., Jaffray, E. G., Ibrahim, A. F. M., Tatham, M. H., and Hay, R. T.
952 (2014) Proteome-wide identification of SUMO2 modification sites. *Sci Signal* **7**, rs2
953 39. Hendriks, I. A., Lyon, D., Young, C., Jensen, L. J., Vertegaal, A. C., and Nielsen, M. L.
954 (2017) Site-specific mapping of the human SUMO proteome reveals co-modification
955 with phosphorylation. *Nat Struct Mol Biol* **24**, 325-336
956 40. Wolf, D., and Goff, S. P. (2007) TRIM28 Mediates Primer Binding Site-Targeted
957 Silencing of Murine Leukemia Virus in Embryonic Cells. *Cell* **131**, 46-57
958 41. Gerstein, M. B., Kundaje, A., Hariharan, M., Landt, S. G., Yan, K. K., Cheng, C., Mu, X.
959 J., Khurana, E., Rozowsky, J., Alexander, R., Min, R., Alves, P., Abyzov, A., Addleman,
960 N., Bhardwaj, N., Boyle, A. P., Cayting, P., Charos, A., Chen, D. Z., Cheng, Y., Clarke,
961 D., Eastman, C., Euskirchen, G., Frietze, S., Fu, Y., Gertz, J., Grubert, F., Harmanci, A.,
962 Jain, P., Kasowski, M., Lacroute, P., Leng, J. J., Lian, J., Monahan, H., O'Geen, H.,
963 Ouyang, Z., Partridge, E. C., Patacsil, D., Pauli, F., Raha, D., Ramirez, L., Reddy, T. E.,
964 Reed, B., Shi, M., Slifer, T., Wang, J., Wu, L., Yang, X., Yip, K. Y., Zilberman-Schapira,
965 G., Batzoglou, S., Sidow, A., Farnham, P. J., Myers, R. M., Weissman, S. M., and
966 Snyder, M. (2012) Architecture of the human regulatory network derived from
967 ENCODE data. *Nature* **489**, 91-100

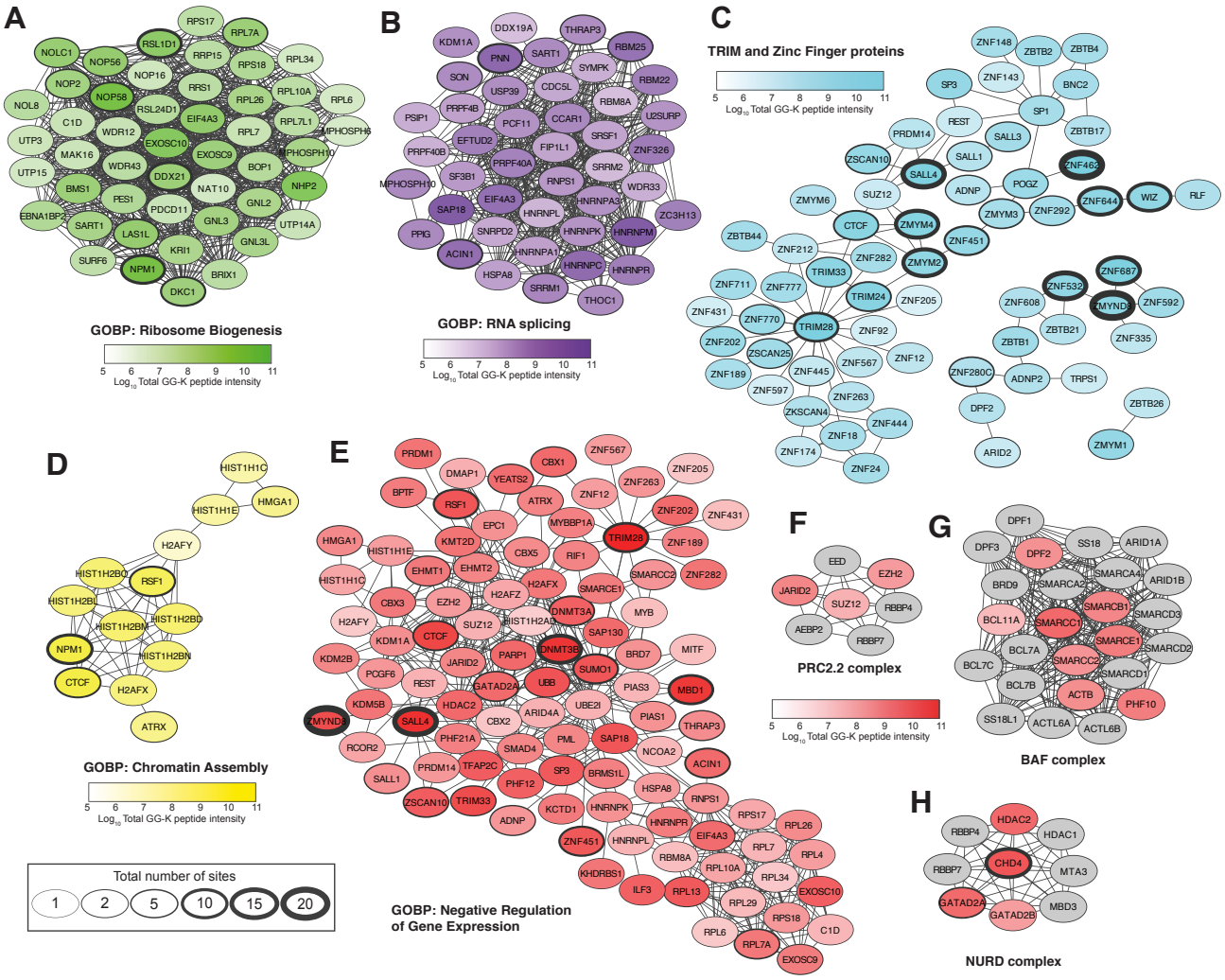
968 42. Yang, B. X., El Farran, C. A., Guo, H. C., Yu, T., Fang, H. T., Wang, H. F., Schlesinger, S.,
969 Seah, Y. F., Goh, G. Y., Neo, S. P., Li, Y., Lorincz, M. C., Tergaonkar, V., Lim, T. M.,
970 Chen, L., Gunaratne, J., Collins, J. J., Goff, S. P., Daley, G. Q., Li, H., Bard, F. A., and
971 Loh, Y. H. (2015) Systematic identification of factors for provirus silencing in
972 embryonic stem cells. *Cell* **163**, 230-245
973 43. Schmidt, N., Domingues, P., Golebiowski, F., Patzina, C., Tatham, M. H., Hay, R. T.,
974 and Hale, B. G. (2019) An influenza virus-triggered SUMO switch orchestrates co-
975 opted endogenous retroviruses to stimulate host antiviral immunity. *Proceedings of*
976 *the National Academy of Sciences* **116**, 17399-17408
977 44. Tie, C. H., Fernandes, L., Conde, L., Robbez-Masson, L., Sumner, R. P., Peacock, T.,
978 Rodriguez-Plata, M. T., Mickute, G., Gifford, R., Towers, G. J., Herrero, J., and Rowe,
979 H. M. (2018) KAP1 regulates endogenous retroviruses in adult human cells and
980 contributes to innate immune control. *EMBO Rep* **19**
981 45. Oleksiewicz, U., Gladych, M., Raman, A. T., Heyn, H., Mereu, E., Chlebanowska, P.,
982 Andrzejewska, A., Sozanska, B., Samant, N., Fak, K., Auguscik, P., Kosinski, M.,
983 Wroblewska, J. P., Tomczak, K., Kulcenty, K., Ploski, R., Biecek, P., Esteller, M., Shah,
984 P. K., Rai, K., and Wiznerowicz, M. (2017) TRIM28 and Interacting KRAB-ZNFs Control
985 Self-Renewal of Human Pluripotent Stem Cells through Epigenetic Repression of Pro-
986 differentiation Genes. *Stem Cell Reports* **9**, 2065-2080
987 46. Neyret-Kahn, H., Benhamed, M., Ye, T., Le Gras, S., Cossec, J. C., Lapaquette, P.,
988 Bischof, O., Ouspenskaia, M., Dasso, M., Seeler, J., Davidson, I., and Dejean, A. (2013)
989 Sumoylation at chromatin governs coordinated repression of a transcriptional
990 program essential for cell growth and proliferation. *Genome Res* **23**, 1563-1579
991 47. Niskanen, E. A., Malinen, M., Sutinen, P., Toropainen, S., Paakinaho, V., Vihervaara,
992 A., Joutsen, J., Kaikkonen, M. U., Sistonen, L., and Palvimo, J. J. (2015) Global
993 SUMOylation on active chromatin is an acute heat stress response restricting
994 transcription. *Genome Biol* **16**, 153
995 48. Seifert, A., Schofield, P., Barton, G. J., and Hay, R. T. (2015) Proteotoxic stress
996 reprograms the chromatin landscape of SUMO modification. *Sci Signal* **8**, rs7
997 49. Liu, H. W., Zhang, J., Heine, G. F., Arora, M., Gulcin Ozer, H., Onti-Srinivasan, R.,
998 Huang, K., and Parvin, J. D. (2012) Chromatin modification by SUMO-1 stimulates the
999 promoters of translation machinery genes. *Nucleic Acids Res* **40**, 10172-10186
1000 50. Theurillat, I., Hendriks, I. A., Cossec, J. C., Andrieux, A., Nielsen, M. L., and Dejean, A.
1001 (2020) Extensive SUMO Modification of Repressive Chromatin Factors Distinguishes
1002 Pluripotent from Somatic Cells. *Cell Rep* **32**, 108146
1003 51. Gareau, J. R., Reverter, D., and Lima, C. D. (2012) Determinants of small ubiquitin-like
1004 modifier 1 (SUMO1) protein specificity, E3 ligase, and SUMO-RanGAP1 binding
1005 activities of nucleoporin RanBP2. *J Biol Chem* **287**, 4740-4751
1006 52. O'Shea, J. P., Chou, M. F., Quader, S. A., Ryan, J. K., Church, G. M., and Schwartz, D.
1007 (2013) pLogo: a probabilistic approach to visualizing sequence motifs. *Nat Methods*
1008 **10**, 1211-1212
1009 53. Kozlowski, L. P. (2016) IPC - Isoelectric Point Calculator. *Biol Direct* **11**, 55
1010

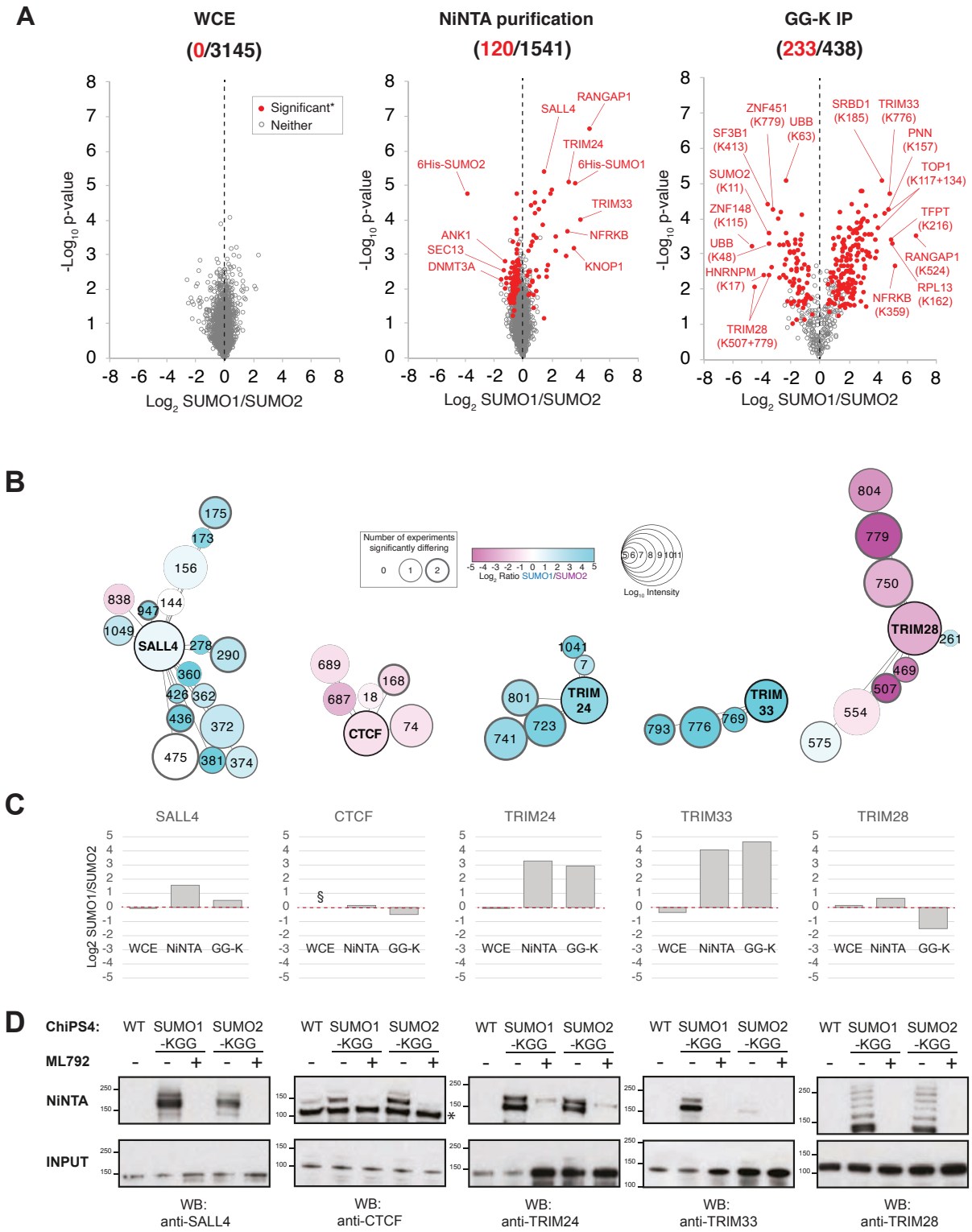

A
Day 0 CTRL
ML792 400 nM
DMSO
Day 1
Day 2
B
DMSO
ML792
Cell density [cm-2]
1.5x10^5
1.0x10^5
0.5x10^5
0
0
24
48
Time [h]
*
D
ML792 [h]
0 24 48
ML792 [h]
0 24 48
250
150
100
75
50
37
25
20
15
SUMO1
SUMO2
NANOG
37
SOX2
OCT4
37
50
Actin
C
24h
48h
DMSO
ML792
Normalized to mode
100
80
60
40
20
0
0
100k
200k
PI-A
E
ML792 0h
ML792 24h
ML792 48h
Relative mRNA expression
1.5
1.0
0.5
0.0
**
****
ns
ns
***
****
ns
ns
NANOG
OCT4
KLF4
SOX2

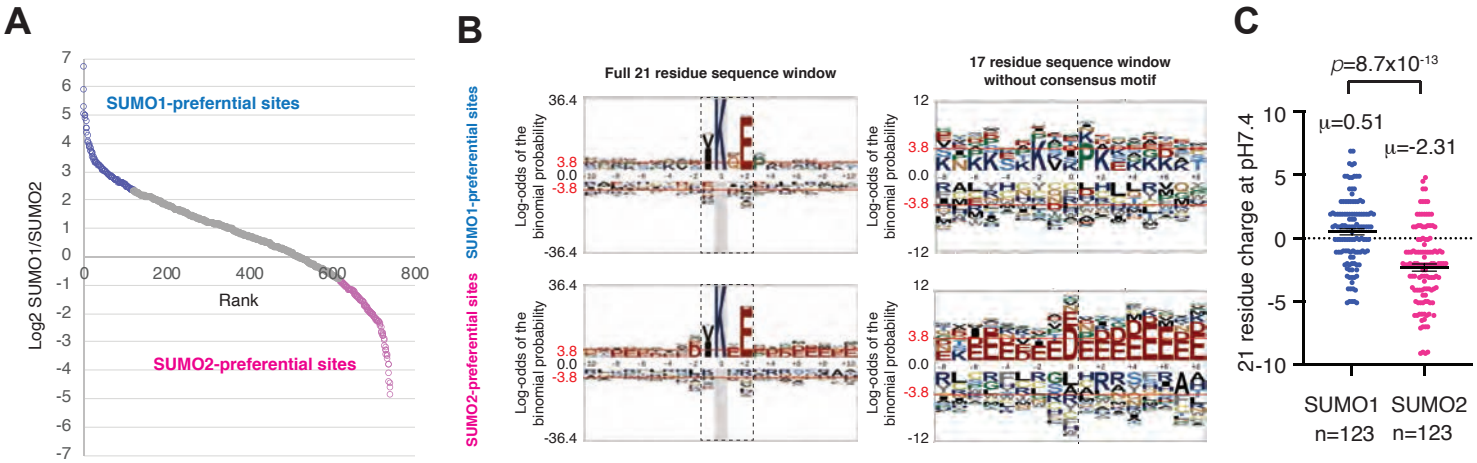

A
CTRL
DMSO 48 h
ML792 48 h
PC1
Area of nuclei (µm²)
Nuclei form factor
Nuclei: FITC texture correlation
B
DAPI
NANOG
SUMO1
DAPI
NOP58
SUMO2
DMSO 48 h
ML792 48 h
C
Pluripotency markers
(DNMT3B, PODXL, UTF1, BRIX1, SALL4, DPPA4, POU5F1, SOX2, CD9, GRB7, GAL, DPPA2, GDF3, TDGF1)
p<0.0001
p= 0.0034
% 0h intensity
0h
24h
48h
ML792 [h]
D
APBEC3B & OPTN
ARG2& GUF1
PRAME
ACTN4
CA8
EXOC7
TNFRSF10B
PCBP2
HMOX1
HNRNPK
JADE1
PRSS8
FILIP1
UHRF1BP1
MND1
Log2 48h/0h
Log2 24h/0h
Core histones
Histone H2A type 1A
Histone H2A.V/Z
Histone H2B type 1B/1J/1O/2E/3B
Histone H2B type 1C/D/E/F/G/H/I/K/L/M/N/2F
Histone H3
Histone H2A
Histone H2A type 2A/C
Histone H3.2
Histone H2A type 1D/H/J
Histone H4
Histone H2A type 1B/C/E/3
Histone H3
Histone H2A type 2B
Linker histones
Histone H1.4
Histone H1.5
Histone H1.0
Histone H1.1
Histone H1.2
Histone H1x
Histone H1.3


A
mCherry
HHHHHH SUMO1 KGG mCherry
HHHHHH SUMO1 KGG Substrate
HHHHHH SUMO2 KGG mCherry
HHHHHH SUMO2 KGG Substrate
Cleaved by endogenous SUMO protease(s)
Cleaved by LysC to leave GG adduct on substrates
B
CAGG promoter
6His-SUMO1-T95K
mCherry
CAGG promoter
6His-SUMO2-T90K
mCherry
piggyBac ITRs
SENPs cleavage
C
WT
SUMO1-KGG
SUMO2-KGG
Normalized to Mode
mCherry
D
ChiPS4 WT
6His-SUMO1 KGG(-mCherry)
6His-SUMO2 KGG(-mCherry)
Replicate
MW (kDa)
Slice A
Slice B
Slice C
Slice D
Conjugated SUMO region
Unconjugated SUMO region
20ug per lane whole cell protein extracts
12% NuPAGE with MOPS running buffer
Coomassie stained
E
Exogenous SUMO1 peptides
IADNHTPKELGMEEEDVIEVYQEQK
ELGMEEEDVIEVYQEQK
Exogenous SUMO2 peptides
FRFDGQPINETDTPAQLEMEDEDTIDVFQQQK
FDGQPINETDTPAQLEMEDEDTIDVFQQQK
mCherry peptides
(20 peptides)
Sum peptide intensity Slices A-D
Cell type
WT S1 S2
F
Common SUMO1 peptides
LKVIGQDSSEIHFK
VIGQDSSEIHFK
QGVPMNSLR
FLFEGQR
Common SUMO2 peptides
EGVKTENDHINLK
TENDHINLK
VAGQDGSVVQFK
% Average WT peptide intensity Slices A-C
Cell type
WT S1 S2
G
WCE PCA (3863 proteins)
6His-SUMO1
6His-SUMO2
WT
Component 1 (21.5%)
Component 2 (17.6%)
H
6His-SUMO1
VIM
VGF
FAM118B
REEP6
UBE2E3
DPPA5
-log10 p-value
log2 SUMO1/WT
SERPINB9
6His-SUMO2
MAGEA4
VGF
FAM118B
EHBP1L1
log2 SUMO2/WT

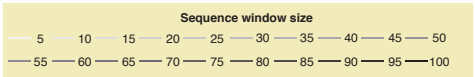

A
Known sites*
(821)
New
sites
(155)
Others
(131)
DNMT3B (8)
ZNF532 (8)
SALL4 (5)
ZMYND8 (5)
ZNF106 (5)
DDX21 (4)
ZNF462 (4)
ZSCAN10 (4)
ZNF687 (3)
C
ChiPS4
48h treatment
DMSO
ML792
250
150
100
50
WB:
anti-TRIM28
WB:
anti-TRIM24
WB:
anti-CTCF
WB:
anti-DNMT3b
WB:
anti-SALL4
WB:
anti-TRIM33
WB: anti-tubulin
WB:
anti-Lamin A/C
B
Total GGK peptide intensity
0.E+00
2.E+10
4.E+10
GTF2I (12)
SUMO3 (2)
ZNF462 (22)
TRIM28 (8)
ZNF687 (13)
SALL4 (17)
DNMT3B (12)
ZBTB9 (5)
MBD1 (8)
ZMYM4 (13)
BRD8 (5)
TRIM24 (5)
MGA (19)
TOP1 (5)
SUMO2 (2)
SAFB2 (11)
L3MBTL2 (4)
NPM1 (9)
CTCF (5)
NOP58 (6)
ZNF106 (26)
TOP2A (13)
HIST1H2BC (2)
RREB1 (3)
ZMYND8 (22)
TRIM33 (3)
NPIPB2 (1)
ZNF644 (12)
ZMYM2 (13)
WIZ (11)
ZNF532 (16)
UBB (3)
TCOF1 (2)
BEND3 (6)
HNRNPM (2)
SAP18 (1)
RSF1 (6)
SUMO1 (6)
CHD4 (11)
CD3EAP (2)
POLR3D (5)
ZNF451 (6)
SRBD1 (3)
SP3 (2)
RPL13 (2)
NACC1 (2)
PBRM1 (3)
DNM1 (2)
RANGAP1 (1)
ZMYM1 (2)
Sum GG-K peptide intensity
Others (178)
Negative regulation
of gene
expression (142)
Zinc-finger,
C2H2 type (62)
Chromatin
Assembly (17)
Ribosome
Biogenesis (50)
RNA splicing (44)
D
GOBP: RNA splicing
(44 proteins - 3.8 e-16)
GOBP: Negative regulation
of gene expression
(140 proteins - 4.3 e-42)
GOBP: Ribosome
Biogenesis
(50 proteins
- 6.0 e-27)
GOBP: Chromatin Assembly
(17 proteins - 3.9 e-6)
Pfam: Zinc-finger,
C2H2 type
(62 proteins - 6.5 e-18)


A
GOBP: Ribosome Biogenesis
Log10 Total GG-K peptide intensity
B
GOBP: RNA splicing
Log10 Total GG-K peptide intensity
C
TRIM and Zinc Finger proteins
Log10 Total GG-K peptide intensity
D
GOBP: Chromatin Assembly
Log10 Total GG-K peptide intensity
E
GOBP: Negative Regulation of Gene Expression
F
PRC2.2 complex
Log10 Total GG-K peptide intensity
G
BAF complex
H
NURD complex
Total number of sites
1
2
5
10
15
20

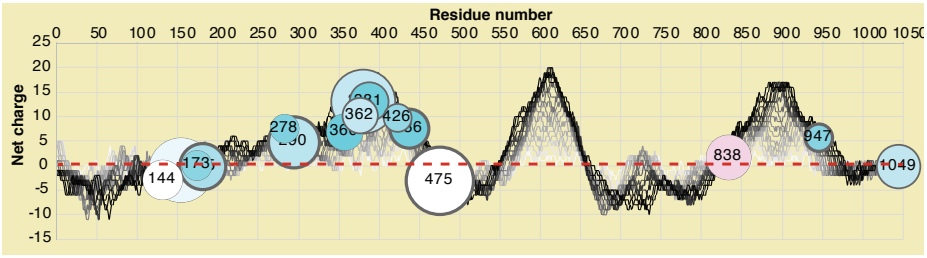

A
WCE
(0/3145)
NiNTA purification
(120/1541)
GG-K IP
(233/438)
Significant*
Neither
-Log10 p-value
Log2 SUMO1/SUMO2
6His-SUMO2
ANK1
SEC13
DNMT3A
SALL4
RANGAP1
TRIM24
6His-SUMO1
TRIM33
NFRKB
KNOP1
ZNF451 (K779)
UBB (K63)
SF3B1 (K413)
SUMO2 (K11)
ZNF148 (K115)
UBB (K48)
HNRNPM (K17)
TRIM28 (K507+779)
SRBD1 (K185)
TRIM33 (K776)
PNN (K157)
TOP1 (K117+134)
TFPT (K216)
RANGAP1 (K524)
RPL13 (K162)
NFRKB (K359)
B
Number of experiments significantly differing
Log2 Ratio SUMO1/SUMO2
Log10 Intensity
SALL4
CTCF
TRIM 24
TRIM 33
TRIM28
C
SALL4
CTCF
TRIM24
TRIM33
TRIM28
Log2 SUMO1/SUMO2
WCE
NiNTA
GG-K
D
ChiPS4:
WT
SUMO1 -KGG
SUMO2 -KGG
ML792
NiNTA
INPUT
WB: anti-SALL4
WB: anti-CTCF
WB: anti-TRIM24
WB: anti-TRIM33
WB: anti-TRIM28



D

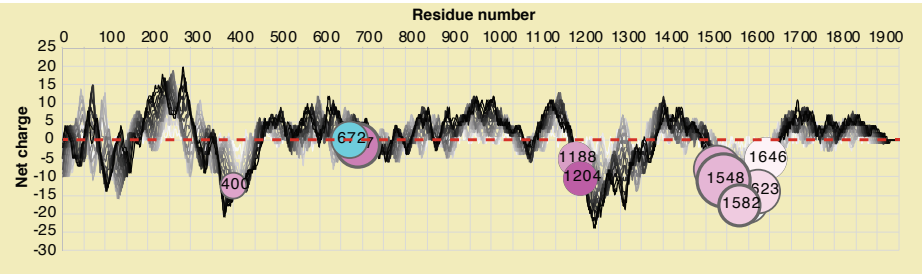


SALL4



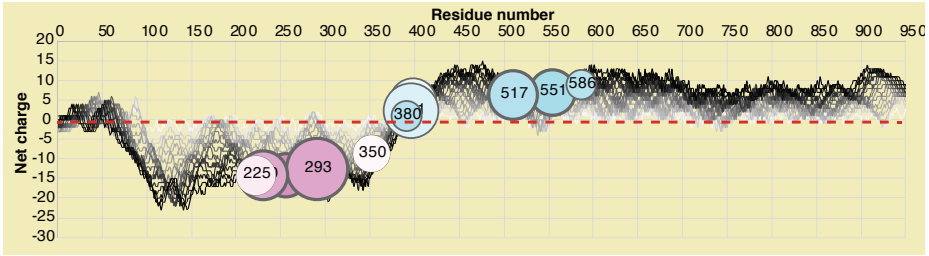
SUMOylation site	Window
SALL4-K0144	RENGGSSEDMKEKPDAESVVY
SALL4-K0156	KPDAESVVYLKTETALPPTPQ
SALL4-K0173	PTPQDISYLAKGKVANTNVTL
SALL4-K0175	PQDISYLAKGKVANTNVTLQA
SALL4-K0278	VSAAVALLSQKAGSQGLSLDA
SALL4-K0290	GSQGLSLDALKQAKLPHANIP
SALL4-K0360	TVALDTSKKGKGKPPNISAVD
SALL4-K0362	ALDTSKKGKGKPPNISAVDVK
SALL4-K0372	KPPNISAVDVKPKDEAALYKH
SALL4-K0374	PNISAVDVKPKDEAALYKHKC
SALL4-K0381	VKPKDEAALYKHKCKYCSKVF
SALL4-K0426	GHRFTTKGNLKVHFHRHPQVK
SALL4-K0436	KVHFHRHPQVKANPQLFAEFQ
SALL4-K0475	IDEPSLSLDSKPVLVTTSVGL
SALL4-K0838	TFIRAPPTYVKVEVPGTFVGP
SALL4-K0947	NTMALLGTDGKRVSEIFPKEI
SALL4-K1049	HQFPHFLEENKIAVS_____

CHD4



SUMOylation site	Window
CHD4-K0400	CEKEGIQWEAKEDNSEGEEIL
CHD4-K0672	GKKLKKVKLRKLERPPETPTV
CHD4-K0687	PETPTVDPTVKYERQPEYLDA
CHD4-K1188	GLGSKTGSMSKQELDDILKFG
CHD4-K1204	ILKFGTEELFKQEATDGGGDN
CHD4-K1541	APVPPAEDGIKIEENSLKEEE
CHD4-K1548	DGIKIEENSLKEEESIEGEKE
CHD4-K1582	QAPAPASEDEKVVVEPPEGEE
CHD4-K1600	GEEKVEKAEVKERTEEPMETE
CHD4-K1623	GAADVEKVEEKSAIDLTPIVV
CHD4-K1646	KEEKKEEEEKKEVMLQNGETP

SAFB



SUMOylation site	Window
SAFB2-K0225	KILDILGETCKSEPVKEESSE
SAFB2-K0230	LGETCKSEPVKEESSELEQPF
SAFB2-K0252	QDTSSVGPDRKLAEEEDLFDS
SAFB2-K0293	SKADSLLAVVKREPAEQPGDG
SAFB2-K0350	EAPSPEARDSKEDGRKFDFDA
SAFB2-K0380	ESSTSEGADQKMSSFKEEKDI
SAFB2-K0391	MSSFKEEKDIKPIIKDEKGRV
SAFB2-K0395	KEEKDIKPIIKDEKGRVGSGS
SAFB2-K0517	SVDRHHSVEIKIEKTVIKKEE
SAFB2-K0551	KKEEKDQDELKPGPTNRSRVT
SAFB2-K0586	KSKGEPVISVKTTSRSKERSS