Transient suppression of SUMOylation in embryonic stem cells generates embryo-like structures

Jack-Christophe Cossec, Tatiana Traboulsi, Sébastien Sart, Yann Loe-Mie, Manuel Guthmann, Ivo Hendriks, Ilan Theurillat, Michael Nielsen, Maria-Elena Torres-Padilla, Charles N. Baroud, et al.

▶ **To cite this version:**

Jack-Christophe Cossec, Tatiana Traboulsi, Sébastien Sart, Yann Loe-Mie, Manuel Guthmann, et al.. Transient suppression of SUMOylation in embryonic stem cells generates embryo-like structures. Cell Reports, 2023, 42 (4), pp.112380. 10.1016/j.celrep.2023.112380 . pasteur-04081183

HAL Id: pasteur-04081183
https://pasteur.hal.science/pasteur-04081183

Submitted on 25 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

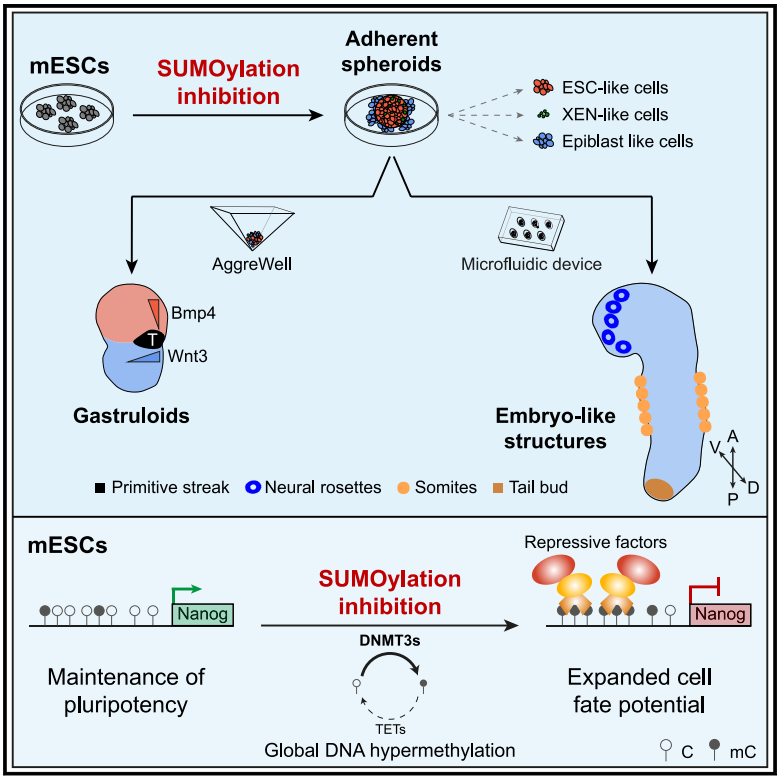
Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

Cell Reports

Article

Transient suppression of SUMOylation in embryonic stem cells generates embryo-like structures

Graphical abstract



Highlights

- SUMOylation levels display cell-to-cell heterogeneity in early mouse embryos
- Chemical SUMOylation inhibition in ESCs generates self-assembled adherent spheroids
- Spheroid cells self-organize into multi-axial embryo-like structures
- Loss of SUMOylation triggers genome-wide DNA hypermethylation and Nanog repression

Authors

Jack-Christophe Cossec, Tatiana Traboulsi, Sébastien Sart, ..., Maria-Elena Torres-Padilla, Charles N. Baroud, Anne Dejean

Correspondence

jcossec@pasteur.fr (J.-C.C.), anne.dejean@pasteur.fr (A.D.)

In brief

Cossec et al. show that chemical inhibition of SUMOylation in mouse embryonic stem cells produces self-assembled embryo-like structures with characteristics resembling a head and a trunk. Suppression of SUMOylation, a general chromatin barrier to cell fate transitions, increases genome-wide DNA methylation, supporting the emergence of cell diversity.

Cossec et al., 2023, Cell Reports *42*, 112380
April 25, 2023 © 2023 The Author(s).
https://doi.org/10.1016/j.celrep.2023.112380

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Transient suppression of SUMOylation in embryonic stem cells generates embryo-like structures

Jack-Christophe Cossec,[1,2,11,*] Tatiana Traboulsi,[1,2,11] Sébastien Sart,[3,4,11] Yann Loe-Mie,[1,2,5] Manuel Guthmann,[6] Ivo A. Hendriks,[7] Ilan Theurillat,[1,2,8,10] Michael L. Nielsen,[7] Maria-Elena Torres-Padilla,[6,9] Charles N. Baroud,[3,4] and Anne Dejean[1,2,12,*]

[1]Nuclear Organization and Oncogenesis Unit, Department of Cell Biology and Infection, Equipe Labellisée Ligue Nationale Contre le Cancer, Institut Pasteur, Université Paris Cité, 75015 Paris, France
[2]INSERM, U993, 75015 Paris, France
[3]LadHyX, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, 91120 Palaiseau, France
[4]Physical Microfluidics and Bioengineering Unit, Department of Genomes and Genetics, Institut Pasteur, 75015 Paris, France
[5]Institut Pasteur, Université Paris Cité, Bioinformatics and Biostatistics HUB, 75015 Paris, France
[6]Institute of Epigenetics and Stem Cells, Helmholtz Zentrum München, 81377 München, Germany
[7]Proteomics Program, Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, 2200 Copenhagen, Denmark
[8]Sorbonne Université, Collège Doctoral, 75005 Paris, France
[9]Faculty of Biology, Ludwig-Maximilians-Universität, 81377 München, Germany
[10]Present address: Max Delbrück Center for Molecular Medicine (MDC), 10115 Berlin, Germany
[11]These authors contributed equally
[12]Lead contact
*Correspondence: jcossec@pasteur.fr (J.-C.C.), anne.dejean@pasteur.fr (A.D.)
https://doi.org/10.1016/j.celrep.2023.112380

SUMMARY

Recent advances in synthetic embryology have opened new avenues for understanding the complex events controlling mammalian peri-implantation development. Here, we show that mouse embryonic stem cells (ESCs) solely exposed to chemical inhibition of SUMOylation generate embryo-like structures comprising anterior neural and trunk-associated regions. HypoSUMOylation-instructed ESCs give rise to spheroids that self-organize into gastrulating structures containing cell types spatially and functionally related to embryonic and extraembryonic compartments. Alternatively, spheroids cultured in a droplet microfluidic device form elongated structures that undergo axial organization reminiscent of natural embryo morphogenesis. Single-cell transcriptomics reveals various cellular lineages, including properly positioned anterior neuronal cell types and paraxial mesoderm segmented into somite-like structures. Transient SUMOylation suppression gradually increases DNA methylation genome wide and repressive mark deposition at *Nanog*. Interestingly, cell-to-cell variations in SUMOylation levels occur during early embryogenesis. Our approach provides a proof of principle for potentially powerful strategies to explore early embryogenesis by targeting chromatin roadblocks of cell fate change.

INTRODUCTION

Lowering chromatin barriers to cell fate change has been widely reported to favor reversion of somatic cells and embryonic stem cells (ESCs) to pluripotent and totipotent-like stages, respectively.[1–5] Conversely, little is known about whether manipulating these barrier functions could work in the other, "forward" direction, and, if so, to what extent. Post-translational modification by the ubiquitin-related SUMO protein, through targeting a large number of chromatin-associated proteins,[6,7] acts as a major regulator of chromatin function that controls wide transcriptional programs.[8–12] We and others have shown that decreasing global cellular SUMOylation facilitates cell fate change regardless of type (reprogramming to pluripotency *in vitro* and *in vivo*,[1,13,14] *trans*-differentiation, and terminal differentiation[13,15]), establishing SUMOylation as a general guardian of cell identity. Moreover, suppression of SUMOylation in mouse ESCs is sufficient to favor their cycling into a two-cell-like (2C-L) state.[13,16,17] We hypothesized that disrupting SUMO's barrier functions through repeated waves of hypoSUMOylation in mouse ESCs would expand cell diversity beyond the 2C-L state.

RESULTS

The SUMO pathway is highly dynamic during early embryogenesis

We first explored possible dynamic changes in the SUMO pathway during early embryonic development *in vivo*. There are three functional paralogs of SUMO in mammals, SUMO1, SUMO2, and SUMO3, of which the latter two (collectively termed

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

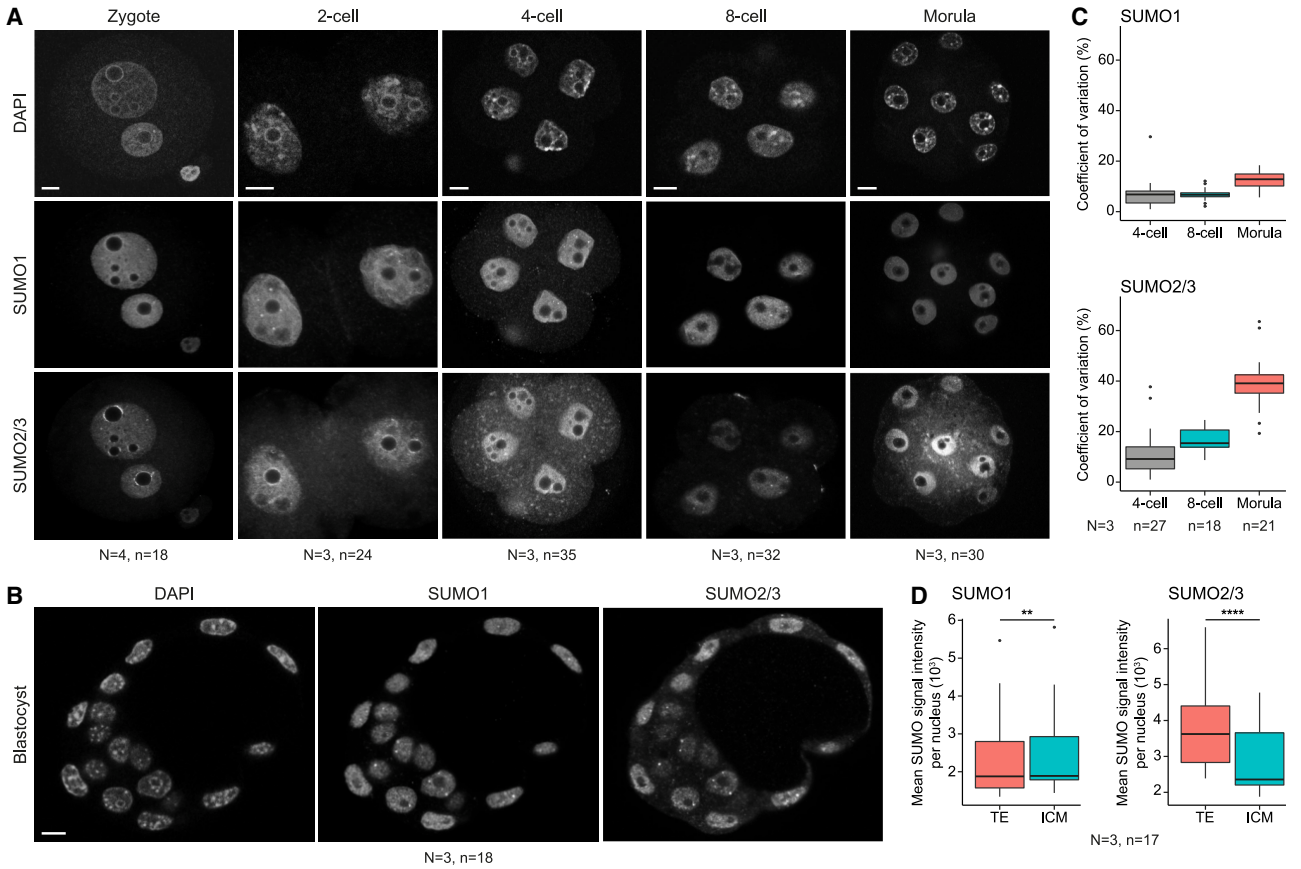


Figure 1. SUMO2/3 displays cell-to-cell heterogeneity at different stages of mouse pre-implantation development

(A and B) Representative single confocal sections of mouse embryos immunostained for SUMO1 and SUMO2/3. Additional examples are shown in Figure S1A.

(C) Coefficient of variation of SUMO intensity per nucleus for each embryo. Embryos with cells in mitosis were excluded from the quantification.

(D) Mean SUMO intensity per nucleus for inner cell mass (ICM) and trophectoderm (TE) for each blastocyst.

**p < 0.01, ****p < 0.0001 (two-tailed paired-sample Wilcoxon test). Scale bars, 10 μm. See also Figure S1.

SUMO2/3) are nearly identical. The SUMO2 paralog is the dominant isoform at early embryonic stages, accounting for about 80% of total SUMO and, in contrast to SUMO1 and SUMO3, is essential for embryonic development, indicating that SUMO2 conjugation indeed prevails in early embryos.[18,19] Identification and comparative analysis of the repertoire of endogenous SUMO2/3 substrates present in each cell type of the different developmental stages by mass spectrometry is so far not feasible, given the large amount of material necessary (10^8–10^9 cells). We thus set out to monitor possible dynamic changes of SUMO protein expression and performed immunostaining for SUMO1 and SUMO2/3 from the zygote to the early blastocyst stage in mice (Figures 1A, 1B, and S1A). The specificity of SUMO1 and SUMO2/3 antibodies has been validated previously.[20] SUMO1 and SUMO2/3 are detectable in cell nuclei at all stages with a diffuse distribution together with accumulation in PML (promyelocytic leukemia) nuclear bodies (Figure S1B). Moreover, SUMO2/3 was enriched at the nucleoli precursors (nucleolar-like bodies) at the zygote and 2-cell stages, consistent with previous work.[21] Remarkably, and in contrast to SUMO1, which showed a rather homogeneous intra-embryonic distribution up to the morula stage, SUMO2/3 displayed large cell-to-cell heterogeneity, which was maximal in the morula (Figures 1A, 1C, and S1A). Simultaneous staining for SUMO1 and SUMO2/3, as performed here, rules out that the heterogeneous SUMO2/3 abundance between cells in the same embryo is due to a technical artifact. In ESCs, 92% of all SUMO2/3 is conjugated to its protein substrates.[7] Moreover, upon ML-792 treatment, when SUMOylation inhibition is almost complete, and the pool of free SUMO2/3 expected to be maximal, we observed a very faint nuclear diffuse signal for unconjugated SUMO2/3, as assessed by immunostaining of blastocysts (Figure S1C). Thus, variations in SUMO2/3 signal intensity mainly reflect variations in conjugated SUMO2/3 and hence represent a reasonable proxy for monitoring changes in global SUMOylation status at the single-cell level. Interestingly, the nuclear intensity of SUMO2/3 also showed an asymmetric lineage distribution at the early blastocyst stage with significantly lower levels in the pluripotent inner cell mass (ICM) than in the trophectoderm (Figures 1B, 1D, and S1A). The opposite trend, albeit less pronounced, was observed for SUMO1. Of note, SUMO immunostaining signals relative to DAPI intensity showed that variations in SUMO2/3 conjugation levels are independent of DNA content

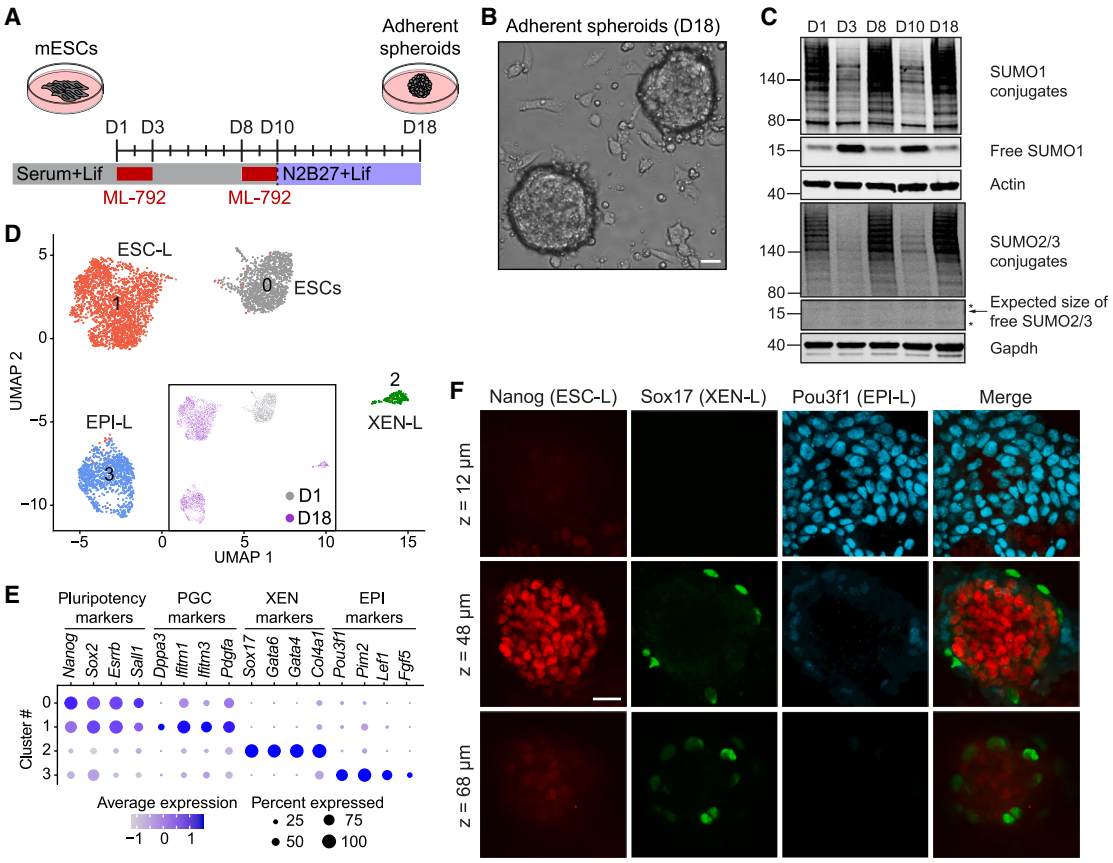


Figure 2. HypoSUMOylation-instructed ESCs generate adherent spheroids

(A) Protocol schematic for generation of adherent spheroids from mouse ESCs.

(B) Representative image of spheroids obtained on day 18.

(C) Immunoblots of SUMO1 and SUMO2/3. Asterisks indicate non-specific bands.

(D) UMAP (uniform manifold approximation and projection) plot of 4,707 cells from day 1 and day 18. Cells are colored by their cluster annotation. The inset shows cells colored by their time point. XEN-L, extraembryonic endoderm-like; EPI-L, epiblast-like.

(E) Gene expression of cluster markers. The size of each dot reflects the percentage of cells in a cluster in which the gene is detected, and the color intensity reflects the average expression level within each cluster. PGC, primordial germ cell.

(F) Confocal sections at three levels of a spheroid after immunostaining for the different cell types. Similar results were obtained in 3 independent experiments. Scale bars, 25 µm. See also Figures S2 and S3 and Table S1.

(Figure S1D). In addition, analysis of previously published single-cell RNA sequencing (scRNA-seq) data generated from mouse embryos revealed highly dynamic expression patterns of the various components of the SUMO machinery across the different developmental stages (Figure S1E). These data strongly suggest a profound rewiring of the repertoire of SUMO target networks during pre-implantation development.

Our analyses reveal that cell-to-cell heterogeneity in SUMOylation levels, particularly for SUMO2/3, occurs at different stages of mouse pre-implantation development, suggesting that this may facilitate cell fate specification *in vivo*.

SUMOylation inhibition induces ESC plasticity

We next assessed the impact of SUMOylation suppression on ESC fate determination. Complete loss of SUMOylation induces apoptosis of ICM cells and embryonic lethality at the early post-implantation stage in mice.[22] To circumvent the cytotoxic effects of prolonged loss of SUMOylation, we subjected ESCs to successive pulses of ML-792, a highly selective inhibitor of the SUMO E1 enzyme.[23] We found that two rounds of 48 h of ML-792 treatment, followed by a shift to N2B27 + Lif medium, yielded large adherent spheroid structures with 100% efficacy on day 18 (Figures 2A and 2B). We confirmed the drop of SUMO1 and SUMO2/3 conjugates on day 3 and day 10, corresponding to the end of the first and second treatment, respectively, whereas inhibitor withdrawal fully restored the level of SUMOylation on day 8 and day 18 (Figure 2C). scRNA-seq revealed that, although day 1 ESCs form a homogeneous population (cluster 0), day 18 spheroids are composed of three distinct cell types (Figure 2D). Using known marker genes and comparisons with a scRNA-seq dataset from peri-implantation embryos,[24] we annotated cluster 2 as extraembryonic endoderm like (XEN-L) and cluster 3 as epiblast like (EPI-L) and found strong similarity of these two clusters to the corresponding cell types in the peri-implantation

embryo (Figures 2E, S2A, and S2B; Table S1). Cluster 1 was enriched for pluripotency markers and overlapped with untreated control ESCs grown under similar conditions (serum + Lif and switch to N2B27 + Lif); we thus referred to this cluster as ESC like (ESC-L) (Figures 2E, S2A, and S2C). Uniform manifold approximation and projection (UMAP) visualization of cell populations obtained in day 18 spheroids compared with those observed in peri-implantation embryos confirmed the identity of the three clusters (Figure S2D). Of note, ESC-L cells strongly expressed a small subset of primordial germ cell (PGC) genes (*Dppa3*, *Ifitm1*, *Ifitm3*, and *Pdgfa*), whereas other *bona fide* PGC markers (*Prdm1* and *Dazl*) were not expressed (Figures 2E and S2A). Importantly, two rounds of SUMOylation inhibition, the switch to N2B27 medium, and the presence of Lif were all necessary for spheroid formation and co-emergence of the three cell types (Figure S2E). Spheroids were formed with a similar efficacy when the interval between the two ML-792 treatments was prolonged, as tested up to 15 days (Figure S2F).

Immunostaining for the three cell types present in the spheroids revealed a 3D organization, with an ESC-L core resting on a basal layer of EPI-L cells and surrounded by XEN-L cells (Figure 2F). We isolated the three cell types by flow cytometry to investigate their interdependence for spheroid formation (Figures S3A–S3C). Although ESC-L cells alone were able to divide, remaining phenotypically stable, sorted fractions of XEN-L and EPI-L cells did not survive on their own (Figures S3D and S3E). Interestingly, mixing equal amounts of ESC-L and EPI-L cells was sufficient to recover spheroid morphology, indicating that XEN-L cells are dispensable for this phenotype (Figure S3E). Cells dissociated from day 18 spheroids were unable to form alkaline phosphatase-positive colonies and failed to activate retinoic acid-responsive genes (Figures S3F and S3G), demonstrating the lack of pluripotency of all spheroid cell types.

Thus, sequential waves of SUMOylation inhibition in ESCs give rise to three cell types capable of self-assembly into adherent spheroids.

HypoSUMOylation-instructed ESCs generate gastruloids

Previous studies have demonstrated that combining ESCs, XENs, and trophoblast stem cells (TSCs) generates embryo-like structures (ELSs) closely recapitulating gastrulation.[25,26] We therefore hypothesized that crosstalk between spheroid cell types, under specific culture conditions, might mimic morphogenetic events akin to early embryogenesis. Transferring 100 spheroid cells to a non-adherent microwell (AggreWell) resulted in elongated structures after 3 days in 83% of wells (Figures 3A and 3B; Video S1). When grown under similar conditions, ESCs that had not been pre-treated with ML-792 failed to elongate and rather formed spherical aggregates in suspension (Figure S4A). Similar aggregates that were unable to elongate were observed when spheroid cells were placed in AggreWells in the absence of Lif (Figure S4B). For the sake of clarity, we will hereafter call the elongated structures obtained with our protocol as AggreWell gastruloids (A-gastruloids). In addition to the 3 cell types detected in spheroids, scRNA-seq analysis of A-gastruloids revealed new cell populations derived from EPI-L cells (Figure 3C). Comparing their transcriptional signatures with a dataset from post-implantation embryos[27] allowed us to identify cell types that correlated with primitive streak (cluster 4), nascent mesoderm (cluster 5), neuromesodermal progenitors (NMPs; cluster 7), neuroepithelium (cluster 8), and a very small number of cells with definitive endoderm identity (cluster 6) (Figures 3C, 3D, S4C, and S4D; Table S1). A more detailed analysis revealed that the transcriptome of the A-gastruloids was most similar to the transcriptome of mouse embryos at embryonic day 6.75 (E6.75) (Figure 3E). Expectedly, none of the new cell clusters listed above were isolated from aggregates of untreated control ESCs (Figure S4A), which formed a single transcriptomic state overlapping with ESC-L cells (Figure S4E).

Immunostaining of the A-gastruloids revealed spontaneous organization with discrete ESC-L- and EPI-L-derived compartments echoing the proximal-distal axis of the post-implantation embryo[28] (Figure 3F). Moreover, the polarized positioning of T-(Brachyury)-positive cells at the border between EPI-L and ESC-L cells recapitulated the symmetry-breaking event that establishes the anteroposterior (A/P) body axis around E6 in mice[28] (Figures 3F and S4F). Remarkably, although lacking key TSC markers such as *Elf5*, *Rhox5*, and *Cdx2* (Table S1), ESC-L cells displayed an extraembryonic ectoderm signature and expressed high *Bmp4* levels (Figures S4D and S4G). In addition, blocking BMP signaling decreased *T* induction (Figure S4H). This suggests that ESC-L cells partly mimic the TSC-derived extraembryonic ectoderm by secreting Bmp4 to activate *Wnt3* expression in EPI-L cells (Figure S4G), contributing to *T* expression and primitive streak establishment.[29] Moreover, the downregulation of epithelial genes (*Cdh1* and *Sox3*) and induction of EMT markers (*Snai1*, *Foxc1*, and *Fn1*) (cluster 5) suggest that *T*-expressing EPI-L cells have undergone epithelial-to-mesenchymal transition (EMT) to form the mesoderm (Figures S4I and S4J). Strikingly, an NMP signature (cluster 7) was prematurely detected in A-gastruloids (Figures 3C and 3E). This was also true for the neuroepithelium (cluster 8), as exemplified by the midbrain-hindbrain marker *En1* expressed on the opposite side of the primitive streak (*in situ* hybridization chain reaction [HCR][30]) (Figure 3G). These findings indicate that our structures are partially desynchronized while maintaining strict regional identities.

Together, these results show that self-organized structures generated from spheroids undergo gastrulating events that resemble those observed in the post-implantation embryo.

Microfluidics-derived ELSs

Microfluidic systems have recently been used to improve multicellular self-organization in controlled environments.[31–33] In an attempt to expand the developmental potential of the spheroids, we implemented a custom-made droplet microfluidic platform optimized for ESC culture (Figures S5A–S5D). Dissociated cells from day 18 spheroids were encapsulated at a density of 120 cells per droplet. Structures grown in this device, hereafter called fluidic gastruloids (F-gastruloids) (Figure 4A), showed a higher degree of axial elongation over time compared with A-gastruloids (Figures 4B and S5E). When grown under similar conditions, control ESCs that had not been pre-treated with ML-792 merely formed spherical aggregates (Figure S5F). To get dynamic insight into the A/P polarization in F-gastruloids,

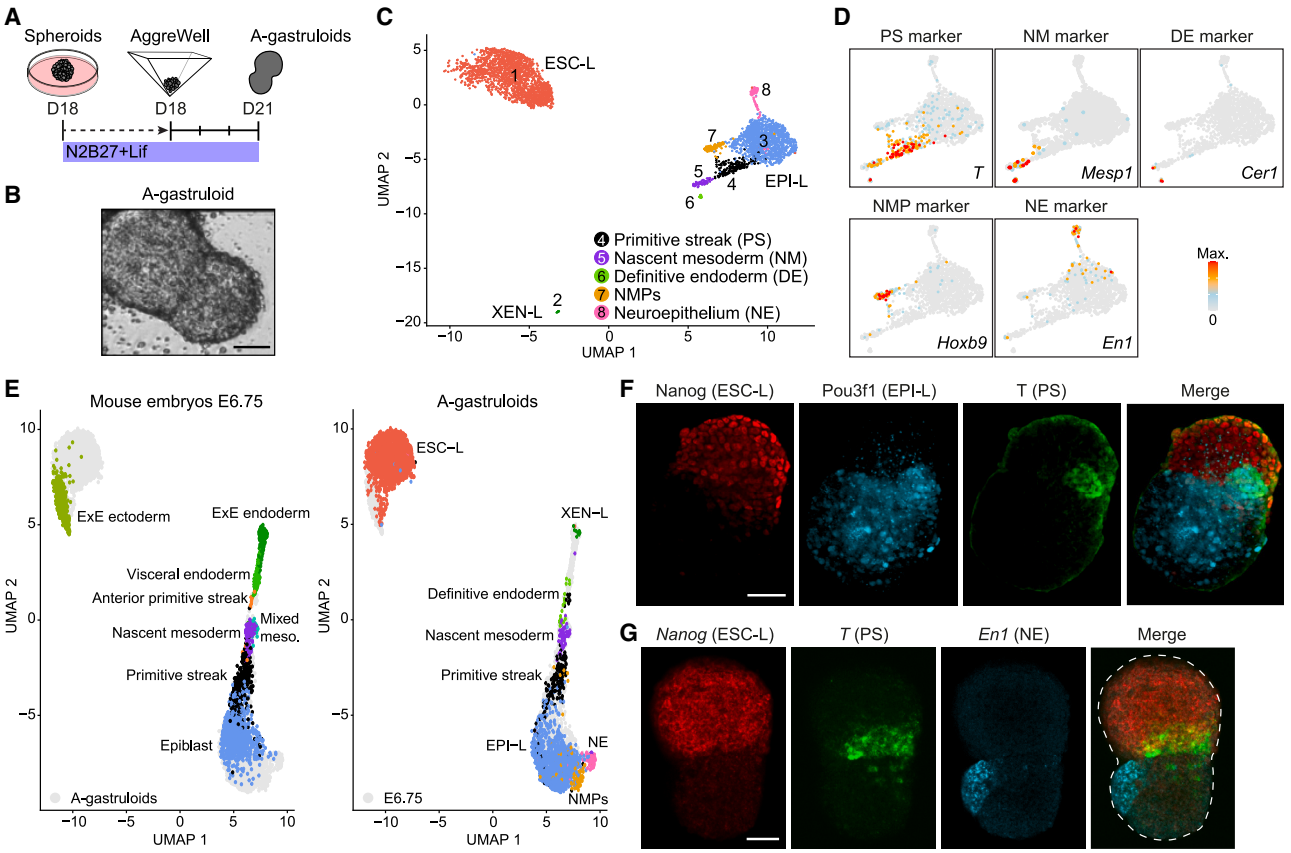


Figure 3. Spheroid cells self-organize into gastruloids under non-adherent cell culture conditions

(A) Protocol schematic for generation of AggreWell gastruloids (A-gastruloids) from dissociated spheroid cells.

(B) Representative image of an A-gastruloid obtained on day 21.

(C) UMAP plot of 3,478 cells isolated from A-gastruloids. Cells are colored by their cluster annotation. NMP, neuromesodermal progenitor.

(D) Magnification of clusters 3–8 of the UMAP plot from (C), showing the expression of key cluster markers.

(E) UMAP plots showing the overlap of 2,075 cells isolated from E6.75 mouse embryos[27] and 3,478 cells isolated from A-gastruloids. Cells are colored by their cell type annotation.

(F and G) Whole-mount immunostaining (F) and *in situ* hybridization (HCR[30]) staining (G) of A-gastruloids (confocal imaging). Similar results were obtained in 3 independent experiments.

Scale bars, 50 μm. See also Figure S4, Table S1, and Video S1.

we used a reporter ESC line to measure spatial and temporal evolution of Sox1 (neuroectoderm) and T (mesoderm) expression. Cells expressing T progressively moved to one pole (likely the posterior end), and the F-gastruloids acquired a neuroectoderm fate, measured by an expanding Sox1-positive domain (likely the anterior end) opposite the T-expressing region (Figure 4C; Video S2; Figure S5G). Immunostaining for Pax6 (neuroectoderm), Foxa2 (definitive endoderm), and T in F-gastruloids cultured for 5 days (F5-gastruloids) confirmed tissue-specific patterning, with Foxa2-positive cells forming an epithelium surrounding a lumen, reminiscent of an embryonic digestive tract (Figure S5H). Thus, F-gastruloids are composed of embryonic cell types from the three germ layers, which are properly organized along the A/P axis.

To improve morphogenetic features in these structures, we embedded F4-gastruloids into 20% Matrigel for 3 days. In 74% of the cases, the resulting structures (called Matrigel ELSs [M-ELSs]) elongated further, reaching a length of more than 0.5 mm (Figures 4B and S5I). Remarkably, M-ELSs adopted a characteristic curvature together with enlargement and thickening at one pole evoking a head-and-trunk morphology (Figure 4B; Video S3; Figures S5J and S5K). We could confirm the reproducibility of spheroid formation as well as gastruloid and M-ELS generation by substituting ML-792 for TAK-981, another highly specific SUMOylation inhibitor[34] (Figures S5L–S5O).

Transcriptional analyses demonstrate embryo-like developmental programs

To further characterize the identity of cells present in the structures, we performed a scRNA-seq analysis in F2- and F5-gastruloids as well as in M-ELSs (Figure 5A). Comparing the 18 resulting clusters with *in vivo* mouse transcriptomics datasets from gastrulation to early organogenesis[27,35] confirmed the presence of cell populations derived from the three germ layers, including ectodermal lineages, mesodermal subtypes, and the gut endoderm (Figures 5A and S6A–S6C; Table S1). Across the three

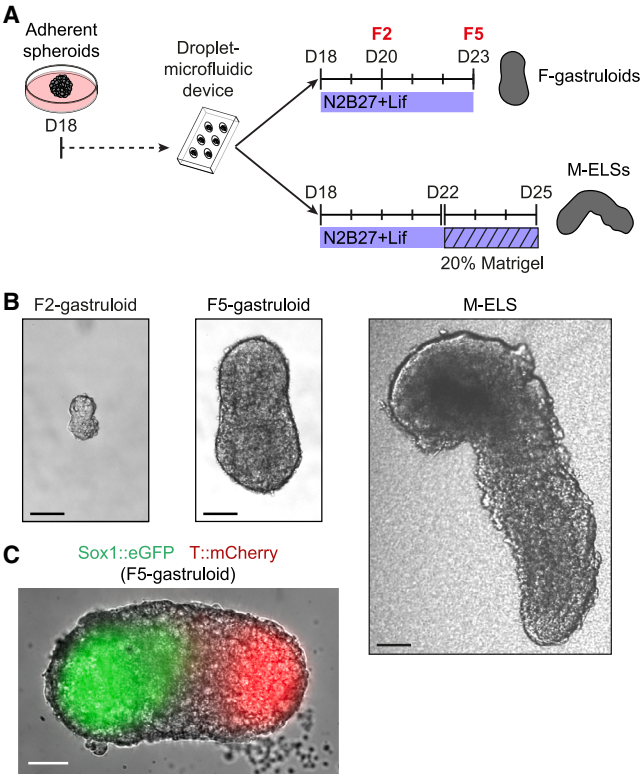


Figure 4. Spheroid cells form ELSs after droplet microfluidic device culture and Matrigel embedding

(A) Protocol schematic for generation of fluidic gastruloids (F-gastruloids) and Matrigel-ELSs (M-ELSs) from dissociated spheroid cells.

(B) Representative images of structures obtained after 2 or 5 days of culture in the droplet microfluidic device (F2- or F5-gastruloids) or after 4 days in the device and 3 days in Matrigel (M-ELSs).

(C) Wide-field image acquired by epifluorescence microscopy of a representative F5-gastruloid obtained from a Sox1:EGFP-T:mCherry ESC line.

Scale bars, 100 μm. See also Figure S5 and Videos S2 and S3.

time points studied, progenitor cells progressively decreased, while more differentiated mesodermal and neural cell types gradually increased, as observed for their *in vivo* counterparts (Figure S6D).

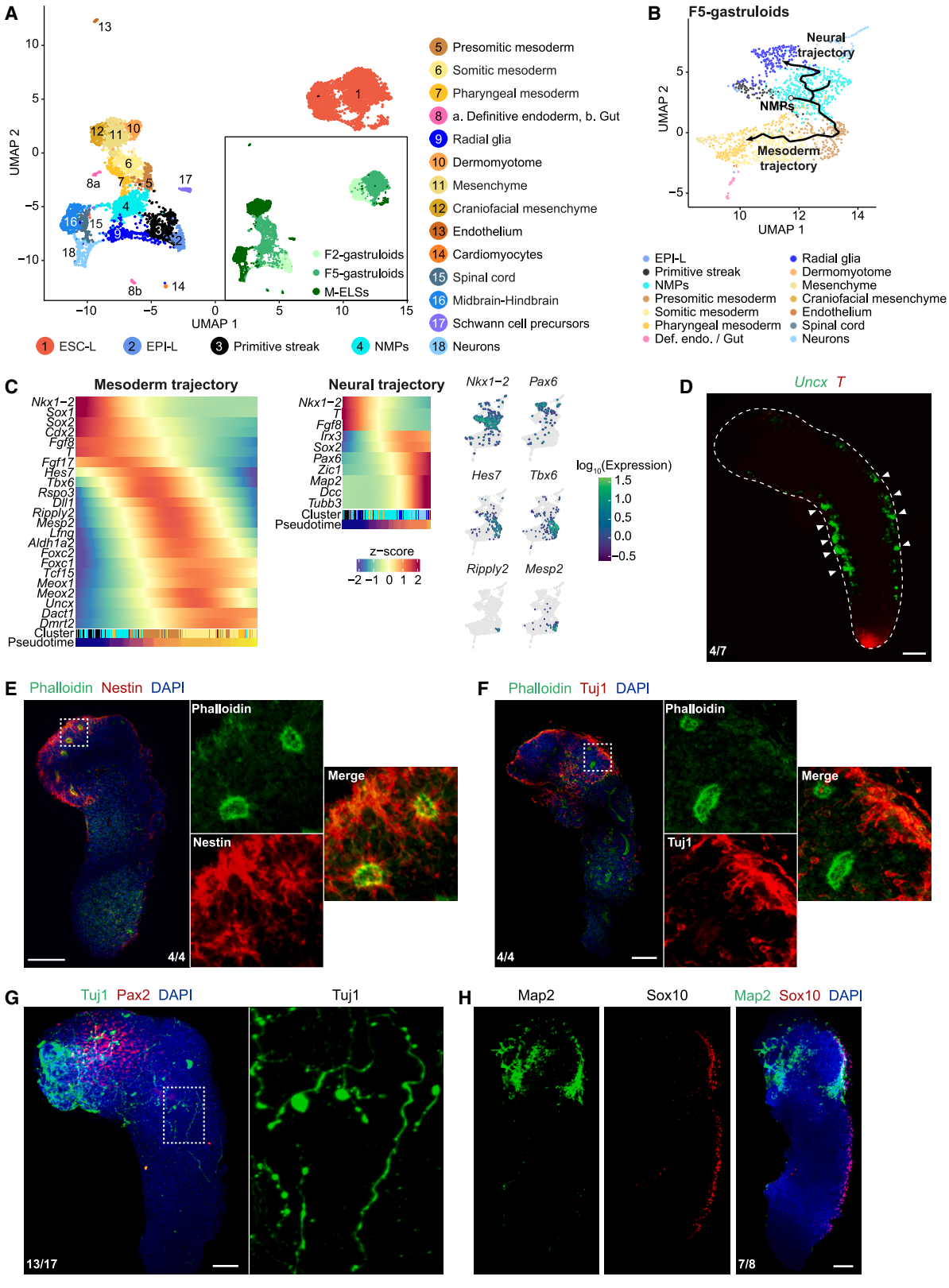
During mammalian trunk development, the bipotent NMPs give rise to spinal cord and paraxial mesoderm derivatives, with Sox2 and T acting antagonistically to specify these two cell fates, respectively.[36] Interestingly, we identified a large cluster (cluster 4), essentially composed of F5-gastruloid cells, with putative NMPs co-expressing *Sox2/T* (Figures 5A, S6A, S6D, and S7A). Cell fate trajectories inferred by pseudotime analysis in F5-gastruloid cells revealed neural and mesodermal developmental routes originating from the NMPs (Figure 5B). Moreover, gene expression dynamics along pseudotime echoes *in vivo* somitic differentiation (Figure 5C). Accordingly, large clusters corresponded to several mesodermal derivatives, including presomitic mesoderm with anterior and posterior identities, somitic mesoderm, dermomyotome, and mesenchymal lineages (Figures 5A and S7B). To visualize putative somitic organization, we performed HCR staining for the specific marker *Uncx* along with *T* and found an average of 3–4 bilateral somite-like segments in over 81% of the elongated M-ELSs (Figures 5D and S7C–S7E), suggesting that these structures undergo somitogenesis. Phalloidin staining revealed proper apicobasal polarity, a feature of epithelial somites (Figure S7F). Moreover, a *T*-labeled region was found at one pole (likely the posterior end) of the M-ELSs, consistent with *T* expression in the growth zone of the tail bud in natural embryos (Figures 5D and S7C). Focusing next on the neural lineage, pseudotime analysis showed progressive expression of neuronal differentiation genes (Figure 5C). In addition, a cluster displaying a spinal cord signature was identified in the M-ELSs (Figures 5A and S7B). These data suggest that the putative NMPs acquired mesodermal and neural fates, reflecting the two differentiation trajectories observed in the embryo.

Additional clusters corresponding to neuroectodermal cell populations were present, such as radial glia and different types of neurons, including inhibitory and excitatory interneuron cell types (Figures 5A, S6B, and S7G). Cluster annotation further revealed a putative midbrain-hindbrain signature, whereas we did not find evidence of expression of forebrain-specific genes (Figure S7H). Co-localization of the midbrain-hindbrain transcription factor Pax2 and the hindbrain marker Gbx2 at the anterior tip of the M-ELSs (Figure S7I) supported anterior neural tissue patterning. Moreover, polarized neural rosette structures, positive for phalloidin and ZO1, were identified at one end (likely the anterior end) of the M-ELSs, consisting of a central lumen surrounded by Nestin-positive neural precursor-like cells (Figures 5E and S7J). The rosette-forming progenitor cells robustly co-expressed the neuroepithelial transcription factors Sox1 and Sox2 (Figure S7K). Conversely, postmitotic neurons along with neurite-like outgrowths were found at the periphery, as shown by Tuj1 immunostaining (Figure 5F). In addition, neurite-like extensions of postmitotic neurons were also observed in the postoccipital region (Figures 5G and 5H). Together, these data indicate the presence of brain tissue along with formation of neurogenic niches at the anterior end of the M-ELSs.

Embryonic dorsoventral (D/V) patterning is orchestrated by balanced instructive signals such as Sonic hedgehog (Shh) and BMPs, which induce ventral and dorsal cell identities, respectively. Cluster identification revealed the absence of notochord, a major source of Shh, and the lack of Shh-dependent ventral cell lineages (spinal motor neurons and sclerotome progenitors). In contrast, presumed dorsal dermomyotome was identified (Figure 5A). In addition, we found a cluster associated with Schwann cell precursors (Figure S6B) and confirmed the presence of these Sox10-expressing cells all along the A/P axis on the convex side of the M-ELSs (Figure 5H). These data are consistent with the localization of Sox10-positive neural crest cells in dorsal root ganglia.[37] In conclusion, while displaying correct A/P patterning, M-ELSs present a predominantly dorsalized embryonic phenotype, likely because of deficient Shh signaling.

SUMOylation inhibition induces global DNA hypermethylation

To gain mechanistic insight into how successive bursts of SUMOylation inhibition in ESCs ultimately lead to generation of structures mimicking some key aspects of early mammalian



(legend on next page)

development, we extended scRNA-seq analysis performed under day 1 to day 3, 8, and 10 conditions (Figure 2A). Clustering analysis revealed a total of 6 different cell states, indicating that treatment with ML-792 induced significant changes in the transcriptome (Figure 6A; Table S1). Clusters 2 and 3 were enriched in pluripotency markers and, together with ESCs (cluster 1), corresponded to the largest clusters (Figure S8A). Other clusters comprised 2C-L cells (cluster 4) and XEN-L cells (cluster 6), whereas cluster 5 did not display any typical signature. Interestingly, analysis across the 4 time points revealed transitioning from ESCs toward new cell states back and forth along with the two-round ML-792 treatment (Figure 6A). These data indicate that cell fate changes triggered by transient loss of SUMOylation are largely reversible in ESCs prior to spheroid formation. However, cells on day 8 recovered a pseudo-ESC transcriptome (cluster 2) rather than the typical ESC program (cluster 1), with 45 genes being differentially expressed (15 up, 30 down) (Table S1). In line with this, day 8 cells showed lower self-renewal capacity than ESCs (Figure S8B). Similarly, whereas the first wave of SUMOylation inhibition (day 3) yielded a large number (35%) of 2C-L cells, the second (day 10) resulted in weaker (19%) induction (Figure S8C), suggesting that the first wave somehow imprints a cellular memory to ESCs that impedes further emergence of 2C-L cells. This finding is in agreement with hyper-repression of the 2C-L cell program on day 8 compared with day 1, as assessed by bulk RNA-seq (Figure S8D). The transition to a 2C-L state was dispensable for formation of spheroids. Indeed, *Dppa2/4* double knockout ESCs, which are unable to convert into 2C-L cells,[38] were still prone to generate adherent spheroids as well as A/F-gastruloids and M-ELSs following transient SUMOylation inhibition (Figure S8E). We verified that treatment with ML-792 of *Dppa2/4* double knockout ESCs failed to restore a 2C-L state (Figure S8F).

To better assess the early impact of SUMOylation loss, we focused on the largest cluster, cluster 3, which emerges at the end of each of the two ML-792 treatments (days 3 and 10). Differential gene expression analysis comparing hypoSUMOylated cells (cluster 3) with ESCs (clusters 1/2) revealed 539 differentially expressed transcripts (317 up, 222 down) (Table S1). Ontology analysis for dysregulated genes showed a strong enrichment in genes involved in p53 signaling and metabolic pathways (Figure S8G). Interestingly, among the top family genes differentially expressed in hypoSUMOylated cells were components of the DNA methylation machinery, with the *de novo* DNA methyltransferase genes *Dnmt3a*, *Dnmt3b*, and *Dnmt3l* being up-regulated, whereas the DNA methylation oxidase genes *Tet1* and *Tet2* were down-regulated (Figure 6B, Table S1; Figure S8H). The variations in expression imposed by each round of ML-792 treatment were confirmed at the protein level (Figure 6C). DNA methylation is essential for mammalian development, and Dnmt3a and Dnmt3b play a crucial role in remethylating the genome following pre-implantation demethylation.[39] Although ESCs can be established and maintained in the absence of DNA methylation, differentiation initiation is almost entirely blocked.[40] To determine whether transient excess in Dnmt3a/3b/3l coupled with decreased Tet1/2 may affect the global pattern of DNA methylation, we performed a methylome analysis (enzymatic Methyl-seq) on days 1, 3, 8, and 10. Remarkably, genome-wide profiling revealed a gradual increase in the methylated CG fraction from day 1 to day 10, with SUMOylation recovery between day 3 and day 8 being unable to reverse this process (Figure 6D). Interestingly, in developing mouse embryos, the global increase in DNA methylation observed from E4.0–E6.5 similarly correlates with a gradual up-regulation of Dnmt3 proteins.[41,42] Together, these findings strongly suggest that the *de novo* hypermethylation induced by hypoSUMOylation in ESCs is directly linked to the activities of the DNA methylation machinery. Of note, ESCs incubated with the DNA methyltransferase inhibitor 5-azacytidine between the two ML-792 treatments retained their undifferentiated morphology and failed to generate adherent spheroids (Figure S8I). Along the same line, female ESCs (LF2) and male ESCs (R1) cultured in 2i medium, both globally hypomethylated,[43,44] are unable to form spheroids (Figures S8J and S8K).

Thus, cyclic chemical inhibition of SUMOylation in ESCs triggers a progressive increase in DNA methylation genome wide, a process reminiscent of that observed between E4.0 and E6.5 *in utero*.

Cumulative repressive marks at *Nanog*

The increase in CG methylation between day 1 and day 10 affected all classes of genomic elements (Figure S9A). Analysis of the most highly differentially methylated regions on day 10 compared with day 1 ($\log_2$FC > 2 or < −2) identified 3,080 regions that were almost exclusively hypermethylated (3,077 of 3,080) and predominantly spanned enhancer elements (Figure 7A; Table S2). Notably, a large fraction of ESC super-enhancers[45] and neighboring regions gained DNA methylation, where it correlated with decreased expression of a subset of associated genes, including the pluripotency genes *Nanog*, *Esrrb*, and *Tbx3* (Figures 7B, 7C, and S9B). We have shown previously that the main role of SUMOylation in controlling cellular plasticity

Figure 5. ELSs with an anterior neural region and a somite-associated domain

(A) UMAP plot of 7,123 cells isolated from F-gastruloids and M-ELSs. Cells are colored by their cluster annotation. The inset shows cells colored by structure type.

(B) UMAP of 3,428 cells from F5-gastruloids colored by their cluster annotation, with neural and mesoderm trajectories inferred from pseudotime analysis.

(C) Heatmaps show scaled expression of genes involved in neural or mesodermal development in the F5-gastruloid cells ordered by pseudotime as defined in (B). UMAP plots on the right show expression of the NMP gene *Nkx1-2*, the neural tube gene *Pax6*, the posterior presomitic mesoderm genes *Hes7* and *Tbx6*, and the anterior presomitic mesoderm genes *Ripply2* and *Mesp2*.

(D) Representative image of *in situ* whole-mount HCR staining for *Uncx* and *T* on M-ELSs (confocal imaging). White arrowheads indicate somite-like segments. Additional examples are shown in Figure S7C.

(E and F) Immunostaining on sections of M-ELSs (confocal imaging).

(G and H) Whole-mount immunostaining of M-ELSs (confocal imaging).

Scale bars, 100 μm. See also Figures S6 and S7 and Table S1.

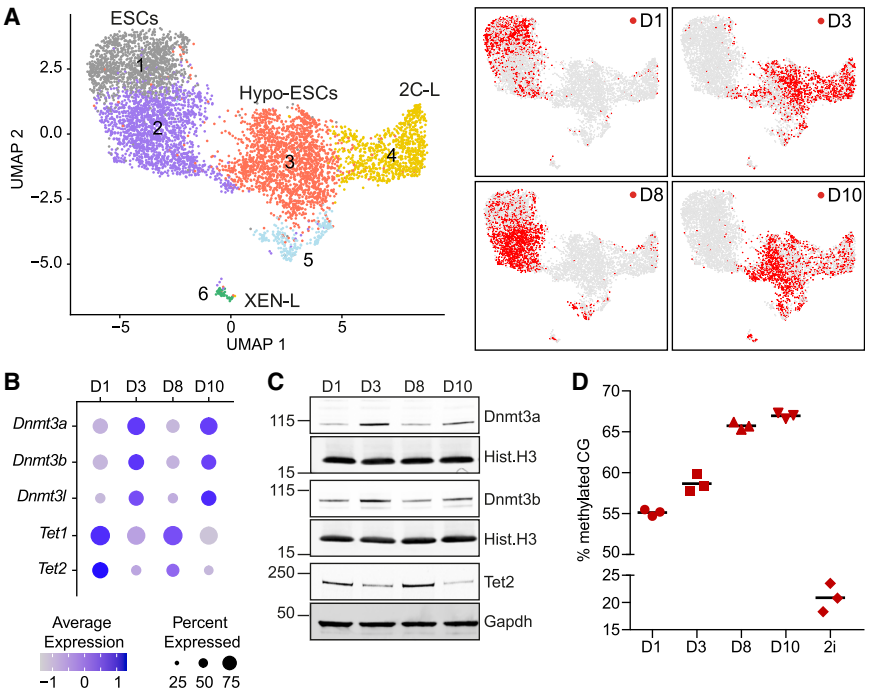


Figure 6. HypoSUMOylation triggers gradual DNA hypermethylation in ESCs

(A) Left: UMAP plot of 6,120 cells isolated on days 1, 3, 8, and 10 of the protocol (Figure 2A). Cells are colored by their cluster annotation. Right: UMAPs highlighting cells from each time point. 2C-L, 2-cell stage-like; hypo-ESCS, hypoSUMOylated ESCs.

(B) Gene expression of the DNA methylation machinery components across the 4 time points (pooled scRNA-seq data).

(C) Immunoblots of Dnmt3a, Dnmt3b, and Tet2 and their respective loading controls.

(D) Increase of the whole-genome methylated CG fraction from day 1 to day 10. ESCs in 2i culture medium were used as a hypomethylated control. See also Figure S8 and Table S1.

takes place at the chromatin level.[13] To assess whether the cellular memory imposed by the first wave of hypoSUMOylation is associated with possible rewiring of the SUMO chromatin landscape following recovery, we performed a comparative chromatin immunoprecipitation sequencing (ChIP-seq) analysis for SUMO1 and SUMO2/3 on day 8 versus day 1. Comparison of the profiles on day 1 or day 8 revealed a strong correlation between the two paralogs for each condition (Figure S9C), in agreement with our previous work.[13] In total, 12,222 peaks were common to SUMO1 and SUMO2/3 (Figure S9D). The higher number of peaks for SUMO1 compared with SUMO2/3 (31,312 and 12,942, respectively) is most likely attributable to the lower efficacy of the SUMO2/3 antibody used in this study relative to SUMO1. Remarkably, among the 12,222 common peaks, differential analysis between day 8 and day 1 identified significant chromatin landscape alterations for SUMO1 (408 peaks up and 312 down) and SUMO2/3 (908 peaks up and 183 down) (Figure 7D; Table S2).

To gain insight into the changes imposed by the first pulse of SUMOylation inhibition in re-establishment of the SUMO genomic profiles after recovery, we used a mass spectrometry (MS)-based proteomics approach to compare the set of endogenous SUMO2/3 substrates on day 1 and day 8. In total, we identified 1,296 SUMO2/3 sites mapping to 568 SUMO2/3 target proteins (Table S2). Strikingly, no significant change was detected in the global SUMOylomes between the day 1 and day 8 conditions (Figure S9E). These data indicate complete recovery of the SUMO2/3 proteome soon after acute loss of SUMOylation, highlighting a previously unappreciated robustness of the SUMOylation/deSUMOylation equilibrium. Thus, relocation of pre-existing chromatin SUMO substrates toward new sites on the genome, rather than rewiring of the SUMO proteome, likely accounts for the reorganization of the SUMO chromatin landscape as observed on day 8.

We next focused the analysis on the fraction of SUMO1 and SUMO2/3 peaks that were simultaneously altered (hereafter called SUMO peaks) because they showed a higher level of variation compared with those uniquely altered for SUMO1 ($p = 2.85e^{-13}$) or SUMO2/3 ($p = 2.59e^{-27}$). A large fraction of reduced SUMO peaks was enriched at the transposable element L1Md repeat family (54 of 75) (Figures 7E, S9F, and S9G). In contrast, the motif for the zinc-finger protein Zfp57 was the top-scoring predicted site at induced SUMO peaks (164 of 179) (Figure 7E). Zfp57 plays a key role in maintaining DNA methylation imprints that silence genes depending on parental origin.[46] Accordingly, imprinted genes were found to be over-represented in genes associated with induced SUMO peaks (Figures 7D and S9H). Among the KRAB zinc-finger proteins, Zfp57 has the unique ability to bind methylated DNA and, in turn, to drive Kap1-mediated deposition of the H3K9me3 repressive mark in ESCs.[46,47] Moreover, SUMOylation, in part by triggering recruitment of Setdb1 to SUMOylated Kap1,[48] is essential to maintain proper H3K9me3 levels genome wide in ESCs.[13] To investigate the impact of transient loss of SUMOylation on the binding of Zfp57 and Kap1 following recovery, we determined the ChIP-seq profiles of these two factors on day 1 and day 8 (Table S2). We observed that the vast majority of induced SUMO peaks (93.8%) colocalized with Zfp57- and/or Kap1-induced peaks (Figure 7F). In line with this observation, induced Zfp57 peaks were associated with a higher gain in DNA methylation compared with other peaks (Figure 7G). Together with our finding that Kap1 is one of the most heavily SUMOylated proteins in our proteomics analyses (rank 9 in Table S2), these data identified Kap1 as a, if not the major, SUMO substrate undergoing genomic relocation after SUMOylation suppression recovery via recruitment of Zfp57 to hypermethylated regions.

Aside from these imprinted genes, *Nanog* appeared among the most dramatically SUMO-enriched loci on day 8 relative to day 1 (Figure 7D). Nanog is a key component of the pluripotency

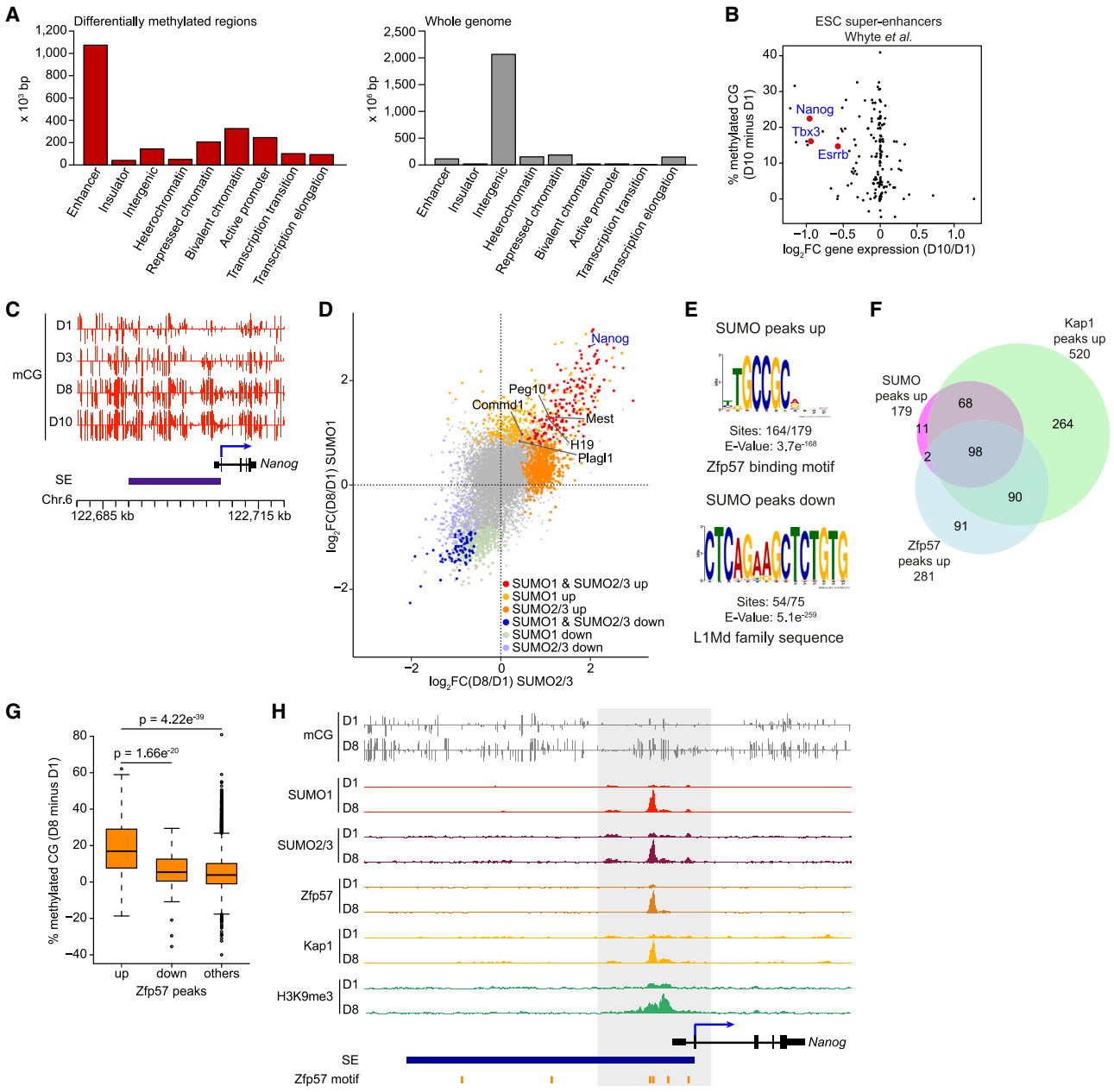


Figure 7. Reorganization of the SUMO chromatin landscape after recovery from SUMOylation suppression induces repression of the *Nanog* locus

(A) Left: Number of base pairs of the 3,080 differentially methylated regions (day 10 vs. day 1) overlapping with 9 pre-defined chromatin (ChromHMM) states.[64] Right: Whole-genome distribution of the 9 chromatin states.

(B) Scatterplot of the difference in percent methylated CG at ESC super-enhancers[45] versus $\log_2$ fold change in mRNA levels of the closest gene between day 10 and day 1.

(C) Increase of DNA methylation at the *Nanog* locus from day 1 to day 10. SE, super-enhancer.

(D) Plot comparing the $\log_2$ fold change of counts for SUMO1 and SUMO2/3 at common SUMO peaks on day 8 vs. day 1. Changes of SUMO levels were considered significant when the adjusted p value was inferior to 0.05. Imprinted genes are indicated in black.

(E) DNA motif enrichment identified at SUMO peaks that are significantly up or down on day 8 vs. day 1.

(F) Venn diagram comparing induced peaks for SUMO, Zfp57, and Kap1 on day 8 vs. day 1.

(G) Boxplots of the difference in percent methylated CG at Zfp57 peaks between day 8 and day 1 (two-tailed t test).

(H) Chromatin profiles of SUMO1, SUMO2/3, Zfp57, Kap1, and H3K9me3 at the *Nanog* SE on day 1 and day 8. DNA methylation is as in (C).

See also Figure S9 and Table S2.

regulatory circuit that enables ESC self-renewal.[49,50] Intriguingly, ESCs in culture display cell-to-cell variability in *Nanog* expression, with low Nanog-expressing cells showing predisposition to differentiate toward extraembryonic endoderm and somatic fates.[49,51,52] Inspection of *Nanog* revealed that induced SUMO peaks localize ~2 kb upstream of the promoter, a region that contains three Zfp57 binding sites and undergoes DNA hypermethylation on day 8 (Figures 7C and 7H). A dramatic increase in Zfp57 binding was visible at the −2 kb region on day 8, which correlated with recruitment of Kap1, H3K9me3 deposition, and decreased *Nanog* expression (Figures 7H, S9I, and S9J). Of note, treatment with 5-azacytidine between days 3 and 8 restored normal *Nanog* expression (Figure S9I).

Hence, together with global DNA hypermethylation, accumulation of methylated DNA and H3K9me3 repressive marks at *Nanog* in hypoSUMOylation-instructed ESCs likely contribute to impaired pluripotency and expanded cell fate potential.

DISCUSSION

To date, two main strategies had been employed to build ELSs: by subjecting ESCs to various developmental signaling molecules[53–60] or by assembling extraembryonic stem cells and ESCs.[25,26] A major advance in modeling post-implantation development was achieved recently when two groups reported that co-aggregation of ESCs with induced TSCs and XEN cells, grown in a dynamic culture system, leads to generation of almost perfect mimics of E8.5 mouse embryos.[61–63] However, although remarkable, the low efficiency (0.1%–0.5%) and heterogeneity may be limiting factors. Here, we show that lowering chromatin barriers to cell fate change enables formation of ELSs that recapitulate critical features of the post-implantation embryo. Specifically, transient chemical inhibition of SUMOylation in murine ESCs generates self-organized ELSs with morphological and transcriptional head- and trunk-like entities. Crosstalk between the three spheroid cell types generated from hypoSUMOylation-instructed ESCs is sufficient to induce gastrulating structures that undergo symmetry breaking, germ layer specification, and axial organization and are capable of further elongation to acquire a more advanced embryo-like architecture. Notably, although lacking extraembryonic ectoderm, gastruloids are able to form a primitive streak. Moreover, our model is highly scalable and adaptable, particularly because day 18 spheroids are stable after freeze/thaw cycles, allowing generation of ELSs in only 7 days with one culture medium. The added value of the droplet microfluidic platform in boosting lineage diversity requires further investigation to determine how this system contributes to a physicochemical microenvironment that favors axial elongation. With 74% efficiency, our model, although not perfect, is highly reproducible and paves the way for exploring the potential of combining SUMOylation inhibitors with existing approaches, including recently developed *in vitro* culture platforms.[61–63]

How cells coordinate complex transcriptional programs to confer proper identity remains a challenging question in cell biology. SUMOylation, by dynamically marking multiple transcription factors and chromatin regulators, provides a mechanism to orchestrate concerted transcriptional control of cell fate. Accordingly, inhibiting SUMOylation could be viewed as compromising the robustness of transcriptional programs underlying cell identity, in part by altering genome-wide DNA methylation and H3K9me3 heterochromatin, notably at the Nanog locus, favoring the emergence of cell diversity. Expanded cell specification together with the inherent self-assembly capacity of ESCs and/or their derivatives for pattern formation would enable development of structures resembling natural embryos. We present several lines of evidence indicating that levels of the core components of the SUMO machinery fluctuate during pre-implantation development, starting with the levels of conjugated SUMO2/3, indicative of changes in the global cellular SUMOylation status. We show that SUMO2/3, the dominant isoform in early embryos, displays increasing cell-to-cell heterogeneity up to the morula stage. Also, in early blastocysts, SUMO2/3 is expressed heterogeneously, with significantly lower levels in ICM than trophectoderm. Moreover, we show that RNA expression of the core components of the SUMOylation cascade is remarkably dynamic during early embryogenesis, indicative of changing combinatorial enzymatic activities acting on SUMO substrates during this process. Thus, intra-embryonic variations in SUMOylation levels, notably for SUMO2/3, occur during pre-implantation development, and for which the possible impact on rewiring the regulatory circuitry controlling cell identity is likely mirrored by global transient chemical SUMOylation inhibition in ESCs.

Overall, our work lays foundations for exploring epigenetic drugs as potentially powerful tools to control the balance between cell fate robustness and lineage commitment, with the ultimate goal of reconstructing faithful models of mammalian organs and embryos.

Limitations of the study

We show here that transient chemical inhibition of SUMOylation in murine ESCs generates ELSs that recapitulate some key aspects of early mammalian development. However, such hypoSUMOylation-instructed ELSs present several defects, including the absence of PGCs and placental lineages, as well as a predominant dorsalized embryonic phenotype. Studying the abnormal development of ELSs may, however, provide new insight into formation of the axial mesoderm-derived notochord, a major regulator of D/V patterning. Moreover, further refinement of methods (cell culture medium, growth conditions of the structures) could partially correct these defects and enhance ELS developmental potential.

Our data reveal intra-embryonic heterogeneity in SUMOylation levels *in vivo*, particularly for SUMO2/3, suggesting a profound rewiring of the repertoire of SUMO target networks that may facilitate cell fate specification during pre-implantation development. To precisely investigate this point at an in-depth biochemical level, an MS-based proteomics approach would be required to determine and compare the repertoire and SUMOylation levels of endogenous SUMO substrates present in each cell type/tissue at the different developmental stages *in vivo*. This, however, must await significant technological advances to reduce the number of cells needed for the initial SUMO2/3 immunoprecipitation step, given the limited amount of embryo material available.

Suppression of SUMOylation in ESCs drives genome-wide DNA hypermethylation, consistent with what happens during the early stages of cell fate specification in embryos. However, the exact molecular mechanisms that govern the interplay between SUMOylation and DNA methylation remain to be identified. Notably, more studies are needed to define the ways in which SUMO regulates the expression and activity of the DNA methylation machinery.

STAR★METHODS

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

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Cell culture
  - Generating adherent spheroids
  - Culturing AggreWell gastruloids
  - Culturing microfluidic gastruloids and Matrigel-embedded embryo-like structures
- METHOD DETAILS
  - Clonogenic assay
  - Cell sorting of the 3 cell types of D18 spheroids
  - Droplet-microfluidic device design and fabrication
  - Pluripotent stem cell loading and manipulation within immobilized droplets
  - Immunofluorescence of spheroids, gastruloids and embryo-like structures
  - *In situ* hybridization
  - Microscopy
  - Image analysis
  - Embryo collection, culture and ML-792 treatment
  - Immunostaining and confocal microscopy of embryos
  - Single-cell RNA-seq
  - Processing single-cell data
  - Comparison between scRNA-seq clusters and *in vivo* datasets
  - Pseudotime analysis
  - Expression levels of the SUMO machinery components
  - Methyl-seq
  - Processing methyl-seq data
  - ChIP-seq
  - Processing ChIP-seq data
  - Gene enrichment analysis
  - Bulk RNA-seq
  - Processing bulk RNA-seq data
  - Cell lysates preparation for the SUMO2/3 IP
  - Crosslinking of SUMO2/3 antibody to beads
  - Purification of SUMOylated peptides
  - StageTip purification and high-pH fractionation of SUMO-IP samples
  - MS analysis
  - Analysis of MS data
  - Immunoblots
  - Quantitative PCR
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2023.112380.

ACKNOWLEDGMENTS

We greatly thank Laure Bailly-Cuif for suggestions and insightful comments. We acknowledge Jacob-Sebastian Seeler for helpful discussions and re-reading of the manuscript. We acknowledge Heather Marlow, François Spitz, and Baptiste Saudemont for assistance with initial scRNA-seq experiments. We are grateful to David M. Suter (Swiss Federal Institute of Technology, Switzerland) for the Sox1:EGFP-T:mCherry double reporter ESCs and Pablo Navarro (Institut Pasteur, Paris) for the LF2 ES cell line. We thank Andrey Aristov for help with image analysis, Sandrine Schmutz and the Cytometry and Biomarkers Platform (Institut Pasteur, Paris) for cell sorting, Juliana Pipoli Da Fonseca and the Biomics Platform (Institut Pasteur, Paris) for scRNA sequencing, as well as Samy Gobaa and the Biomaterials and Microfluidics Platform (Institut Pasteur, Paris) for microfluidic chip manufacturing. Sequencing for ChIP-seq/RNA-seq/Methyl-seq was performed by the GenomEast platform, a member of the "France Génomique"consortium (ANR-10-INBS-0009). This work was supported by grants from European Research Council (AdG SUMiDENTITY), Agence Nationale de la Recherche (ANR-19-CE12-0011-01), Odyssey, and LNCC (Equipe labellisée) (to A.D.); Région Île-de-France DIM-ELICIT (to C.N.B.); the Helmholtz Association and German Research Foundation (DFG) Project-ID 213249687 (SFB 1064) (to M.-E.T.-P.); and the Novo Nordisk Foundation (NNF14CC0001) and the European Union Horizon 2020 (EPIC-XS-823839) (to M.L.N.).

AUTHOR CONTRIBUTIONS

J.-C.C. and A.D. conceived and designed the study. J.-C.C., T.T., and S.S. performed most of the experiments and analyzed the data. S.S. and C.N.B. designed the droplet microfluidic platform. Y.L.-M. performed the bioinformatics analysis. M.G. performed the *in vivo* experiments, and M.G. and M.-E.T.-P. interpreted the data. I.A.H. performed the SUMO mass spectrometry experiments, and I.A.H. and M.L.N. analyzed the data. I.T. contributed to initial experiments. J.-C.C., T.T., S.S., and A.D. wrote the manuscript with input from all co-authors.

DECLARATION OF INTERESTS

J.-C.C., T.T., S.S., C.N.B., and A.D. are designated as inventors of the patent application WO 2023/002057 A2 covering the aspects of the *in vitro* generation of organized 3D cell structures and the microfluidic device described in the manuscript.

Received: August 12, 2022
Revised: February 8, 2023
Accepted: March 26, 2023
Published: April 15, 2023

REFERENCES

1. 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. https://doi.org/10.1038/nature15749.
2. Mazid, M.A., Ward, C., Luo, Z., Liu, C., Li, Y., Lai, Y., Wu, L., Li, J., Jia, W., Jiang, Y., et al. (2022). Rolling back human pluripotent stem cells to an

eight-cell embryo-like stage. Nature *605*, 315–324. https://doi.org/10.1038/s41586-022-04625-0.

3. Onder, T.T., Kara, N., Cherry, A., Sinha, A.U., Zhu, N., Bernt, K.M., Cahan, P., Marcarci, B.O., Unternaehrer, J., Gupta, P.B., et al. (2012). Chromatin-modifying enzymes as modulators of reprogramming. Nature *483*, 598–602. https://doi.org/10.1038/nature10953.
4. Rodriguez-Terrones, D., Gaume, X., Ishiuchi, 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. https://doi.org/10.1038/s41588-017-0016-5.
5. Yang, M., Yu, H., Yu, X., Liang, S., Hu, Y., Luo, Y., Izsvák, Z., Sun, C., and Wang, J. (2022). Chemical-induced chromatin remodeling reprograms mouse ESCs to totipotent-like stem cells. Cell Stem Cell *29*, 400–418.e13. https://doi.org/10.1016/j.stem.2022.01.010.
6. Hendriks, I.A., and Vertegaal, A.C.O. (2016). A comprehensive compilation of SUMO proteomics. Nat. Rev. Mol. Cell Biol. *17*, 581–595. https://doi.org/10.1038/nrm.2016.81.
7. Theurillat, I., Hendriks, I.A., Cossec, J.-C., Andrieux, A., Nielsen, M.L., and Dejean, A. (2020). Extensive SUMO modification of repressive chromatin factors distinguishes pluripotent from somatic cells. Cell Rep. *32*, 108146. https://doi.org/10.1016/j.celrep.2020.108146.
8. Cubeñas-Potts, C., and Matunis, M.J. (2013). SUMO: a multifaceted modifier of chromatin structure and function. Dev. Cell *24*, 1–12. https://doi.org/10.1016/j.devcel.2012.11.020.
9. 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. https://doi.org/10.1101/gr.154872.113.
10. 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. https://doi.org/10.1186/s13059-015-0717-y.
11. 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. https://doi.org/10.1126/scisignal.aaa2213.
12. Decque, A., Joffre, O., Magalhaes, J.G., Cossec, J.C., Blecher-Gonen, R., Lapaquette, P., Silvin, 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. https://doi.org/10.1038/ni.3342.
13. Cossec, J.-C., Theurillat, I., Chica, C., Búa Aguín, S., Gaume, X., Andrieux, A., Iturbide, A., Jouvion, G., Li, H., Bossis, G., et al. (2018). SUMO safeguards somatic and pluripotent cell identities by enforcing distinct chromatin states. Cell Stem Cell *23*, 742–757.e8. https://doi.org/10.1016/j.stem.2018.10.001.
14. 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 Rep. *6*, 704–716. https://doi.org/10.1016/j.stemcr.2016.02.004.
15. Baik, H., Boulanger, M., Hosseini, M., Kowalczyk, J., Zaghdoudi, S., Salem, T., Sarry, J.-E., Hicheri, Y., Cartron, G., Piechaczyk, M., and Bossis, G. (2018). Targeting the SUMO pathway primes all-*trans* retinoic acid–induced differentiation of nonpromyelocytic acute myeloid leukemias. Cancer Res. *78*, 2601–2613. https://doi.org/10.1158/0008-5472.CAN-17-3361.
16. Yan, Y.-L., Zhang, C., Hao, J., Wang, X.-L., Ming, J., Mi, L., Na, J., Hu, X., and Wang, Y. (2019). DPPA2/4 and SUMO E3 ligase PIAS4 opposingly regulate zygotic transcriptional program. PLoS Biol. *17*, e3000324. https://doi.org/10.1371/journal.pbio.3000324.
17. Sheban, D., Shani, T., Maor, R., Aguilera-Castrejon, A., Mor, N., Oldak, B., Shmueli, M.D., Eisenberg-Lerner, A., Bayerl, J., Hebert, J., et al. (2022). SUMOylation of linker histone H1 drives chromatin condensation and restriction of embryonic cell fate identity. Mol. Cell *82*, 106–122.e9. https://doi.org/10.1016/j.molcel.2021.11.011.
18. Zhang, F.-P., Mikkonen, L., Toppari, J., Palvimo, J.J., Thesleff, I., and Jänne, O.A. (2008). Sumo-1 function is dispensable in normal mouse development. Mol. Cell Biol. *28*, 5381–5390.
19. Wang, L., Wansleeben, C., Zhao, S., Miao, P., Paschen, W., and Yang, W. (2014). SUMO2 is essential while SUMO3 is dispensable for mouse embryonic development. EMBO Rep. *15*, 878–885. https://doi.org/10.15252/embr.201438534.
20. Garvin, A.J., Lanz, A.J., and Morris, J.R. (2022). SUMO monoclonal antibodies vary in sensitivity, specificity, and ability to detect types of SUMO conjugate. Sci. Rep. *12*, 21343. https://doi.org/10.1038/s41598-022-25665-6.
21. Liu, Z., Tardat, M., Gill, M.E., Royo, H., Thierry, R., Ozonov, E.A., and Peters, A.H. (2020). SUMOylated PRC1 controls histone H3.3 deposition and genome integrity of embryonic heterochromatin. EMBO J. *39*, e103697. https://doi.org/10.15252/embj.2019103697.
22. 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, S1534-5807(05)00389-8 [pii]. https://doi.org/10.1016/j.devcel.2005.10.007.
23. 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. https://doi.org/10.1038/nchembio.2463.
24. Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. Cell Rep. *20*, 1215–1228. https://doi.org/10.1016/j.celrep.2017.07.009.
25. Sozen, B., Amadei, G., Cox, A., Wang, R., Na, E., Czukiewska, S., Chappell, L., Voet, T., Michel, G., Jing, N., et al. (2018). Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. Nat. Cell Biol. *20*, 979–989. https://doi.org/10.1038/s41556-018-0147-7.
26. Zhang, S., Chen, T., Chen, N., Gao, D., Shi, B., Kong, S., West, R.C., Yuan, Y., Zhi, M., Wei, Q., et al. (2019). Implantation initiation of self-assembled embryo-like structures generated using three types of mouse blastocyst-derived stem cells. Nat. Commun. *10*, 496. https://doi.org/10.1038/s41467-019-08378-9.
27. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. Nature *566*, 490–495. https://doi.org/10.1038/s41586-019-0933-9.
28. Graham, S.J.L., and Zernicka-Goetz, M. (2016). The acquisition of cell fate in mouse development. In Current Topics in Developmental Biology (Elsevier), pp. 671–695. https://doi.org/10.1016/bs.ctdb.2015.11.021.
29. Ben-Haim, N., Lu, C., Guzman-Ayala, M., Pescatore, L., Mesnard, D., Bischofberger, M., Naef, F., Robertson, E.J., and Constam, D.B. (2006). The nodal precursor acting via activin receptors induces mesoderm by maintaining a source of its convertases and BMP4. Dev. Cell *11*, 313–323. https://doi.org/10.1016/j.devcel.2006.07.005.
30. Choi, H.M.T., Schwarzkopf, M., Fornace, M.E., Acharya, A., Artavanis, G., Stegmaier, J., Cunha, A., and Pierce, N.A. (2018). Third-generation in situ hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. Development *145*, dev165753. https://doi.org/10.1242/dev.165753.
31. Zheng, Y., Xue, X., Shao, Y., Wang, S., Esfahani, S.N., Li, Z., Muncie, J.M., Lakins, J.N., Weaver, V.M., Gumucio, D.L., and Fu, J. (2019). Controlled

modelling of human epiblast and amnion development using stem cells. Nature *573*, 421–425. https://doi.org/10.1038/s41586-019-1535-2.

32. Sart, S., Tomasi, R.F.-X., Amselem, G., and Baroud, C.N. (2017). Multiscale cytometry and regulation of 3D cell cultures on a chip. Nat. Commun. *8*, 469. https://doi.org/10.1038/s41467-017-00475-x.
33. Sart, S., Tomasi, R.F.-X., Barizien, A., Amselem, G., Cumano, A., and Baroud, C.N. (2020). Mapping the structure and biological functions within mesenchymal bodies using microfluidics. Sci. Adv. *6*, eaaw7853. https://doi.org/10.1126/sciadv.aaw7853.
34. Langston, S.P., Grossman, S., England, D., Afroze, R., Bence, N., Bowman, D., Bump, N., Chau, R., Chuang, B.-C., Claiborne, C., et al. (2021). Discovery of TAK-981, a first-in-class inhibitor of SUMO-activating enzyme for the treatment of cancer. J. Med. Chem. *64*, 2501–2520. https://doi.org/10.1021/acs.jmedchem.0c01491.
35. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.
36. Koch, F., Scholze, M., Wittler, L., Schifferl, D., Sudheer, S., Grote, P., Timmermann, B., Macura, K., and Herrmann, B.G. (2017). Antagonistic activities of Sox2 and brachyury control the fate choice of neuro-mesodermal progenitors. Dev. Cell *42*, 514–526.e7. https://doi.org/10.1016/j.devcel.2017.07.021.
37. Britsch, S., Goerich, D.E., Riethmacher, D., Peirano, R.I., Rossner, M., Nave, K.A., Birchmeier, C., and Wegner, M. (2001). The transcription factor Sox10 is a key regulator of peripheral glial development. Genes Dev. *15*, 66–78. https://doi.org/10.1101/gad.186601.
38. Eckersley-Maslin, M., Alda-Catalinas, C., Blotenburg, M., Kreibich, E., Krueger, C., and Reik, W. (2019). Dppa2 and Dppa4 directly regulate the Dux-driven zygotic transcriptional program. Genes Dev. *33*, 194–208. https://doi.org/10.1101/gad.321174.118.
39. Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220. https://doi.org/10.1038/nrg3354.
40. Jackson, M., Krassowska, A., Gilbert, N., Chevassut, T., Forrester, L., Ansell, J., and Ramsahoye, B. (2004). Severe global DNA hypomethylation blocks differentiation and induces histone hyperacetylation in embryonic stem cells. Mol. Cell Biol. *24*, 8862–8871. https://doi.org/10.1128/MCB.24.20.8862-8871.2004.
41. Auclair, G., Guibert, S., Bender, A., and Weber, M. (2014). Ontogeny of CpG island methylation and specificity of DNMT3 methyltransferases during embryonic development in the mouse. Genome Biol. *15*, 545. https://doi.org/10.1186/s13059-014-0545-5.
42. Zhang, Y., Xiang, Y., Yin, Q., Du, Z., Peng, X., Wang, Q., Fidalgo, M., Xia, W., Li, Y., Zhao, Z.A., et al. (2018). Dynamic epigenomic landscapes during early lineage specification in mouse embryos. Nat. Genet. *50*, 96–105. https://doi.org/10.1038/s41588-017-0003-x.
43. Zvetkova, I., Apedaile, A., Ramsahoye, B., Mermoud, J.E., Crompton, L.A., John, R., Feil, R., and Brockdorff, N. (2005). Global hypomethylation of the genome in XX embryonic stem cells. Nat. Genet. *37*, 1274–1279. https://doi.org/10.1038/ng1663.
44. Leitch, H.G., McEwen, K.R., Turp, A., Encheva, V., Carroll, T., Grabole, N., Mansfield, W., Nashun, B., Knezovich, J.G., Smith, A., et al. (2013). Naive pluripotency is associated with global DNA hypomethylation. Nat. Struct. Mol. Biol. *20*, 311–316. https://doi.org/10.1038/nsmb.2510.
45. 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. https://doi.org/10.1016/j.cell.2013.03.035.
46. Li, X., Ito, M., Zhou, F., Youngson, N., Zuo, X., Leder, P., and Ferguson-Smith, A.C. (2008). A maternal-zygotic effect gene, Zfp57, maintains both maternal and paternal imprints. Dev. Cell *15*, 547–557. https://doi.org/10.1016/j.devcel.2008.08.014.
47. Shi, H., Strogantsev, R., Takahashi, N., Kazachenka, A., Lorincz, M.C., Hemberger, M., and Ferguson-Smith, A.C. (2019). ZFP57 regulation of transposable elements and gene expression within and beyond imprinted domains. Epigenet. Chromatin *12*, 49. https://doi.org/10.1186/s13072-019-0295-4.
48. Yang, B.X., El Farran, C.A., Guo, H.C., Yu, T., Fang, H.T., Wang, H.F., Schlesinger, S., Seah, Y.F.S., Goh, G.Y.L., Neo, S.P., et al. (2015). Systematic identification of factors for provirus silencing in embryonic stem cells. Cell *163*, 230–245. https://doi.org/10.1016/j.cell.2015.08.037.
49. Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643–655. https://doi.org/10.1016/S0092-8674(03)00392-1.
50. Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631–642. https://doi.org/10.1016/S0092-8674(03)00393-3.
51. Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230–1234. https://doi.org/10.1038/nature06403.
52. Kalmar, T., Lim, C., Hayward, P., Muñoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Martinez Arias, A. (2009). Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biol. *7*, e1000149. https://doi.org/10.1371/journal.pbio.1000149.
53. Kagawa, H., Javali, A., Khoei, H.H., Sommer, T.M., Sestini, G., Novatchkova, M., Scholte Op Reimer, Y., Castel, G., Bruneau, A., Maenhoudt, N., et al. (2022). Human blastoids model blastocyst development and implantation. Nature *601*, 600–605. https://doi.org/10.1038/s41586-021-04267-8.
54. Li, R., Zhong, C., Yu, Y., Liu, H., Sakurai, M., Yu, L., Min, Z., Shi, L., Wei, Y., Takahashi, Y., et al. (2019). Generation of blastocyst-like structures from mouse embryonic and adult cell cultures. Cell *179*, 687–702.e18. https://doi.org/10.1016/j.cell.2019.09.029.
55. Yu, L., Wei, Y., Duan, J., Schmitz, D.A., Sakurai, M., Wang, L., Wang, K., Zhao, S., Hon, G.C., and Wu, J. (2021). Blastocyst-like structures generated from human pluripotent stem cells. Nature *591*, 620–626. https://doi.org/10.1038/s41586-021-03356-y.
56. Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A.-C., Lutolf, M.P., Duboule, D., and Arias, A.M. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. Nature *562*, 272–276. https://doi.org/10.1038/s41586-018-0578-0.
57. van den Brink, S.C., Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivié, J., Baillie-Johnson, P., Nichols, J., Sonnen, K.F., Martinez Arias, A., and van Oudenaarden, A. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. Nature *582*, 405–409. https://doi.org/10.1038/s41586-020-2024-3.
58. Veenvliet, J.V., Bolondi, A., Kretzmer, H., Haut, L., Scholze-Wittler, M., Schifferl, D., Koch, F., Guignard, L., Kumar, A.S., Pustet, M., et al. (2020). Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. Science *370*, eaba4937. https://doi.org/10.1126/science.aba4937.
59. Xu, P.-F., Borges, R.M., Fillatre, J., de Oliveira-Melo, M., Cheng, T., Thisse, B., and Thisse, C. (2021). Construction of a mammalian embryo model from stem cells organized by a morphogen signalling centre. Nat. Commun. *12*, 3277. https://doi.org/10.1038/s41467-021-23653-4.
60. Girgin, M.U., Broguiere, N., Mattolini, L., and Lutolf, M.P. (2021). Gastruloids generated without exogenous Wnt activation develop anterior neural tissues. Stem Cell Rep. *16*, 1143–1155. https://doi.org/10.1016/j.stemcr.2021.03.017.
61. Amadei, G., Handford, C.E., Qiu, C., De Jonghe, J., Greenfeld, H., Tran, M., Martin, B.K., Chen, D.-Y., Aguilera-Castrejon, A., Hanna, J.H., et al. (2022). Synthetic embryos complete gastrulation to neurulation and

organogenesis. Nature *610*, 143–153. https://doi.org/10.1038/s41586-022-05246-3.

62. Lau, K.Y.C., Rubinstein, H., Gantner, C.W., Hadas, R., Amadei, G., Stelzer, Y., and Zernicka-Goetz, M. (2022). Mouse embryo model derived exclusively from embryonic stem cells undergoes neurulation and heart development. Cell Stem Cell *29*, 1445–1458.e8. https://doi.org/10.1016/j.stem.2022.08.013.
63. Tarazi, S., Aguilera-Castrejon, A., Joubran, C., Ghanem, N., Ashouokhi, S., Roncato, F., Wildschutz, E., Haddad, M., Oldak, B., Gomez-Cesar, E., et al. (2022). Post-gastrulation synthetic embryos generated ex utero from mouse naive ESCs. Cell *185*, 3290–3306.e25. https://doi.org/10.1016/j.cell.2022.07.028.
64. Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B., and Brockdorff, N. (2017). hnRNPK recruits PCGF3/5-PRC1 to the xist RNA B-repeat to establish polycomb-mediated chromosomal silencing. Mol. Cell *68*, 955–969.e10. https://doi.org/10.1016/j.molcel.2017.11.013.
65. 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. https://doi.org/10.1126/science.1245316.
66. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, 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. https://doi.org/10.1186/gb-2008-9-9-r137.
67. Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359. https://doi.org/10.1038/nmeth.1923.
68. 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. https://doi.org/10.1186/s13059-014-0550-8.
69. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
70. Korsunsky, I., Millard, N., Fan, J., Slowikowski, K., Zhang, F., Wei, K., Baglaenko, Y., Brenner, M., Loh, P.R., and Raychaudhuri, S. (2019). Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods *16*, 1289–1296. https://doi.org/10.1038/s41592-019-0619-0.
71. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386. https://doi.org/10.1038/nbt.2859.
72. Krueger, F., and Andrews, S.R. (2011). Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics *27*, 1571–1572. https://doi.org/10.1093/bioinformatics/btr167.
73. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; Genome Project data processing subgroup (2009). The sequence alignment/map format and SAMtools. Bioinformatics *25*, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
74. Li, Q., Brown, J.B., Huang, H., and Bickel, P.J. (2011). Measuring reproducibility of high-throughput experiments. Ann. Appl. Stat. *5*, 1752–1779. https://doi.org/10.1214/11-AOAS466.
75. Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. Bioinformatics *27*, 1696–1697. https://doi.org/10.1093/bioinformatics/btr189.
76. Alhamdoosh, M., Ng, M., Wilson, N.J., Sheridan, J.M., Huynh, H., Wilson, M.J., and Ritchie, M.E. (2017). Combining multiple tools outperforms individual methods in gene set enrichment analyses. Bioinformatics *33*, 414–424. https://doi.org/10.1093/bioinformatics/btw623.
77. Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. Nat. Biotechnol. *26*, 1367–1372. https://doi.org/10.1038/nbt.1511.
78. Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. J. Proteome Res. *10*, 1794–1805. https://doi.org/10.1021/pr101065j.
79. Perez-Riverol, Y., Bai, J., Bandla, C., García-Seisdedos, D., Hewapathirana, S., Kamatchinathan, S., Kundu, D.J., Prakash, A., Frericks-Zipper, A., Eisenacher, M., et al. (2022). The PRIDE database resources in 2022: a hub for mass spectrometry-based proteomics evidences. Nucleic Acids Res. *50*, D543–D552. https://doi.org/10.1093/nar/gkab1038.
80. Deluz, C., Friman, E.T., Strebinger, D., Benke, A., Raccaud, M., Callegari, A., Leleu, M., Manley, S., and Suter, D.M. (2016). A role for mitotic bookmarking of SOX2 in pluripotency and differentiation. Genes Dev. *30*, 2538–2550. https://doi.org/10.1101/gad.289256.116.
81. Semrau, S., Goldmann, J.E., Soumillon, M., Mikkelsen, T.S., Jaenisch, R., and van Oudenaarden, A. (2017). Dynamics of lineage commitment revealed by single-cell transcriptomics of differentiating embryonic stem cells. Nat. Commun. *8*, 1096. https://doi.org/10.1038/s41467-017-01076-4.
82. Tomasi, R.F.-X., Sart, S., Champetier, T., and Baroud, C.N. (2020). Individual control and quantification of 3D spheroids in a high-density microfluidic droplet array. Cell Rep. *31*, 107670. https://doi.org/10.1016/j.celrep.2020.107670.
83. Torres-Padilla, M.E., and Zernicka-Goetz, M. (2006). Role of TIF1alpha as a modulator of embryonic transcription in the mouse zygote. J. Cell Biol. *174*, 329–338. https://doi.org/10.1083/jcb.200603146.
84. Jaitin, D.A., Kenigsberg, E., Keren-Shaul, H., Elefant, N., Paul, F., Zaretsky, I., Mildner, A., Cohen, N., Jung, S., Tanay, A., and Amit, I. (2014). Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. Science *343*, 776–779. https://doi.org/10.1126/science.1247651.
85. Keren-Shaul, H., Kenigsberg, E., Jaitin, D.A., David, E., Paul, F., Tanay, A., and Amit, I. (2019). MARS-seq2.0: an experimental and analytical pipeline for indexed sorting combined with single-cell RNA sequencing. Nat. Protoc. *14*, 1841–1862. https://doi.org/10.1038/s41596-019-0164-4.
86. Haghverdi, L., Lun, A.T.L., Morgan, M.D., and Marioni, J.C. (2018). Batch effects in single-cell RNA-sequencing data are corrected by matching mutual nearest neighbors. Nat. Biotechnol. *36*, 421–427. https://doi.org/10.1038/nbt.4091.
87. Picard toolkit (2019) (GitHub repository: Broad Institute).
88. 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. https://doi.org/10.1016/j.celrep.2014.10.018.
89. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics *26*, 841–842. https://doi.org/10.1093/bioinformatics/btq033.
90. Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.
91. Hendriks, I.A., Lyon, D., Su, D., Skotte, N.H., Daniel, J.A., Jensen, L.J., and Nielsen, M.L. (2018). Site-specific characterization of endogenous SUMOylation across species and organs. Nat. Commun. *9*, 2456. https://doi.org/10.1038/s41467-018-04957-4.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Cd31(Pecam1)-FITC	Thermo Fisher Scientific	Cat# 11-0311-81; RRID: AB_465011
Cd140a(Pdgfra)-PE	Thermo Fisher Scientific	Cat# 12-1401-81; RRID: AB_657615
Cd24a-APCeFluor780	Thermo Fisher Scientific	Cat# 47-0242-82; RRID: AB_10853172
Oct4	Millipore	Cat# MAB4419; RRID: AB_1977399
Nanog	Thermo Fisher Scientific	Cat# 14-5761-80; RRID: AB_763613
Sox17	R and D Systems	Cat# AF1924; RRID: AB_355060
Pou3f1	Atlas Antibodies	Cat# HPA073824; RRID: AB_2686635
T	R and D Systems	Cat# AF2085; RRID: AB_2200235
T	Abcam	Cat# ab209665; RRID: AB_2750925
Pax6	Abcam	Cat# ab78545; RRID: AB_1566562
FoxA2	Cell Signaling Technology	Cat# 8186; RRID: AB_10891055
Pax2	Thermo Fisher Scientific	Cat# 71–6000; RRID: AB_2533990
Tuj1	BioLegend	Cat# 801201; RRID: AB_2313773
Map2	Sigma-Aldrich	Cat# M4403; RRID: AB_477193
Sox10	Abcam	Cat# ab264405; RRID: AB_2927647
Gbx2	R and D Systems	Cat# AF4638; RRID: AB_2109492
H3K4me3	Active Motif	Cat# 39159; RRID: AB_2615077
H3K9me3	Abcam	Cat# ab8898; RRID: AB_306848
H3K27me3	Millipore	Cat# 07–449; RRID: AB_310624
Zfp57	Abcam	Cat# ab45341; RRID: AB_946192
Kap1	Abcam	Cat# ab10483; RRID: AB_297222
SUMO-1	Abcam	Cat# ab32058; RRID: AB_778173
SUMO-2/3	Abcam	Cat# ab81371; RRID: AB_1658424
SUMO-2/3	Abcam	Cat# ab3742; RRID: AB_304041
Actin	Sigma	Cat# A1978; RRID: AB_476692
Gapdh	Cell Signaling Technology	Cat# 2118, RRID: AB_561053
Dnmt3a	Abcam	Cat# ab2850; RRID: AB_303355
Dnmt3b	Abcam	Cat# ab2851; RRID: AB_303356
Tet2	Cell Signaling Technology	Cat# 36449; RRID: AB_2799102
Histone H3	Abcam	Cat# ab24834; RRID: AB_470335
Snai1	Cell Signaling Technology	Cat# 3879; RRID: AB_2255011
Nestin	Millipore	Cat# MAB353; RRID: AB_94911
PML	Millipore	Cat# MAB3738; RRID: AB_2315177
Ssea1	BD Biosciences	Cat# 560120; RRID: AB_11151898
ZO1	Thermo Fisher Scientific	Cat# 40–2200; RRID: AB_2533456
Sox2	Millipore	Cat# AB5603; RRID: AB_2286686
Alexa Fluor 488 donkey anti-rat	Thermo Fisher Scientific	Cat# A-21208; RRID: AB_2535794
Alexa Fluor 546 donkey anti-goat	Thermo Fisher Scientific	Cat# A-11056; RRID: AB_2534103
Alexa Fluor 488 goat-anti-mouse	Thermo Fisher Scientific	Cat# A-11001; RRID: AB_2534069
Alexa Fluor 488 donkey-anti-goat	Thermo Fisher Scientific	Cat# A32814, RRID: AB_2762838
Alexa Fluor 647 donkey-anti-rabbit	Thermo Fisher Scientific	Cat# A-31573; RRID: AB_2536183
Alexa Fluor 488 Goat anti-Rabbit	Thermo Fisher Scientific	Cat# A32731; RRID: AB_26332800
Alexa Fluor 555 Goat anti-Rabbit	Thermo Fisher Scientific	Cat# A32732; RRID: AB_2633281
Alexa Fluor 647 Goat anti-Rabbit	Thermo Fisher Scientific	Cat# A32733, RRID: AB_2633282

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Alexa Fluor 555 goat-anti-mouse	Thermo Fisher Scientific	Cat# A-21424, RRID: AB_141780
Alexa Fluor 647 goat-anti-mouse	Thermo Fisher Scientific	Cat# A32728, RRID: AB_2633277
Chemicals, peptides, and recombinant proteins		
Lif	Miltenyi Biotec	Cat# 130-099-895
PD0325901	Miltenyi Biotec	Cat# 130-103-923
CHIR99021	Miltenyi Biotec	Cat# 130-103-926
Retinoic acid	Merck (Sigma-Aldrich)	Cat# R2625
5-azacytidine	Abcam	Cat# ab142744
ML-792	Takeda Pharmaceuticals	N/A
ML-792	Clinisciences	Cat# HY-108702
Trypsin-EDTA	GIBCO	Cat# 25300-054
GlutaMAX	GIBCO	Cat# 35050-038
NEM-NEAA	GIBCO	Cat# 11140-050
TRIzol	Thermo Fisher Scientific	Cat# 15596018
StemPro Accutase	Gibco	Cat# A111-05-01
Propidium Iodide	Thermo Fisher Scientific	Cat# P3566
DMH1	Merck (Sigma-Aldrich)	Cat# 203646
Matrigel	Corning	Cat# 354234
Calcein AM	Thermo Fisher Scientific	Cat# C3100MP
Exonuclease I	New England Biolabs	Cat# M0293S
Formaldehyde	Thermo Fisher Scientific	Cat# 28908
Lysyl Endopeptidase	Wako	Cat# 125-05061
Protein G Agarose beads	Roche	Cat# 11243233001
Endoproteinase Asp-N Sequencing Grade	Roche	Cat# 11420488001
TAK-981	Clinisciences	Cat# HY-111789
T4 RNA Ligase 1	New England Biolabs	Cat# M0204S
KAPA HiFi HotStart ReadyMix	Roche	Cat# 7958935001
Critical commercial assays		
Alkaline Phosphatase solution kit	Merck (Sigma-Aldrich)	Cat# AB0300
HiScribe T7 High Yield RNA Synthesis Kit	New England Biolabs	Cat# E2040S
NEBNext Ultra II Non-Directional RNA Second Strand Synthesis Module	New England Biolabs	Cat# E6111S
Quick-DNA Midiprep Plus Kit	Zymo research	Cat# D4075
NEBNext Enzymatic Methyl-seq Kit	New England Biolabs	Cat# E7120S
ChIP-IT kit	Active Motif	Cat# 53040
MicroPlex Library Preparation Kit v2	Diagenode	Cat# C05010012
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems	Cat# 4368814
Deposited data		
Raw MS data (SUMOylome)	This study	ProteomeXchange: PXD031548
scRNA-seq, RNA-seq, ChIP-seq, Methyl-seq	This study	GEO: GSE180489
scRNA-seq of mouse early gastrulation	Mohammed et al.[24]	GEO: GSE100597
scRNA-seq of mouse gastrulation and early organogenesis	Pijuan-Sala et al.[27]	ArrayExpress: E-MTAB-6967; https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-6967
scRNA-seq of mouse organogenesis	Cao et al.[35]	GEO: GSE119945
Experimental models: Cell lines		
R1 mouse ES cells	Cossec et al.[13]	N/A
MERVL:tdTomato Dppa2/4 double KO ES-E14 cells	Laboratory of W.Reik	N/A
Sox1:eGFP-T:mCherry double reporter CGR8 mouse ES cells	Laboratory of D. Suter	N/A

(*Continued on next page*)

Continued		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Mouse ES-D3 cells	ATCC	Cat# CRL-1934
Mouse ES-LF2 cells	Laboratory of P. Navarro	N/A
Mouse embryonic fibroblast feeder	This study	N/A
Experimental models: Organisms/strains		
Mouse: C57BL/6J	Institut Pasteur	N/A
Mouse: CD1	Helmholtz Zentrum	N/A
Oligonucleotides		
RT-qPCR primers	Eurofins	Table S2
HCR probe: *Nanog*	Molecular Instruments	NM_001289828
HCR probe: *T*	Molecular Instruments	NM_009309.2
HCR probe: *En1*	Molecular Instruments	NM_010133.2
HCR probe: *Uncx*	Molecular Instruments	NM_013702.3
Software and algorithms		
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/
MarkDuplicates tool of Picard	N/A	https://broadinstitute.github.io/picard/
MACS2	Zhang et al.[66]	N/A
bedtools	N/A	https://bedtools.readthedocs.io/en/latest/
bowtie2	Langmead and Salzberg[67]	N/A
DESeq2	Love et al.[68]	N/A
Seurat 4R	Hao et al.[69]	N/A
Harmony	Korsunsky et al.[70]	N/A
Monocle 3	Trapnell et al.[71]	N/A
Bismark pipeline	Krueger and Andrews[72]	N/A
SAMtools	Li et al.[73]	N/A
IDR	Li et al.[74]	N/A
MEME-ChIP	Machanick and Bailey[75]	N/A
EGSEA	Alhamdoosh et al. [76]	N/A
MaxQuant, version 1.5.3.30	Cox and Mann; Cox et al.[77,78]	N/A

RESOURCE AVAILABILITY

Lead contact

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

Materials availability

This study did not generate new unique reagents.

Data and code availability

- All sequencing datasets produced in this study have been deposited at the Gene Expression Omnibus (GEO) and are publicly available as of the date of publication. The accession number is listed in the key resources table. The mass spectrometry proteomics data have been deposited at the ProteomeXchange Consortium via the PRIDE[79] partner repository and are publicly available as of the date of publication. The dataset identifier is listed in the key resources table. This paper analyzes existing, publicly available data. These accession numbers for the datasets are listed in the key resources table.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this work paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Cell culture

Mouse ES-R1 cells[13] were used for most experiments and were maintained in serum+Lif medium (KnockOut DMEM supplemented with 15% ES cell qualified FBS, 1% GlutaMAX, 1% MEM non-essential amino acids, 1% penicillin-streptomycin, 0.1 mM 2-mercaptoethanol, 10 ng/mL Lif (Miltenyi Biotec #130-099-895)) on gelatine-coated plates in a humidified incubator (37°C, 5% CO_2). The Sox1eGFP–TmCherry double reporter CGR8 mouse ES cells (gift from David M. Suter, Swiss Federal Institute of Technology, Lausanne, Switzerland[80]) and the female mouse ES-LF2 cells (gift from Pablo Navarro, Institut Pasteur, Paris, France) were maintained on Mitomycin C-treated mouse embryonic fibroblast feeder cells in serum+Lif medium. The MERVLtdTomato–Dppa2/4 DKO mouse ES-E14 cells (gift from Wolf Reik, Babraham Institute, Cambridge, UK[38]) were cultured in serum+Lif medium on gelatine-coated plates. Mouse ES-D3 cells (from ATCC) were used for droplet-microfluidic platform validation experiments and were cultured in ESLIF medium as previously described.[57]

For 2i culture, ES-R1 cells were maintained in N2B27 + Lif medium (1:1 Neurobasal medium and Advanced DMEM/F-12, 1% N-2 supplement, 2% B-27 supplement, 0.05% Bovine albumin fraction V, 1% GlutaMAX, 1% penicillin-streptomycin, 0.1 mM 2-mercaptoethanol, 10 ng/mL Lif) supplemented with 1 μm PD0325901 (Miltenyi Biotec #130-103-923) and 3 μm CHIR99021 (Miltenyi Biotec #130-103-926).

For retinoic acid differentiation assays, cells were seeded in 6-well plates in medium without Lif (serum medium for D1, N2B27 medium for D18) and treated with 1 μM of Retinoic Acid (Sigma #R2625). Medium was refreshed every day for a total of 5 days of treatment. Cells were collected every 24 h for RNA extractions.

For 5-azacytidine experiments, cells were seeded in 6-well plates in serum+Lif medium and treated with 0.1 μM of 5-azacytidine (Abcam #ab142744). Medium was refreshed every day for a total of 5 days of treatment.

Generating adherent spheroids

ESCs were dissociated with Trypsin-EDTA and plated (2.5 million cells in 100 mm dish) in serum+Lif medium supplemented with 2.5 μM ML-792 (Gift from Takeda Pharmaceuticals International Co.; CliniSciences #HY-108702) or 0.1 μM TAK-981 (CliniSciences # HY-111789). Medium was replaced the next day to refresh treatment. After 48 h of treatment, plates were rinsed twice with PBS then cells were allowed to recover in serum+Lif. 5 days after the end of the first round, cells were similarly counted and plated for a second wave of ML-792 or TAK-981. After 48 h of treatment, plates were rinsed twice with PBS then cells were allowed to recover in N2B27 + Lif medium. 6 to 8 days later, cells form three-dimensional adherent spheroids, which can be maintained in culture or frozen for subsequent use (cell stocks in ES cell qualified FBS with 10% DMSO). Spheroid morphology and properties are conserved for up to one month, during which time they can be used to generate A-gastruloids, F-gastruloids and M-ELSs.

Culturing AggreWell gastruloids

1 mL of anti-adherence rinsing solution (StemCell Technologies #07010) was added to each well of an AggreWell800 plate (StemCell Technologies #34815). Plate was centrifuged for 5 min at 2,000 rpm then incubated 30 min at room temperature in a tissue culture hood. Wells were washed twice with 2 mL of PBS, then 500 μL of N2B27 + Lif medium were added and plate was stored in a humidified incubator (37°C, 5% CO_2) until use.

Spheroids were briefly dissociated with Trypsin-EDTA and 30,000 cells in 1 mL of N2B27 + Lif were seeded in each well (~100 cells per microwell). Plate was incubated (37°C, 5% CO_2) for 3 days to obtain A-gastruloids.

For BMP inhibitor experiments, 500 nM of BMP inhibitor II DMH1 (Sigma #203646) was added at seeding and A-gastruloids were collected after 3 days for RNA extraction.

Culturing microfluidic gastruloids and Matrigel-embedded embryo-like structures

Spheroids were briefly dissociated with Trypsin-EDTA and a suspension of 168,000 cells in 10 mL of N2B27 + Lif was prepared (~120 cells per 7 μL droplet). Cells were loaded into the droplet-microfluidic device as described above and incubated for 5 days (37°C, 5% CO_2) to obtain F5-gastruloids.

Structures were recovered from the droplet-microfluidic device after 4 days by flipping the chip at a 90° angle while flushing pure FC-40. The oil was separated from the aqueous phase containing the gastruloids by filtration on a PTFE membrane (ThermoFisher Scientific #F2517-9). 1–4 gastruloids were seeded in each well of low adhesion flat-bottom 96-well plates (Corning #3474) in 100 μL of N2B27 + Lif medium containing 20% Matrigel (Corning #354234). Plates were incubated for 2 to 3 days (37°C, 5% CO_2) to promote elongation of the embryo-like structures.

METHOD DETAILS

Clonogenic assay

500 cells (D1 and D8 in serum+Lif medium; D18 in N2B27 + Lif) were seeded in 6-well plates. After 6 days, cells were fixed with 4% paraformaldehyde for 5 min and washed twice with PBS. Colonies were stained using Alkaline Phosphatase Blue Membrane Solution

Kit (Sigma #AB0300) for 15 min in the dark. Wells were washed once with PBS and left to dry. Plates were scanned and colonies were counted using ImageJ.

Cell sorting of the 3 cell types of D18 spheroids

Cell surface markers specific to each spheroid cell type were extracted from the list of scRNA-seq cluster markers (Table S1). The markers chosen for XEN-L and EPI-L cells were previously validated.[81]

Spheroids were briefly dissociated with StemPro Accutase (Gibco #A111-05-01) and cells were resuspended in PBS with 3% ES cell qualified FBS. Cells were incubated for 30 min at 4°C with the following antibodies: Cd31(Pecam1)-FITC (1:100, Invitrogen #11-0311-81), Pdgfra(Cd140)-PE (1:100, Invitrogen #12-1401-81), Cd24a-APCeFluor780 (1:100, Invitrogen #47-0242-82). Cells were washed twice with PBS–3% ES cell qualified FBS then resuspended in PBS–3% ES cell qualified FBS. Propidium iodide (1 μg/mL, Invitrogen #P3566) was added to cell suspension before transferring sample to a cell strainer cap tube. Cells were analyzed on a BD FACSAria III Cell Sorter (BD Biosciences). Cell fractions were collected in PBS–3% ES cell qualified FBS then divided for RNA extraction and resuspension in N2B27 + Lif medium for re-plating.

Droplet-microfluidic device design and fabrication

The molds to fabricate the chips were designed using Fusion 360 (Autodesk). The molds were patterned with 81 traps on the top of the culture chamber (Figure S5A for the dimensions). The drop dimension was increased, compared with previous designs,[33,82] to allow for the culture of mESC aggregates. In parallel, the trap dimensions and shapes were optimised to function at this size range, in which both gravitational and capillary effects are important (Bond number of order unity). In particular, we modified our previous designs by placing the anchors in the device ceiling and rounding their corners. As a result, gravity and capillarity combine together to insert the droplet into the trap and to keep it anchored over long periods. The molds were also equipped with four oblique rails in order to guide the drops evenly within the culture chamber. The molds were 3D printed using a ClearV4 resin (Formlabs) and an SLA 3D printer (Form3, Formlabs). The molds were filled with a mixture of PDMS (SYLGARD, Dow) base and a curing agent at a ratio of 1:10 (about 50–60 mL per chip), before being placed in an oven set at 65°C for at least 4 h. After curing, the PDMS was separated from the molds and then plasma treated (Cute, Femto Science Inc.) for two rounds of 40 s. It was then bonded to a 75 × 50 mm glass slide (Corning #2947) and placed in an oven set at 80°C for at least 2 h. Finally, the chips were rendered fluorophilic by treating them with Novec 1720 (3M) and heating them at 110°C for three rounds of 30 min each.

Pluripotent stem cell loading and manipulation within immobilized droplets

An Upchurch cross-junction (PEEK, low pressure, 1/16 compression size) was used to form 7 μL plugs flowed at 1,000 μL/min using syringe pumps (neMESYS, Cetoni) and each containing 120–300 pluripotent stem cells (PSCs) in their culture medium. The aqueous plugs were separated by 6 μL plugs of fluorinated oil (FC-40, 3M) containing a fluorogenic surfactant (RAN biotechnologies) at a concentration of 0.5% v/v, and were also flowed at 1,000 μL/min. The chips were placed at an angle of 45° from the horizontal to allow gravity to act as a driving force for droplet motion. The drops were then spontaneously captured in the bottom part of the traps, thus preventing other drops from being anchored in the same traps at this stage. The drops then spontaneously moved by gravity to the top part of the traps over 3–5 min, leaving empty the bottom part of the traps. The chips were then placed in a humidified incubator (37°C, 5% CO_2) to allow PSC culture for long time periods.

The performance of the droplet-microfluidic platform to promote PSC aggregation and expansion, and maintain expression of pluripotency markers was compared to standard 96-well plates. Briefly, about 300 mES-D3 were encapsulated into drops containing ESLIF medium, while the same cell number was seeded into 96-well plates in 100 μL of ESLIF medium per well. The kinetics of cell aggregation and proliferation were monitored by imaging.

The level of expression of pluripotency marker Ssea1 was quantified by flow cytometry, using an LSR-Fortessa (BD Biosciences), and by labeling the cells with mouse AlexaFluor647-conjugated anti-Ssea1 antibody (1:100, BD Biosciences # 560120).

The level of expression of pluripotency marker Oct4 was analyzed by imaging after methanol fixation and *in situ* immunolabelling using a mouse anti-Oct4 antibody (1:100, Millipore #MAB 4419), which was revealed using an Alexa Fluor 488 conjugated goat anti-mouse IgG1 (1:100, Invitrogen #A21121).

Immunofluorescence of spheroids, gastruloids and embryo-like structures

Spheroids were seeded in 8-well chamber slides (Ibidi #80826) at a density of 40,000 cells/300 μL. 48 h later, cells were fixed in 4% paraformaldehyde for 8 min, permeabilized in 0.2% Triton-PBS for 20 min and incubated in blocking buffer (10% BSA, 5% serum, 0.1% Triton-PBS) for 2 h at room temperature. For A-gastruloids, the structures were collected from AggreWells and fixed in 4% paraformaldehyde for 15 min. Then, the structures were permeabilized in 0.5% Triton-PBS for 20 min and incubated in blocking buffer (10% BSA, 5% serum, 0.1% Triton-PBS) for 3 h at room temperature. Primary antibodies were diluted in 1% BSA, 0.1% Triton-PBS and incubated overnight at 4°C with the spheroids or the A-gastruloids. After 3 washes of 10 min in PBS, the structures were incubated with the secondary antibodies diluted at 1:400 for at least 1 h at room temperature then washed with PBS 3 times for 10 min.

To immunolabel the F5-gastruloids and the M-ELS, the samples were first fixed with 4% paraformaldehyde for 2 h at 4°C. The structures were then incubated overnight at 4°C in PBSFT (5% FBS and 0.5% Triton X-100 in PBS). The primary antibodies were

diluted in PBSFT and incubated with the samples overnight at 4°C on an orbital rocker. After 3 washes with PBSFT, the samples were incubated overnight with a solution of 1:100 diluted secondary conjugated antibody containing 0.2 mM DAPI (ThermoFisher Scientific #R37606) at 4°C on an orbital rocker. After washing with PBS, the samples were cleared using RapiClear 1.52 (Sunjin lab), following the manufacturer's instructions. The specificity of the primary antibodies was verified by incubating the samples with the secondary antibody alone. Under these conditions, an absence of fluorescent signal validated the specificity of the primary antibodies.

Immunostaining on cryosections was performed as previously described.[33] Elongated M-ELSs were selected and fixed with 4% paraformaldehyde for 2 h at 4°C. After washing with PBS, M-ELSs were incubated overnight with a 30% sucrose (Sigma) solution at 4°C. M-ELSs were then placed in inclusion molds (7 × 7 × 5 mm, Simport). The sucrose solution was exchanged to Optimal Cutting Temperature (O.C.T) medium (Tissue Tek). The samples were frozen using a cooling bath (70% ethanol containing dry ice) and stored at −80°C. The day of the experiment, the O.C.T blocks were cut at 20 μm using a cryostat (CM3050 S, Leica). The cryosections were placed on glass slides (SuperFrost Plus Adhesion, ThermoFisher Scientific), dried at 37°C, and rehydrated using PBS. The samples were permeabilized on the slides for 5 min with a solution of 0.5% Triton X-100 in PBS, and blocked with a 5% FBS solution for 30 min. The primary antibodies were diluted in a 1% FBS solution and incubated with the samples for 4 h at 25°C. After 3 washes with PBS, the samples were incubated for 1.5 h with a solution of 1:100 diluted secondary conjugated antibody containing 0.2 mM DAPI (ThermoFisher Scientific #R37606). Lastly, the slides were mounted in mounting medium containing DAPI (Fluoromount-G, Invitrogen). A-gastrulo sections were performed with this same protocol.

The antibodies used were rat anti-Nanog (1:300, ThermoFisher Scientific #14-5761-80), goat anti-Sox17 (1:100, R&D Systems #AF1924), rabbit anti-Pou3f1 (1:100, Sigma #HPA073824), goat anti-T (1:100, R&D Systems #AF2085) or rabbit anti-T (1:100, Abcam #ab209655), rabbit anti-Snai1 (1:100, Cell Signaling #3879S), mouse anti-Pax6 (1:100, Abcam #ab78545), rabbit anti-Foxa2 (1:100, Cell Signaling #D56D6), mouse anti-Nestin (1:100, Millipore #MAB353), mouse anti-Tuj1 (1:100, Biolegend #8186), rabbit anti-Pax2 (1:100, Invitrogen #71–6000), mouse anti-Map2 (1:100, Sigma-Aldrich #M4403), rabbit anti-Sox10 (1:100, Abcam #ab264405), goat anti-Gbx2 (1:100, R&D systems #AF4638), rabbit anti-ZO1 (1:100, ThermoFisher Scientific #40–2200), mouse anti-Sox2 (1:100, Millipore #AB5603), Alexa Fluor 488 donkey anti-rat (Invitrogen #A21208), Alexa Fluor 546 donkey anti-goat (Invitrogen #A11056), Alexa Fluor 647 donkey anti-rabbit (Invitrogen #A31573), Alexa Fluor 488 goat-anti-mouse (Invitrogen # A11001), Alexa Fluor 488 donkey-anti-goat (Invitrogen # A32814) as well as Alexa Fluor 488 Phalloidin (1:400, Invitrogen #A12379).

In situ hybridization

A-gastruloids were collected 4 days after cell seeding in AggreWells and fixed for 7 h in 4% paraformaldehyde at 4°C before dehydration in methanol. No proteinase K incubation was performed after rehydration. M-ELSs were collected 2 days after Matrigel embedding and fixed overnight in 4% paraformaldehyde at 4°C before dehydration in methanol. M-ELSs were incubated for 10 min with proteinase K (10 μg/mL) after rehydration. *In situ* whole mount HCR V3 was performed as previously described[30] using reagents from Molecular Instruments. Briefly, each condition (up to 20 A-gastruloids or 6 M-ELSs) was incubated in 1 mL of probe hybridization buffer for 5 min at room temperature and 30 min at 37°C before incubation with 2 pmol of each probe in 500 μL of probe hybridization buffer overnight at 37°C. The next day, samples were washed 4 × 15 min with 1 mL probe wash buffer at 37°C, and 2 × 5 min with 1mL 5X SSC-Tween at room temperature, then incubated in 1 mL amplification buffer for 5 min at room temperature. A mixture of 30 pmol of each hairpin (individually snap cooled beforehand) in 500 μL of amplification buffer was added to samples for an overnight incubation at room temperature in the dark. The next day, samples were washed 2 × 5 min, 2 × 30 min, 1 × 5 min with 1 mL 5X SSC-Tween at room temperature in the dark then stored at 4°C before imaging. Accession numbers for HCR probes used were *Nanog* (NM_001289828.1, hairpin B1), *T* (NM_009309.2, hairpin B3), *En1* (NM_010133.2 hairpin B4), *Uncx* (NM_013702.3, hairpin B1). Hairpins B1 were labeled with Alexa Fluor 488 or Alexa Fluor 546, hairpin B3 with Alexa Fluor 647 and hairpin B4 with Alexa Fluor 488.

Microscopy

The images were acquired using a motorized microscope (Ti or Ti 2, Eclipse, Nikon), equipped with a CMOS (complementary metal-oxide semiconductor) camera (ORCA-Flash4.0, Hamamatsu). Widefield imaging was performed by illuminating the samples with a fluorescence light-emitting diode source (Spectra X, Lumencor), while for spinning disc confocal imaging the samples were illuminated with lasers (W1, Yokogawa). The images were taken with a 10× objective with a 4-mm working distance (extra-long working distance) and a 0.45 numerical aperture (NA) (Plan Apo λ, Nikon). Z sections were taken every 2 μm.

For widefield live imaging, the samples were imaged using a Muvicyte (PerkinElmer) equipped with a 10× objective with a 10-mm working distance and 0.30 NA (UPlanFL N, Olympus), which was placed in a humidified incubator (37°C, 5% CO_2). Images were acquired in brightfield every 30 min for the M-ELSs. For the fluorescent reporter cell line Sox1eGFP–TmCherry, cells were cultured in phenol red-free N2B27 + Lif medium and images were acquired every 3 h.

Image analysis

The brightfield and fluorescent images were analyzed with a Python custom image analysis algorithm. Briefly, the gastruloids and M-ELSs were first detected from the brightfield images by edges detection. They were then centered and aligned along their major

axis, which enabled to measure their major (*a*) and minor (*b*) axis length, area, convex area and perimeter. The eccentricity, aspect ratio, circularity, solidity and radius of curvature were calculated as follows:

$$Eccentricity = \sqrt{\left(1 - \frac{a^2}{b^2}\right)}$$

$$Aspect\ ratio = \frac{a}{b}$$

$$Circularity = 4\pi \times \frac{Area}{Perimeter^2}$$

$$Solidity = \frac{Area}{Convex\ Area}$$

The radius of curvature was calculated as the radius of the circle fitting the major axis of the M-ELSs. To demonstrate the head-trunk organization of M-ELS, the brightfield images were threshold (Otsu's method) and the width was measured along the major axis. The positions along the major axis were then normalized to the total length of the major axis. The values were grouped into a specific number of bins (ranging from 0 to 1, with 0 being the edge of the head and 1 the edge of the end of the trunk of M-ELSs).

To overlap the shapes of elongated M-ELSs, brightfield images were threshold (Otsu's method) and the boundaries of the binary images were outlined. The outlines were aligned along the major axis of each M-ELS and centered along their centers of mass. The median geometric profile of the M-ELSs was calculated and overlaid on the aligned shapes.

The fluorescent images were segmented using an automatically calculated threshold (Otsu's method). Then, the segments corresponding to F-gastruloids were oriented along their major axis, according to their red fluorescent signal (i.e. mCherry, TRITC). To quantify the time evolution of the structural organization within Sox1eGFP–TmCherry fluorescent reporter F-gastruloids, the area of the mCherry and eGFP signals was measured for every time point.

For immunofluorescence images, the length of the major or minor axis was normalized for each gastruloid. Then, the images were segmented along the selected axis into a specific number of bins (ranging from −0.5 to 0.5, with 0 being the center of the gastruloid), for which the average fluorescent signal of each channel was measured.

The representative confocal fluorescent images are presented as a maximum z-intensity projection. For *Uncx* staining, the fluorescent images were pre-treated with a median filter of a radius of 4 pixels and a maximum z-intensity projection of the consecutive frames containing *Uncx* fluorescent signal is presented.

Embryo collection, culture and ML-792 treatment

Animal experiments were carried out in compliance with local regulations (Government of Upper Bavaria). For the SUMO1, SUMO2/3 and PML immunostainings, CD1 female mice (4–8 weeks old) were mated with CD1 male mice (3–6 months old) and zygotes, 2-cell, 4-cell, 8-cell, morula and blastocyst embryos were collected at 18 h, 42 h, 53 h, 60 h, 67 h and 88 h post-coitum. For the SUMOylation inhibition experiments with ML-792, embryos were collected from 5–6-week-old F1 (C57BL/6J×CBA/H) females mated with F1 males (3–6 months old) after hormonal induction with pregnant mare serum gonadotropin (5 IU, Intervet) and human chorionic gonadotropin (hCG; 7.5 IU, Intervet) 46–48 h later. Embryos were collected at 98 h post-hCG injection for the early to late blastocyst treatment (Figure S1C). Embryos were cultured for 16 h with 0.2% DMSO or 20 M of ML-792 in K-modified simplex optimized medium (KSOM) microdrops under paraffin oil at 37°C, 5% CO_2.

Immunostaining and confocal microscopy of embryos

Embryos were fixed as previously described.[83] Briefly, the zona pellucida was removed with Acid Tyrode solution, followed by two washes in PBS and fixation for 20 min in 4% paraformaldehyde, 0.04% Triton, 0.3% Tween 20, 0.2% sucrose at 37°C. Embryos were then washed with PBS and permeabilized with 0.05% Triton X-100 for 20 min. After permeabilization, embryos were washed three times in PBST (0.1% Tween 20 in PBS), quenched in 2.6 mg/mL freshly prepared ammonium chloride, washed three times in PBST, blocked for 3–4 h at 4°C in blocking solution (3% BSA in PBST) and incubated with primary antibodies in blocking solution. Antibodies used were as follows: anti-SUMO1 (1:125, Abcam #ab32058), anti-SUMO2/3 (1:125, Abcam #ab81371), anti-PML (1:500, Merck MAB3738). After overnight incubation at 4°C, embryos were washed three times in PBST, blocked and incubated for 3 h at room temperature in blocking solution containing secondary antibodies labeled with Alexa fluorophores (1:500, Invitrogen #A32731, #A32732, #A32733, #A11029, #A21424 and #A32728). After washing, embryos were mounted in Vectashield containing DAPI (Vector Laboratories).

Confocal microscopy was performed using a plan-Apo 63x NA 1.4 oil immersion objective on a TCS SP8 inverted confocal microscope (Leica). Z sections were taken every 0.3–1 μm. Image analysis was performed using Imaris software (Bitplane). For all experiments, acquisition parameters were set to obtain fluorescence intensity signals in the linear range of hybrid detectors. These detectors have negligible detector noise and linearly amplify incoming photons into photoelectrons, thereby allowing direct counting of emitted photons, provided the detector is not saturated, which we verify for each acquisition. Hence, given identical acquisition settings, the recovered fluorescence signal accurately reflects the amount of antigen present in the sample. The quantifications are subsequently performed in 3D reconstructions based on Z section acquisitions for which the DAPI is used to segment the whole volume of each individual nucleus. All the fluorescence signal (and thus measured photons) contained within the nucleus is then calculated as the sum of intensities throughout the 3D volume for each nucleus. For the analysis of SUMO levels with regards to DNA content, the sum of intensities of each SUMO channel was divided by the sum of intensities of the DAPI channel based on DAPI segmentation. The coefficient of variation (CV) of the mean SUMO1 and SUMO2/3 intensity is then calculated for each embryo as the standard deviation divided by the mean. Embryos with cells in mitosis were excluded from the quantification.

Single-cell RNA-seq

Cells were dissociated with Trypsin-EDTA and resuspended in PBS. Propidium iodide (1 μg/mL, Invitrogen #P3566) and Calcein AM (1.5 μg/mL, Invitrogen #C3100MP) were added to cell suspension before transferring sample to a cell strainer cap tube. Cells were sorted into 384-well cell capture plates using a BD FACSAria III Cell Sorter (BD Biosciences) to collect live cells and sort only singlets. Plates were snap frozen on dry ice and stored at −80°C until further processing. All single cell libraries were prepared with the same conditions and reagents using the MARS-seq protocol as previously described.[84] Briefly, a Bravo Automated Liquid Handling Platform (Agilent) was used to reverse transcribe (Invitrogen #18080085) mRNA into cDNA with an oligonucleotide containing both the unique molecule identifiers (UMIs) and cell barcodes. Unused oligonucleotides were removed by Exonuclease I (New England Biolabs #M0293S) treatment. cDNAs were pooled (each pool containing half of a 384-well plate) for second strand synthesis (New England Biolabs #E6111S) and *in vitro* transcription amplification (New England Biolabs #E2040S). DNA template was removed (Invitrogen #AM2238) before fragmenting (Invitrogen #AM8740) and ligating (New England Biolabs #M0204S) resulting RNA to an oligo containing the pool barcode and Illumina sequences. Finally, RNA was reverse transcribed (Agilent Technologies #600107) and libraries were amplified (Roche #7958935001). Libraries were quantified with a Qubit 2.0 (Invitrogen) and their size distribution was determined by a 4200 TapeStation System (Agilent Technologies). Finally, libraries were pooled at equimolar concentration and sequenced on an Illumina NextSeq500, in 8 sequencing runs, using high-output 75 cycles v2.5 kits (Illumina #20024906).

Processing single-cell data

The mouse genome GRCm38.p6 (mm10) with the gencode annotation M23 was used for all sequencing analyses (https://www.gencodegenes.org/mouse/release_M23.html).

The MARS-seq2.0 pipeline[85] was used to produce count tables. The Seurat 4 R package[69] was used for normalization, dimension reduction and clustering. A manual iterative strategy was used to exclude cell libraries with low complexities. Briefly, all libraries (cells and empty control wells) in the count matrix were run through a standard Seurat workflow from count data to cluster computation (50 PCA dimensions to generate the neighbors graph and UMAP computation). Empty wells and poor-quality cells usually clustered together and manual inspection allowed removal of clusters with low UMIs (inferior in mean ∼1,000 UMIs). This process was repeated until no low UMIs cluster remained. Cells with mitochondrial gene expression fractions greater than 2.5% were also excluded. Batch effects due to sequencing runs performed on different days were removed using the Harmony package[70] in Seurat.

Cluster markers were computed with the 'FindAllMarkers' Seurat function using the default parameters (except for the only.pos argument set to True, to only list genes upregulated in each cluster). Markers were considered significant if their adjusted p value was inferior to 0.05.

Comparison between scRNA-seq clusters and *in vivo* datasets

Count tables for our scRNA-seq data and for the published datasets were filtered to keep only the common annotations (20,358 for comparison with the E4.5–6.5 data[24]; 28,859 for comparison with the E6.5–8.5 data[27]; 21,054 for comparison with the E9.5–13.5 data[35]). Cluster markers were computed for these annotation sets with the 'FindAllMarkers' Seurat function using the default parameters (except for the only.pos argument set to True, to only list genes upregulated in each cluster). For the D18 spheroid clusters comparison to *in vivo* data, the 'FindMarkers' Seurat function was used to compute genes differentially expressed between the XEN-L (#2) and EPI-L (#3) clusters. The list of genes upregulated in each cluster was compared to the lists of differentially expressed genes between the epiblast and the primitive endoderm/visceral endoderm at E4.5, E5.5, E6.5.[24] For the A-gastruloids, F-gastruloids and M-ELSs, the lists of cluster markers computed with the 'FindAllMarkers' Seurat function were compared to the markers identified for the different embryonic cell types defined in previously published mouse embryo scRNA-seq datasets.[27,35] Common genes, with a $\log_2$-transformed fold change superior to 1.01 and an adjusted p value inferior to 0.01, were found and significance was assigned using a binomial test as previously described.[57]

To overlay our cells with the *in vivo* data, the Seurat objects for the stage/condition of interest were merged. Cluster computation was performed (50 PCA dimensions to generate the neighbors graph and UMAP computation) and batch effects were removed using FastMNN[86] (200 features).

Pseudotime analysis

Cells from the F5-gastruloid condition were selected and cluster computation was performed as described above. This F5-gastruloid UMAP was then used to determine trajectories for pseudotime with the Monocle3 R package[71] using the 'learn_graph' function with default parameters. The pseudotime start was chosen in the NMPs cluster.

For heatmap representation, cells were ordered by their pseudotime, and a cubic smoothing spline (R function 'smooth.spline') was applied on the gene expression matrix. The z-scores of the smoothed values were plotted.

Expression levels of the SUMO machinery components

The scRNA-seq dataset generated by Deng et al.[65] was used. Count tables were downloaded from GSE45719. Counts were normalized with the Seurat function 'NormalizeData'. Stages and genes of interest were selected for plotting.

Methyl-seq

DNA was extracted and purified from 2 million cells for each condition with the *Quick*-DNA Midiprep Plus Kit (Zymo Research D4075) following manufacturer's instructions. DNA was quantified with a NanoDrop ND-1000 (ThermoFisher Scientific). The NEBNext Enzymatic Methyl-seq Kit (New England Biolabs #E7120S) was used to prepare libraries for detection of 5-mC and 5-hmC. 200 ng of DNA from each sample were sheared to 275 bp fragments with an E220 Focused-ultrasonicator (Covaris) with the following settings: Duty Factor, 10% – Peak Incident Power, 175 W – Cycles per burst, 200 – Duration, 100 s. Fragment size was validated by a 4200 TapeStation System (Agilent Technologies). The NEBNext Enzymatic Methyl-seq Kit workflow was then followed, using the sodium hydroxide option for the denaturation step. The size distribution and concentration of the libraries was determined by TapeStation. The libraries were sequenced on an Illumina HiSeq4000 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.

Processing methyl-seq data

Methyl-seq data were processed with the Bismark pipeline[72] using bowtie2 aligner[67] with the default parameters. Biological triplicates were merged and CG sites with at least 5 reads were kept for downstream analyses. Methylated and unmethylated CG sites were counted within predetermined windows (bin or interval) and a binomial test was used to compare different timepoints or regions. In a second approach, bins of 500 bp were compiled across the whole genome with a step of 200 bp. Next, replicates were classified as ''early'' (D1 and D3 replicates) or ''late'' (D8 and D10 replicates) with an inclusion criterion of at least 20 reads per bin in all combined conditions and a minimum of 1 read per time point. ''Early'' versus ''late'' conditions were compared using a binomial generalized linear model (logit) as follows: glm(data = test, cbind(test$meth, test$unmeth) ~ test$condi, family = binomial(link = "logit")) where ''condi'' is the status of the sample (''early'' or ''late''). This model was computed for each bin and a Bonferroni correction of 5% was applied to the p value. Significant bins were extracted corresponding to 22,670 genomic regions (overlapping significant bins were merged). 3,080 regions were obtained after filtering the 22,670 regions based on a $\log_2$FC > 2 or < −2 between D10 and D1(Table S2). Allocation to pre-defined chromatin states was as in Pintacuda et al.[64] (https://github.com/guifengwei/ChromHMM_mESC_mm10). The categories ''Enhancer'', ''Weak Enhancer'' and ''Strong Enhancer'' were merged into a single category ''Enhancer''. ESC super-enhancers genome coordinates were taken from Whyte et al.[45] The liftOver webtool from the UCSC website (https://genome.ucsc.edu/cgi-bin/hgLiftOver) was used to convert the mm9 track bed file to an mm10 bed file.

ChIP-seq

Cells at D1, D3, D8 and D10 were fixed for 10 min at room temperature in culture medium with 1% formaldehyde (ThermoFisher Scientific #28908). Formaldehyde was then quenched with glycine (125 mM final). Cells were washed in ice-cold PBS. The extracted chromatin was sonicated with a Bioruptor Pico (Diagenode) until chromatin fragments reached a size of 200–400 base pairs (30 s ON, 30 s OFF, 6 cycles), as assayed by electrophoresis through agarose gels. Immunoprecipitation, reversal of cross-linking and DNA purification were performed using ChIP-IT kit (Active Motif #53040). Antibodies against SUMO1 (Abcam #ab32058), SUMO2/3 (Abcam #ab3742), H3K4me3 (Active Motif #39159), H3K9me3 (Abcam #ab8898), H3K27me3 (Millipore #07–449), Zfp57 (Abcam #ab45341), Kap1 (Abcam #ab10483) were used for ChIP-seq. 50 ng of spike-in chromatin (Active Motif #53083) and 2 μg of spike-in antibody (Active Motif #61686) were added to normalize the signal between ChIP-seq experimental samples. ChIP-seq libraries were prepared using Microplex Library Preparation kit V2 (Diagenode #C05010012) 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 (200–400 bp) and cleaned-up using AMPure XP beads (Beckman Coulter #A63881). Prior to analyses, DNA libraries were checked for quality and quantified using a 2100 Bioanalyzer (Agilent). The libraries were sequenced on an Illumina HiSeq4000 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.

Processing ChIP-seq data

Libraries were aligned using bowtie2[67] with default parameters on mouse and fly genomes together. All alignments were filtered on MAPQ (mapping quality value) 30 with SAMtools.[73] Libraries were deduplicated with the Picard toolkit.[87] The number of reads mapped on the fly genome was used as a spike-in value to downsample libraries as previously described.[88] Peak calling was performed with MACS2[66] with default parameters.

For the SUMO1 ChIP-seq data, low coverage peaks were filtered out. The pileup values were extracted from the MACS2 output, transformed using $\log_{10}$ and scaled. Peaks with a scaled $\log_{10}$ value inferior to -0.5 were filtered out from each replicate. This corresponded approximately to the 33% quantile (-0.52, -0.53, -0.48, -0.55, for D1 rep1, D1 rep2, D8 rep1, D8 rep2, respectively).

An irreproducible discovery rate (IDR)[74] of 0.1 was used to filter out irreproducible peaks. For each histone mark or SUMO1 ChIP-seq dataset, the IDR validated peaks from all time points were merged using the bedtools merge function.[89]

A differential analysis was performed for the SUMO1 ChIP-seq data by counting the number of reads for each peak in each downsampled replicate with the featureCounts program.[90] The produced matrix was analyzed with the DESeq2 R package,[68] using a size factor of 1 for the 4 libraries (2 rep D1, 2 rep D8). Changes of SUMO1 levels between D1 and D8 were considered significant if the adjusted p value was inferior to 0.05. For SUMO2/3, Zfp57 and Kap1 the same procedures were used except the spike-in part. For the differential analysis by DESeq2 the size factors were computed by the program.

For motif enrichment analysis, a 400 bp window centered on the local maximum coverage for each peak was first identified. The MEME-ChIP webtool[75] was used with default parameters. H3K4me3 and H3K27me3 ChIP-seq data generated in this study were used to classify the transcription start sites (TSSs). TSSs were classified as "inactive" in the absence of both peaks, "active" when marked only by H3K4me3, "repressed" when marked only by H3K27me3 and "bivalent" when having both H3K4me3 and H3K27me3. The class of TSSs was attributed for SUMO1, SUMO2/3, Zfp57 and Kap1 peaks overlapping the 1 kb neighborhood centered on any TSS. SUMO1, SUMO2/3, Zfp57 and Kap1 peaks were annotated with the following priority: TSS, exon, intron, intergenic with respect to the gencode M23 transcript annotations.

Gene enrichment analysis

The EGSEA R package[76] was used for gene list enrichment with Gene Ontology term (GO term), pathways (KEGG, Biocarta) or curated gene list (mSigDB) with the egsea.ora function (Over-representation Analysis).

Bulk RNA-seq

Total RNA was purified by Trizol extraction and RNA was analyzed on a BioAnalyzer Nano chip (Agilent). If the RNA integrity number was superior to 8, samples were used for subsequent analyses. RNA concentration was quantified with a Qubit (Invitrogen). Total RNA-seq libraries were generated from 500 ng of total RNA using TruSeq Stranded Total RNA Library Prep Gold kit and TruSeq RNA Single Indexes kits A and B (Illumina), according to manufacturer's instructions. Briefly, cytoplasmic and mitochondrial ribosomal RNA (rRNA) were removed using biotinylated, target-specific oligos combined with Ribo-Zero rRNA removal beads. Following purification, the depleted RNA was fragmented into small pieces using divalent cations at 94°C for 2 min. Cleaved RNA fragments were then copied into first strand cDNA using reverse transcriptase and random primers followed by second strand cDNA synthesis using DNA Polymerase I and RNase H. Strand specificity was achieved by replacing dTTP with dUTP during second strand synthesis. 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 s at 98°C, 30 s at 60°C, 30 s at 72°C] x 12 cycles; 5 min at 72°C). Surplus PCR primers were further removed by purification using AMPure XP beads (Beckman Coulter #A63881) and the final cDNA libraries were checked for quality and quantified using capillary electrophoresis. The libraries were sequenced on an Illumina HiSeq4000 sequencer as paired-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.

Processing bulk RNA-seq data

FastQC (Version 0.11.2) was run using the following arguments –nogroup –casava to produce base quality, base sequence content and duplicated reads. FastQ-Screen (Version 0.5.1) was run using the following arguments: –subset 10000000 –aligner bowtie –bowtie '-p 2′x. In order to avoid PCR amplification biases in read quantification, duplicated reads were removed using the MarkDuplicates tool of Picard. The differential expression analysis of DESeq2 was applied on the filtered replicates.[68]

Cell lysates preparation for the SUMO2/3 IP

Untreated control cells (D1) or cells having recovered from ML-792 treatment (D8) were cultured following the protocol described in Figure 2A. 200 million cells per replicate (7 × 150 mm dishes) were washed twice with ice-cold PBS and collected on ice by scraping. Cells were centrifuged at 1,000*g* for 5 min at 4°C, and cell pellets were resuspended in 2 mL of ice-cold PBS per 100 μL of pellet. A 100 μL aliquot of cell suspension was collected, spun down at 1,000*g* for 5 min at 4°C, after which the cells were lysed in 100 μL of SNTBS buffer (2% SDS, 1% NP40, 150 mM NaCl, 50 mM Tris-HCl pH 7.0) and snap frozen in liquid nitrogen to serve as an input control. The rest of the cell suspension was spun down at 1,000*g* for 5 min at 4°C, after which the cells were vigorously lysed in 5 mL of guanidine lysis buffer (6 M guanidine hydrochloride, 50 mM Tris-HCl pH 8.5) and snap frozen in liquid nitrogen. Samples

were stored at −80°C until processing. In essence, sample preparation and SUMO-IP for native and endogenous mass spectrometry (MS) analysis were performed as described previously.[91] Lysates were thawed at room temperature, after which they were supplemented with 5 mM chloroacetamide (CAA) and 5 mM Tris(2-carboxyethyl)phosphine (TCEP). Samples were homogenized via sonication using a microtip sonicator, at 20 W using two 10 s pulses, and afterward cleared by centrifugation at 4,250*g*. Endoproteinase Lys-C (Wako) was added to samples in a 1:200 enzyme-to-protein ratio (w/w). Digestion was performed overnight, still, and at room temperature. Digested samples were diluted with three volumes of 50 mM ammonium bicarbonate (ABC), and a second round of overnight digestion was performed by addition of Lys-C in a 1:200 enzyme-to-protein ratio. Digests were acidified by addition of trifluoroacetic acid (TFA) to a final concentration of 0.5%, after which they were centrifuged at 4,250*g* and at 4°C for 30 min. Clarified digests were carefully decanted into clean 50 mL tubes, after which peptides were purified using C8 Sep-Pak cartridges (Waters) according to the manufacturer's instructions. Sep-Pak cartridges with 500 mg C8 sorbent were used, with one cartridge used per sample. Unrelated small peptides were washed off using 5 mL of 25% ACN in 0.1% TFA. SUMOylated peptides were eluted using 4 mL of 45% ACN in 0.1% TFA. SepPak elutions were collected in 50 mL tubes with small holes punctured into the caps, and then frozen for 2 h at −80°C. Deep-frozen samples were lyophilized to dryness for 120 h, with the pressure target set at 0.004 mbar and the condenser coil at −90°C.

Crosslinking of SUMO2/3 antibody to beads

300 μL of Protein G Agarose beads (Roche) were used to capture 100 μL of SUMO2/3 antibody (8A2, from Abcam #ab81371; ~5–10 μg/μL antibody). All washing and handling steps were followed by centrifugation of the beads at 500*g* for 3 min in a swing-out centrifuge with delayed deceleration and careful aspiration of buffers, to minimize loss of beads. Beads were pre-washed 3 times with ice-cold PBS, after which the antibody was added, and the tube was completely filled with ice-cold PBS. Beads and antibody were incubated at 4°C on a rotating mixer for 2 h, and subsequently washed 3 times with ice-cold PBS. Crosslinking of the antibody to the beads was achieved by addition of 1.2 mL of 0.2 M sodium borate, pH 9.0, which was freshly supplemented with 20 mM dimethyl pimelimidate. Crosslinking was performed for 30 min at room temperature on a rotating mixer, after which the crosslinking step was repeated once. Next, SUMO-IP beads where washed twice with ice-cold PBS, twice with 0.1 M glycine pH 2.8, and three times with ice-cold PBS, after which all beads were pooled in a single 1.5 mL tube and stored until use at 4°C in PBS supplemented with 10 mM sodium azide.

Purification of SUMOylated peptides

Lyophilized peptides were dissolved in 5 mL ice-cold SUMO-IP buffer (50 mM MOPS, 10 mM Na_2HPO_4, 50 mM NaCl, buffered at pH 7.2). Samples were clarified by centrifugation at 4,250*g* for 30 min at 4°C in a swing-out centrifuge with delayed deceleration. Samples were transferred to new tubes, after which 35 μL SUMO-IP beads were added per sample. Samples were incubated at 4°C for 4 h in a rotating mixer, after which the beads were washed twice with ice-cold SUMO-IP buffer, twice with ice-cold PBS, and twice with ice-cold MQ water. Upon each first wash with a new buffer, beads were transferred to a clean 1.5 mL LoBind tube (Eppendorf). To minimize loss of beads, all centrifugation steps were performed at 500*g* for 3 min at 4°C in a swing-out centrifuge with delayed deceleration. Elution of SUMO peptides from the beads was performed by addition of 75 μL of ice-cold 0.15% TFA, and performed for 30 min while standing still on ice, with gentle mixing every 10 min. The elution of the beads was repeated once, and both elutions were cleared through 0.45 μm spin filters (Millipore) by centrifuging at 12,000*g* for 1 min at 4°C. The two elutions from the same samples were pooled after clarification. Next, samples were pH-neutralized by addition of 10 μL 1 M Na_2HPO_4, and allowed to warm up to room temperature. Second-stage digestion of SUMOylated peptides was performed with 250 ng of Endoproteinase Asp-N (Roche). Digestion was performed overnight, at 30°C and shaking at 300 rpm, after which samples were frozen at −80°C until further processing.

StageTip purification and high-pH fractionation of SUMO-IP samples

Preparation of StageTips and high-pH fractionation of SUMO-IP samples on StageTip, was performed essentially as described previously.[7,91] Quad-layer StageTips were prepared using four punch-outs of C18 material (Sigma-Aldrich, Empore SPE Disks, C18, 47 mm). StageTips were equilibrated using 100 μL of methanol, 100 μL of 80% ACN in 200 mM ammonium, and two times 75 μL 50 mM ammonium. Samples were thawed out, and supplemented with 10 μL 200 mM ammonium, just prior to loading them on StageTip. The StageTips were subsequently washed twice with 150 μL 50 mM ammonium, and afterward eluted as six fractions (F1-6) using 80 μL of 4, 7, 10, 13, 17, and 25% ACN in 50 mM ammonium. All fractions were dried to completion in LoBind tubes, using a SpeedVac for 3 h at 60°C, after which the dried peptides were dissolved using 11 μL of 0.1% formic acid.

MS analysis

All samples were analyzed using an EASY-nLC 1200 system (ThermoFisher Scientific) coupled to an Orbitrap Exploris 480 mass spectrometer (ThermoFisher Scientific). Samples were analyzed as two technical replicates, with 5 μL of sample injected per run. The two technical methods for analysing the samples are from here on referred to as "Standard" and "Sensitive" methods. Separation of peptides was performed using 15-cm columns (75 μm internal diameter) packed in-house with ReproSil-Pur 120 C18-AQ 1.9 μm beads (Dr. Maisch). Elution of peptides from the column was achieved using a gradient ranging from buffer A (0.1% formic acid) to buffer B (80% acetonitrile in 0.1% formic acid), at a flow of 250 nL/min. Gradient length was 65 or 50 min per sample, including ramp-up and wash-out, with an analytical gradient of 40 or 30 min, for the Standard and Sensitive methods, respectively. The buffer B

ramp for the analytical gradient was as follows: F1: 13–24%, F2: 14–27%, F3-5: 15–30%, F6: 17–32%. The analytical column was heated to 40°C using a column oven, and ionization was achieved using a NanoSpray Flex NG (ThermoFisher Scientific) with the spray voltage set at 2 kV, an ion transfer tube temperature of 275°C, and an RF funnel level of 50%. Full scan range was set to 500-1,500 m/z with a precursor selection window of 550-1,500 m/z, MS1 resolution to 120,000, MS1 AGC target to "200" (=2,000,000 charges), and MS1 maximum injection time to "Auto". Monoisotopic Precursor Selection (MIPS) was enabled in "Peptide" mode, without relaxation of restrictions in case insufficient precursors were observed. Precursor intensity threshold was set to 80,000 (Normal) or 40,000 (Sensitive). Precursors with charges 2–6 (Normal) or 3–6 (Sensitive) were selected for fragmentation using an isolation width of 1.3 m/z, and fragmented using higher-energy collision disassociation (HCD) with normalized collision energy of 25. Precursors were excluded from re-sequencing by setting a dynamic exclusion of 40 s (Normal) or 30 s (Sensitive), and with an exclusion mass tolerance of 15 ppm MS2 resolution was set to 60,000, MS2 AGC target to "200" (=200,000 charges), and MS2 maximum injection time to 250 ms (Normal) or 500 ms (Sensitive). Data-dependent scans were limited to 3 (Normal) or 2 (Sensitive) per duty cycle.

Analysis of MS data

All MS RAW data was analyzed using the freely available MaxQuant software, version 1.5.3.30.[77,78] All data was processed in a single computational run, with exceptions to default MaxQuant settings specified below. For generation of the theoretical spectral library, the mouse FASTA database was downloaded from Uniprot on the 14th of February, 2020. The mature sequence of SUMO2 was inserted in the database to allow for detection of free SUMO. *In silico* digestion of theoretical peptides was performed with Lys-C, Asp-N, and Glu-N, allowing up to 8 missed cleavages. Variable modifications used were protein N-terminal acetylation, methionine oxidation, peptide N-terminal pyroglutamate, and Lys SUMOylation, with a maximum of 3 modifications per peptide. The SUMO mass remnant was defined as described previously[91]; DVFQQQTGG, H60C41N12O15, monoisotopic mass 960.4301, neutral loss b7-DVFQQQT, diagnostic mass remnants [b2-DV, b3-DVF, b4-DVFQ, b5-DVFQQ, b6-DVFQQQ, b7-DVFQQQT, b9-DVFQQ QTGG, QQ, FQ, FQQ]. Label-free quantification was enabled, with "Fast LFQ" disabled. Maximum peptide mass was set to 6,000 Da. Stringent MaxQuant 1% FDR filtering was applied (default), and additional automatic filtering was ensured by setting the minimum delta score for modified peptides to 20, with a site decoy fraction of 2%. Second peptide search was disabled. Matching between runs was enabled, with a match time window of 1 min and an alignment window of 20 min. For protein quantification, the same variable modifications were included as for the peptide search. To further minimize false-positive discovery, additional manual filtering was performed at the peptide level. All modified peptides were required to have a localization probability of >75%, be supported by diagnostic mass remnants, be absent in the decoy database, and have a delta score of >100 in case SUMO modification was detected on a peptide C-terminal lysine not preceding an aspartic acid or glutamic acid. Multiply-SUMOylated peptides were accepted with an Andromeda score of >100 and a delta score of >40. SUMO target proteins were derived from the "proteinGroups.txt" file, and all post-filtering SUMO sites were manually mapped. Only proteins containing at least one SUMO site were considered as SUMO target proteins, and other putative SUMO target proteins were discarded.

Immunoblots

Cells were collected and directly lysed in Laemmli buffer (Bio-Rad #161–0747). Proteins were quantified using Pierce 660 nm Protein Assay (ThermoFisher Scientific #22662) according to manufacturer's instructions. Equal amounts of proteins were loaded on gels and good equilibration of the different samples was assessed by Ponceau staining after membrane transfer. Antibodies against SUMO1 (1:1000, Abcam #ab32058), Actin (1:4000, Sigma #A1978), SUMO2/3 (1:1000, Abcam #ab81371), Gapdh (1:1000, Cell Signaling #2118), Dnmt3a (1:1000, Abcam #ab2850), Histone H3 (1:5000, Abcam #ab24834), Dnmt3b (1:1000, Abcam #ab2851), Tet2 (1:1000, Cell Signaling #36449S) were used according to standard protocols and suppliers' recommendations.

Quantitative PCR

cDNA was generated with a High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems #4368814) from 500 ng to 2 μg of total RNA purified by Trizol extraction. Quantitative real-time PCR analysis was performed with SYBR Green PCR master mix (Applied Biosystems #4309155) and the primer sets indicated in Table S2 using cDNA. Quantitative real-time PCR analysis was performed on a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) or a QuantStudio 6 Flex Real-Time PCR System (Applied Biosystems).

QUANTIFICATION AND STATISTICAL ANALYSIS

Graphical presentation and statistical analysis were performed by Graphpad Prism 9 and R version 4. Data with error bars represent means ± standard error of mean (S.E.M.). Boxplots show median (horizontal line inside the box), first and third quartiles of data (lower and upper hinges) and values no further or lower than 1.5*distance between the first and third quartiles (upper and lower whisker). Statistical tests used are indicated in the figure legends and different levels of statistical significance were denoted by p values (*$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$).