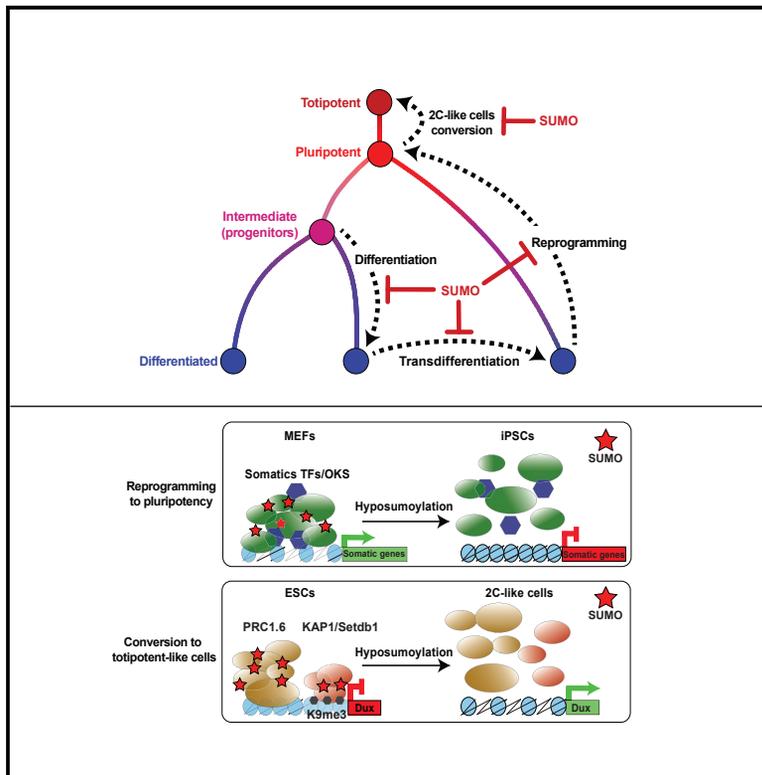


Cell Stem Cell

SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States

Graphical Abstract



Authors

Jack-Christophe Cossec,
Ilan Theurillat, Claudia Chica, ...,
Jacob-Sebastian Seeler,
Maria-Elena Torres-Padilla,
Anne Dejean

Correspondence

anne.dejean@pasteur.fr

In Brief

SUMO functions as a tether on distinct chromatin types in MEFs and ESCs to stabilize occupancy of protein substrate complexes on key cell-identity genes, thus preserving the somatic and pluripotent states.

Highlights

- SUMO acts on chromatin to maintain cellular identity
- SUMO impairs somatic enhancer inactivation early during iPSC reprogramming
- Loss of SUMO converts ESCs into a 2C-like state by releasing PRC1.6 from the Dux locus
- Loss of SUMO in ESCs leads to genome-wide loss of H3K9me3-dependent heterochromatin

SUMO Safeguards Somatic and Pluripotent Cell Identities by Enforcing Distinct Chromatin States

Jack-Christophe Cossec,^{1,2,11} Ilan Theurillat,^{1,2,3,11} Claudia Chica,^{4,12} Sabela Búa Aguín,^{5,6,12} Xavier Gaume,^{7,10} Alexandra Andrieux,^{1,2} Ane Iturbide,⁷ Gregory Jouvion,⁸ Han Li,^{5,6} Guillaume Bossis,⁹ Jacob-Sebastian Seeler,^{1,2} Maria-Elena Torres-Padilla,⁷ and Anne Dejean^{1,2,13,*}

¹Nuclear Organization and Oncogenesis Unit, Equipe Labellisée Ligue Nationale Contre le Cancer, Institut Pasteur, 75015 Paris, France

²INSERM, U993, 75015 Paris, France

³Sorbonne Université, Collège Doctoral, 75005 Paris, France

⁴Bioinformatics and Biostatistics Hub – C3BI, USR 3756 Institut Pasteur & CNRS, 75015 Paris, France

⁵Cellular Plasticity and Disease Modelling Unit, Institut Pasteur, 75015 Paris, France

⁶CNRS UMR3738, 75015 Paris, France

⁷Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München, München, Germany

⁸Experimental Neuropathology Unit, Institut Pasteur, 75015 Paris, France

⁹Institut de Génétique Moléculaire de Montpellier, University of Montpellier, CNRS, Montpellier, France

¹⁰Present address: Cellular Reprogramming and Oncogenesis Unit, CRCL, 69008 Lyon, France

¹¹These authors contributed equally

¹²These authors contributed equally

¹³Lead Contact

*Correspondence: anne.dejean@pasteur.fr

<https://doi.org/10.1016/j.stem.2018.10.001>

SUMMARY

Understanding general principles that safeguard cellular identity should reveal critical insights into common mechanisms underlying specification of varied cell types. Here, we show that SUMO modification acts to stabilize cell fate in a variety of contexts. Hyposumoylation enhances pluripotency reprogramming *in vitro* and *in vivo*, increases lineage transdifferentiation, and facilitates leukemic cell differentiation. Suppressing sumoylation in embryonic stem cells (ESCs) promotes their conversion into 2-cell-embryo-like (2C-like) cells. During reprogramming to pluripotency, SUMO functions on fibroblastic enhancers to retain somatic transcription factors together with Oct4, Sox2, and Klf4, thus impeding somatic enhancer inactivation. In contrast, in ESCs, SUMO functions on heterochromatin to silence the 2C program, maintaining both proper H3K9me3 levels genome-wide and repression of the Dux locus by triggering recruitment of the sumoylated PRC1.6 and Kap/Setdb1 repressive complexes. Together, these studies show that SUMO acts on chromatin as a glue to stabilize key determinants of somatic and pluripotent states.

INTRODUCTION

Post-translational modification by SUMO is an essential regulatory mechanism of protein function that is involved in most chal-

lenges faced by eukaryotic cells (Flotho and Melchior, 2013; Seeler and Dejean, 2017). Sumoylation is a reversible process composed of an E1 (Aos1/Uba2), an E2 (Ubc9), several E3 enzymes, and desumoylases such as SENPs. There are three functional paralogs of SUMO in mammals, SUMO-1, SUMO-2, and SUMO-3, of which the latter two (collectively termed SUMO-2) are nearly identical. Like ubiquitin, SUMO-2—but not SUMO-1—can form polymeric chains on numerous substrates. In many instances, sumoylation of proteins was found to mediate novel interactions with other proteins containing SUMO interaction motifs (SIMs). Among the nearly 7,000 different SUMO substrates identified so far, transcription factors (TFs) and chromatin proteins represent the largest class of targets (Cubeñas-Potts and Matunis, 2013). Recent genome-wide studies have revealed that SUMO functions as an instructive chromatin-associated mark that plays a pivotal role in the transcriptional response to external cues (Liu et al., 2012; Neyret-Kahn et al., 2013). Notably stresses like heat shock (Niskanen et al., 2015; Seifert et al., 2015), inflammation (Decque et al., 2016), or oncogene-induced senescence (Neyret-Kahn et al., 2013) profoundly alter the SUMO chromatin landscape, driving important transcriptional changes.

The hierarchical interplay between TFs, chromatin regulators and coordinated chromatin states plays a key role in maintaining cellular identity (Apostolou and Hochedlinger, 2013; Vierbuchen and Wernig, 2012). Whereas numerous studies have led to remarkable insight into the roles of DNA and histone modifications in both reprogramming and transdifferentiation models, the contribution of post-translational modifications of proteins by other proteins in these processes remains largely unknown. In this study, we exploited various *in vitro* and *in vivo* cell-fate conversion systems to investigate the role of sumoylation in cellular plasticity. We demonstrate that SUMO functions as a key regulator of cell fate and that it occupies highly distinctive

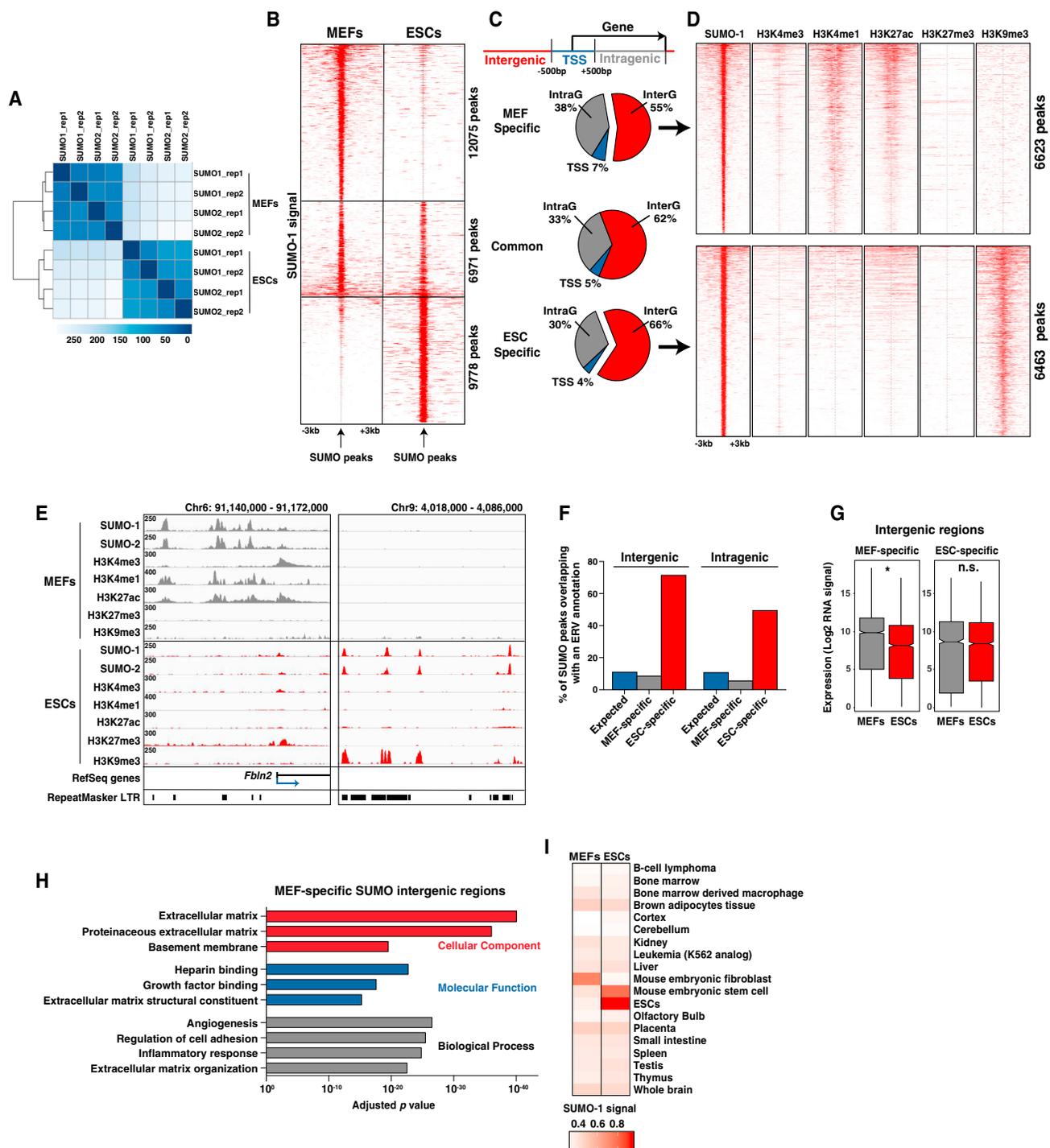


Figure 1. Extensive Divergence of the SUMO Chromatin Landscapes between MEFs and ESCs

(A) Complete-linkage clustering of SUMO-1 and SUMO-2 ChIP-seq. Rep1, replicate1; Rep2, replicate2; r , 0.91 in MEFs, r , 0.88 in ESCs, Spearman correlation test.

(B) Spatial distribution of SUMO-1 ChIP-seq signal around the SUMO peaks in MEFs and ESCs.

(C) Proportion of SUMO peaks assigned to TSSs, intergenic (InterG), and intragenic (IntraG) regions as defined in the schema above.

(D) Association of SUMO-1 with histone marks in intergenic regions. Comparison of tag densities around the MEF-specific (upper panels) and ESC-specific (lower panels) SUMO peaks.

(E) Screen shots of SUMO-1, SUMO-2, and histone marks at the *Fbn2* locus (left) and at an ERV-rich region (right) in MEFs and ESCs.

(F) Proportion of SUMO peaks overlapping with an ERV annotation.

(legend continued on next page)

chromatin types in somatic and pluripotent cells. Mechanistically, SUMO operates, in large part, from active enhancers in mouse embryonic fibroblasts (MEFs) and heterochromatin in embryonic stem cells (ESCs) to resist TF-induced reprogramming to pluripotency and spontaneous conversion to totipotent-like states, respectively. On chromatin, SUMO acts as a molecular stabilizer of critical protein group substrates required for maintenance of cell identity, suggesting that manipulating sumoylation could be an effective way to modulate cell fate.

RESULTS

SUMO Chromatin Landscapes Strongly Differ between MEFs and ESCs

To explore the role of SUMO on chromatin in cell-fate determination, we first compared the profiles of sumoylated proteins on the genome in differentiated and pluripotent stem cells. Chromatin immunoprecipitation coupled to DNA sequencing (ChIP-seq) was performed in MEFs and ESCs for SUMO-1 and SUMO-2 along with five characteristic histone marks. The specificity of the SUMO antibodies was confirmed by the nearly complete absence of SUMO signals on chromatin from sumoylation-devoid *Ubc9*^{-/-} MEFs (Demarque et al., 2011) (Figures S1A and S1B). Comparison of the SUMO-1 and SUMO-2 profiles revealed a high degree of overlap between the two SUMO paralogs within each of the two cell types (Figure 1A). In contrast, both SUMO-1 and SUMO-2 exhibited strikingly different patterns when comparing MEFs to ESCs (Figures 1A and 1B). Peak calling was applied on merged SUMO-1 and SUMO-2 signals (henceforth referred to as SUMO) in each cell type. Differential analysis identified 6,971 common peaks, whereas 12,075 were unique to MEFs and 9,778 specific for ESCs (Figures 1B and S1C), with a predominant distribution in distal (inter- and intragenic) regions in all three cases (Figure 1C).

Comparison of the SUMO profiles with histone marks revealed that SUMO at transcription start sites (TSSs) associates with active promoters in MEFs (H3K4me3, H3K27ac) and with a small subset of repressed (H3K9me3) or poised promoters in ESCs (H3K4me3, H3K27me3) (Figure S1D). In MEFs, SUMO at promoters marks genes involved in translation (Figure S1E), consistent with our previous studies in human fibroblasts (Neyret-Kahn et al., 2013). Intriguingly, in ESCs, promoters enriched in SUMO associate, almost exclusively, with germ cell functions (Figure S1F), suggesting a role for sumoylation in the regulation of the germline transcriptional program.

In distal regions, SUMO correlated with drastically different chromatin types in the two cell types. Whereas in MEFs, SUMO is mainly associated with active enhancers (H3K27ac and H3K4me1), in ESCs, SUMO is primarily present in H3K9me3 heterochromatin (Figures 1D, S1D, and S1G). Representative examples are shown in Figure 1E. Consistent with a previous report (Yang et al., 2015), we detected a strong association of SUMO with H3K9me3-marked genomic endogenous

retroviruses (ERVs) in ESCs, with 50%–72% of distal SUMO peaks overlapping with ERVs (Figures 1F and S1H).

We next combined ChIP-seq with RNA-seq data in MEFs and ESCs. Genes assigned to MEF-specific SUMO peaks in distal regions were significantly more expressed in MEFs than in ESCs (Figures 1G and S1I) and were strongly enriched for fibroblast-associated processes (Figures 1H and S1J). In contrast, the nearest genes associated with ESC-specific SUMO peaks showed similar expression in the two cell types (Figures 1G and S1I), and ontology analysis did not reveal any particular enrichment. Consistent with the above data, the SUMO signal in MEFs was higher in MEF-specific super-enhancers relative to those defined for other cell types (Figures 1I and S1K). In ESCs, a small fraction of SUMO overlaps with H3K27ac and H3K4me1 in distal regions (5%–7%) (Figure S1G). The ESC-specific super-enhancers were densely loaded with SUMO in ESCs (Figure 1I), including those associated with the key pluripotent genes *Oct4*, *Sox2*, *Klf4*, and *Nanog* (Figure S1L). Of note, at the *Nanog* locus, the most prominent SUMO peak overlaps with a region highly enriched in H3K9me3.

Thus, whereas sumoylated proteins in MEFs predominantly associate with active enhancers linked to the fibroblast identity, in ESCs, they are strongly associated with heterochromatin-enriched genomic ERVs, yet present at a subset of pluripotency-specific regulatory elements.

Ubc9 Deficiency Enhances Reprogramming to Pluripotency *In Vitro* and *In Vivo*

To then probe the role of SUMO in cell plasticity, we first lowered the levels of sumoylation during reprogramming of MEFs to induced pluripotent stem cells (iPSCs) using both genetic and small-molecule approaches. We treated MEFs derived from reprogrammable “i4F” mice—that carry a doxycycline-inducible transgene expressing *Oct4*, *Klf4*, *Sox2*, c-Myc (OKSM) (Abad et al., 2013)—with a short hairpin RNA (shRNA) targeting *Ubc9* (shUbc9) or a control shRNA (shCtrl). Knockdown (KD) of *Ubc9* led to a slight decrease in global sumoylation and clear appearance of free SUMO (Figure S2A). Hyposumoylation increased iPSC formation (Figures 2A and 2B), a finding in line with the recent identification of *Ubc9* and SUMO-2 as top hits in shRNA screens for enhanced reprogramming of fibroblasts to iPSCs (Borkent et al., 2016; Cheloufi et al., 2015). In *Ubc9* KD cells, the level of the pluripotency factor *Nanog* was massively increased in comparison to control cells (Figure 2C). Interestingly, we observed a gradual increase in the global SUMO-2 patterns, followed by the appearance of very-high-molecular-weight species (Figure 2C). These polymodified species, which were not observed for SUMO-1, were highly sensitive to *Ubc9* KD when compared to other SUMO conjugates. To then validate our findings, we used ML-792, a highly selective small-molecule inhibitor of the SUMO E1 enzyme (He et al., 2017). Chemical manipulation of sumoylation similarly increased reprogramming efficiency (Figures 2A, 2B, and S2B). The maximal enhancing

(G) Expression of genes located within 30 kb from MEF- and ESC-specific intergenic SUMO peaks.

(H) Ontology analysis of genes located within 50 kb from MEF-specific intergenic SUMO peaks.

(I) SUMO-1 signal enrichment in MEFs and ESCs averaged over SEs of different cell types.

See Figure S1.

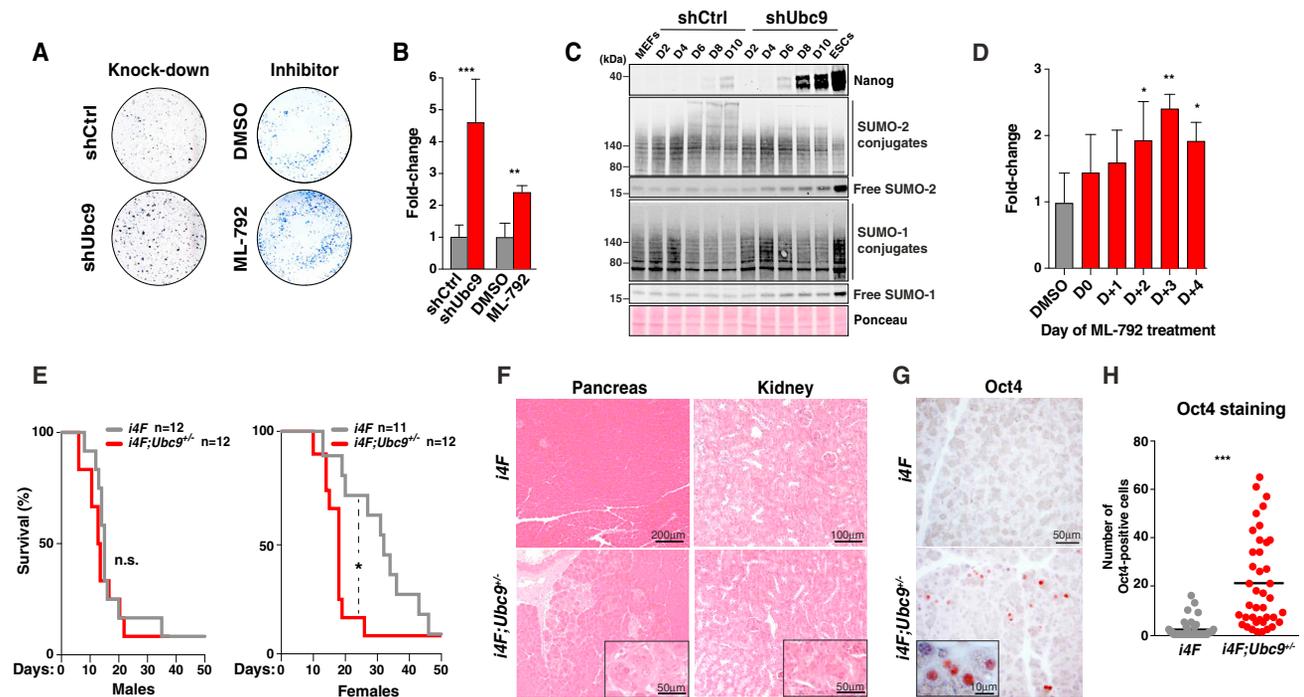


Figure 2. Ubc9 is a Barrier to Reprogramming to Pluripotency In Vitro and In Vivo

(A) Alkaline phosphatase (AP)-positive iPSC colonies after OKSM induction in reprogrammable MEFs in combination with control or Ubc9 shRNA (left) or treated with ML-792 or vehicle (right). (B) Quantification of data shown in (A). shRNA: n = 5; inhibitor: n = 4. (C) Immunoblot of Nanog, SUMO-2, and SUMO-1 during reprogramming at the indicated days in combination with shCtrl or shUbc9 as compared to MEFs and ESCs. (D) Quantification of AP-positive iPSC colonies treated or not with ML-792 at the indicated day of reprogramming, n = 4. (E) Survival curves of *i4F* and *i4F;Ubc9^{+/-}* males (left) and females (right) after doxycycline treatment. Log-rank test. (F) H&E staining of pancreas and kidney of *i4F* and *i4F;Ubc9^{+/-}* females after doxycycline treatment. (G) Oct4 staining in pancreases of *i4F* and *i4F;Ubc9^{+/-}* females after doxycycline treatment. (H) Quantification of the number of Oct4-positive cells as in (G), n = 4. Error bars indicate mean + SD. See Figure S2.

effect was visible when cells were treated on day 3 of reprogramming (Figure 2D), indicating that hyposumoylation should occur early in the process.

We had previously shown that complete lack of Ubc9 in mice is embryonic lethal due to rapid death of inner cell mass cells (Nacerddine et al., 2005). To test the possibility that sumoylation may affect reprogramming *in vivo*, we crossed *i4F* mice with *Ubc9^{+/+}* and *Ubc9^{+/-}* mice, the latter displaying a weak decrease in global sumoylation (Figure S2C) but otherwise no overt phenotype. Treatment for 2.5 weeks with low doxycycline (0.2 mg/mL) led to earlier death of *i4F* males than *i4F* females (Figure S2D), indicating that males are more prone to reprogramming *in vivo*, in agreement with a recent report (Mosteiro et al., 2017). We then evaluated *Ubc9* heterozygous mice in the *i4F* background. Although no noticeable difference was visible when comparing *i4F;Ubc9^{+/-}* to *i4F* males, remarkably, *i4F;Ubc9^{+/-}* females showed significantly shortened survival with respect to their *i4F* counterparts (Figure 2E). Whereas most of the *i4F* females (7/11) had developed teratomas (Figure S2E), the lifespan of *i4F;Ubc9^{+/-}* females was too short to measure teratoma formation. Rather, we observed that tissue architecture of several organs was highly dysplastic in these mice at their time of death

(5/8) (Figure S2F), a finding predictive of full subsequent reprogramming (Mosteiro et al., 2016). In line with these observations, analysis of mice sacrificed earlier in the process, after 1 week treatment with high doxycycline (1 mg/mL), revealed that *i4F;Ubc9^{+/-}* females showed multifocal lesions in the pancreas and the kidney. In contrast, pancreatic and kidney tissue architecture was largely unaffected in *i4F* females (Figure 2F). We then explored *in situ* reprogramming in this setting by looking at the expression of Oct4 in the pancreas. Whereas the pancreases of *i4F* mice showed a few Oct4-positive cells, the pancreatic tissues of *i4F;Ubc9^{+/-}* female mice displayed numerous Oct4-positive reprogrammed cells, which were enriched in dysplastic areas (Figures 2G and 2H). Altogether, these data suggest that, although the loss of a single *Ubc9* allele leads to no particular phenotype in normal conditions, it becomes critical during cellular reprogramming *in vivo*, favoring this process.

Hyposumoylation Drives Early Extinction of the Fibroblastic Transcriptional Program during Reprogramming to iPSCs

To gain insight into the molecular mechanisms by which sumoylation inhibition enhances iPSC formation, we generated the

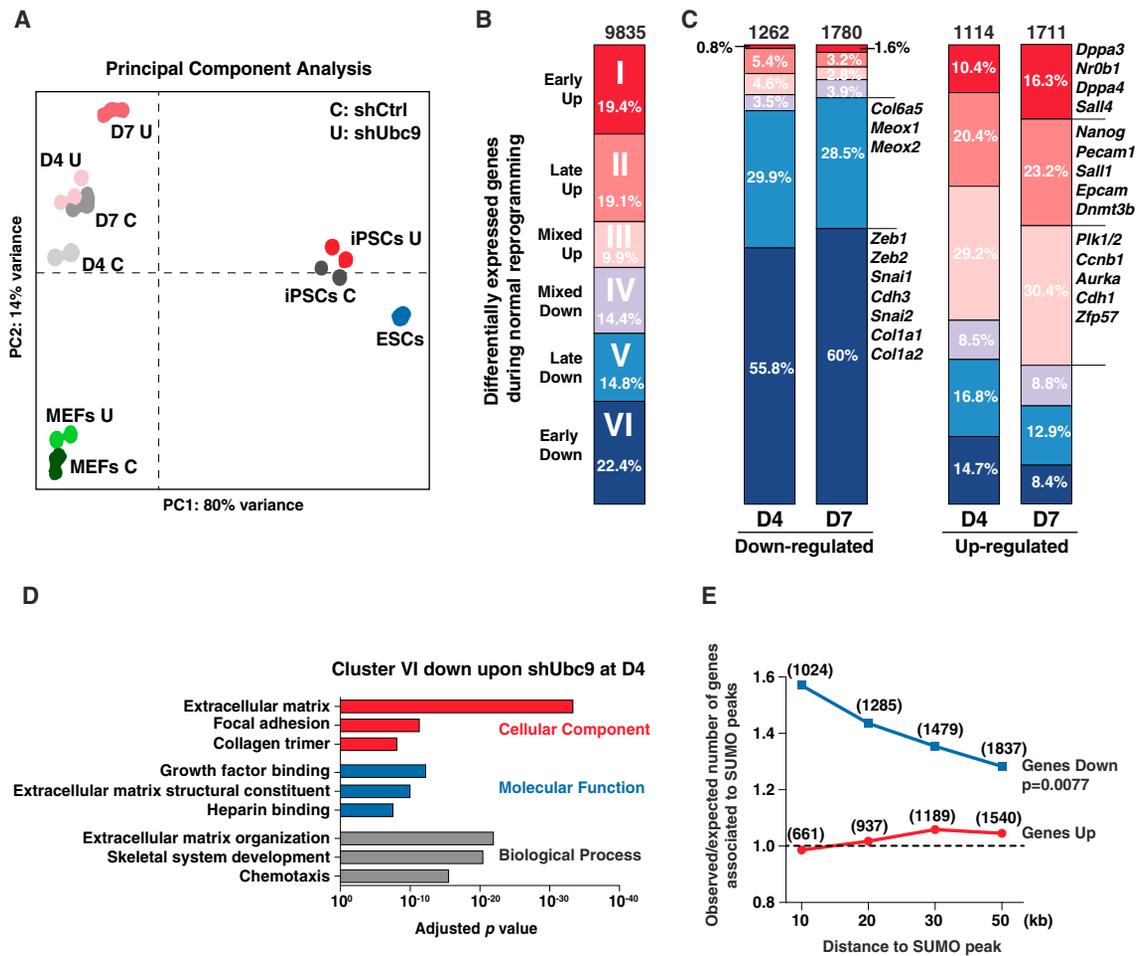


Figure 3. Impaired Sumoylation Facilitates Extinction of the Fibroblastic Transcriptional Program during iPSC Generation

(A) Principal-component analysis (PCA) of gene expression data of the indicated cell populations. Samples are colored according to their time point, MEFs at day 0 (MEFs), day 4 (D4), and day 7 (D7) of reprogramming, iPSCs and ESCs, and their condition, shCtrl (C) and shUbc9 (U). $n = 2-3$ cell cultures for each condition. (B) Gene expression categories (I–VI) clustered according to common expression changes during reprogramming in normal (shCtrl) conditions, as defined in Figure S3A. Number on top indicates the total number of genes.

(C) Distribution of genes down- and upregulated by Ubc9 KD at D4 and D7 of reprogramming in the 6 clusters as defined in (B). Numbers on top indicate the total number of genes in the different clusters.

(D) Ontology analysis of genes downregulated at D4 of reprogramming upon shUbc9 and belonging to cluster VI.

(E) Number of up- and downregulated genes at D7 of reprogramming in the vicinity of distal MEF-specific SUMO peaks relative to the expected number of genes that would randomly fall in the same genomic interval, χ^2 test. Dashed line corresponds to the expected ratio. Numbers in bracket indicate the absolute number of genes.

See Figure S3 and Tables S1 and S2.

expression profiles of shCtrl or shUbc9-treated reprogrammable MEFs at different time points during the process (days 0, 4, and 7 and iPSCs), and of untreated ESCs. Principal-component analysis of the different populations revealed that the patterns of shUbc9 cells at day 4 and control cells at day 7 clustered together (Figure 3A), indicating that hyposumoylation accelerates the reprogramming process. We then determined categories of genes showing changes in expression during the normal reprogramming process across 4 different stages from the starting state (MEFs) to the pluripotent state (ESCs). We identified a total of 9,835 genes with dynamic expression in shCtrl cells, thereafter referred to as reprogramming-specific genes, that were subdivided into six groups according to similar pat-

terns (Figures 3B and S3A; Table S1). Expectedly, upregulated genes (clusters I and II) were enriched in pluripotency genes, whereas downregulated genes (clusters V and VI) mainly contained genes associated with fibroblast functions (Figures S3A and S3B).

Suppression of sumoylation dramatically altered gene expression, with roughly the same number of genes being up- or downregulated at day 4 (2,542 and 2,491 genes, respectively) and day 7 (3,117 and 3,292 genes, respectively) (Table S2). Nearly 50% of these genes belong to clusters I–VI, indicating that hyposumoylation profoundly impacts the reprogramming-specific gene expression program. Of note, the effect of Ubc9 depletion on transcription was minimal in MEFs without OKSM induction

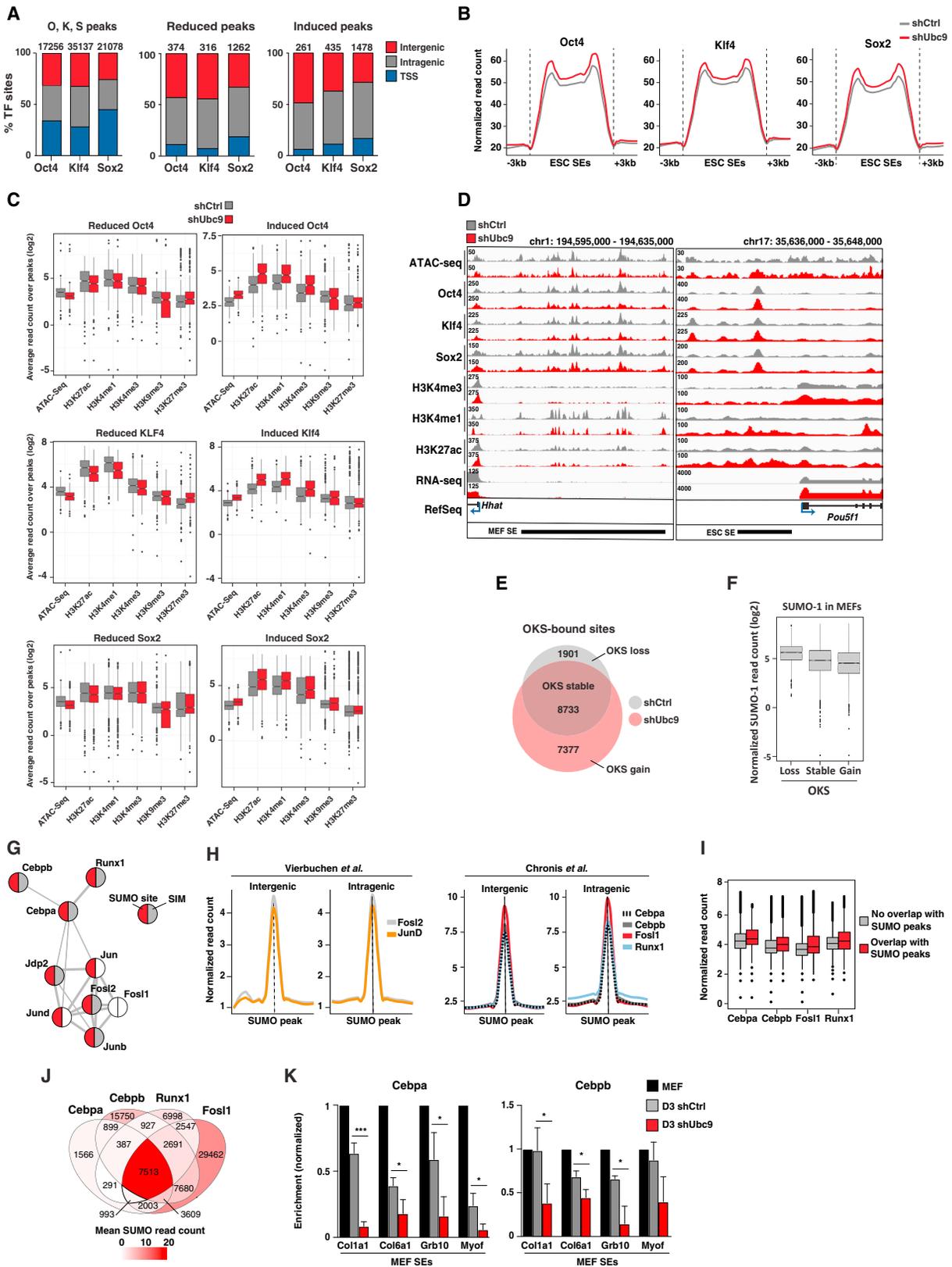


Figure 4. Ubc9 Depletion Affects OKS and TF Substrate Occupancy of Somatic Enhancers during Reprogramming to Pluripotency (A) Proportion of O, K, and S peaks assigned to TSSs, intra- and intergenic regions (left). Distribution of reduced and induced O, K, S peaks upon Ubc9 KD (right). Numbers on top indicate the number of peaks. (legend continued on next page)

(116 differentially expressed genes) (Figure S3C; Table S2). Focusing on reprogramming-specific genes, we assigned the genes whose expression was up- or downregulated by Ubc9 KD to the six clusters as defined during the normal reprogramming process (Figure 3C). Genes of clusters V and VI were remarkably overrepresented among the genes downregulated by Ubc9 depletion (Figure 3C) and were strongly associated with fibroblast features (Figures 3D and S3D). These results were validated on a subset of mesenchymal markers (Figure S3E). In contrast, intersection of the set of genes upregulated by shUbc9 with the set of reprogramming-specific genes identified cluster III as the relatively more affected gene category (Figure 3C). The upregulated genes of this cluster were mainly involved in cell-cycle regulation (Figure S3F), a finding likely related to the acceleration of cell-cycle progression during reprogramming to iPSCs (Polo et al., 2012). In addition, genes upregulated by Ubc9 depletion at day 7 include key pluripotency-associated genes such as *Nanog*, *Sall1*, *Sall4*, *Dppa3*, *Dppa4*, and *Nr0b1* (Figure 3C), only a small fraction of which (e.g., *Sall4* and *Nr0b1*) were readily seen as differentially expressed at day 4 (Table S2). Thus, hyposumoylation impacts the early stage of OKSM-induced reprogramming mainly through massive and rapid downregulation of the fibroblast gene expression program, while inducing a small subset of pluripotency-specific genes.

We next investigated the relationship between genes dysregulated by Ubc9 KD during reprogramming and SUMO occupancy on chromatin in MEFs. Only a small proportion (4%) of up- or downregulated genes were enriched in SUMO at their TSSs, excluding a major direct effect of SUMO at promoters (Figure S3G). In sharp contrast, distal SUMO peaks in MEFs were strongly associated with downregulated genes, whereas no significant association could be detected between distal SUMO sites and upregulated genes (Figure 3E).

Together with our findings above that SUMO is prominently enriched at active enhancers in MEFs, these results suggest that maintenance of the somatic transcriptional program is the main direct function of SUMO on chromatin during reprogramming.

Hyposumoylation Switches OKS Enhancer Occupancy Early in Reprogramming

It was reported recently that, early in reprogramming, OKS predominantly bind active somatic enhancers and are redistributed to pluripotency enhancers at later stages. Somatic TFs, presumably guided by OKS, are depleted from fibroblast enhancers to immediately redistribute toward new sites contributing, together with other mechanisms, to the rapid extinction of the somatic

transcriptional program (Chronis et al., 2017). We thus assessed the impact of Ubc9 depletion on O, K, and S binding as well as histone mark levels by performing ChIP-seq analysis at day 4 of reprogramming in shUbc9 or shCtrl cells. In addition, to compare chromatin accessibility in these two conditions, we performed assay for transposase accessible chromatin sequencing (ATAC-seq). In shCtrl cells, O, K, and S were evenly distributed in TSSs and intra- and intergenic regions (Figure 4A). Comparison of the O, K, and S profiles at day 4 with those obtained for SUMO and H3K27ac in MEFs revealed that regions co-occupied by the three reprogramming factors contain a large fraction of SUMO-marked enhancers that were active in fibroblasts (Figure S4A). In addition, both ESC and MEF super-enhancers were strongly enriched in O, K, and S (Figure S4B).

We then focused on the O, K, or S peaks that were significantly increased (“induced peaks”) or decreased (“reduced peaks”) in Ubc9 KD compared to control cells (Table S3). Comparable numbers of peaks were induced or reduced for each of the three factors over the genome (Figure 4A). Genes associated with reduced peaks revealed an overall prevalence of fibroblast-specific genes and genes downregulated during the mesenchymal to epithelial transition such as *Snai1*, *Twist1*, *Zeb2*, and *Meox2* (Figure S4C; Table S4). In contrast, gained sites were associated with early-activated epithelial genes (*Cdh1*, *Epcam*, *Cldn1*, and *Cldn4*) and pluripotency-specific genes *Oct4*, *Sox2*, and *Sall1* (Figure S4C; Table S4). In line with this, a noticeable global increase in O, K, and S binding was detected upon averaging all ESC super-enhancers (Figure 4B). We next plotted ATAC-seq and histone mark signals over O, K, and S peaks in the corresponding regions (Figure 4C). Sites depleted in either of the three factors showed decreased ATAC-seq signal together with a reduction in H3K27ac and H3K4me1 concomitant to elevation of H3K27me3. This indicates that hyposumoylation-induced loss in O, K, or S is associated with chromatin closing and reduced levels of active chromatin at enhancer regions. A similar decrease in H3K27ac and H3K4me1 levels was also visible in shUbc9 cells when restricting the analysis to MEF super-enhancers (Figures S4D and 4D). Conversely, enriched sites displayed more ATAC-seq signals and increase in H3K27ac, H3K4me1, and H3K4me3 levels (Figure 4C). This indicates that gained O, K, or S sites align with more open and more active chromatin at enhancer elements, as exemplified at the *Oct4* locus (Figure 4D).

Given that cooperative binding of O, K, and S is critical for both somatic enhancer silencing and pluripotency enhancer selection, we next focused on sites co-occupied by O, K, and S. To analyze

(B) O, K, and S signals over ESC SEs in shUbc9 and shCtrl cells.

(C) ATAC-seq and histone mark signals at reduced and induced O, K, and S peaks in shUbc9 and shCtrl cells.

(D) Screen shots of ATAC-seq, RNA-seq, O, K, S, and histone mark ChIP-seq signals at the *Hhat* and *Oct4* loci in shUbc9 and shCtrl cells.

(E) Venn diagram comparing OKS co-bound regions in shUbc9 and shCtrl conditions.

(F) Boxplots of SUMO-1 signal in MEFs at OKS sites that were gained, lost, or maintained in shUbc9 versus shCtrl cells at D4.

(G) STRING visualization of the protein interaction network of Cebpa, Cebpb, Runx1, and the AP-1 complex. Red semi-circle indicates the presence of a SUMO site; gray semi-circle indicates the presence of a validated or predicted SUMO interaction motif (SIM).

(H) Metaplots of Fosl2, JunD, Cebpa, Cebpb, Fosl1, and Runx1 signal intensities centered on the MEF-specific SUMO peaks in distal regions.

(I) Distribution of the average signal for Cebpa, Cebpb, Fosl1, and Runx1 among peaks enriched or not in SUMO in MEFs.

(J) Venn diagram showing the intersection between Cebpa, Cebpb, Fosl1, and Runx1 binding sites in MEFs. The intensity of the color corresponds to the average SUMO signal over the corresponding binding sites.

(K) ChIP-qPCR experiments showing Cebpa and Cebpb binding to representative MEF SEs in reprogrammable MEFs at D0 or D3 upon shUbc9 or shCtrl, n = 3. Error bars indicate mean + SD. See Figure S4 and Tables S3, S4, and S5.

OKS co-binding events, the sets of identified O, K, and S peaks were used and regions that are recognized by one, two, or three of these factors in shCtrl and shUbc9 cells were counted. The number of OKS-bound sites increased from ~11,000 to ~16,000 upon suppression of sumoylation (Figure S4E; Table S5). More importantly, the repertoires of sites co-occupied by the three factors differed significantly between Ubc9 KD and control cells, with many sites that were gained (“OKS gain,” 7,377), whereas others were lost (“OKS-loss,” 1,901) (Figure 4E). In accordance with the predominant location of SUMO on somatic enhancers in MEFs, we found that the OKS sites that are lost were markedly enriched in SUMO in MEFs when compared to those maintained or gained upon Ubc9 depletion (Figure 4F). Thus, decreasing sumoylation levels on fibroblast enhancers leads to the coordinated release of the three factors away from these elements early in the reprogramming process.

To exclude that the molecular changes induced by Ubc9 KD would result from a sole acceleration of the reprogramming process, we analyzed the impact of loss of sumoylation in a non-reprogramming context. To this aim, we performed ATAC-seq on Ubc9^{-/-} MEFs. We identified 1,865 and 1,447 ATAC-seq peaks that were down- or upregulated in Ubc9^{-/-} MEFs when compared to Ubc9^{+/+} MEFs, respectively (Figure S4F). Remarkably, the regions that showed chromatin closing upon Ubc9 loss correspond to regions that become less accessible during the normal reprogramming process, and even more closed upon Ubc9 KD in this latter context (Figures S4G and S4H). These data strongly suggest that moderate hyposumoylation achieved by shUbc9 on its own, though only weakly impacting gene expression, may facilitate chromatin closing at active enhancers, thus reinforcing the OKSM-induced loss of chromatin accessibility at these regions.

SUMO Favors Co-occupancy of Somatic TF Substrates on Fibroblastic Enhancers

In the context of DNA repair, SUMO was reported to act synergistically by modifying multiple proteins in a given DNA-bound complex, likely functioning as a glue to increase complex stability (Psakhye and Jentsch, 2012). To assess whether a similar situation could apply for transcription at fibroblastic enhancers, we focused on Fos1, Cebpa, Cebpb, and Runx1 somatic TFs given their key role in the resistance to reprogramming to iPSCs (Chronis et al., 2017). These TFs, which are part of a protein-protein interaction network that is almost exclusively composed of SUMO substrates possessing SIMs, precisely colocalize with SUMO at MEF-specific distal regions (Figures 4G and 4H). Remarkably, sites bound by these TFs show significantly higher occupancy of the factor when SUMO is present (Figure 4I). Reciprocally, sites co-occupied by the four factors show dramatically higher levels of SUMO than those occupied by one, two, or three of the four factors (Figure 4J). Altogether, these data strongly suggest that SUMO favors the co-occupancy of the fibroblastic TFs on DNA, likely by potentiating their physical interaction. To then directly assess the impact of hyposumoylation on the binding of these TFs, we performed local ChIP for Cebpa and Cebpb, two bona fide SIM-containing SUMO substrates (Hendriks and Vertegaal, 2016), on a set of fibroblastic super-enhancers. At day 3 of reprogramming, both TFs displayed decreased occupancy of these regions and depletion of

Ubc9 further drastically reduced the levels of Cebpa and Cebpb binding to these elements (Figure 4K).

Taken together, these data indicate that SUMO prevents fibroblastic enhancer inactivation early in reprogramming by stabilizing OKS and somatic TF substrates on these elements, presumably by fostering the interactions between these factors that otherwise show low affinity for each other on their own.

Hyposumoylation Facilitates Transdifferentiation and Directed Differentiation

To then assess whether SUMO can impact cell identity in a more general manner, we analyzed the effect of hyposumoylation in transdifferentiation systems. We first examined direct conversion of MEFs into induced neurons via overexpression of Brn2a, Ascl1, and Myt1l (Vierbuchen et al., 2010) (Figures 5A and S5A). Immunofluorescence for neuronal markers confirmed that MEFs could efficiently transdifferentiate into mature neurons (Figure S5B). Knockdown of Ubc9 induced a 2-fold increase in the number of cells positively stained for the neuronal marker Tuj1 as determined at day 5 (Figures 5B and 5C), together with an upregulation of the transcripts for Tuj1 and the other neuronal marker NSE (Figure 5D). Next, we analyzed the effect of knockdown of Ubc9 on the reprogramming of pre-B cells into macrophages with C/EBP α (Bussmann et al., 2009) (Figures 5E and S5A). Though sumoylation depletion slightly impaired downregulation of the pre-B Cd19 marker (Figure 5F), the vast majority of the cells showed an increase in typical hallmarks of granulocytic differentiation after 3 days such as activated Mac1 (Cd11b) and increase in cell granularity and phagocytic activity (Figures 5G–5I), together with characteristic morphological changes (Figure S5C). Of note, a mixed population of cells expressing lower and higher Mac1 levels than control cells was observed at earlier time points, that resolved into highly Mac1-expressing cells after 3 days (Figure 5G). Thus, hyposumoylation can facilitate the transition of one type of differentiated cells into another specialized cell type.

We then investigated whether sumoylation may regulate cell differentiation using the model of retinoic acid (RA)-induced differentiation of HL-60 human leukemic promyelocytes into granulocytes (Figures 5J and S5A). Transduction of HL-60 cells with two independent Ubc9 shRNAs (#1 and #2) achieving weak knockdown of Ubc9 (Figure S5D) significantly increased their differentiation by RA, as assayed by the expression of Cd11b and appearance of cytosolic granules (Figures 5K, 5L, and S5E). Similar experiments performed with a strong Ubc9 shRNA (#3) resulted in a block in terminal differentiation (Figures S5D, 5K, and 5L). Moreover, in keeping with the slower proliferation rate of differentiated versus undifferentiated cells, hyposumoylation strongly synergizes with RA to block HL-60 proliferation (Figure 5M). These data indicate that mild hyposumoylation potentiates the differentiating and anti-proliferative effects of RA while potent inhibition is detrimental to the differentiation process. We next investigated a second model of differentiation in which immature myeloblasts undergo terminal neutrophil differentiation, a process accompanied by expression of a lysosome (Lys)-GFP reporter (Sykes et al., 2016) (Figure S5A). Treatment with ML-792 significantly increased terminal differentiation in a time- and dose-dependent manner as assayed by the activation of Lys-GFP and expression of Itgam, another

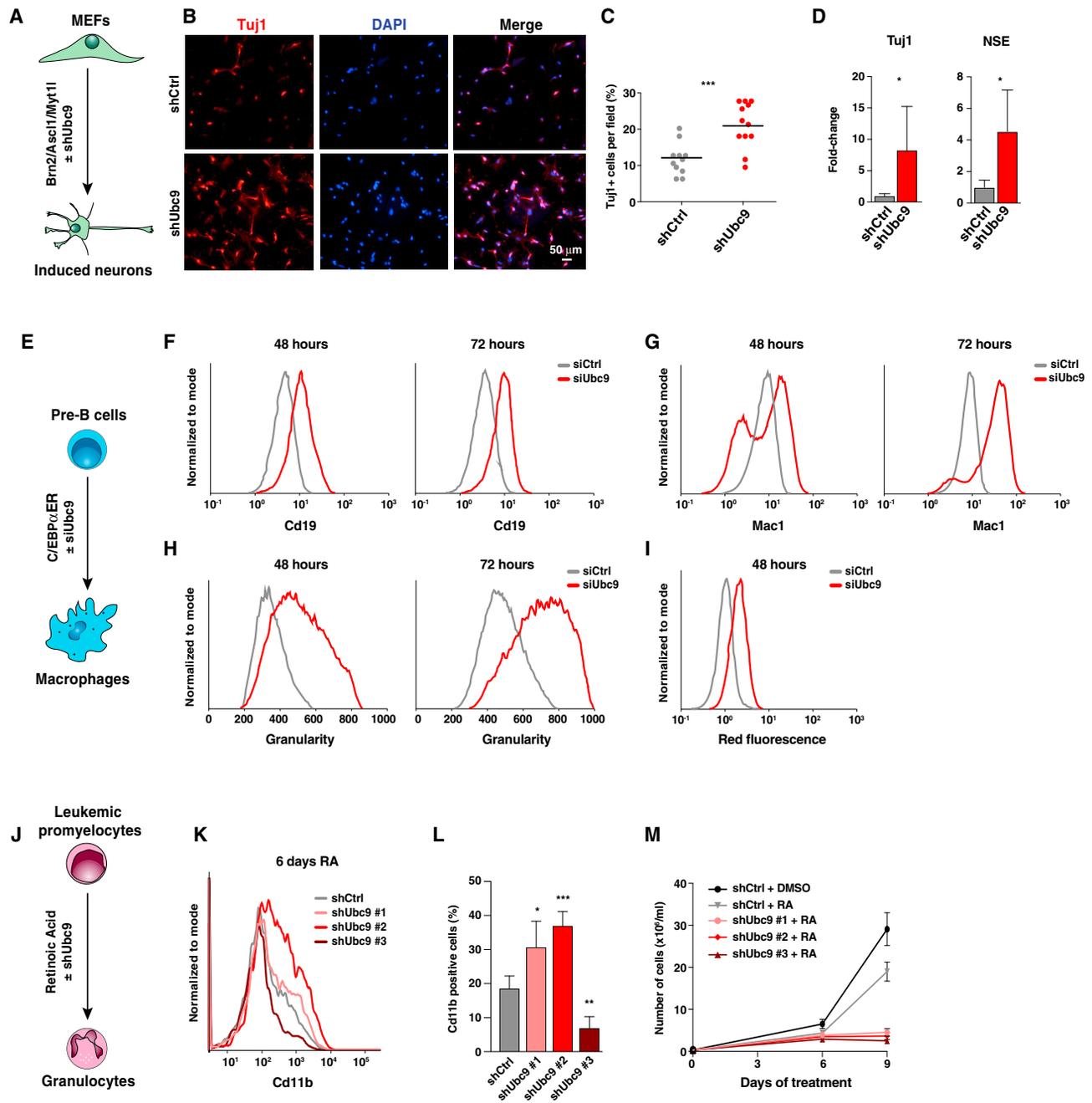


Figure 5. Hyposumoylation Enhances Cell-Fate Transitions

(A) Transdifferentiation model of MEFs into neurons.

(B) Tuj1 immunostaining after 5 days of Brn2, Ascl1, and Myt11 induction with shUbc9 or shCtrl.

(C) Percentage of Tuj1 positive cells as related to (B). 11 and 12 fields were analyzed.

(D) Relative quantification of Tuj1 and NSE mRNA levels in shCtrl- or shUbc9-treated MEFs at D5 of transdifferentiation.

(E) Transdifferentiation model of pre-B cells into macrophages.

(F–I) Flow cytometry analysis of Cd19 (F), Mac1 (G), granularity (H), and phagocytic activity (I) signals upon transdifferentiation of pre-B cells into macrophages at indicated time points after treatment with siUbc9 or siCtrl. Representative examples, $n = 3$.

(J) Differentiation model of leukemic HL-60 cells into granulocytes.

(K) Flow cytometry analysis of Cd11b expression for cells transduced with 3 independent Ubc9 shRNAs (#1, #2, #3) or shCtrl and treated with RA. Representative example, $n = 3$.

(L) Quantification of Cd11b positive cells, $n = 3$.

(M) Cell proliferation of HL-60 cells as indicated, $n = 3$.

Error bars indicate mean + SD. See Figure S5.

neutrophil-associated marker (Figures S5F–S5H). Interestingly, ML-792 alone was able to initiate neutrophil differentiation, albeit to a low degree (Figure S5I).

Collectively, these data indicate that, in addition to favoring reprogramming to pluripotency, inhibition of sumoylation also enhances lineage conversion and cellular differentiation.

Loss of Sumoylation in ESCs Induces the Emergence of 2C-like Cells

In contrast to its predominant location on active enhancers in MEFs, in ESCs, SUMO is strongly associated with heterochromatin-enriched genomic ERVs (Figures 1D and 1F). Interestingly, several repeated elements, in particular, major satellite repeats and MERVL, are spontaneously reactivated in a rare subpopulation of ESCs called 2C-like cells since they resemble the 2-cell stage embryo (Macfarlan et al., 2012). These cells, like 2-cell stage embryos, show decondensed chromatin (Ishiiuchi et al., 2015) and express MERVL-initiated chimeric transcripts that encode part of 2-cell-specific proteins (Macfarlan et al., 2012). Using two different small interfering RNAs (siRNAs) targeting Ubc9, we depleted sumoylation in ESCs stably expressing a MERVL::eGFP reporter (Figure S6A). Knockdown of Ubc9 led to a 8- to 12-fold increase in the number of eGFP-positive cells, to a level comparable to that obtained by depleting Chaf1, the p150 subunit of CAF-1 used here as a positive control (Ishiiuchi et al., 2015) (Figures 6A and 6B). RNA fluorescence *in situ* hybridization (RNA-FISH) confirmed that activation of the MERVL::eGFP reporter was associated with the upregulation of endogenous MERVL expression (Figures 6C and 6D). Moreover, like 2-cell-stage embryos, 2C-like cells induced through Ubc9 KD lack both Oct4 protein expression and chromocenter organization (Figure 6E). Interestingly, suppression of other components of the SUMO machinery such as the E1 enzyme Uba2 subunit or the SUMO-2 and SUMO-3 paralogs similarly increased the conversion of ESCs to 2C-like cells, albeit to a lesser extent for SUMO-3 (Figures 6F and S6A). In contrast, depletion in SUMO-1 failed to induce 2C-like cells, indicating some striking functional divergence between the different SUMO paralogs.

To further evaluate the extent to which sumoylation inhibits the conversion to 2C-like cells, we performed RNA-seq in ESCs upon Ubc9 KD. Sumoylation suppression led to the upregulation of 55 subfamilies of repetitive elements including MERVL, whereas 7 subfamilies of repeats were downregulated (Figure 6G; Table S6). These data are consistent with a recent report describing an important role for sumoylation in provirus silencing (Yang et al., 2015). In addition, 1,655 genes were upregulated upon hyposumoylation, whereas 1,150 genes were downregulated (Figure 6H; Table S6). Interestingly, genes induced after Ubc9 KD were physically closer to upregulated ERVs when compared to the rest of the genes in the genome (Figure 6I), a type of co-regulation that is reminiscent of 2-cell-stage embryos. Accordingly, a significant subset of genes were commonly upregulated in Ubc9-depleted ESCs and in endogenous 2C-like cells (Ishiiuchi et al., 2015) (Figure S6B). Among the activated genes in the Ubc9 KD datasets were prominent 2C-like genes such as *Dux*, *Zscan4b/c/d/f*, *Tdp03/4*, *Zfp352*, and *Eif1a*-like genes (*Gm2016*, *Gm5662*...), with *Dux*, a major inducer of the 2-cell state (De Iaco et al., 2017; Hendrickson et al., 2017), being among the most highly upregulated gene (Figure 6H). We

observed a similar upregulation of 2C-like genes and repeats, including the major satellite repeats, in Ubc9-, SUMO-2-, and SUMO-3-, but not SUMO-1-depleted ESCs (Figure S6C). The set of Ubc9 KD-induced genes were significantly enriched in the gene subset specific to 2-cell stage embryos in comparison with other early embryonic stages (Deng et al., 2014) (Figure 6J). Analysis of RNA-seq data for KD of Aos1, Uba2, Ubc9, and SUMO-2 in ESCs (Yang et al., 2015) revealed a similar pattern of enrichment (Figure 6J), thus indicating that several members of the sumoylation machinery can regulate 2C-like reprogramming. Finally, treatment of ESCs with ML-792 was able to induce the 2C-like transcriptional program in a dose-dependent manner, as assessed on typical 2C genes and repeats such as *Dux*, *Zscan4*, and MERVL (Figures 6K and S6D).

Together, these results show that suppression of sumoylation induces the conversion of ESCs into 2C-like cells, indicating an important role for SUMO in stabilizing the pluripotent state.

SUMO Maintains H3K9me3 Heterochromatin Integrity in ESCs

Developmental progression is accompanied by a gradual deposition of the heterochromatin H3K9me3 mark, which constitutes a potent barrier to cell-fate conversions (Becker et al., 2016). Notably, large Suv39h1/2-dependent H3K9me3 domains, called reprogramming resistant regions (RRRs), are involved in the low efficiency of somatic cell nuclear transfer and are efficiently transcribed in the 2-cell stage embryos (Matoba et al., 2014). We observed that depleting Ubc9 in ESCs led to their massive upregulation (Figure 7A), indicating that SUMO plays a key role in RRR silencing. In addition to its repressive activity at RRRs, H3K9me3 has been shown to mediate ERV silencing in ESCs. In this case, H3K9me3 deposition is dependent, in large part, on the Kap1/Setdb1 complex (Matsui et al., 2010). In agreement with the predominant location of SUMO in H3K9me3-marked ERVs in ESCs, these genomic regions were highly enriched in Setdb1 (Figure S7A). Local ChIP analysis at ERVs neighboring the Ubc9 KD-induced 2C-like genes *Gm21319*, *Zscan4*, and *Tdp03* revealed drastic reduction in H3K9me3 and Setdb1 levels upon hyposumoylation (Figure 7B). Remarkably, 2C-like genes upregulated upon Ubc9 KD were part of large genomic regions, which became heavily transcribed upon hyposumoylation as determined by RNA-seq (Figure S7A), indicating a long-range silencing effect of SUMO in regions encompassing 2C-like genes. Strikingly, suppression of sumoylation in ESCs led to a systematic ~30%–50% decrease in the global levels of H3K9me3 as assessed by western blotting, whereas no significant change was observed for the other repressive mark H3K27me3 (Figure 7C). In addition, ChIP-seq analysis revealed a dramatic reduction at nearly all mappable H3K9me3 peaks upon ablation of sumoylation (Figure 7D).

SUMO Regulates *Dux* by Tethering PRC1.6 and Kap1/Setdb1 Complexes onto the *Dux* Locus

To gain further mechanistic insight into how SUMO inhibits conversion to 2C-like states, we employed a combined siRNA approach to assess the functional interaction of sumoylation with key 2C-like regulator pathways. We first focused on *Dux*, a TF recently shown to act upstream and activate the 2C-specific genes and ERVs (De Iaco et al., 2017; Hendrickson et al., 2017).

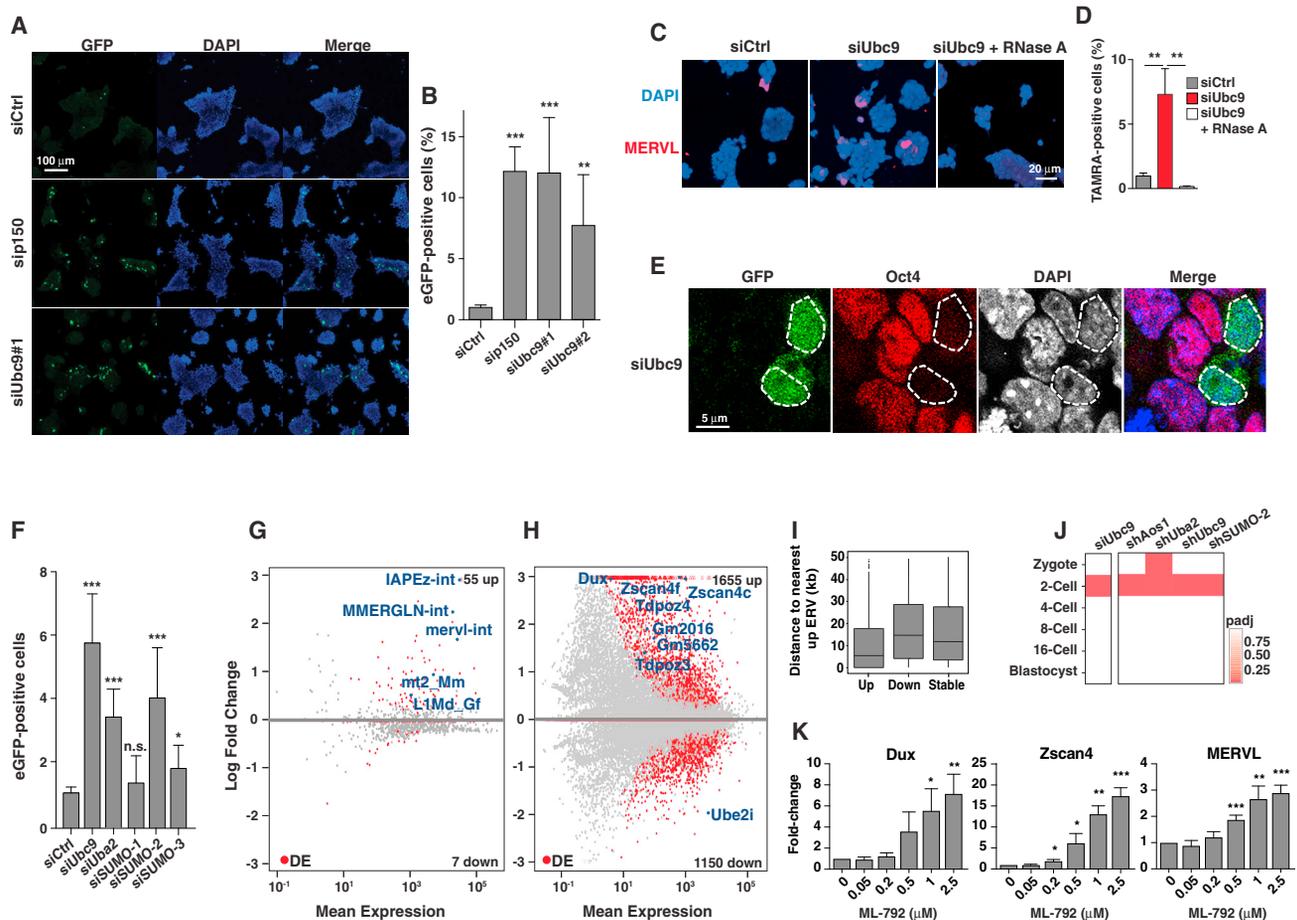


Figure 6. Loss of Sumoylation Enhances Conversion of ESCs to 2C-like Cells

(A) Immunostaining for GFP in MERVL::eGFP ESC line after 2 days of p150 or Ubc9 KD. (B) Quantification of GFP-positive cells by flow cytometry 2 days after transfection of Ctrl, p150, Ubc9#1, and Ubc9#2 siRNAs, $n = 3$. (C) RNA-FISH using a MERVL probe labeled with tetramethylrhodamine (TAMRA) in ESCs after 2 days of Ubc9 KD treated or not with RNase A (used as a negative control). (D) Quantification of TAMRA positive cells after transfection of siCtrl and siUbc9, $n = 3$. (E) Immunostaining for Oct4 and GFP in MERVL::eGFP ESCs after Ubc9 KD. (F) Quantification of GFP-positive cells by flow cytometry 2 days after transfection of Ctrl, Ubc9, Uba2, SUMO-1, SUMO-2, and SUMO-3 siRNAs, $n = 3$. (G and H) MA plots displaying differentially expressed repetitive elements (G) and genes (H) in ESCs after 2 days of siUbc9 in comparison to siCtrl. Some key transcripts are highlighted in blue. (I) Average distance between up-, downregulated, and stable genes and the closest upregulated long terminal repeat (LTR) element upon hyposumoylation. (J) Comparison between genes expressed in specific developmental stages and genes upregulated upon KD of Ubc9 or various members of the sumoylation machinery, Fisher's exact test. (K) Upregulation of *Dux*, *Zscan4*, and *MERVL* transcripts in ESCs treated with ML-792 for 2 days at the indicated concentrations, $n = 3$. Error bars indicate mean + SD. See [Figure S6](#) and [Table S6](#).

Depletion of *Dux* drastically reduced the 2C-like inductive effect of Ubc9 KD, indicating that the effect of hyposumoylation is largely dependent on *Dux* (Figure 7E). We next examined the relationship between sumoylation and the recently identified epigenetic barriers to 2C-like cell emergence (Rodriguez-Terrones et al., 2018). Whereas co-suppression of Ubc9 together with components of the EP400-TIP60 complex (Ep400, Dmap1, or Tip60) or replication-associated proteins (Rif1, Chaf1b, or Usp7) further enhanced the conversion to 2C-like cells, remarkably, simultaneous suppression of Ubc9 and subunits of the repressive PRC1.6 complex (Ring1b, Pcgf6, Mga, Max, or Rybp) showed no additive effect, suggesting that su-

moylation and PRC1.6 function through related pathways (Figure 7F).

Kap1 and H3K9me3 were recently shown to be enriched at the *Dux* gene (De Iaco et al., 2017; Percharde et al., 2018). However, how the *Dux* locus is regulated remains largely unknown. Our finding that SUMO, PRC1.6, and *Dux* functionally interact prompted us to examine whether SUMO and PRC1.6 were present on *Dux*. Analysis of ChIP-seq data revealed significant enrichment of SUMO together with the PRC1.6 components Ring1b, Max, and Rybp and the PRC1-catalyzed modification H2AK119ub, as well as H3K9me3, at the *Dux* gene in ESCs (Figure 7G). Genome-wide analysis of SUMO-marked regions in

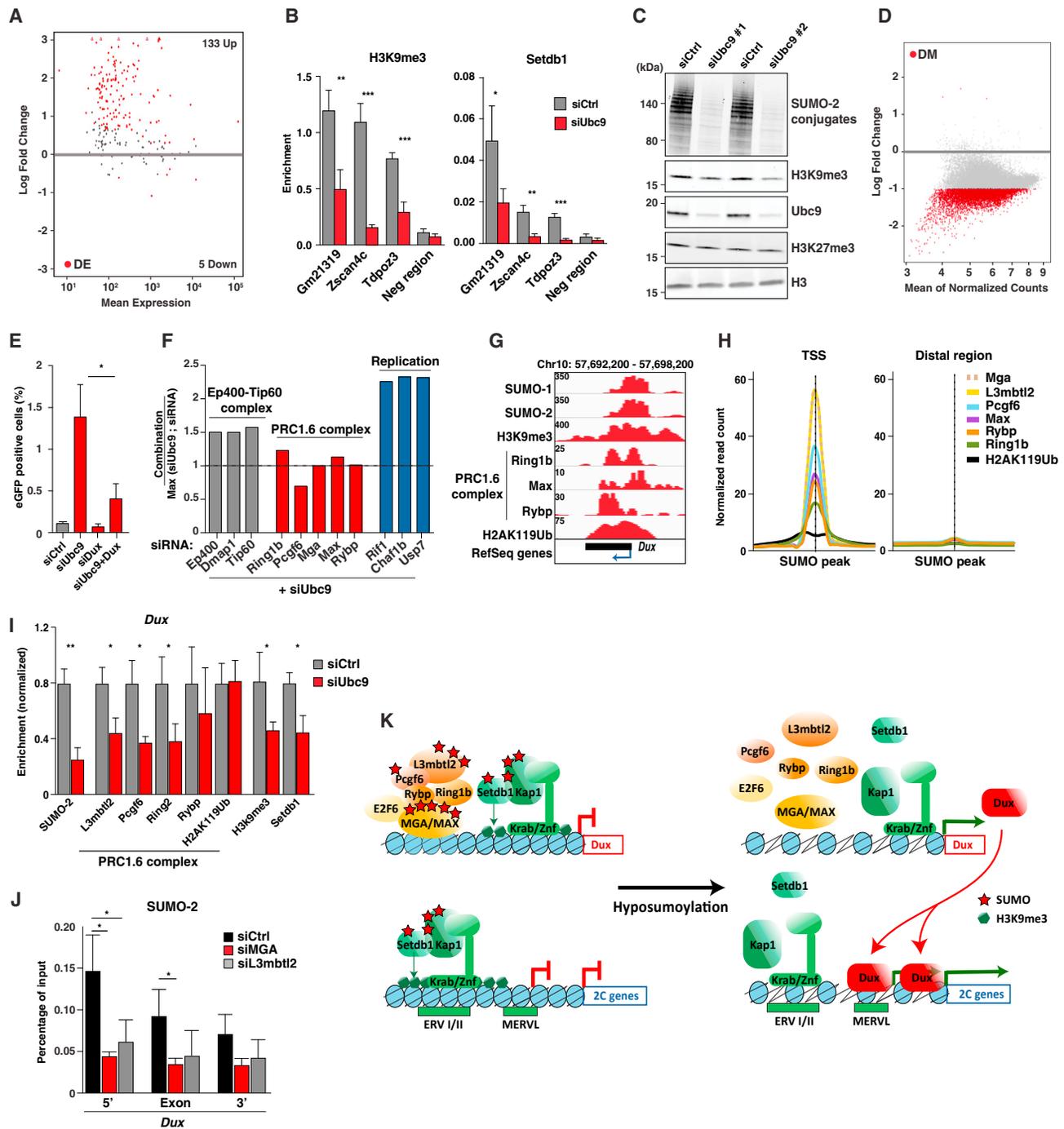


Figure 7. Sumoylation Depletion Leads to Loss of H3K9me3 Heterochromatin and Release of PRC1.6 and Setdb1 from *Dux* in ESCs

(A) MA plot displaying differentially expressed (DE) RRRs after 2 days of siUbc9 in comparison to siCtrl.
 (B) Local ChIP for H3K9me3 and Setdb1 at ERVs neighboring notable 2C-like genes in ESCs after 2 days of siUbc9 in comparison to siCtrl, n = 3.
 (C) Immunoblots for the indicated proteins in ESCs after 2 days of siUbc9 in comparison to siCtrl. Two independent siRNAs against Ubc9 were used.
 (D) MA plot displaying differentially marked (DM) H3K9me3 peaks in siUbc9 versus siCtrl ESCs.
 (E) Percentage of GFP-positive cells in MERV1::eGFP ESC line transfected with the indicated siRNAs, n = 4.
 (F) Combinatorial additive effects of siUbc9 together with one of the indicated siRNAs on the percentage of GFP-positive cells, n = 4.
 (G) Screen shot of the *Dux* locus in ESCs. ChIP-seq tracks were taken from our data (SUMO and H3K9me3) or published datasets (see STAR Methods).
 (H) Metaplots of the indicated PRC1.6 components and H2AK119Ub centered on ESC-specific SUMO peaks.
 (I) Enrichment (normalized) for the indicated proteins in ESCs after 2 days of siUbc9 in comparison to siCtrl. n = 3.
 (J) Percentage of input for the indicated proteins in ESCs after 2 days of siUbc9 in comparison to siCtrl. n = 3.
 (K) Schematic of the *Dux* locus showing the transition from a SUMOylated state (top) to a hypoSUMOylated state (bottom) after siUbc9 treatment, leading to loss of H3K9me3 and release of PRC1.6 and Setdb1.

(legend continued on next page)

ESCs revealed a strong enrichment in PRC1.6 subunits at TSSs indicating that the colocalization between SUMO and PRC1.6 is not restricted to *Dux* (Figure 7H). Notably, germline genes, which constitute primary targets of SUMO (Figure S1F) and PRC6.1 (Stielow et al., 2018) at ESC promoters, showed clear co-enrichment of these proteins at individual genes (Figure S7B).

Given that most subunits of the PRC1.6 complex are SUMO-2 substrates, of which Mga and L3mbtl2 are heavily sumoylated (Hendriks and Vertegaal, 2016), we tested the possibility that sumoylation may regulate the recruitment of PRC1.6 onto *Dux*. Knockdown of Ubc9 significantly reduced the binding of L3mbtl2, Pcgf6, Ring1b, and Rypb on the *Dux* region (Figure 7I). In addition, sumoylation inhibition resulted in concomitant reduction of H3K9me3 levels and Setdb1 binding. We did not observe a decrease of H2AK119ub levels, potentially due to short time span of our KD experiments, which may not enable efficient removal of the H2AK119ub mark. We hypothesized that sumoylation of Mga, a protein that is known to be essential for loading the entire PRC1.6 complex on its targets (Stielow et al., 2018), may play an important role in SUMO-dependent tethering of PRC1.6 onto *Dux*. Importantly, when Mga was depleted, the levels of SUMO-2 binding on *Dux* were considerably reduced (Figure 7J). A strong decrease was similarly observed upon L3mbtl2 suppression. Removal of SUMO-2 was also detected at germline genes upon Mga or L3mbtl2 suppression (Figure S7C). Collectively, these data demonstrate that sumoylation is crucial for genomic targeting or stabilization of the PRC1.6 complex onto the *Dux* gene, and that Mga and L3mbtl2 are likely to be important SUMO-2 substrates for efficient PRC1.6 loading.

Together with our data above, these data demonstrate that SUMO, operating in large part from heterochromatin in ESCs, functions at two different levels to prevent the spontaneous reversion to a 2C-like state. SUMO acts both by maintaining the integrity of H3K9me3 genome-wide and by silencing the *Dux* locus through co-recruitment of the two distinct PRC1.6 and Kap/Setdb1 repressive complexes (Figure 7K).

DISCUSSION

In this work, we show that lowering sumoylation facilitates reprogramming to pluripotency *in vitro* and *in vivo*, transdifferentiation, directed differentiation, and conversion of pluripotent to totipotent-like cells, suggesting that SUMO acts as a general barrier to cell-fate transitions. Mechanistically, we show that SUMO functions as a tether on distinct chromatin types in MEFs and ESCs to stabilize occupancy of key protein substrate complexes, thus preserving the somatic and pluripotent states.

In MEFs, SUMO is mainly associated with active enhancers and suppression of sumoylation, in concert with OKSM, leads to extensive silencing of the fibroblast transcriptional program early in reprogramming, together with inducing a small number of pluripotency-related genes. Reprogramming is associated

with a switch of OKS from MEF to pluripotency enhancers, driving inactivation of MEF enhancers, in part through OKS-guided depletion of somatic TFs (Chronis et al., 2017). The O, K, and S proteins are sumoylated, with SUMO-deficient Klf4 enhancing reprogramming, albeit with a low efficiency (Tahmasebi et al., 2013). Furthermore, a large number of somatic TFs are SUMO modified (Hendriks and Vertegaal, 2016), including Runx1, Cebpa, and Cebpb, all of which are released from fibroblastic enhancers upon reprogramming (Chronis et al., 2017). Sumoylation has been shown most often to target whole protein complexes rather than unique substrates, such that SUMO-deficient mutants of individual targets usually lack overt phenotypes (Psakhye and Jentsch, 2012). As exemplified in DNA repair, SUMO, through individual modifications of proteins exhibiting low affinity for each other, acts as a glue via SUMO-SIM interactions to physically stabilize whole complexes (Psakhye and Jentsch, 2012). We provide several lines of evidence that a similar scenario applies in reprogramming in which SUMO acts synergistically on OKS and somatic TFs to confer stability and robustness to MEF regulatory elements, thus preventing their inactivation. We show that (1) sumoylation depletion promotes the redistribution of OKS from somatic to pluripotency enhancers, (2) SUMO favors the co-occupancy of somatic TFs on active enhancers, and (3) suppression of sumoylation leads to the release of these factors from somatic enhancers, as exemplified by Cebpa and Cebpb. It was proposed recently that multiple TFs form phase-separated multi-molecular assemblies on enhancers with potential important impact on gene activity (Hnisz et al., 2017). Moreover, SUMO-SIM interactions were recently proposed to participate in liquid-like droplet formation associated with PML nuclear body assembly (Banani et al., 2016). It will be interesting to test whether such a phase-separation mechanism may apply for SUMO in creating functional sub-compartments associated with important genomic regulatory elements, thus contributing to cell-identity gene control.

In sharp contrast to its strong enrichment on somatic enhancers in MEFs, SUMO is prominently associated with heterochromatin in ESCs, indicating that SUMO on chromatin is highly patterned according to cell identity. We find here that suppression of sumoylation promotes the spontaneous conversion of ESCs to 2C-like cells and show that SUMO preserves pluripotency through two distinct mechanisms. First, hyposumoylation activates transcription of broad genomic regions that are highly enriched in Setdb1-dependent H3K9me3-marked proviral elements, as well as reexpression of Suv39h1-mediated H3K9me3-silenced RRRs. More generally, SUMO loss strongly reduces global H3K9me3 levels on chromatin. These data suggest that SUMO functions in ESCs by stabilizing H3K9me3 heterochromatin genome-wide. Sumoylation of KAP1 was reported to trigger the recruitment of Setdb1 to ERVs (Ivanov et al., 2007; Yang et al., 2015). Moreover, Suv39h1-stimulated sumoylation of HP1 α reportedly plays an important role in heterochromatin formation (Maison et al., 2016). Thus, KAP1 and HP1 α are likely

(I) Local ChIP for the indicated proteins and histone marks at the *Dux* locus, n = 3.

(J) Local ChIP for SUMO-2 at the *Dux* locus, n = 3.

(K) Model for the functions of SUMO in suppression of 2C-like cell emergence. Error bars indicate mean + SD. See Figure S7.

to be among the highly relevant SUMO substrates implicated in SUMO-dependent maintenance of proper H3K9me3 levels. Second, we uncovered a key role for sumoylation in regulating *Dux* expression. We found that the repressive PRC1.6 complex is present together with Setdb1 on the *Dux* locus and that their recruitment is dependent on sumoylation. Similar to what we observed at somatic enhancers in MEFs, these findings underscore the instrumental role of sumoylation in tethering multi-protein complexes to key loci to preserve the pre-existing transcriptional program. Moreover, we present evidence that the two heavily sumoylated Mga and L3mbtl2 proteins are major SUMO-2 substrates within the multi-subunit PRC1.6 complex. Future work should seek insight into the sequential dynamics of sumoylation/desumoylation at essential loci, such as *Dux*, during the transition from totipotency to pluripotency and, more generally, advance our knowledge on how chromatin modifier complexes can be regulated at their targets, a long-standing question in the chromatin field.

Finally, our findings that small-molecule inhibitors of sumoylation enhance reprogramming to stemness opens new avenues for the use of these compounds in regenerative medicine. Moreover, reactivation of endogenous differentiation programs in cancer cells, by forcing the maturation process, proved to be an alternative promising approach to tumor cell killing in cancer treatment. The finding that moderate hyposumoylation facilitates RA-induced differentiation and proliferation arrest of leukemic cells paves the way toward further exciting progress in cancer therapy.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **CONTACT FOR REAGENT AND RESOURCE SHARING**
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Animals
 - Cell culture
- **METHOD DETAILS**
 - Cell reprogramming
 - *In vivo* reprogramming
 - Induced neuronal transdifferentiation
 - Pre-B to macrophage transdifferentiation
 - HL-60 to granulocyte differentiation
 - HoxA9-ER-Lys-GFP cell differentiation
 - May-Grunwald-Giemsa staining
 - siRNA transfection
 - RNA-FISH
 - Lentiviral particles preparation
 - Immunoblots
 - Immunofluorescence
 - Quantitative PCR
 - RNA-seq
 - ChIP-seq
 - ATAC-seq
 - Quality control of sequencing data
 - ChIP-seq mapping and peak calling
 - ChIP-seq differential analysis

- ATAC-seq analysis
- RNA-seq mapping and quantification
- RNA-seq differential analysis
- Visualization of sequence data
- Integrative bioinformatics analysis
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
- **DATA AND SOFTWARE AVAILABILITY**

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and seven tables and can be found with this article online at <https://doi.org/10.1016/j.stem.2018.10.001>.

ACKNOWLEDGMENTS

We greatly thank M. Serrano (IRB, Barcelona) for the i4F mice and for generously sharing unpublished results. We thank T. Graf (CRG, Barcelona) for the pre-B C10 cell line, D. Sykes (MGH, Cambridge) for the Lys-GFP-ER-HoxA9 cell line, and M. Ouspenskaia and M. Dasso (NIH, Bethesda, MD) for the SUMO-2 antibody. We acknowledge J. Jachowicz (IES, Munich) for help with RNA-FISH and P-F. Roux for bioinformatic expertise. Sequencing for RNA-seq/ChIP-seq was performed by the IGBMC Sequencing platform (ANR-10-INBS-0009) and for ATAC-seq by the Institut Pasteur Transcriptome and Epigenome platform. This work was supported by grants from European Research Council (Advanced Research Grant 294341), LNCC (Equipe labellisée), INCa (n°2017-1 PL BIO-05-IP-1), and Odyssey to A.D. and from Association pour la Recherche sur le Cancer (ARC) (PJA 20161205028) and Revive (Laboratoire d'Excellence Revive, Investissement d'Avenir; ANR-10-LABX-73) to H.L. J.-C.C. was supported by INCa, I.T. by MESR and FRM, and X.G. by the Helmholtz Association.

AUTHOR CONTRIBUTIONS

J.-C.C., I.T., and A.D. planned the studies. J.-C.C. and I.T. performed most of the experiments and analyzed the data. C.C. performed the bioinformatic analysis, and G.J., the histological analysis. A.A. contributed to the mice studies, X.G. and A.I. contributed to the 2C-like analyses, S.B.A. and H.L. performed the Ubc9 KD *in vitro* reprogramming and the iNeuron experiments, and G.B. performed the HL60 experiments. J.-S.S. produced the viruses and provided intellectual support. M.-E.T.-P. designed research and interpreted the data. J.-C.C., I.T., and A.D. wrote the paper with input from all co-authors.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 26, 2018

Revised: September 10, 2018

Accepted: October 1, 2018

Published: October 25, 2018

REFERENCES

- Abad, M., Mosteiro, L., Pantoja, C., Cañamero, M., Rayon, T., Ors, I., Graña, O., Megias, D., Domínguez, O., Martínez, D., et al. (2013). Reprogramming *in vivo* produces teratomas and iPS cells with totipotency features. *Nature* 502, 340–345.
- Apostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. *Nature* 502, 462–471.
- Banani, S.F., Rice, A.M., Peeples, W.B., Lin, Y., Jain, S., Parker, R., and Rosen, M.K. (2016). Compositional Control of Phase-Separated Cellular Bodies. *Cell* 166, 651–663.
- Beck, S., Lee, B.K., Rhee, C., Song, J., Woo, A.J., and Kim, J. (2014). CpG island-mediated global gene regulatory modes in mouse embryonic stem cells. *Nat. Commun.* 5, 5490.

- Becker, J.S., Nicetto, D., and Zaret, K.S. (2016). H3K9me3-dependent heterochromatin: Barrier to cell fate changes. *Trends Genet.* **32**, 29–41.
- Borkent, M., Bennett, B.D., Lackford, B., Bar-Nur, O., Brumbaugh, J., Wang, L., Du, Y., Fargo, D.C., Apostolou, E., Cheloufi, S., et al. (2016). A serial shRNA screen for roadblocks to reprogramming identifies the protein modifier SUMO2. *Stem Cell Reports* **6**, 704–716.
- Bussmann, L.H., Schubert, A., Vu Manh, T.P., De Andres, L., Desbordes, S.C., Parra, M., Zimmermann, T., Rapino, F., Rodriguez-Ubreva, J., Ballestar, E., and Graf, T. (2009). A robust and highly efficient immune cell reprogramming system. *Cell Stem Cell* **5**, 554–566.
- Cheloufi, S., Elling, U., Hopfgartner, B., Jung, Y.L., Murn, J., Ninova, M., Hubmann, M., Badeaux, A.I., Euong Ang, C., Tenen, D., et al. (2015). The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528**, 218–224.
- Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative binding of transcription factors orchestrates reprogramming. *Cell* **168**, 442–459.
- Cubeñas-Potts, C., and Matunis, M.J. (2013). SUMO: A multifaceted modifier of chromatin structure and function. *Dev. Cell* **24**, 1–12.
- De Iaco, A., Planet, E., Coluccio, A., Verp, S., Duc, J., and Trono, D. (2017). DUX-family transcription factors regulate zygotic genome activation in placental mammals. *Nat. Genet.* **49**, 941–945.
- Decque, A., Joffre, O., Magalhaes, J.G., Cossec, J.C., Blecher-Gonen, R., Lapaquette, P., Silvín, A., Manel, N., Joubert, P.E., Seeler, J.S., et al. (2016). SUMOylation coordinates the repression of inflammatory and anti-viral gene-expression programs during innate sensing. *Nat. Immunol.* **17**, 140–149.
- Demarque, M.D., Nacerddine, K., Neyret-Kahn, H., Andrieux, A., Danenberg, E., Jouvion, G., Bomme, P., Hamard, G., Romagnolo, B., Terris, B., et al. (2011). Sumoylation by Ubc9 regulates the stem cell compartment and structure and function of the intestinal epithelium in mice. *Gastroenterology* **140**, 286–296.
- Deng, Q., Ramsköld, D., Reinius, B., and Sandberg, R. (2014). Single-cell RNA-seq reveals dynamic, random monoallelic gene expression in mammalian cells. *Science* **343**, 193–196.
- Dobin, A., and Gingeras, T.R. (2015). Mapping RNA-seq reads with STAR. *Curr. Protoc. Bioinformatics* **51**, 11–19.
- Flotho, A., and Melchior, F. (2013). Sumoylation: A regulatory protein modification in health and disease. *Annu. Rev. Biochem.* **82**, 357–385.
- He, X., Riceberg, J., Soucy, T., Koenig, E., Minissale, J., Gallery, M., Bernard, H., Yang, X., Liao, H., Rabino, C., et al. (2017). Probing the roles of SUMOylation in cancer cell biology by using a selective SAE inhibitor. *Nat. Chem. Biol.* **13**, 1164–1171.
- Hendrickson, P.G., Doráis, J.A., Grow, E.J., Whiddon, J.L., Lim, J.W., Wike, C.L., Weaver, B.D., Pflueger, C., Emery, B.R., Wilcox, A.L., et al. (2017). Conserved roles of mouse DUX and human DUX4 in activating cleavage-stage genes and MERVL/HERVL retrotransposons. *Nat. Genet.* **49**, 925–934.
- Hendriks, I.A., and Vertegaal, A.C. (2016). A comprehensive compilation of SUMO proteomics. *Nat. Rev. Mol. Cell Biol.* **17**, 581–595.
- Hnisz, D., Shrivinas, K., Young, R.A., Chakraborty, A.K., and Sharp, P.A. (2017). A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13–23.
- Ishiyoshi, T., Enriquez-Gasca, R., Mizutani, E., Bošković, A., Ziegler-Birling, C., Rodriguez-Terrones, D., Wakayama, T., Vaquerizas, J.M., and Torres-Padilla, M.E. (2015). Early embryonic-like cells are induced by downregulating replication-dependent chromatin assembly. *Nat. Struct. Mol. Biol.* **22**, 662–671.
- Ivanov, A.V., Peng, H., Yurchenko, V., Yap, K.L., Negorev, D.G., Schultz, D.C., Psulkowski, E., Fredericks, W.J., White, D.E., Maul, G.G., et al. (2007). PHD domain-mediated E3 ligase activity directs intramolecular sumoylation of an adjacent bromodomain required for gene silencing. *Mol. Cell* **28**, 823–837.
- Kundu, S., Ji, F., Sunwoo, H., Jain, G., Lee, J.T., Sadreyev, R.I., Dekker, J., and Kingston, R.E. (2017). Polycyclic repressive complex 1 generates discrete compacted domains that change during differentiation. *Mol. Cell* **65**, 432–446.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25.
- Liu, H.W., Zhang, J., Heine, G.F., Arora, M., Gulcin Ozer, H., Onti-Srinivasan, R., Huang, K., and Parvin, J.D. (2012). Chromatin modification by SUMO-1 stimulates the promoters of translation machinery genes. *Nucleic Acids Res.* **40**, 10172–10186.
- Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550.
- Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* **487**, 57–63.
- Maison, C., Bailly, D., Quivy, J.P., and Almouzni, G. (2016). The methyltransferase Suv39h1 links the SUMO pathway to HP1 α marking at pericentric heterochromatin. *Nat. Commun.* **7**, 12224.
- Matoba, S., Liu, Y., Lu, F., Iwabuchi, K.A., Shen, L., Inoue, A., and Zhang, Y. (2014). Embryonic development following somatic cell nuclear transfer impeded by persisting histone methylation. *Cell* **159**, 884–895.
- Matsui, T., Leung, D., Miyashita, H., Maksakova, I.A., Miyachi, H., Kimura, H., Tachibana, M., Lorincz, M.C., and Shinkai, Y. (2010). Proviral silencing in embryonic stem cells requires the histone methyltransferase ESET. *Nature* **464**, 927–931.
- Matsumura, Y., Nakaki, R., Inagaki, T., Yoshida, A., Kano, Y., Kimura, H., Tanaka, T., Tsutsumi, S., Nakao, M., Doi, T., et al. (2015). H3K4/H3K9me3 bivalent chromatin domains targeted by lineage-specific DNA methylation pauses adipocyte differentiation. *Mol. Cell* **60**, 584–596.
- Mosteiro, L., Pantoja, C., Alcazar, N., Marión, R.M., Chondronasiou, D., Rovira, M., Fernandez-Marcos, P.J., Muñoz-Martin, M., Blanco-Aparicio, C., Pastor, J., et al. (2016). Tissue damage and senescence provide critical signals for cellular reprogramming in vivo. *Science* **354**, 354.
- Mosteiro, L., Pantoja, C., de Martino, A., and Serrano, M. (2017). Senescence promotes in vivo reprogramming through p16(INK4a) and IL-6. *Aging Cell*. Published online December 27, 2017. <https://doi.org/10.1111/ace1.12711>.
- Nacerddine, K., Lehembre, F., Bhaumik, M., Artus, J., Cohen-Tannoudji, M., Babinet, C., Pandolfi, P.P., and Dejean, A. (2005). The SUMO pathway is essential for nuclear integrity and chromosome segregation in mice. *Dev. Cell* **9**, 769–779.
- Neyret-Kahn, H., Benhamed, M., Ye, T., Le Gras, S., Cossec, J.C., Lapaquette, P., Bischof, O., Ouspenskaia, M., Dasso, M., Seeler, J., et al. (2013). Sumoylation at chromatin governs coordinated repression of a transcriptional program essential for cell growth and proliferation. *Genome Res.* **23**, 1563–1579.
- Niskanen, E.A., Malinen, M., Sutinen, P., Toropainen, S., Paakinaho, V., Vihervaara, A., Joutsen, J., Kaikkonen, M.U., Sistonen, L., and Palvimo, J.J. (2015). Global SUMOylation on active chromatin is an acute heat stress response restricting transcription. *Genome Biol.* **16**, 153.
- Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E., and Guenther, M.G. (2014). Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell Rep.* **9**, 1163–1170.
- Percharde, M., Lin, C.J., Yin, Y., Guan, J., Peixoto, G.A., Bulut-Karslioglu, A., Biechele, S., Huang, B., Shen, X., and Ramalho-Santos, M. (2018). A LINE1-Nucleolin partnership regulates early development and ESC identity. *Cell* **174**, 391–405.
- Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632.
- Psakhye, I., and Jentsch, S. (2012). Protein group modification and synergy in the SUMO pathway as exemplified in DNA repair. *Cell* **151**, 807–820.
- Rodriguez-Terrones, D., Gaume, X., Ishiyoshi, T., Weiss, A., Kopp, A., Kruse, K., Penning, A., Vaquerizas, J.M., Brino, L., and Torres-Padilla, M.E. (2018). A molecular roadmap for the emergence of early-embryonic-like cells in culture. *Nat. Genet.* **50**, 106–119.

- Rose, N.R., King, H.W., Blackledge, N.P., Fursova, N.A., Ember, K.J., Fischer, R., Kessler, B.M., and Klose, R.J. (2016). RYBP stimulates PRC1 to shape chromatin-based communication between Polycomb repressive complexes. *eLife* 5. Published online October 5, 2016. <https://doi.org/10.7554/eLife.18591>.
- Seeler, J.S., and Dejean, A. (2017). SUMO and the robustness of cancer. *Nat. Rev. Cancer* 17, 184–197.
- Seifert, A., Schofield, P., Barton, G.J., and Hay, R.T. (2015). Proteotoxic stress reprograms the chromatin landscape of SUMO modification. *Sci. Signal.* 8, rs7.
- Stielow, B., Finkernagel, F., Stiewe, T., Nist, A., and Suske, G. (2018). MGA, L3MBTL2 and E2F6 determine genomic binding of the non-canonical Polycomb repressive complex PRC1.6. *PLoS Genet.* 14, e1007193.
- Sykes, D.B., Kfoury, Y.S., Mercier, F.E., Wawer, M.J., Law, J.M., Haynes, M.K., Lewis, T.A., Schajnovitz, A., Jain, E., Lee, D., et al. (2016). Inhibition of dihydroorotate dehydrogenase overcomes differentiation blockade in acute myeloid leukemia. *Cell* 167, 171–186.
- Tahmasebi, S., Ghorbani, M., Savage, P., Yan, K., Gocevski, G., Xiao, L., You, L., and Yang, X.J. (2013). Sumoylation of Krüppel-like factor 4 inhibits pluripotency induction but promotes adipocyte differentiation. *J. Biol. Chem.* 288, 12791–12804.
- Vierbuchen, T., and Wernig, M. (2012). Molecular roadblocks for cellular reprogramming. *Mol. Cell* 47, 827–838.
- Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035–1041.
- Vierbuchen, T., Ling, E., Cowley, C.J., Couch, C.H., Wang, X., Harmin, D.A., Roberts, C.W.M., and Greenberg, M.E. (2017). AP-1 transcription factors and the BAF complex mediate signal-dependent enhancer selection. *Mol. Cell* 68, 1067–1082.
- Wei, Y., Zhang, S., Shang, S., Zhang, B., Li, S., Wang, X., Wang, F., Su, J., Wu, Q., Liu, H., and Zhang, Y. (2016). SEA: A super-enhancer archive. *Nucleic Acids Res.* 44, D172–D179.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319.
- Yang, B.X., El Farran, C.A., Guo, H.C., Yu, T., Fang, H.T., Wang, H.F., Schlesinger, S., Seah, Y.F., Goh, G.Y., Neo, S.P., et al. (2015). Systematic identification of factors for provirus silencing in embryonic stem cells. *Cell* 163, 230–245.
- Zhang, Y., Liu, T., Meyer, C.A., Eickhout, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
SUMO-1	Abcam	Cat#ab32058; RRID:AB_778173
SUMO-2/3	Abcam	Cat#ab81371; RRID:AB_1658424
SUMO-2/3	Laboratory of M. Dasso	N/A
Ubc9	Abcam	Cat#ab75854; RRID:AB_1310787
Nanog	Cell Signaling	Cat#8822S; RRID:AB_11220237
H3K4me3	Millipore	Cat#17-614; RRID:AB_11212770
H3K4me1	Abcam	Cat#ab8895; RRID:AB_306847
H3K27ac	Abcam	Cat#ab4729; RRID:AB_2118291
H3K9me3	Abcam	Cat#ab8898; RRID:AB_306848
H3K27me3	Millipore	Cat#07-449; RRID:AB_310624
Oct3/4 (ChIP)	Santa Cruz	Cat#sc8628; RRID:AB_653551
Oct4 (Immunohistochemistry)	Santa Cruz	Cat#sc9081; RRID:AB_2167703
Klf4	Santa Cruz	Cat#sc20691; RRID:AB_669567
Sox2	Millipore	Cat#AB5603; RRID:AB_2286686
IgG	Cell Signaling	Cat#2729S; RRID:AB_1031062
CD11b-PE	Miltenyi Biotec	Cat#130-091-240; RRID:AB_244271
CD11b-APC	Miltenyi Biotec	Cat#130-110-554; RRID:AB_2654667
CD19-APC	BD Bioscience	Cat#550992; RRID:AB_398483
Tuj1	Biologend	Cat#845502; RRID:AB_2566589
β -actin	Sigma	Cat#A1978; RRID:AB_476692
H3	Abcam	Cat#ab24834; RRID:AB_470335
Setdb1	Proteintech	Cat#11231-1-AP; RRID:AB_2186069
Cebpa	Diagenode	Cat#C15410225; RRID:AB_2737367
Cebpb	Abcam	Cat#ab32358; RRID:AB_726796
Rybp/DEDAF	Sigma	Cat#PRS2227; RRID:AB_1847589
L3mbtl2	Active Motif	Cat#39569; RRID:AB_2615062
Pcgf6	Proteintech	Cat#24103-1-AP; RRID:AB_2737369
Ring1b	Abcam	Cat#ab101273; RRID:AB_10711495
H2AK119Ub	Cell Signaling	Cat#8240; RRID:AB_10891618
Spike-in Antibody	Active Motif	Cat#61686; RRID:AB_2737370
Chemicals, Peptides, and Recombinant Proteins		
β -Estradiol	Calbiochem	Cat#3301
ML-792	Takeda Oncology	N/A
Doxycycline	Sigma	Cat#D9891
Retinoic acid	Sigma	Cat#R2625
IL3	Peprtech	Cat#213-13
M-CSF	Peprtech	Cat#3115-02
LIF	Miltenyi Biotec	Cat#130-099-895
KSR	Invitrogen	Cat#10828028
Micrococcal nuclease	Cell signaling	Cat#10011
Tn5 Transposase	illumina	Cat#FC-121-1030
CHIR99021	Miltenyi Biotec	Cat#130-103-926
PD0325901	Miltenyi Biotec	Cat#130-103-923
LIVE/DEAD Fixable Aqua Dead Cell Stain Kit	Thermo Scientific	Cat#L34957

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Spike-in Chromatin	Active Motif	Cat#53083
Latex beads amine-modified polystyrene fluorescent red	Sigma	Cat#L2778
Critical Commercial Assays		
ChIP-IT kit	Active Motif	Cat#53040
Phosphatase alkaline	Sigma	Cat#AB0300
HiPerfect transfection reagent	QIAGEN	Cat#301705
Lipofectamine 2000	Thermo Scientific	Cat#11668-019
Power SYBR Green PCR Master Mix	Thermo Scientific	Cat#4367659
High-Capacity cDNA Reverse Transcription Kit	Thermo Scientific	Cat#4368814
Histofine Simple Stain MAX PO	Microm Microtech	Cat#414141F
MicroPlex Library Preparation Kit v2	Diagenode	Cat#C05010014
Deposited Data		
ChIP-seq, RNA-seq, ATAC-seq	This study	GEO: GSE99009; GEO: GSE115842
Experimental Models: Cell Lines		
R1 mouse ESCs	Laboratory of H. Li	N/A
Mouse EGFP-MERVL ESCs	Ishiuchi et al., 2015	N/A
iPS cells	This study	N/A
HL-60 cells	DSMZ	Cat#ACC3; RRID:CVCL_0002
Pre-B C10 cells	Bussmann et al., 2009	N/A
Lys-GFP/HoxA9-ER cells	Sykes et al., 2016	N/A
Mouse embryonic fibroblasts <i>Ubc9^{Flox/-}/CreERT2</i>	Decque et al., 2016	N/A
Mouse embryonic fibroblasts <i>Ubc9^{+/-}</i>	Demarque et al., 2011	N/A
Mouse embryonic fibroblasts from <i>i4F</i> mouse	Abad et al., 2013	N/A
Experimental Models: Organisms/Strains		
Mouse: <i>i4F</i>	Abad et al., 2013	N/A
Mouse: <i>Ubc9^{+/-}</i>	Demarque et al., 2011	N/A
Oligonucleotides		
ChIP-qPCR primers	Eurofins	Table S7
RT-qPCR primers	Eurofins	Table S7
ATAC-seq barcode primers	Eurofins	Table S7
shRNA	Sigma	Table S7
siRNA	Dharmacon	Table S7
Software and Algorithms		
ImageJ software	N/A	https://imagej.nih.gov/ij/
FlowJo software	N/A	https://www.flowjo.com
Graphpad – Prism software	N/A	https://www.graphpad.com
IGV software	N/A	http://software.broadinstitute.org/software/igv/
fastqc tool	N/A	http://www.bioinformatics.babraham.ac.uk/projects/fastqc/
fastq-mcf.	N/A	https://github.com/ExpressionAnalysis/ea-utils/blob/wiki/FastqMcf.md
MarkDuplicates tool of Picard	N/A	https://broadinstitute.github.io/picard/
phantompeakqualtools	N/A	https://github.com/kundajelab/phantompeakqualtools
clusterProfiler	N/A	https://guangchanyu.github.io/clusterProfiler/
MACS2	Zhang et al., 2008	N/A
bedtools	N/A	https://bedtools.readthedocs.io/en/latest/
bowtie2	Langmead and Salzberg, 2012	N/A
STAR	Dobin and Gingeras, 2015	N/A

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DESeq2	Love et al., 2014	N/A
SIM prediction	N/A	http://elm.eu.org/
Other		
Setdb1 ChIP-seq	Matsumura et al., 2015	GEO: GSE7342
Max ChIP-seq	Beck et al., 2014	GEO: GSE48666
Rybp ChIP-seq	Rose et al., 2016	GEO: GSE83094
Ring1b ChIP-seq	Kundu et al., 2017	GEO: GSE89949
H2AK119Ub ChIP-seq	Kundu et al., 2017	GEO: GSE89949
Fosl2 ChIP-seq	Vierbuchen et al., 2017	GEO: GSE83295
JunD ChIP-seq	Vierbuchen et al., 2017	GEO: GSE83295
Cebpa ChIP-seq	Chronis et al., 2017	GEO: GSE90895
Cebpb ChIP-seq	Chronis et al., 2017	GEO: GSE90895
Runx1 ChIP-seq	Chronis et al., 2017	GEO: GSE90895
Fosl1 ChIP-seq	Chronis et al., 2017	GEO: GSE90895
Embryonic stages RNA-seq	Deng et al., 2014	GEO: GSE45719
Super enhancer archive	Wei et al., 2016	http://sea.edbc.org/
ESC super enhancer list	Whyte et al., 2013	N/A

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for reagents may be directed and will be fulfilled by the lead contact, Anne Dejean (anne.dejean@pasteur.fr).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Animals

All animal studies were conducted under animal study protocols approved by the French Ministry of Research (Cetea 160065). All mice were bred at the Pasteur Institute animal facility under specific pathogen-free conditions and housed in a 12-hour light/12-hour dark cycle conditions. Reprogrammable (i4F) C57BL/6 mice were kindly provided by Manuel Serrano (Spanish National Cancer Research Centre, Madrid, Spain) ([Abad et al., 2013](#)). *Ubc9* heterozygous (*Ubc9^{+/-}*) and *Ubc9^{Fl/-}* C57BL/6 mice were previously described ([Demarque et al., 2011](#)). *In vivo* reprogramming experiments were performed with mice of both sexes and from 2 to 4 months of age.

Cell culture

Primary MEFs were grown in complete medium: DMEM + Glutamax (GIBCO) medium supplemented with 10% FBS and penicillin/streptomycin. 293T cells were maintained and expanded in complete medium. EGFP-MERVL ESCs were maintained in 2i medium: DMEM + Glutamax supplemented with 15% FCS, LIF, 0.1 mM 2-mercaptoethanol, non-essential amino acid, 3 μM CHIR9902, 1 μM PD0325901 and penicillin/streptomycin on gelatin-coated plates as previously described ([Ishiiuchi et al., 2015](#)). R1 ES cells were maintained in KSR (knock-out serum replacement medium, GIBCO): DMEM + Glutamax supplemented with 15% KSR, LIF, 0.1 mM, 2-mercaptoethanol, non-essential amino acid and penicillin/streptomycin on gelatin-coated plates. Individual iPSCs clones were isolated and expanded in KSR medium over Mitomycin C-treated fibroblasts plated on gelatin-coated plates (feeder cells). Pre-B (C10, gift from T. Graf) and HL-60 cells were maintained in RPMI (GIBCO) medium supplemented with 10% of FBS heat inactivated for 30 minutes at 56°C, 50 μM 2-mercaptoethanol, and penicillin/streptomycin. *HoxA9-ER/Lys-GFP* cells were maintained in RPMI supplemented with 10% FBS cells, 0.5 μM estradiol, penicillin/streptomycin and 2% conditioned media generated from a CHO cell line that secretes SCF. Cell line identity was validated by routine PCR analysis and mycoplasma detection tests were conducted routinely to ensure mycoplasma free conditions throughout the study.

METHOD DETAILS

Cell reprogramming

Primary MEFs for reprogramming experiments were isolated at 13.5 dpc. Pregnant reprogrammable i4F females were sacrificed by cervical dislocation, male and female embryos were removed and the fetal liver and head were excised. The rest of embryonic tissue

was minced and incubated with 0.25% trypsin-EDTA (Sigma-Aldrich) for 15 min at 37 degrees. MEFs were expanded and frozen at passage 1 for subsequent experiments. For doxycycline-induced cell reprogramming, 2.5×10^5 *i4F* MEFs were seeded in triplicates on 6-well plates. 24h later, cells were infected twice with control or Ubc9 shRNA lentivirus containing-medium in the presence of polybrene (8 μ g/ml). Cells were kept in KSR medium with 1 μ g/ml of doxycycline. For reprogramming upon treatment with the sumoylation inhibitor ML-792, cells were plated and treated for 8 hours at the indicated time points with 1 μ M of ML-792. To assess reprogramming efficiency, plates were stained with alkaline phosphatase (AP detection kit, Sigma-Aldrich) 10 days post-induction and the number of colonies was quantified using ImageJ.

In vivo reprogramming

We generated the *Ubc9* heterozygous reprogrammable mouse line *i4F;Ubc9^{+/-}* by crossing the reprogrammable mouse line *i4F-A* (Abad et al., 2013) to our *Ubc9^{+/-}* mouse line (Demarque et al., 2011). To induce *in vivo* reprogramming doxycycline was administered at 0.2 mg/mL for 2.5 weeks or at 1 mg/ml for 1 week in the drinking water supplemented with 7.5% sucrose. For the survival experiments, mice were sacrificed when reaching the humane endpoints as defined by the Pasteur ethics committee. Operators were blind to the animal's genotype. Tissue samples were fixed in 10% formalin and paraffin embedded. 5 μ m sections were cut and stained with hematoxylin and eosin. For immunohistochemistry, tissues were rehydrated in PBS, saturated with 3% BSA (Sigma Aldrich) and permeabilized with 0.5% Triton X-100 (Sigma Aldrich). Oct4 primary antibody (SC9081, Santa cruz) was incubated overnight at 4°C (dilution: 1/300). Primary antibody was visualized with the Histofine Simple Stain MAX-PO (414141F, Microm Microtech) and color was developed with 3-Amino-9-EthylCar-bazole (Sigma). Cell counting was done by operators blind to the animal's genotype or treatment.

Induced neuronal transdifferentiation

Briefly, 2×10^5 Rosa26-rtTA-expressing MEFs were plated in poly-L-ornithine-coated 6 wells and 24h after, infected 3 times during 2 consecutive days with control or Ubc9 shRNA lentivirus containing-medium. To induce neuronal transdifferentiation cells were infected overnight with *Ascl1*, *Brn2* and *Myt1l* doxycycline inducible lentiviruses. Cells were kept for 48h in complete medium with doxycycline (2 μ g/ml) and subsequently switched to B27-supplemented N3 medium (GIBCO) with doxycycline. 5 days after, cells were harvested for RNA isolation or fixed for Beta 3 Tubulin (Tuj1) (1/100, BioLegend 845502) immunofluorescence.

Pre-B to macrophage transdifferentiation

Pre-B cells were seeded at a density of 7×10^4 cells/mL in culture medium and transfected with siCtrl or siUbc9 using lipofectamine 2000 according to manufacturer protocol. The next day transdifferentiation was induced by supplementing the medium with 10 ng/ml IL3, 10 ng/ml M-CSF and 100 nM β -estradiol. For antibody staining, cells were incubated in PBS for 30 minutes with a LIVE/DEAD Fixable Aqua Dead Cell Stain (1/1000) to mark dead cells then incubated with directly PE conjugated anti-CD11b and directly APC conjugated anti-CD19 antibodies (1/100) for 30 min in PBS and fixed for 5 minutes in 4% para-formaldehyde. For phagocytosis assay, 2×10^5 cells were incubated with red fluorescent latex beads (10 beads per cell) for 30 min at 37°C in culture medium. Cells were then collected and incubated in PBS for 30 minutes with a LIVE/DEAD Fixable Aqua Dead Cell Stain (1/1000) and fixed for 5 minutes in 4% para-formaldehyde. Flow-cytometry was performed on a MACSQuant Analyzer and results analyzed with Flowjo software.

HL-60 to granulocyte differentiation

HL-60 cells were infected with PLK0 lentiviruses expressing a control shRNA (Sigma, SHC002) or 3 different shRNA targeting Ubc9 (#shRNA1: NM_003345.3-194S1C1, #shRNA2: NM_003345.3-538S21C1, #shRNA3: NM_003345.3-545S1C1). Cells were selected with puromycin (1 μ g/ml) for 3 weeks. Then, cells were seeded at 3×10^5 cells/ml and treated with vehicle (DMSO) or retinoic acid (Sigma, R2625) 1 μ M for 6 days. Cells were splitted every 3 days. For cytometry analysis 10^5 cells were labeled with CD11b-APC antibody for 30 minutes and analyzed on a Fortessa cytometer (Beckman Coulter). CD11b expression analysis were performed using FlowJow software after gating on the population of living cells (FSC/SSC). For proliferation analysis, living cells (between 10 and 16 μ m diameter) were counted every 3 days using a Z2 Particle Counter (Beckman Coulter).

HoxA9-ER-Lys-GFP cell differentiation

Differentiation was induced by seeding *HoxA9-ER/Lys-GFP* cells at density of 1.7×10^5 cells/mL in culture medium without estradiol. Cells were collected at indicated time points for RT-qPCR or flow-cytometry experiments. For flow cytometry cells were incubated in PBS for 30 minutes with a LIVE/DEAD Fixable Aqua Dead Cell Stain (1/1000). Flow-cytometry was performed on a MACSQuant Analyzer and results analyzed with Flowjo software.

May-Grunwald-Giemsa staining

Pre-B and HL-60 smears were incubated for 3 min with the May-Grunwald stain (Cat#63590 Sigma) followed by 15 min incubation with the 1:6 Giemsa:distilled water solution (Cat#GS500, Sigma). After washes, coverslips were mounted and images were collected on an inverted Zeiss observer2.1 Apotome.2 using the Plan Apo Chromat 40X objective.

siRNA transfection

ESCs and Pre-B cells were transfected with siRNA (Table S7; Rodriguez-Terrones et al., 2018) using HiPerfect or Lipofectamine transfection reagent following manufacturer's protocol. After 2 days, cells were collected for flow-cytometry, immunoblot and RNA-extraction. FACS Calibur (BD Biosciences) was used to quantify the population of MERVL GFP-positive cells.

RNA-FISH

RNA-FISH for MERVL was carried out in wild-type mouse ESCs after transfection with control or Ubc9 siRNA. RNA-FISH was performed as described elsewhere (Ishuchi et al., 2015). Cells were cultured on coverslips coated with laminin-511 (BioLamina). Cot1 DNA was omitted in the hybridization buffer. Probe for MERVL corresponding to the 250 bp fragment in the gag-coding region was prepared by PCR with TAMRA-labeled dATP. RNase A (1 mg/ml in PBS) treatment was performed at 37°C for 30 min after the permeabilization step, as a negative control. Images were collected on an inverted TCS SP5 confocal microscope (Leica).

Lentiviral particles preparation

8.5 million of 293T cells were plated and immediately transfected using calcium-based transfection with constructs encoding different small hairpins (Table S7). The next day medium was changed. 2 days after transfection, a total of 4 collections of supernatants was done by collecting twice a day the supernatants further filtered through a 0.45 μ m filter. Supernatants were aliquoted, frozen at -80°C and directly added to the culture medium for different experiments.

Immunoblots

Cells were collected and directly lysed in Laemmli buffer. Proteins were quantified following manufacturer's instructions using Pierce solution supplemented with Ionic Detergent Compatibility Reagent (Thermo Scientific). Equal amounts of proteins were loaded for immunoblotting and good equilibration of the different samples was assessed by Ponceau staining after membrane transfer. Antibodies against SUMO-1 (1/1000, Abcam Ab32058), SUMO-2 (1/1000, Abcam, Ab81371), Ubc9 (1/1000, Abcam, Ab75857), Nanog (1/1000, Cell Signaling, 8822), H3K9me3 (1/1000, Abcam, Ab8898), H3K27me3 (1/1000, Millipore, 07-449), H3 (1/1000, Abcam, Ab24834) and Actin (1/5000, Sigma, A1978) were used according to standard protocols and supplier's recommendations.

Immunofluorescence

Cells were cultured on coverslips coated with gelatin and fixed with 2% or 4% paraformaldehyde in PBS for 10 min at room temperature. After the fixation, cells were permeabilized with 0.2% Triton X-100 in PBS for 10 min at room temperature and incubated in blocking buffer (3% BSA in PBS) for 30 min at room temperature. Cells were incubated with primary antibodies diluted in the blocking buffer for 1–1.5 h at room temperature and washed with PBS-T (PBS containing 0.02% Triton X-100) three times. Cells were then incubated with Alexa-conjugated secondary antibodies (Life Technologies) diluted in the blocking buffer for 1–1.5 h at room temperature, washed with PBS-T three times and mounted in Vectashield solution (Vector Labs). Images were collected on an inverted TCS SP5 confocal microscope (Leica).

Quantitative PCR

cDNA was generated from 1 μ g total RNA purified by Trizol extraction with a High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Quantitative real-time PCR analysis was performed with Power SYBR Green master mix (Thermo Scientific) and the primer sets (Table S7) using cDNA or genomic DNA (local ChIP). Quantitative real-time PCR was performed on a CFX96 PCR system (Bio-Rad).

RNA-seq

Total RNA was purified by Trizol extraction, clean-up was done using RNeasy kit (QIAGEN) and RNA was analyzed on a BioAnalyzer Nano chip (Agilent). If the RNA integrity number (RIN) was superior to 8, samples were used for subsequent analyses. RNA concentration was quantified with Qubit (Thermo Fisher Scientific). Libraries of template molecules suitable for strand specific high throughput DNA sequencing were created using "TruSeq Stranded Total RNA with Ribo-Zero Gold Prep Kit" (# RS-122-2301, Illumina). Briefly, starting with 300 ng of total RNA, the first step involved the removal of cytoplasmic and mitochondrial ribosomal RNA (rRNA) using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal beads. Following purification, the RNA was fragmented into small pieces using divalent cations under elevated temperature. The cleaved RNA fragments were copied into first strand cDNA using reverse transcriptase and random primers, followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. The double stranded cDNA fragments were blunted using T4 DNA polymerase, Klenow DNA polymerase and T4 PNK. A single 'A' nucleotide was added to the 3' ends of the blunt DNA fragments using a Klenow fragment (3' to 5' exo minus) enzyme. The cDNA fragments were ligated to double stranded adapters using T4 DNA ligase. The ligated products were enriched by PCR amplification (30 s at 98°C; [10 sec at 98°C, 30 sec at 60°C, 30 sec at 72°C] x 12 cycles; 5 min at 72°C). Then surplus PCR primers were removed by purification using AMPure XP beads (Agencourt Biosciences Corporation). Final cDNA libraries were checked for quality and quantified using 2100 Bioanalyzer (Agilent). The libraries were sequenced on Illumina HiSeq 4000 sequencer as paired-end 100 base reads following Illumina's instructions. Image analysis and base calling were performed using RTA 2.7.3 and bcl2fastq 2.17.1.14.

ChIP-seq

Cells were fixed for 10 min at room temperature in culture medium with formaldehyde (1% final concentration). Formaldehyde was then quenched with glycine (125 mM final). Cells were washed in ice cold PBS. For histone marks and SUMO, the extracted chromatin was sonicated with a Bioruptor Pico (Diagenode) until chromatin fragments reached a size of 200–500 base pairs (30 sec ON, 30 sec OFF, 8 cycles), as assayed by electrophoresis through agarose gels. For TFs, the extracted chromatin was digested by the micrococcal nuclease for 15 min at 37°C, (#10011, Cell signaling) and sonicated to break nuclear membrane with a Bioruptor Pico (30 sec ON, 30 sec OFF, 2 cycles). Immunoprecipitation, reversal of cross-linking and DNA purification were performed using ChIP-IT kit (#53040, Active Motif). Polyclonal antibodies against SUMO-2/3 (provided by M. Dasso), SUMO-1 (Abcam, Ab32058), H3K4me3 (Millipore, #17-614), H3K4me1 (Abcam, Ab8895), H3K27ac (Abcam, Ab4729), H3K9me3 (Abcam, Ab8898), H3K27me3 (Millipore, 07-449), Oct3/4 (Santa Cruz Biotechnology, sc8628), Sox2 (Millipore, AB5603), Klf4 (Santa Cruz Biotechnology, sc20691), Setdb1 (Proteintech, 11231-1-AP) and IgG (Cell Signaling, 2729) were used for ChIP. 50 ng of spike-in chromatin and 2 µg of spike-in antibody (Active Motif) were added to normalize without bias the H3K9me3 ChIP-seq from siCtrl and siUbc9 ESC samples.

ChIP-seq libraries were prepared using Microplex Library Preparation kit V2, (C05010014, Diagenode) following the manufacturer's protocol (V2 02.15) with some modifications. Briefly, in the first step, 10 ng of double-stranded ChIP enriched DNA or input DNA was repaired to yield molecules with blunt ends. In the next step, stem-loop adaptors with blocked 5' ends were ligated to the 5' end of the genomic DNA, leaving a nick at the 3' end. In the third step, the 3' ends of the genomic DNA were extended to complete library synthesis and Illumina-compatible indexes were added through a high-fidelity amplification. In an additional step, the libraries were size selected (200–400bp) and cleaned-up using Agencourt AMPure XP beads (#A63881, Beckman). Prior to analyses DNA libraries were checked for quality and quantified using a 2100 Bioanalyzer (Agilent). The libraries were sequenced on Illumina HiSeq 4000 sequencer as single-end 50 base reads following Illumina's instructions. Image analysis and base calling were performed using RTA 2.7.3 and bcl2fastq 2.17.1.14.

ATAC-seq

50,000 cells were spun down and resuspended in cold lysis buffer (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP40) to isolate nuclei. Then, nuclei were treated with 2.5 µL Tn5 Transposase (#FC-121-1030, Illumina) for 30 min at 37°C. Elute transposed DNA was PCR-amplified using barcode primers (Table S7). After purification, libraries were validated using DNA bioanalyzer assay (Agilent) and Qubit measurement (Thermo Fisher Scientific). Libraries were sequenced using paired-end sequencing on the Illumina Hi-seq 2500 platform (100 cycles). ATAC-seq were done in triplicates for each condition.

Quality control of sequencing data

The quality of every library was determined using the fastqc tool (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Reads were subsequently trimmed and adapters clipped using the fastq-mcf. (<https://github.com/ExpressionAnalysis/ea-utils/blob/wiki/FastqMcf.md>). Only reads with none of the known high-throughput sequencing adapters, longer than 25 base pairs, with a mean quality score above 30 and maximum 1 N-call were kept.

ChIP-seq mapping and peak calling

High quality single end reads of both immunoprecipitated (IP) and input libraries were mapped to the *Mus musculus* reference genome (mm9, NCBI Build 37) using bowtie1 (Langmead et al., 2009). Single-end reads mapping uniquely with no more than 2 mismatches in the best alignment stratum were kept for further analysis. In order to avoid PCR amplification biases in read quantification, duplicated reads were removed using the MarkDuplicates tool of Picard (<https://broadinstitute.github.io/picard/>). Blacklisted regions likely to give false positives in the peak calling step were removed with bedtools (<https://bedtools.readthedocs.io/en/latest/>) and the degree of IP fragment clustering was assessed by estimating the cross correlation scores with phantompeakqualtools (<https://github.com/kundajelab/phantompeakqualtools>). Peak calling was performed with MACS2 (Zhang et al., 2008) with default parameters for regular peak calling and using the input library for the estimation of the significance of peak enrichment. Consistency of peak calling among biological replicates was determined using the Irreproducible Discovery Rate (IDR) approach (<https://sites.google.com/site/anshulkundaje/projects/idr>). The final set of peaks was obtained using the IDR threshold recommended by ENCODE to select the top reproducible peaks called with the pooled IP libraries.

This procedure was applied to the two biological replicates of (1) H3K27ac, H3K4me1, H3K4me3, H3K27me3, H3K9me3, SUMO-1, SUMO-2 in MEFs and ESCs. (2) H3K27ac, H3K4me1, H3K4me3, H3K27me3, H3K9me3, Oct4, Klf4, Sox2 (O, K, S) at D4 during reprogramming in MEFs infected with shCtrl and shUbc9. Peaks were labeled according to their position with respect to the UCSC mm9 transcript annotations. (1) Transcription start site (TSS): for peaks overlapping the 1kb neighborhood centered in any TSS. (2) Intragenic: for peaks overlapping any transcript, excluding the 500bp belonging to the TSS neighborhood. (3) Intergenic: for all other peaks.

ChIP-seq differential analysis

Complete-linkage clustering of SUMO-1 and SUMO-2 ChIP-seq samples was done using the Euclidean distance over the matrix of read counts. The set of peaks used to compare the marking/binding between two conditions correspond to the union of the IDR-selected peaks found in both conditions. Mark or binding intensity was estimated by counting the number of IP reads overlapping these peaks per condition in each replicate. Read counts were transformed into continuous values using the voom method allowing

the use of the limma and empirical Bayes pipeline for the differential analysis. The cyclic local regression (loess) normalization was used to remove the variation due to differences in library size. The following comparisons were performed. (1) MEFs versus ESCs for SUMO-1 and SUMO-2. (2) shCtrl versus shUbc9 at D4 for O, K, S. A spiked-in normalization was used for the comparison of H3K9me3 between siCtrl and siUbc9 in ESCs. The normalization factor was calculated as described in (Orlando et al., 2014) and using *Drosophila melanogaster* as the exogenous genome.

ATAC-seq analysis

High quality paired-end reads were mapped to the *Mus musculus* reference genome (mm9, NCBI Build 37) using the end-to-end mode and the very-sensitive parameterization of bowtie2 (Langmead and Salzberg, 2012), keeping concordant pairs even if they dovetail and with a maximum fragment size of 2 kb. In order to avoid PCR amplification biases in read quantification, duplicated reads were removed using the MarkDuplicates tool of Picard. Retained pairs were then classified according to their correspondent fragment size and analysis was restricted to fragment sizes sufficiently sampled in the ATAC-seq library by selecting read pairs coming from fragments between 50 bp and 2 kb. All the downstream analysis was aimed at quantifying the nucleosome free regions and therefore only pairs from fragments of 100 bp or less were kept. This procedure was applied to the three biological replicates of 4D cells in shCtrl and shUbc9 conditions and three biological replicates of MEFs in Ubc9^{+/+} and Ubc9^{-/-} condition.

RNA-seq mapping and quantification

High quality paired-end reads were mapped to the *Mus musculus* reference genome (mm9, NCBI Build 37) using STAR (Dobin and Gingeras, 2015). The 2-pass mapping mode was applied, doing a first pass with usual parameters to detect junctions and use them, together with the annotated junctions, in a second pass. Only the best unique matches were kept. In order to avoid PCR amplification biases in read quantification, duplicated reads were removed using the MarkDuplicates tool of Picard. Read summarization over annotations was done with the featureCounts function of the Rsubread R package requiring both read ends to map. For genes, the in build mm9 annotation was used and reads falling over exons were counted and summarized per gene annotation. For repetitive regions, the mm9 UCSC repeat annotation was used and reads were summarized per repeat family adding up the counts for all annotations belonging to the same repeat family. For reprogramming resistant regions, a manually curated annotation was used (Matoba et al., 2014) and reads were summarized per region. Quality assurance clustering and principal component analysis were performed on the read count matrices to identify outlying replicates and discard them from the differential analysis. This procedure was applied to the three biological replicates of MEFs, D4, D7, iPSCs in shCtrl and shUbc9 conditions; and ESCs in siCtrl and siUbc9 conditions.

RNA-seq differential analysis

The standard differential expression analysis of DESeq2 was applied (Love et al., 2014) on the filtered replicates using a simple experimental design of one factor (time point or condition) with two levels. The following comparisons were performed. (1) MEFs versus D4 versus D7 versus ESCs in shCtrl condition. (2) shCtrl MEFs versus shUbc9 MEFs. (3) D4 shCtrl versus shUbc9. (4) D7 shCtrl versus shUbc9. To define transcriptional gene clusters for reprogramming in shCtrl condition, 9835 genes with significant changes in expression between MEFs, D4, D7 or ESCs in normal condition were identified (adjusted $p < 0.005$ and $\text{abs}(\log_2\text{fold}) > 2$). A matrix including the pooled read counts for the three replicates of D4, D7 and ESCs was normalized with DESeq2 and then clustered using the Camberra distance and the Ward method for agglomeration. 20 clusters were identified that optimize the distinction between well-known markers of fibroblasts and pluripotent cells. The final transcriptional clusters were defined manually by merging together similar transcriptional profiles that are also enriched in the same functional categories.

Visualization of sequence data

For the purpose of visualization, ChIP-seq, ATAC-seq and RNA-seq libraries were merged by pooling the biological replicates. Prior to pooling, quality assurance clustering and principal component analysis were performed to identify and discard outlying replicates. The read coverage of the merged libraries was calculated over non-overlapping windows of 50 bp genome wide and library size normalization was applied to allow comparisons of the read coverage between samples: quantile normalization for the ChIP-seq and ATAC-seq datasets; spiked-in normalization for the H3K9me3 ChIP-seq in ESCs, siCtrl and siUbc9 conditions; DESeq2 size normalization for the RNA-seq datasets. This normalized read coverage was used for all the graphical representations: genome browser views, profiles of mark or transcription factor over super enhancers and average DNA accessibility, mark or transcription factor signal over specific peak subsets. Super-enhancer annotations were taken from the SEA archive (Wei et al., 2016) and from a published list for ESCs (Whyte et al., 2013). ChIP-seq data for Setdb1 (GEO: GSE73432) (Matsumura et al., 2015), Max (GEO: GSE48666) (Beck et al., 2014), Rybp (GEO: GSE83094) (Rose et al., 2016), Ring1b and H2AK119Ub (GEO: GSE89949) (Kundu et al., 2017), Fosl2 and JunD (GEO: GSE83295) (Vierbuchen et al., 2017), Cebpa, Cebpb, Fos1 and Runx1 (GEO: GSE90895) (Chronis et al., 2017) were obtained from GEO.

Integrative bioinformatics analysis

All gene ontology enrichment analyses were performed using the Rpackage clusterProfiler (<https://guangchuangyu.github.io/clusterProfiler/>).

The proportion of SUMO peaks found in MEFs and ESCs that overlap by at least 1bp any ERV annotation was determined using mm9 UCSC repeat annotation. This observed value was compared to the expected proportion estimated using a bootstrap approach. The SUMO peaks coordinates were randomized, keeping the same peak size and the number of overlaps with ERV annotation counted. This process was iterated 100 times to obtain an average overlap frequency between randomized SUMO peaks and ERV annotations.

To compare SUMO peaks and genes differentially expressed upon hypo-SUMOylation at D4 and D7, the number of differentially expressed genes falling within the 10 to 50kb neighborhood of any MEF-specific peak were counted and compared to the genome-wide expected value. The expected number of genes was calculated by multiplying the probability of any gene to fall within the given neighborhood (number of genes falling in the SUMO peak neighborhood / total number of genes (21677)) times the total number of differentially expressed genes. Direct effect on gene expression due to SUMO presence at TSS was ignored by filtering out all genes holding a SUMO peak in the 1kb neighborhood surrounding their TSS.

To analyze OKS co-binding, the final sets of identified O, K, S peaks were used and regions that are recognized by one, two or three of these factors in shCtrl and shUbc9 conditions in 4D cells were counted. In each condition, all possible combinations were found and were defined as the OKS, OK, OS, KS, O, K, S peak kinds. Binding dynamics of these transcription factors was compared between shCtrl and shUbc9 conditions by determining the number of all possible transitions among peak kinds between conditions (see Table S5). For the sake of simplicity, subsequent analyses were focused on the regions showing the most frequent transitions that were respectively called: OKS-stable (17% of regions showing OKS -> OKS transition), OKS-gain (14% of regions showing O, S, K, OS, OK, KS -> OKS transition) and OKS-loss (4% of regions showing OKS -> KS, OK, OS, K, S, O transition). To test the similarity between hyposumoylated ESCs and 2C-like cells, the genes and repeat families significantly up- and downregulated in siUbc9 (adjusted $p < 0.005$ and adjusted $p < 0.01$ respectively) were compared to the ones up- and downregulated in 2C-like cells defined as the ESCs spontaneously reverting to 2C-like cells and previously referred as GFP+ cells in a published study (Ishiyuchi et al., 2015). Genes and repeat families were considered significantly up- and downregulated in 2C-like versus ESCs if the adjusted p was lower than 0.005 and 0.01 respectively.

QUANTIFICATION AND STATISTICAL ANALYSIS

The number of independent experimental replications, the definition of center and precisions measures are reported in the figure legends (n, mean + SD). Statistical analyses were performed using the GraphPad Prism v7 software. Statistical significance was assessed by two-tailed t test except when specified in the figure legends. For boxplots, upper and lower whiskers are defined as respectively $Q3 + 1.5 \times IQR$ and $Q1 - 1.5 \times IQR$ with $Q1$ and $Q3$ being the first and third quartile of the plotted distribution and IQR the inter-quartile range. Data are represented as Mean + SD, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

DATA AND SOFTWARE AVAILABILITY

Alkaline phosphatase-positive cells were quantified using ImageJ software (<https://imagej.nih.gov/ij/download.html>). Flow cytometry profile analyses were performed using Flowjo software (<https://www.flowjo.com>). Genome browser views were generated using Integrative Genomics Viewer (<http://software.broadinstitute.org/software/igv/>). All the statistical analyses were performed using the GraphPad Prism v7 software. The accession numbers for the large-scale datasets generated for this paper are GEO: GSE99009 and GEO: GSE115842.