nature cell biology

Article

https://doi.org/10.1038/s41556-023-01191-z

H3K36 methylation maintains cell identity by regulating opposing lineage programmes

Received: 29 April 2022

Accepted: 19 June 2023

Published online: 17 July 2023

Check for updates

Michael S. Hoetker[1,2,3,4,5,6], Masaki Yagi[1,2,3,4,5,6], Bruno Di Stefano[1,2,3,4,5,6], Justin Langerman[7], Simona Cristea[8,9], Lai Ping Wong[1], Aaron J. Huebner[1,2,3,4,5,6], Jocelyn Charlton[10,11], Weixian Deng[7], Chuck Haggerty[11], Ruslan I. Sadreyev[1,12], Alexander Meissner[5,6,10,11], Franziska Michor[6,8,9,10,13,14], Kathrin Plath[7] & Konrad Hochedlinger[1,2,3,4,5,6]

The epigenetic mechanisms that maintain differentiated cell states remain incompletely understood. Here we employed histone mutants to uncover a crucial role for H3K36 methylation in the maintenance of cell identities across diverse developmental contexts. Focusing on the experimental induction of pluripotency, we show that H3K36M-mediated depletion of H3K36 methylation endows fibroblasts with a plastic state poised to acquire pluripotency in nearly all cells. At a cellular level, H3K36M facilitates epithelial plasticity by rendering fibroblasts insensitive to TGFβ signals. At a molecular level, H3K36M enables the decommissioning of mesenchymal enhancers and the parallel activation of epithelial/stem cell enhancers. This enhancer rewiring is Tet dependent and redirects Sox2 from promiscuous somatic to pluripotency targets. Our findings reveal a previously unappreciated dual role for H3K36 methylation in the maintenance of cell identity by integrating a crucial developmental pathway into sustained expression of cell-type-specific programmes, and by opposing the expression of alternative lineage programmes through enhancer methylation.

The process of cellular differentiation has been relatively well characterized in vivo and in vitro using animal and stem cell models in combination with genome-scale assays[1–8]. By contrast, the mechanisms that maintain and reinforce specialized cell states once differentiation is complete remain less well understood. Resolving this fundamental question is key not only for ensuring the continuous maintenance and functionality of adult tissues but also for dissecting oncogenesis, which is considered an aberration of normal cell states and differentiation programmes[9]. The experimentally induced conversion of somatic cells to induced pluripotent stem (iPS) cells with transcription factors represents a powerful system to interrogate the mechanisms that maintain differentiated cell identities[10]. Induced pluripotency is typically inefficient (1–3%) and lengthy (1–2 weeks), which is thought to be due to redundant mechanisms that safeguard differentiated

[1]Department of Molecular Biology, Massachusetts General Hospital, Boston, MA, USA. [2]Center for Regenerative Medicine, Massachusetts General Hospital, Boston, MA, USA. [3]Cancer Center, Massachusetts General Hospital, Boston, MA, USA. [4]Department of Genetics, Harvard Medical School, Boston, MA, USA. [5]Harvard Stem Cell Institute, Cambridge, MA, USA. [6]Broad Institute of MIT and Harvard, Cambridge, MA, USA. [7]David Geffen School of Medicine, Department of Biological Chemistry, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, Jonsson Comprehensive Cancer Center, Molecular Biology Institute, University of California Los Angeles, Los Angeles, CA, USA. [8]Department of Data Science, Dana-Farber Cancer Institute, Boston, MA, USA. [9]Department of Biostatistics, Harvard T.H. Chan School of Public Health, Boston, MA, USA. [10]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA, USA. [11]Department of Genome Regulation, Max-Planck Institute for Molecular Genetics, Berlin, Germany. [12]Department of Pathology, Harvard Medical School and Massachusetts General Hospital, Boston, MA, USA. [13]The Center for Cancer Evolution, Dana-Farber Cancer Institute, Boston, MA, USA. [14]The Ludwig Center at Harvard, Boston, MA, USA. e-mail: hochedlinger@molbio.mgh.harvard.edu

cell states[11]. Indeed, the perturbation of diverse regulatory factors enhances the generation of iPS cells by increasing cellular plasticity[12]. While histone-modifying enzymes are among these regulators[13–16], their effects on reprogramming are often subtle, and the corresponding genes are typically underrepresented in unbiased screens[14,17,18]. Histone-modifying enzymes are challenging to study in cell fate regulation as they are often essential for cellular survival or compensated by redundant enzymes[19]. Additionally, some histone-modifying enzymes recognize non-histone substrates, complicating their functional analyses[20–23]. Thus, the functional roles of specific histone modifications in the maintenance of cell identity remain incompletely understood.

Lysine-to-methionine (K-to-M) mutations of histone H3 have recently been suggested as powerful tools to probe the physiological roles of specific histone marks in vitro and in vivo[24]. K-to-M mutants function as dominant alleles that block methylation at the respective sites across the genome[25–30]. In contrast to deletions of histone-modifying enzymes, K-to-M mutations function as hypomorphs[31], thus solving toxicity concerns. Indeed, we recently demonstrated that introduction of specific H3 K-to-M mutants into embryonic stem (ES) cells and mice reveals highly specific roles of the targeted histone marks in mammalian pluripotency, tissue homeostasis and tumourigenesis[32].

In this Article, we interrogated the molecular and functional roles of two key active and two key repressive histone modifications on cell identity maintenance by combining specific histone mutants with experimental models of cell fate conversion. We uncovered a striking effect of the H3K36M mutation, which targets H3K36 di- and trimethylation, on various lineage conversion paradigms. By elucidating the underlying mechanisms in the context of induced pluripotency, we demonstrated that H3K36 methylation plays a previously unexplored, dual role in the maintenance of cell identity. Our findings place H3K36 methylation at the nexus of a key extracellular signalling pathway and epigenetic gene regulation during the maintenance of cell state, which will advance our understanding of development, regeneration and cancer.

Results

Key role of H3K36 methylation in cell identity maintenance

To probe the role of specific histone marks in the maintenance of cell identity, we introduced doxycycline (DOX)-inducible lentiviral constructs with either wild-type histone H3.3 (WT) or H3.3 variants carrying specific K-to-M mutations at lysine residues 4, 9, 27 and 36 (K4M, K9M, K27M and K36M) (Fig. 1a,b) into murine embryonic fibroblasts (MEFs) harbouring a DOX-inducible polycistronic cassette encompassing the reprogramming factors Oct4, Klf4, Sox2 and c-Myc (OKSM)[33]. Transduced cultures were treated with DOX for 12 days to induce OKSM as well as WT or mutant histone expression before counting iPS cell colonies on day 15 (Fig. 1b). We detected no iPS cell colonies in K4M-transduced cells, an approximately 3-fold increase of iPS cell colonies in K9M-transduced cells and a similar number of enlarged iPS cell colonies in K27M-transduced cells compared with WT control (Fig. 1c and Extended Data Fig. 1a), confirming and extending previous observations using small interfering RNA (siRNA) suppression of the respective histone-modifying enzymes[13,34–39]. Strikingly, K36M-transduced cells exhibited the most profound phenotype, with wells being entirely overgrown by iPS cell colonies. Although certain H3K36-specific demethylases (Kdm2a/Kdm2b) have previously been implicated in reprogramming[40,41], the reported effects were rather subtle compared with our K36M phenotype, probably owing to compensatory mechanisms. We conclude that disrupted H3K36 methylation most profoundly affects cell fate change during reprogramming.

K36M drives nearly all somatic cells towards pluripotency

We generated transgenic MEFs containing the *Col1a1::tetO-OKSM* allele together with either the *Col1a1::tetO-K36M* allele (OKSM/K36M) or the *Col1a1::tetO-H3.3* allele (OKSM/WT)[32], allowing us to express OKSM and histones in a homogeneous, inducible and reversible manner (Extended Data Fig. 1b). Indeed, K36M and Sox2 were uniformly induced, whereas H3K36me2 and H3K36me3 were uniformly depleted in DOX-treated OKSM/K36M cells (Extended Data Fig. 1c). Mass-spectrometric analysis confirmed the robust depletion of H3K36me1, H3K36me2 and H3K36me3 on canonical and variant histone H3 in OKSM/K36M cells (Extended Data Fig. 1d).

We first determined the minimally required time of OKSM expression to produce stable iPS cells (Fig. 1d). While OKSM/WT cultures gave rise to rare iPS cell colonies after 6–8 days of OKSM expression consistent with previous observations[42], OKSM/K36M cultures were full of iPS cell colonies after as little as 4 days of OKSM expression, indicating a markedly accelerated and more uniform acquisition of pluripotency. We next assessed when K36M expression is required to boost reprogramming (Extended Data Fig. 1e). Expression of K36M in MEFs for 2 days before OKSM induction had a similar effect on iPS cell formation as the co-expression of K36M and OKSM during the first 2 days of reprogramming, suggesting that K36M primes the MEF epigenome for an efficient acquisition of pluripotency (Extended Data Fig. 1f). Importantly, the frequency of iPS cell formation further increased with prolonged K36M expression, implying an independent role at later stages of reprogramming.

We next tracked surface antigens that dynamically change with reprogramming (Fig. 1e). OKSM/WT control cells downregulated the fibroblast-associated marker Thy1 and subsequently upregulated the early pluripotency-associated marker SSEA1 in ~25% of cells by day 6 of reprogramming, paralleling prior observations[42–44] (Fig. 1f,g). By contrast, OKSM/K36M cultures already expressed SSEA1 in ~10% of Thy1$^+$ cells on day 2, and this population further increased to over 75% by day 6. When we examined surface marker combinations that identify more rare reprogramming intermediates poised to produce iPS cells at high efficiency (Fig. 1e)[45], we detected these populations earlier and in a higher fraction of cells in OKSM/K36M versus OKSM/WT cultures (Fig. 1h). In agreement, the vast majority (~85%) of OKSM/K36M cultures upregulated the late pluripotency reporter Oct4–GFP[46] (Fig. 1e,i) between days 4 and 8 of reprogramming while only a minor fraction (~5%) upregulated the reporter in controls, independent of culture conditions (Fig. 1j and Extended Data Fig. 1g). OKSM/K36M cultures reached over 90% Oct4–GFP positivity by day 10 of reprogramming, suggesting that almost every cell had acquired a pluripotent state (Fig. 1j). K36M induction neither affected cell proliferation (Extended Data Fig. 1h) nor cell death (Extended Data Fig. 1i), and we determined that residual Oct4–GFP$^-$ cells at day 10 probably comprise pluripotent cells in which the reporter has been switched off or cells that still have the potential to produce iPS cells (Extended Data Fig. 1j–l).

Critically, the co-expression of OKSM and K36M led to a robust increase in iPS cell formation from mouse granulocyte/macrophage progenitors (GMPs) and keratinocytes (Extended Data Fig. 2a,b) as well as from human fibroblasts, indicating that H3K36 methylation functions as a barrier to reprogramming beyond MEFs (Fig. 1k–m).

Transgene-independent, passaged iPS cells derived from OKSM/WT and OKSM/K36M MEFs were highly similar at the transcriptomic and epigenomic level (Extended Data Fig. 2c–j). Accordingly, both OKSM/WT and OKSM/K36M iPS cells gave rise to differentiating embryoid bodies (Extended Data Fig. 2k,l) and well-differentiated teratomas (Extended Data Fig. 2m). K36M iPS cells also contributed to adult coat colour chimaeras upon blastocyst injections (Extended Data Fig. 2n).

K36M confers epithelial plasticity on cells

Analysis of gene expression data from cultures undergoing reprogramming (days 2–8) revealed that OKSM/K36M intermediates were on an accelerated trajectory compared with OKSM/WT samples, with day 8 samples being closest to established iPS cells (Fig. 2a, filled circles). Notably, in MEFs expressing K36M or WT H3.3 alone, we observed only

https://doi.org/10.1038/s41556-023-01191-z

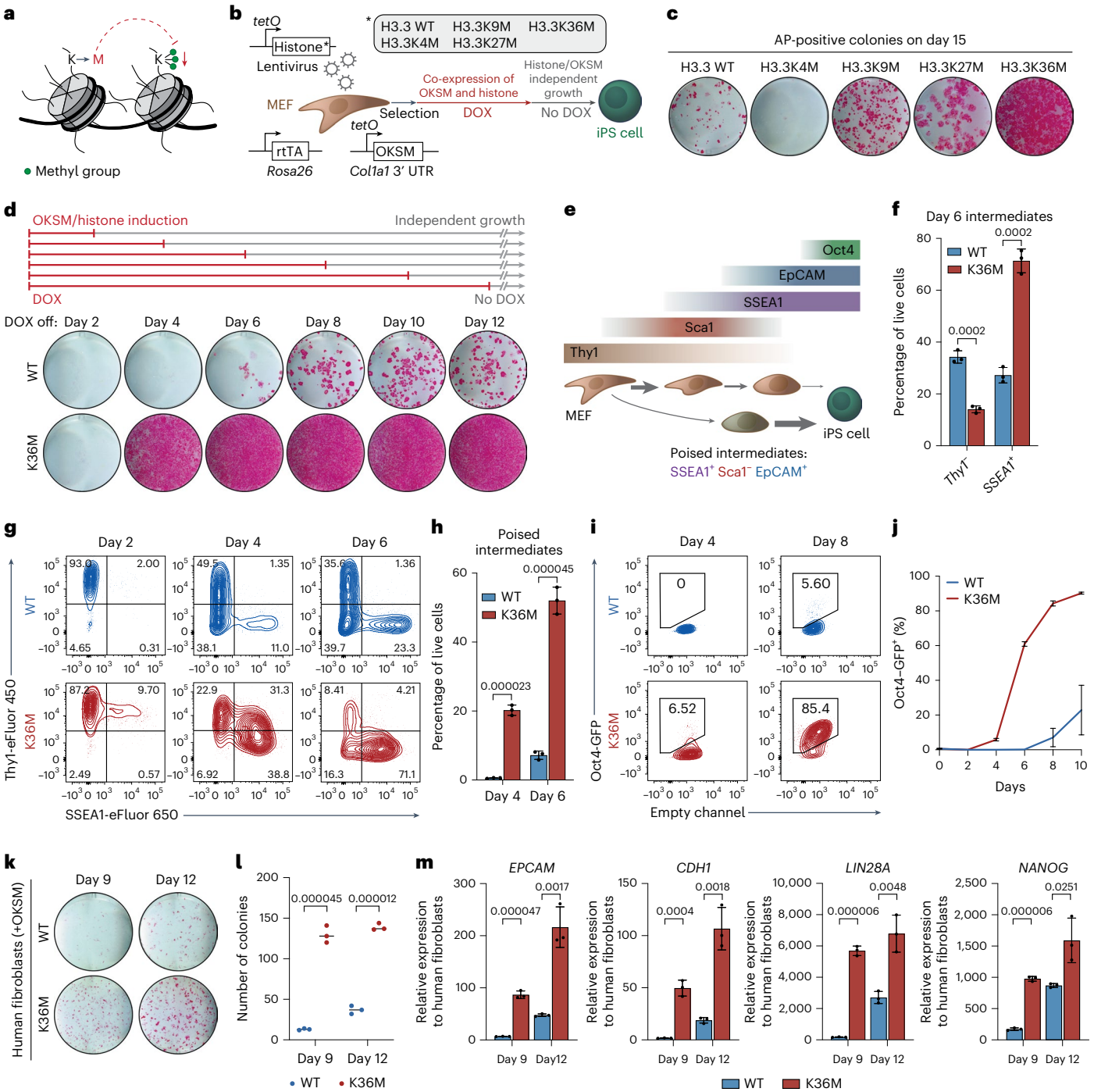


Fig. 1 | Histone mutant analysis reveals crucial role of H3K36 methylation in cell identity maintenance. a, K-to-M mutants of histone H3.3 dominantly block histone methylation at the respective residue across the genome. **b**, WT or mutant histones (K4M, K9M, K27M and K36M) were co-expressed with OKSM in fibroblasts during reprogramming. **c**, Alkaline phosphatase (AP) staining of transgene-independent iPS cell colonies. Transduced MEFs were treated with DOX for 12 days, followed by 3 days of withdrawal. **d**, OKSM transgene dependency assay. MEFs were treated with DOX, ascorbic acid and CHIR99021 as indicated, and iPS cells were scored by AP staining on day 15. **e**, Marker dynamics during MEF to iPS cell reprogramming. **f**, Percentage of Thy1$^-$ and SSEA1$^+$ intermediates on day 6 of reprogramming. *P* values were determined by two-sided unpaired Student's *t*-test, error bars indicate mean ± s.d. ($n = 3$ biologically independent experiments). **g**, Flow cytometry analysis of Thy1 and SSEA1 at early timepoints of reprogramming in WT and K36M cells. **h**, Abundance of intermediate populations that reprogram with high efficiency (SSEA1$^+$/Sca1$^-$/Epcam$^+$) on days 4 and 6 of reprogramming. *P* values were determined by two-sided unpaired Student's *t*-test, error bars indicate mean ± s.d. ($n = 3$ biologically independent experiments). **i**, Flow cytometry using the Oct4–GFP reporter as a readout for successful activation of the endogenous pluripotency network on day 4 and day 8 of reprogramming. **j**, Detection of Oct4–GFP$^+$ cells at different timepoints of reprogramming in WT and K36M samples (ascorbic acid and CHIR99021 condition). Error bars indicate mean ± s.d. ($n = 3$ biologically independent experiments). **k**, AP staining of human reprogramming cultures at the indicated timepoints. Human fibroblasts were transduced with constitutive vectors expressing H3.3 WT or K36M, and OKSM. **l**, Quantification of AP colony counts on days 9 and 12. *P* values were determined by two-sided unpaired Student's *t*-test, $n = 3$ biologically independent experiments. **m**, Quantitative reverse transcription polymerase chain reaction (qRT-PCR) for epithelial and pluripotency-associated genes in human reprogramming cultures on days 9 and 12 of reprogramming. *P* values were determined by two-sided unpaired Student's *t*-test, error bars indicate mean ± s.d. ($n = 3$ biologically independent experiments).

https://doi.org/10.1038/s41556-023-01191-z

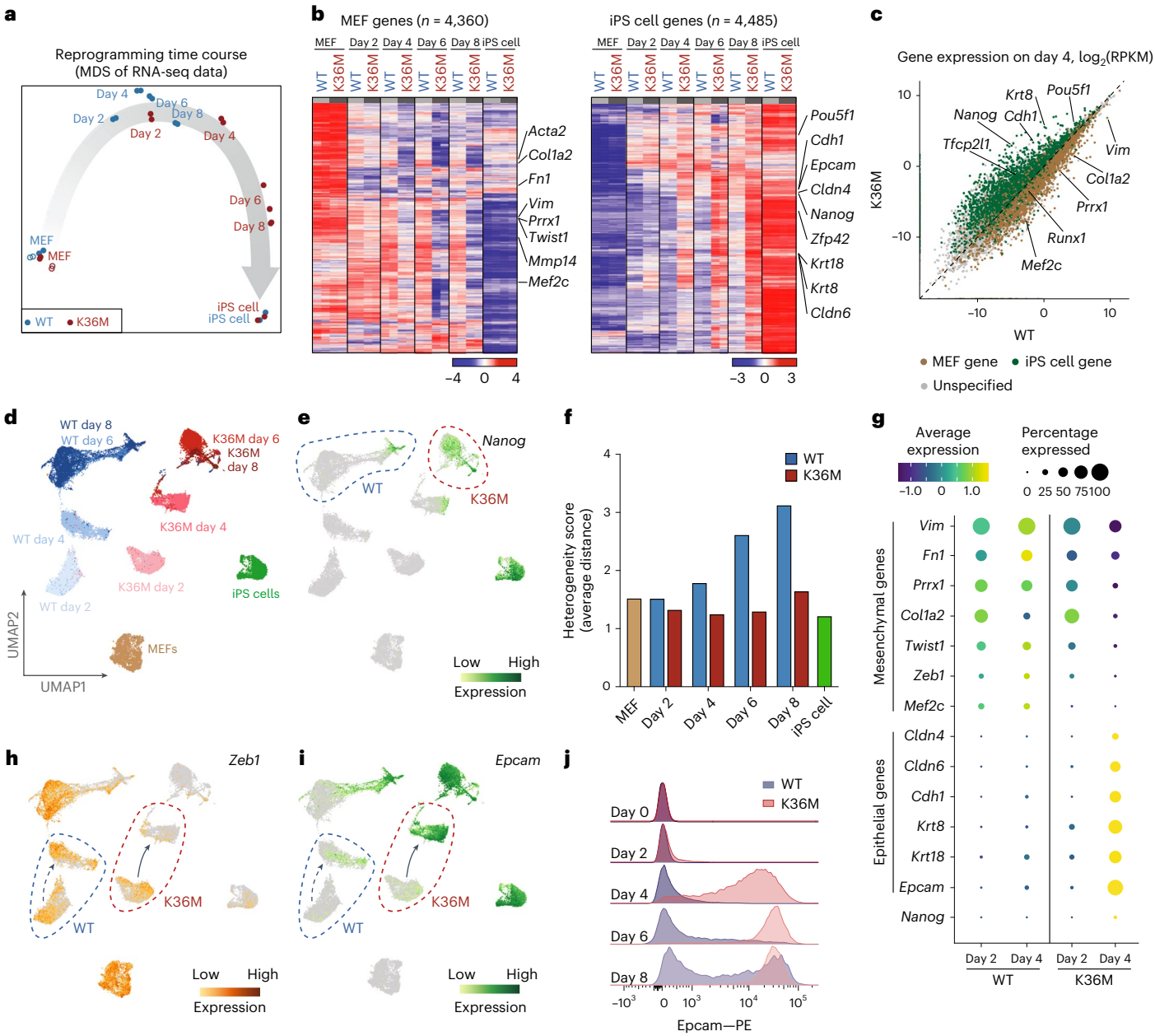


Fig. 2 | K36M endows iPS cell intermediates with transcriptional homogeneity and epithelial plasticity. a, Multidimensional scaling plot of RNA-seq data based on genes with most variable expression among all timepoints and conditions. Filled circles, MEFs expressing both histone genes and OKSM for indicated number of days; open circles, MEFs expressing histone genes alone for 4 days. **b**, Heat maps showing gene expression dynamics in WT and K36M reprogramming intermediates for genes differentially expressed between MEFs and iPS cells ($n = 2$ biologically independent replicates). **c**, Scatter plot showing gene expression differences (RPKM (reads per kilobase per million reads mapped) at $\log_2$ scale) between WT and K36M reprogramming intermediates on day 4. Genes that are upregulated in iPS cells versus MEFs are highlighted in green, genes that are downregulated in iPS cells versus MEFs are highlighted in brown ($n = 2$ biologically independent replicates were integrated for this analysis). **d**, Uniform manifold approximation and projection (UMAP) embedding of single-cell RNA-seq data using MEFs, reprogramming intermediates on days 2, 4, 6 and 8 for WT and K36M, as well as passaged iPS cells. For each indicated group, one sample was encapsulated leading to $n = 38,743$ cells total. **e**, Expression of *Nanog* projected on the same UMAP embedding as shown in **d**. Dashed circles highlight scattered expression of *Nanog* in WT cells (blue) on day 8 compared with homogeneous expression in K36M cells on days 6 and 8 (red). **f**, Quantification of transcriptional heterogeneity of indicated samples as measured by average distance between cells within each sample in the same UMAP embedding as shown in **d**. Median values for all cells within each sample are plotted. **g**, Selected DEGs between WT and K36M samples that distinguish day 2 and day 4 intermediates. **h**, Expression of the mesenchymal regulator *Zeb1* projected on the same UMAP embedding as shown in **d**. Dashed circles highlight d2 and d4 samples for WT (blue) and K36M (red). **i**, Expression of epithelial gene *Epcam* projected on the same UMAP embedding as used in **d**. Dashed circles highlight day 2 and day 4 samples for WT (blue) and K36M (red). **j**, Flow cytometric quantification of Epcam expression in WT versus K36M reprogramming intermediates.

minor effects on gene expression patterns, pointing to a synergistic effect between K36M and OKSM (Fig. 2a, open circles). While day 2 reprogramming intermediates were highly similar between the OKSM/WT and OKSM/K36M conditions, starting at day 4 we observed a more robust silencing of MEF-associated genes and a more robust induction of pluripotency-associated genes in OKSM/K36M cells compared

with OKSM/WT cells (Fig. 2b). In agreement, transcriptional differences between OKSM/WT and OKSM/K36M cells at day 4 were largely attributable to fibroblast genes upregulated in OKSM/WT cells, and pluripotency genes upregulated in OKSM/K36M cells (Fig. 2c).

MEFs expressing OKSM exhibit a substantial degree of heterogeneity with most cells undergoing cell cycle arrest or acquiring alternative fates[47]. To assess how K36M expression impacts these phenotypes, we performed single-cell RNA sequencing (RNA-seq) of OKSM/WT and OKSM/K36M intermediates (Fig. 2d and Extended Data Fig. 3a). In agreement with our analysis of Oct4–GFP cells, we detected expression of the late pluripotency marker *Nanog* in only a rare subset of day 8 cells in OKSM/WT cultures whereas *Nanog* was already expressed in every cell at day 6 in OKSM/K36M cultures (Fig. 2e, dotted circles, and Extended Data Fig. 3b). When we assessed the transcriptional similarity between single cells as a proxy for cellular heterogeneity, we found that OKSM/K36M intermediates exhibited as little heterogeneity as MEFs and established iPS cells, while OKSM/WT reprogramming intermediates exhibited an early, transient increase in heterogeneity, mirroring previous observations[47] (Fig. 2f). Hence, K36M expression effectively neutralizes early reprogramming barriers typical of WT cells and prevents the formation of heterogeneous cell states refractory to iPS cell formation.

Focusing on the earliest stages of reprogramming between day 2 and day 4, we noticed a marked downregulation of mesenchymal markers such as *Col1a2*, *Zeb1* and *Prrx1* and a homogeneous upregulation of epithelial markers such as *Epcam*, *Cdh1* and *Krt8* in OKSM/K36M intermediates (Fig. 2g–i and Extended Data Figs. 3c and 4a,c–e). However, OKSM/WT cells largely maintained a mesenchymal programme and only a small subset of cells expressed epithelial genes at these timepoints. We corroborated these observations with a diffusion model (Extended Data Fig. 3d,e), which shows that K36M cells uniformly progress along a successful reprogramming trajectory defined by *Nanog* and *Cdh1* expression (Extended Data Fig. 3f,g), while the majority of WT cells follow an unproductive path defined by continued expression of the MEF gene *Zeb1* (Extended Data Fig. 3h). Consistently, we observed a homogeneous upregulation of Epcam protein expression specifically in OKSM/K36M cells between days 2 and 4 using fluorescence-activated cell sorting and immunofluorescence analyses (Fig. 2j and Extended Data Fig. 4b). We note that some mesenchymal/epithelial genes (Extended Data Fig. 3i) as well as genetic programmes linked to development and differentiation (Extended Data Fig. 3j,k) were also modulated in K36M MEFs without OKSM expression and probably contributed to reprogramming via the aforementioned priming effect of K36M (Extended Data Fig. 1e,f). Together, these observations suggest that K36M facilitates an efficient switch from a mesenchymal to an epithelial identity, a process previously shown to be important for the acquisition of pluripotency[48,49].

K36M acts downstream of TGFβ and Smad2 but upstream of Zeb1

We next explored a possible molecular relationship between H3K36 methylation and TGFβ signalling as TGFβ ligands and downstream effectors are well-known regulators of mesenchymal identity and antagonists of epithelial identity[50] (Fig. 3a). In OKSM/WT cells, inhibition of TGFβ signalling enhanced the fraction of Epcam$^+$ and Oct4–GFP$^+$ cells, while exposure to recombinant TGFβ abolished both Epcam and Oct4–GFP expression, confirming our previous finding[51] (Fig. 3b,c and Extended Data Fig. 5a,b). However, OKSM/K36M cells were insensitive to either TGFβ inhibition or recombinant TGFβ treatment, suggesting that K36M functions downstream of TGFβ signals (Fig. 3b,c and Extended Data Fig. 5a,b). Consistently, OKSM/K36M cells maintained low levels of mesenchymal genes and high levels of epithelial genes regardless of the presence of recombinant TGFβ or TGFβ inhibitor (Fig. 3d and Extended Data Fig. 5c,d), while TGFβ and TGFβ inhibitor elicited the expected mesenchymal and epithelial biased gene expression changes, respectively, in OKSM/WT cells. Furthermore, the modulation of TGFβ signalling led to a comparable change in phospho-Smad2 levels in OKSM/WT and OKSM/K36M cells, indicating that K36M exerts its effect downstream of this key TGFβ mediator (Fig. 3e).

To define TGFβ/Smad effector genes that are modulated by K36M, we perturbed candidate mesenchymal transcription factors with siRNAs. Suppression of *Snai1*, *Snai2* and *Twist1* levels had minimal effects on Epcam expression, whereas suppression of *Zeb1* levels led to a >5-fold increase of Epcam$^+$ cells in OKSM/WT cells and this effect was neutralized in OKSM/K36M cells (Fig. 3f). We observed a similar, albeit smaller effect when measuring Oct4–GFP expression, suggesting that K36M's effect is in part attributable to *Zeb1* downregulation (Fig. 3g). Strikingly, OKSM/K36M cells overexpressing *Zeb1* failed to upregulate Epcam compared with control (Fig. 3h). These results suggest that *Zeb1* is a downstream effector of H3K36 methylation critical for the maintenance of a mesenchymal state in our system. Since the effects of *Zeb1* suppression on acquiring an epithelial/pluripotent state did not fully recapitulate the effects of K36M expression (Fig. 3g), we tested whether microRNAs (miRNAs) previously implicated in both mesenchymal-to-epithelial transition and pluripotency induction may be involved. Indeed, we observed a robust upregulation of miRNAs that target mesenchymal regulators and favour an epithelial state, including *miR-205* and the *miR-200* family[52,53], as well as of miRNAs that facilitate the acquisition of a pluripotent state, including the *miR-290-295* and *miR-302* families (Fig. 3i and Extended Data Fig. 5e,f)[54,55].

K36M modulates epithelial plasticity in diverse contexts

To test whether disrupted H3K36 methylation impacts mesenchymal/epithelial identity and TGFβ signalling more broadly, we measured the effects of K36M expression during MEF-to-induced neuron conversion using *Brn2*/*Ascl1*/*Myt1l* expression[56], and during MEF-to-myocyte conversion using *MyoD* expression[57]. Consistent with the pro-epithelial effect of K36M during induced pluripotency, its forced expression led to a significant increase in the number of induced neurons from MEFs (Fig. 3j). However, K36M expression substantially reduced the formation of Myh1-positive myotubes from MEFs (Fig. 3k), indicating that a mature myogenic fate cannot be efficiently established from mesenchymal cells in the absence of H3K36 methylation. Notably, K36M expression did not interfere with the reprogramming of MEFs to muscle progenitors using a recently published protocol[58,59], suggesting that H3K36 methylation is dispensable for the acquisition of a self-renewing myogenic state but essential for muscle cell differentiation (Extended Data Fig. 5g).

To investigate whether K36M's effect on epithelial identity is conserved in cell fate transitions that do not involve ectopic transcription factor expression, we induced K36M in ES cells (Extended Data Fig. 5h,i) that were coaxed into gastruloids. This process mimics early stages of gastrulation and comprises an epithelial-to-mesenchymal transition (EMT)[60–62]. Control ES cell cultures readily transitioned from a spherical to a tube-like structure typical of gastruloids, whereas K36M cultures remained spherical and failed to elongate, consistent with a requirement for H3K36-methylation to exit the pluripotent/epithelial state and acquire a differentiated/mesenchymal state (Extended Data Fig. 5j). Supporting these data, K36M also blocked EMT in a monolayer differentiation model from ES cells to pre-somitic mesoderm[63], as inferred from a failure to induce N-Cadherin, *Tbx6* and *Msgn1* expression (Extended Data Fig. 5k). Thus, H3K36 methylation is critical in physiologically relevant models of early development involving an EMT switch.

We finally assessed whether K36M expression modulates TGFβ signalling in a physiological context that does not entail an EMT. Epidermal stem cells represent a classic type of epithelial stem cells that require TGFβ cues for proper differentiation in vivo[64] and in vitro[65]. Thus, we generated epidermal stem cell cultures expressing either H3.3 WT or K36M. Acute withdrawal of TGFβ inhibitor and parallel

addition of recombinant TGFβ led to the immediate flattening and differentiation of WT cultures as determined by the loss of cells expressing the epidermal stem cell marker p63 (Fig. 3l). By contrast, cultures expressing K36M retained p63 expression in a substantial fraction of cells, consistent with the notion that K36M partially phenocopies the effect of TGFβ inhibition.

H3K36me2 and H3K36me3 cooperatively safeguard cell identity

To determine whether K36M's phenotype is driven by the disruption of H3K36me3, H3K36me2 or both marks, we modulated cognate histone methyltransferases/demethylases during reprogramming (Extended Data Fig. 6a). The suppression of individual methyltransferases (*Nsd1*/*Nsd2* for H3K36me2, *Setd2* for H3K36me3) had only subtle effects on the formation of Epcam$^+$ intermediates and iPS cell colonies (Extended Data Fig. 6b). By contrast, the combined suppression of *Nsd1*/*Nsd2* and that of *Nsd1*/*Nsd2*/*Setd2* had the strongest beneficial effect on the formation of Epcam$^+$ cells in WT fibroblasts. Conversely, the overexpression of *Nsd2* in WT cells substantially reduced reprogramming efficiency and this effect was neutralized by K36M expression (Extended Data Fig. 6c,d). Corroborating these results, siRNA-mediated inhibition of the H3K36me2 demethylases *Kdm2a* and *Kdm2b* significantly reduced reprogramming efficiency (Extended Data Fig. 6e–h), whereas knockdown of the H3K36me3 demethylases *Kdm4a* and *Kdm4c* had no effect. Consistently, the overexpression of *Kdm2a*, *Kdm2b* and *Kdm4a* strongly increased reprogramming efficiency while the overexpression of *Kdm4c* had only a subtle beneficial effect (Extended Data Fig. 6i–k). Of relevance, *Kdm4a* was shown to demethylate not only H3K36me3 but also H3K36me2 (ref. 66), suggesting that K36M's effect on reprogramming is largely driven by the depletion of H3K36me2 levels while the parallel depletion of H3K36me3 levels has a supportive role.

K36M alters H3K36me2/3 deposition and gene expression

To understand how K36M facilitates the extinction of a mesenchymal identity and the acquisition of an epithelial/pluripotent identity at a chromatin level, we assessed the epigenome of day 4 reprogramming intermediates. As expected, H3K36me3 was enriched over gene bodies and progressively increased with gene expression levels in OKSM/WT cells, whereas this mark was globally depleted in OKSM/K36M cells with the exception of the most highly expressed genes that retained residual H3K36me3 levels (Fig. 4a,b and Extended Data Fig. 7a–d). Together with our functional data on H3K36me3-specific histone-modifying enzymes, these results suggest that H3K36me3 depletion contributes to the extinction of the somatic programme, while residual H3K36me3 levels at highly expressed genes may facilitate the induction of epithelial and pluripotency programmes.

H3K36me2 signal was distributed over broad intergenic and genic regions including enhancers and promoter-proximal introns in OKSM/WT control cells, confirming and extending previous observations[67,68] (Fig. 4c). OKSM/K36M cells showed genome-wide depletion of this mark across these broad H3K36me2 domains and other genomic features (Fig. 4c and Extended Data Fig. 7e). Notably, H3K36me2-depleted domains in OKSM/K36M cells were associated with a similar number of upregulated (n = 910) and downregulated (n = 1,053) genes associated with epithelial/pluripotent and mesenchymal identity, respectively (Fig. 4d). Supporting the instructive role of H3K36me2 in driving these gene expression changes, the forced expression of *Kdm2a* in OKSM/WT MEFs was sufficient to drive the downregulation of fibroblast genes and the upregulation of epithelial/pluripotency genes (Extended Data Fig. 6l).

K36M decommissions MEF enhancers and activates ES cell enhancers

To understand how K36M-dependent H3K36me2 depletion leads to both gene activation and gene silencing, we mapped the deposition of the repressive H3K27me3 mark known to antagonize H3K36 methylation[68–70]. Only a fraction of H3K36me2-depleted domains gained H3K27me3, while the remaining domains showed little or no H3K27me3 accumulation (Fig. 4e and Extended Data Fig. 8a). Genes located within H3K36me2-depleted domains that gained H3K27me3 were enriched for ontology terms related to signal integration (Extended Data Fig. 8b), in line with our earlier findings linking H3K36 methylation to TGFβ signalling (Fig. 3). Moreover, TGFβ-responsive genes that were downregulated in K36M cells (Fig. 3d) gained H3K27me3 at their promoters consistent with a role of PRC2 in their silencing (Extended Data Fig. 8c). Indeed, when we suppressed the PRC2 components *Ezh2* and *Suz12* using siRNAs in OKSM/K36M cells, we observed a reduction in reprogramming efficiency and a delayed silencing of mesenchymal genes (Extended Data Fig. 8d–h).

Recent studies suggested that H3K36 methylation can have differential effects on transcription depending on the chromatin context[71–74]. We therefore investigated whether H3K36me2 depletion leads to locus-specific effects on regulatory elements during reprogramming (Fig. 4f and Extended Data Fig. 7f). Indeed, a subset of enhancers within H3K36me2-depleted domains exhibited reduced H3K27ac levels and chromatin accessibility in OKSM/K36M cells; these enhancers were enriched for MEF enhancers (Fig. 4f–h). By contrast, enhancers within H3K36me2-depleted domains that gained H3K27ac signal and chromatin accessibility were enriched for pluripotency enhancers (Fig. 4f–h). Gene Ontology analysis of these two enhancer groups showed an overrepresentation of mesenchymal and epithelial terms

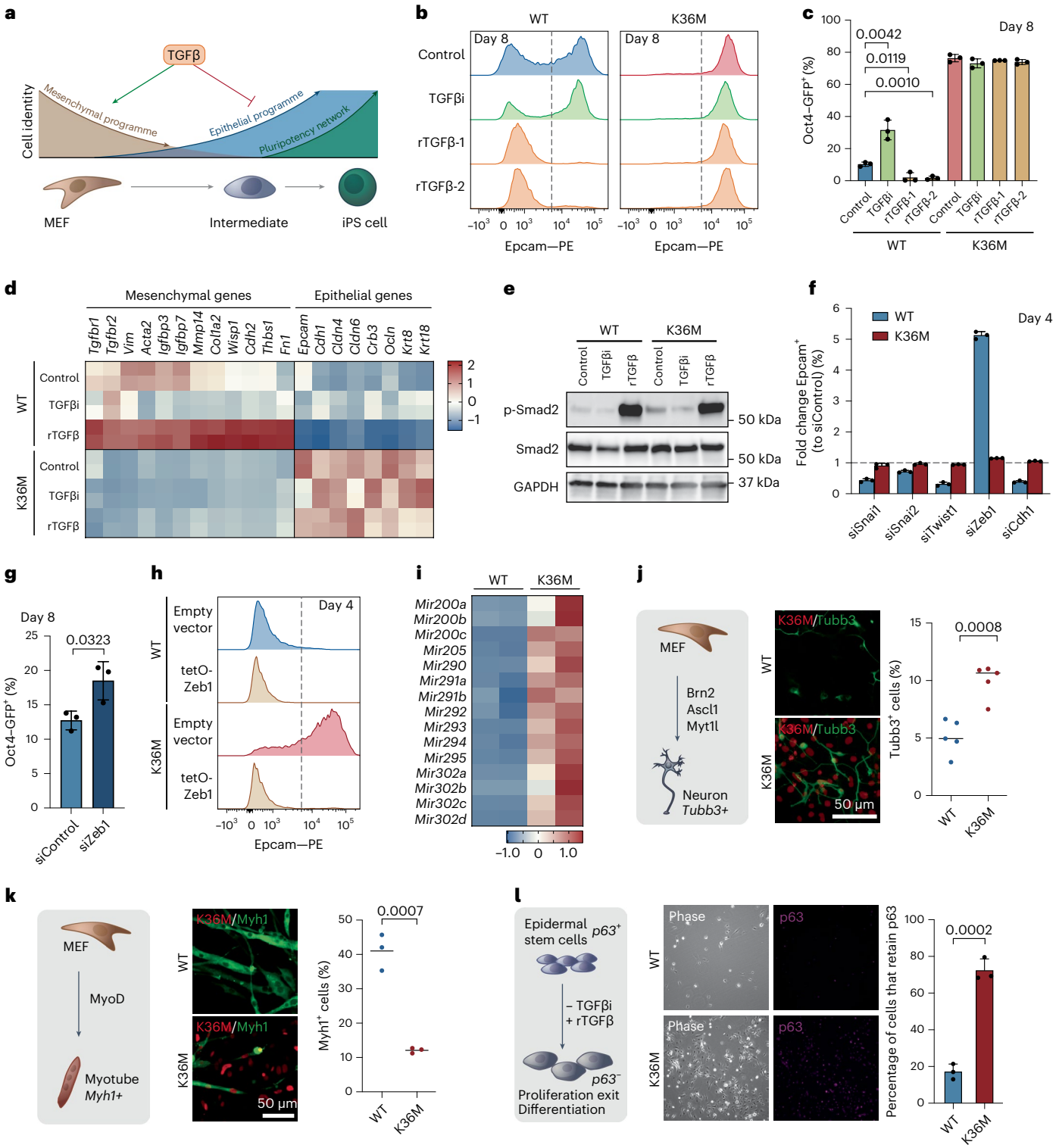
Fig. 3 | K36M acts downstream of TGFβ and Smad2 but upstream of Zeb1. **a**, Effects of TGFβ signalling on reprogramming. **b**, Flow cytometry for Epcam in day 8 reprogramming intermediates for WT and K36M samples. Untreated controls versus cells treated with 250 nM Repsox (TGFβi) or 2.5 ng ml^{-1} recombinant TGFβ-1 or -2 (rTGFβ-1 and rTGFβ-2). **c**, Fraction of Oct4–GFP$^+$ cells in reprogramming cultures treated with TGFβi or rTGFβ-1/ rTGFβ-2 on day 8. *P* values were determined by two-sided unpaired Student's *t*-test, error bars indicate mean ± s.d. (n = 3 independent biological experiments). **d**, Heat map showing gene expression (*z*-score) of mesenchymal and epithelial genes in WT and K36M reprogramming intermediates treated with either TGFβi or rTGFβ-1 (n = 2 biologically independent replicates). **e**, Immunoblot analysis for phospho-Smad2, Smad2 and GAPDH in WT and K36M reprogramming intermediates treated with either TGFβi or rTGFβ-1. Blot is representative of three independent biological experiments. **f**, Epcam expression (fold change) of WT and K36M reprogramming cultures (day 4) transfected with the indicated siRNAs relative to control. Error bars indicate mean ± s.d. (n = 3 independent biological experiments). **g**, Fraction of Oct4–GFP$^+$ cells in day 8 WT samples treated with control siRNA or siRNA targeting *Zeb1*. *P* values were determined by two-sided unpaired Student's *t*-test, error bars indicate mean ± s.d. (n = 3 independent biological experiments). **h**, Flow cytometry for Epcam on day 4 in WT and K36M reprogramming intermediates transduced with empty vector or DOX-inducible *Zeb1*. **i**, Small RNA-seq of day 4 reprogramming intermediates. Heat map of key pro-epithelial and pro-pluripotency miRNAs (n = 2 biologically independent replicates). **j**, MEF-to-neuron conversion. Representative immunofluorescence images (scale bar, 50 µm) and quantification of Tubb3$^+$ neurons (line indicates mean). *P* value determined by two-sided unpaired Student's *t*-test, n = 3 biologically independent experiments. **k**, MEF-to-myotube conversion. Representative immunofluorescence images (scale bar, 50 µm) and quantification of Myh1$^+$ myotubes (line indicates mean). *P* values determined by two-sided unpaired Student's *t*-test, n = 3 biologically independent experiments. **l**, Epidermal stem cells exposed to recombinant TGFβ-1 (10 ng ml^{-1}). Representative images for phase and p63 immunofluorescence. Fraction of p63$^+$ cells per field. *P* value determined by unpaired Student's *t*-test, error bars indicate mean ± s.d. (n = 3 independent biological experiments).

https://doi.org/10.1038/s41556-023-01191-z

(Extended Data Fig. 7h,i), and their genomic locations were enriched for binding sites of regulatory factors linked to mesenchymal (for example, Fosl1 and Cebpb) and epithelial/pluripotent (for example, Oct4 and Nanog) identity, respectively (Extended Data Fig. 7j). Accordingly, promoters proximal to enhancers that lost H3K27ac concomitantly lost H3K4me3 levels and associated genes were downregulated (for example, *Vim* and *Prrx1*), while promoters proximal to enhancers that gained H3K27ac concomitantly increased H3K4me3 levels and associated genes were upregulated (for example, *Cdh1* and *Pou5f1*) (Fig. 4i–k and Extended Data Fig. 7g). Collectively, our results demonstrate that H3K36me2 depletion differentially impacts enhancer activity in OKSM/K36M cells, leading to the parallel induction and repression of opposing lineage programmes, respectively.

K36M rewires transcription factor binding preferences

K36M expression could replace *Oct4* or *Klf4* expression during reprogramming, while *Sox2* expression was essential, suggesting that Sox2 is key for mediating the effects of K36M (Extended Data Fig. 9a). To



https://doi.org/10.1038/s41556-023-01191-z

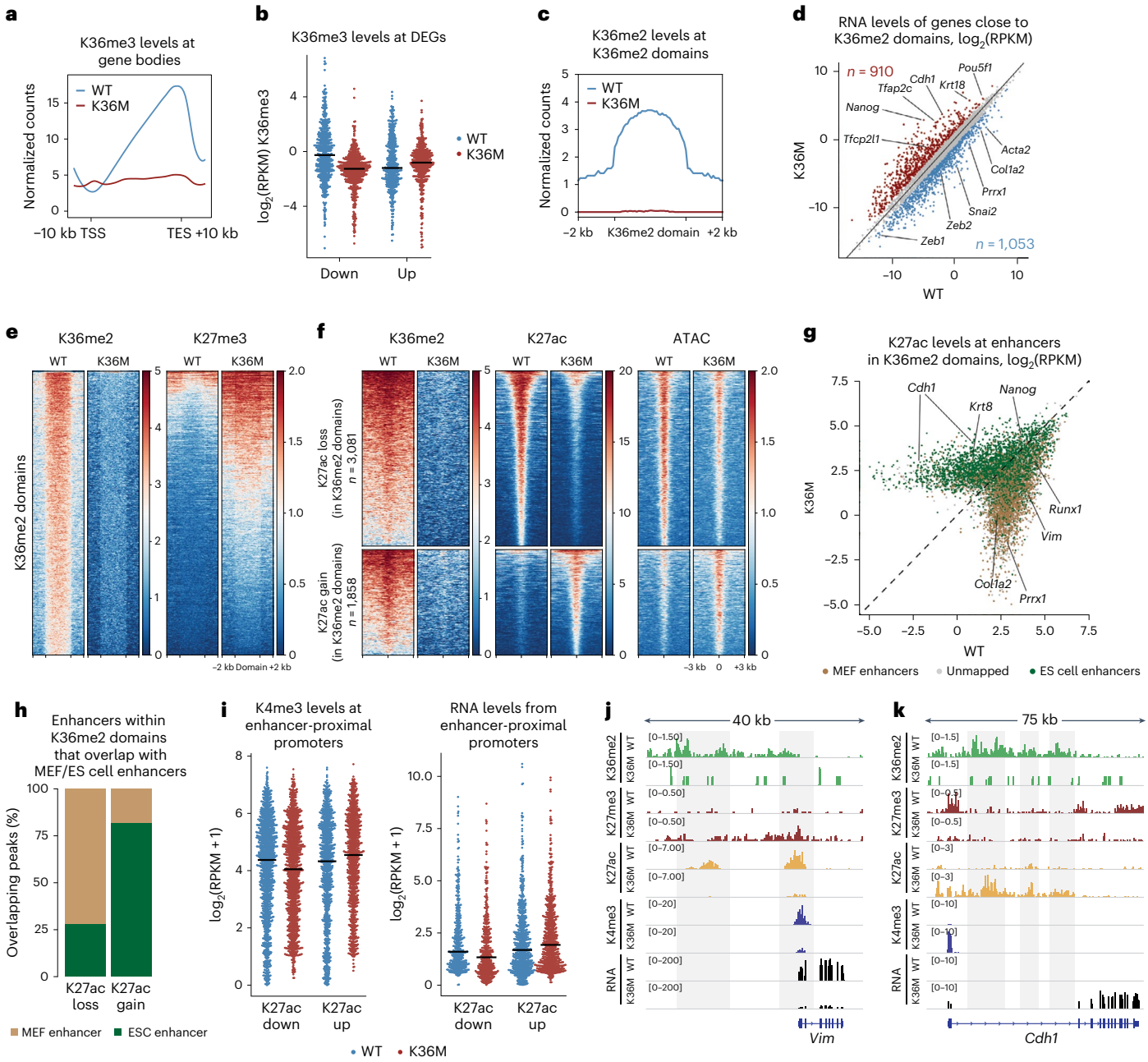


Fig. 4 | K36M decommissions MEF enhancers and activates ES cell enhancers. **a**, Profile plot of mean H3K36me3 density over gene bodies ± 10 kb. TSS, transcription start site, TES, transcription end site ($n = 2$ biologically independent replicates were integrated for this analysis). **b**, H3K36me3 deposition at DEGs ($n = 2{,}068$) for WT (blue) and K36M (red) samples on day 4 of reprogramming. RPKM, reads per kilobase per million mapped reads. Centre line indicates median ($n = 2$ biologically independent replicates were integrated for this analysis). **c**, Profile plot of median H3K36me2 density over H3K36me2 domains ($n = 2$ biologically independent replicates were integrated for this analysis). **d**, Scatter plot showing expression of genes proximal to or overlapping with H3K36me2 domains in WT and K36M samples. Red indicates upregulated and blue indicates downregulated genes in OKSM/K36M versus OKSM/WT cells on day 4 ($n = 2$ biologically independent replicates were integrated for this analysis). **e**, Heat maps showing signal intensities for H3K36me2 and H3K27me3 at H3K36me2 domains including a 2 kb flanking region ($n = 2$ biologically independent replicates were integrated for this analysis). **f**, Heat maps showing signal intensities for H3K36me2, H3K27ac and chromatin accessibility (ATAC-seq) around differentially regulated (>2-fold) enhancers within H3K36me2 domains. Enhancers were called by measuring H3K27ac abundance in regions proximal to H3K36me2 domains excluding promoters ($n = 2$ biologically independent replicates were integrated for this analysis). **g**, Scatter plot showing H3K27ac abundance at enhancers embedded within H3K36me2 domains in WT and K36M reprogramming intermediates ($n = 2$ biologically independent replicates were integrated for this analysis). Colours show overlap with MEF (beige) or ES cell enhancers (green)[75]. **h**, Fraction of differentially regulated (>2-fold) enhancers embedded within H3K36me2 domains and overlapping with MEF (beige) or ES cell enhancers (green) in WT and K36M reprogramming intermediates[75]. **i**, H3K4me3 enrichment at enhancer-proximal promoters ($n = 3{,}687$) in WT and K36M cells on day 4, and of corresponding RNA expression. Centre line indicates median ($n = 2$ biologically independent replicates were integrated for this analysis). **j**,**k**, Representative tracks of mesenchymal gene *Vim* (**j**) and of epithelial gene *Cdh1* (**k**) in WT and K36M reprogramming intermediates on day 4 ($n = 2$ biologically independent replicates). Putative regulatory elements are highlighted in grey. The data in square brackets represent the range of normalized read densities.

https://doi.org/10.1038/s41556-023-01191-z

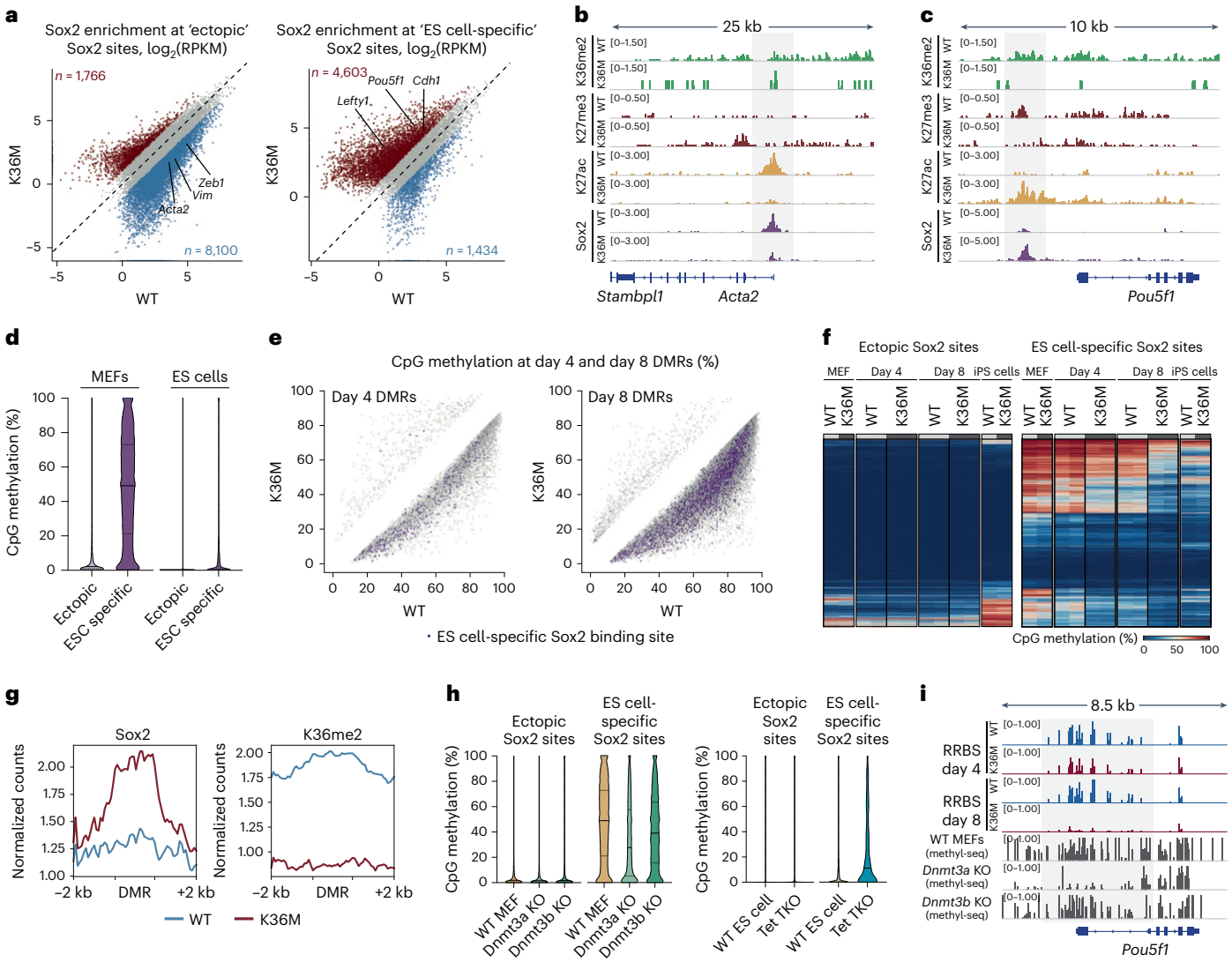


Fig. 5 | K36M-dependent chromatin rewiring alters transcription factor binding preferences. a, Scatter plots showing Sox2 occupancy at sites specific to early reprogramming intermediates ('ectopic', left) and ES cells ('ES cell-specific', right) between WT and K36M reprogramming intermediates ($n = 2$ biologically independent replicates were integrated for this analysis). **b**, Representative tracks showing ectopic Sox2 binding to the *Acta2* promoter ($n = 2$ biologically independent replicates). The data in square brackets represent the range of normalized read densities. **c**, Representative tracks showing ES cell-specific Sox2 binding at the proximal enhancer of *Pou5f1* ($n = 2$ biologically independent replicates). The data in square brackets represent the range of normalized read densities. **d**, DNA methylation at ectopic and ES cell-specific Sox2 binding sites using published methyl-seq data of MEFs ($n = 2$) and RRBS data of ES cells ($n = 2$) (refs. 80,81). **e**, Scatter plots of differentially methylated regions (DMRs) between WT and K36M reprogramming intermediates on day 4 and day 8 ($n = 2$ biologically independent replicates were integrated for this analysis). DMRs overlapping ES cell-specific Sox2 binding sites are highlighted in purple. **f**, Heat maps showing percent DNA methylation at ectopic and ES cell-specific Sox2 binding sites in MEFs, reprogramming intermediates on day 4 and day 8, and passaged iPS cells ($n = 2$ biologically independent replicates for day 4 and day 8 samples, one sample for each genotype in uninduced MEFs and iPS cells). **g**, Profile plots showing enrichment of Sox2 (left) and H3K36me2 (right) over differentially demethylated regions in WT (blue) and K36M (red) reprogramming intermediates on day 4 ($n = 2$ biologically independent replicates were integrated for this analysis). **h**, DNA methylation at ectopic and ES cell-specific Sox2 binding sites in WT, *Dnmt3a* knockout (KO) and *Dnmt3b* KO MEFs[80], as well as in WT and *Tet* triple KO (TKO) ES cells[81] ($n = 2$). **i**, Representative tracks showing DNA methylation at the *Pou5f1* locus in reprogramming intermediates (RRBS) and WT/*Dnmt3a* KO/*Dnmt3b* KO MEFs (methyl-seq)[80] ($n = 2$ biologically independent replicates).

understand how K36M expression collaborates with Sox2 to activate enhancers, we determined Sox2's chromatin binding patterns on day 4 of reprogramming. Genomic Sox2 occupancy was strongly correlated with H3K27ac deposition in OKSM/K36M cells, consistent with the established role of Sox2 as an enhancer-bound factor (Extended Data Fig. 9b). Of relevance, Chronis et al. previously demonstrated that OKS predominantly bind to ectopic targets associated with an active chromatin signature (MEF enhancers/promoters) early in reprogramming before robustly engaging with pluripotency-specific targets (ES cell enhancers) late in reprogramming[75]. Strikingly, we observed a marked, K36M-dependent reduction in Sox2 binding to ectopic targets such as *Acta2*, *Vim* and *Zeb1* and a parallel increase in Sox2 binding to ES cell-specific targets such as *Pou5f1*, *Cdh1* and *Lefty1* (Fig. 5a–c), with H3K36me2 levels being similarly depleted across both types of target (Extended Data Fig. 9c). We confirmed this differential enrichment using an independent set of iPS cell-associated Sox2 sites (Extended Data Fig. 9d). Mirroring the increase in Sox2 signal, H3K27ac and chromatin accessibility were also increased at ES cell-specific

sites in OKSM/K36M cells (Extended Data Fig. 9b,e). These results suggest that K36M-dependent chromatin rewiring reduces the early, inappropriate binding of pluripotency factors to ectopic/somatic targets while it promotes their binding to epithelial/pluripotency targets.

We hypothesized that differential DNA methylation patterns between OKSM/K36M and OKSM/WT cells could underlie the observed Sox2 binding preferences because DNA methylation typically restrains transcription factor binding[76,77] and H3K36me2/3 have been shown to target DNA methylation via Dnmt3a/b (refs. 78,79). Indeed, our analysis of published DNA methylation data and chromatin accessibility patterns suggested that ectopic Sox2 sites are normally hypomethylated in MEFs and ES cells, whereas ES cell-specific Sox2 targets are hypermethylated and inaccessible in MEFs but demethylated and accessible in ES cells (Fig. 5d and Extended Data Fig. 9f). When we applied reduced representative bisulfite sequencing (RRBS) to our OKSM/K36M and OKSM/WT cells, we observed a striking, K36M-specific loss of DNA methylation during reprogramming (Fig. 5e) including at sites that are normally methylated in ES cells/iPS cells (Extended Data Fig. 9g). Consistent with our previous meta-analysis, we found that ectopic Sox2 targets were largely hypomethylated in our MEFs and iPS cells, while over two-thirds of ES cell-specific Sox2 targets were methylated in our MEFs but demethylated in our iPS cells (Fig. 5f). Notably, K36M facilitated a profound and progressive demethylation of ES cell-specific Sox2 targets, reaching levels akin to ES cells at day 8 of reprogramming. In line with this finding, we observed a strong enrichment of Sox2 binding, but a depletion of H3K36me2 deposition, over hypomethylated regions in OKSM/K36M cells at day 4 (Fig. 5e,g) including at the *Cdh1*, *Krt8*, *miR-290*-295 and *Pou5f1* loci (Extended Data Fig. 9h–k). These results suggest that the timely and effective Sox2 binding to ES cell-associated targets is typically restrained by DNA methylation but enabled by K36M expression.

We failed to detect differences in the methylation of ectopic Sox2 sites between WT, *Dnmt3a*$^{-/-}$ and *Dnmt3b*$^{-/-}$ MEFs[80], suggesting that these sites are not regulated by the de novo methyltransferases (Fig. 5h). However, ES cell-specific Sox2 sites showed reduced DNA methylation in *Dnmt3a*$^{-/-}$ MEFs (Fig. 5h,i) and elevated methylation in *Tet1/2/3*-deficient ES cells[81] compared with ectopic Sox2 sites, consistent with the notion that de novo methyltransferases and Tet enzymes compete for targets in ES cells[82]. Methylation differences between enhancers that were more or less active in OKSM/K36M cells relative to OKSM/WT cells may be due in part to differences in CpG density (Extended Data Fig. 9l). Collectively, these findings suggest that H3K36me2 depletion at enhancers leads to an attenuated de novo methylation via Dnmt3a, which in turn facilitates Tet-dependent demethylation and activation of target genes.

K36M-dependent enhancer activation requires Tet activity

To investigate whether Tets play a functional role in the activation of epithelial/pluripotency genes[83–85] in our system, we measured the levels of 5-hydroxymethylcytosine (5hmC), which is the Tet-catalysed product of 5mC. We observed a strong 5hmC signal in OKSM/K36M cells that was absent in the starting MEFs and OKSM/WT intermediates (Fig. 6a). To determine whether this genomic demethylation is required for enhanced iPS cell generation in OKSM/K36M cells, we treated cultures with dimethyloxalylglycine (DMOG) previously shown to inhibit Tet enzymes[58,86,87]. DMOG-treated OKSM/K36M cultures exhibited a reduction of 5hmC levels commensurate with OKSM/WT cultures and untreated MEFs, and it abrogated the demethylation of ES cell-specific Sox2 binding sites and enhancers more active in K36M cells (Fig. 6a and Extended Data Fig. 10a,b). Consequently, DMOG treatment abolished the formation of Epcam$^+$ and Oct4–GFP$^+$ cells in OKSM/K36M cells (Fig. 6b,c). We observed similar results upon suppression of *Tet1*, *Tet2* and *Tet3* using short hairpin RNAs (Extended Data Fig. 10c,d) and confirmed the attenuated upregulation of the ES cell-associated genes *Epcam*, *Pou5f1* and *miR-290* in DMOG-treated OKSM/K36M cells (Fig. 6d and Extended Data Fig. 10e). By contrast, the MEF-associated gene *Vim* continued to be effectively downregulated in the presence of OKSM/K36M and DMOG, underscoring the Tet-independent regulation of mesenchymal genes (Fig. 6d). Moreover, Dnmt3a/b overexpression neutralized the beneficial effect of K36M on reprogramming and led to increased *Cdh1* methylation, supporting the antagonistic roles of Tet and Dnmt3 enzymes in the regulation of epithelial/pluripotency genes (Fig. 6e and Extended Data Fig. 10f,g). Hence, depletion of H3K36 methylation facilitates reprogramming via disrupting the balance between Dnmt3a/b and Tet enzymes at regulatory elements of epithelial/pluripotency genes, leading to their demethylation and activation.

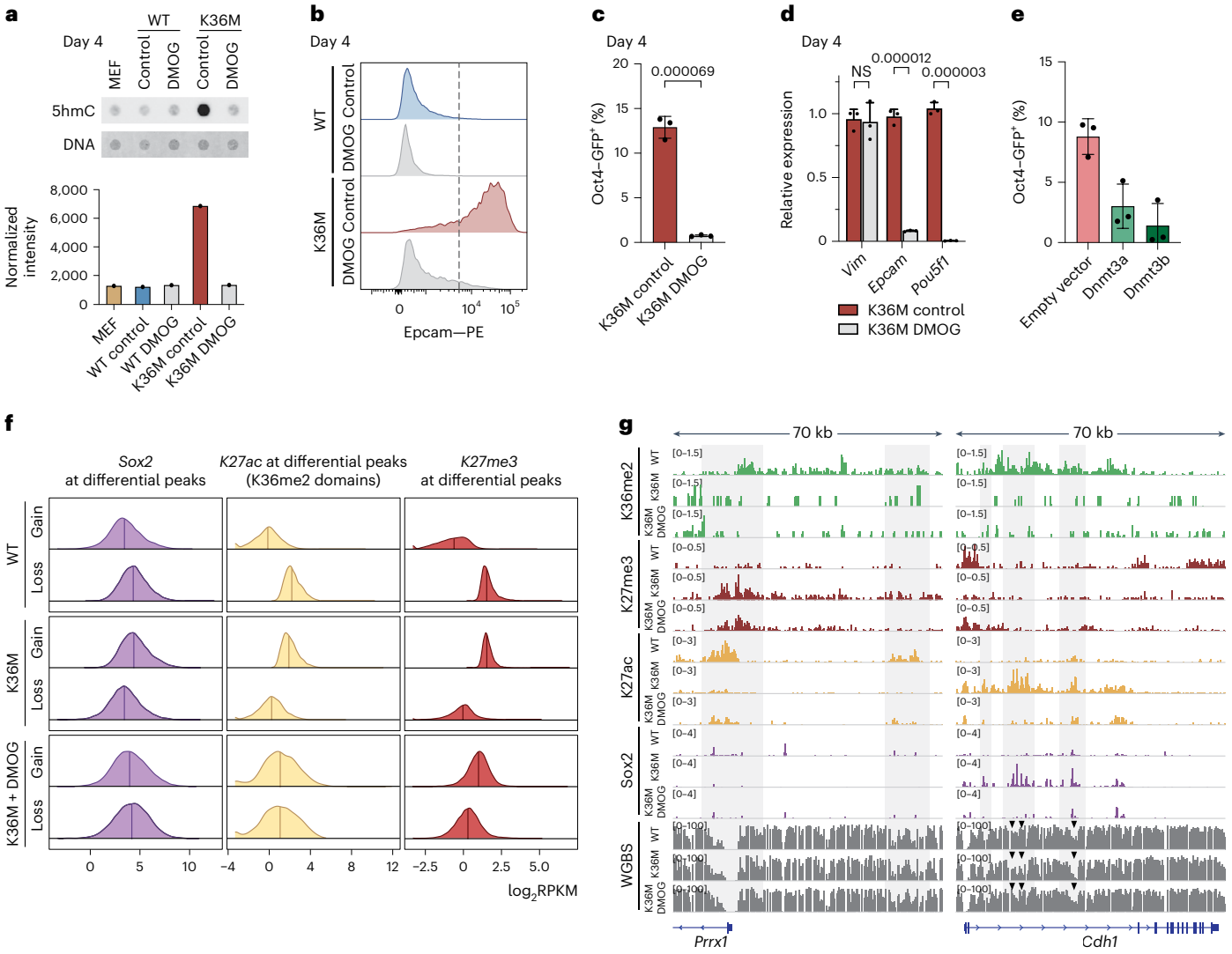
To dissect how impaired DNA demethylation affects regulatory elements in our system, we assessed the epigenome of OKSM/K36M reprogramming intermediates in the presence and absence of DMOG. H3K36me2 was depleted to a similar extent in control and DMOG samples expressing K36M (Extended Data Fig. 10h). Strikingly, DMOG treatment neutralized the K36M-dependent gain of H3K27ac and Sox2 peaks while the K36M-dependent redistribution of H3K27me3 remained largely unaffected (Fig. 6f). For example, at the mesenchymal gene *Prrx1*, we observed efficient depletion of H3K36me2, H3K27ac and Sox2 signal over the enhancer region and a concomitant gain of H3K27me3 over the promoter region in DMOG-treated OKSM/K36M cells (Fig. 6g). However, at the epithelial/pluripotency gene *Cdh1*, we observed that DMOG treatment blocked demethylation and attenuated the increase in H3K27ac and Sox2 binding at the enhancer (Fig. 6g and Extended Data Fig. 10i–k). Together, these results suggest that the inhibition of Tet enzymes

Fig. 6 | DNA demethylation is limiting for K36M-dependent enhancer activation and reprogramming. a, Dot blot assay to quantify 5hmC levels in MEFs as well as day 4 WT and K36M reprogramming intermediates in the presence and absence of DMOG (1 mM). **b**, Flow cytometric quantification of Epcam levels on day 4 of reprogramming in untreated or DMOG-treated WT and K36M intermediates. **c**, Quantification of Oct4–GFP$^+$ cells detected in day 4 K36M reprogramming intermediates in the presence or absence of DMOG. *P* value determined by two-sided unpaired Student's *t*-test, error bars indicate mean ± s.d. ($n = 3$ independent biological experiments). **d**, Relative expression of the fibroblast gene *Vim*, the epithelial gene *Epcam* and the pluripotency gene *Pou5f1* in untreated and DMOG-treated K36M intermediates. Error bars indicate mean ± s.d. ($n = 3$). *P* values were determined by two-sided unpaired Student's *t*-test. **e**, Fraction of Oct4–GFP$^+$ K36M cultures transduced with either an empty vector or DOX-inducible overexpression vectors for *Dnmt3a* and *Dnmt3b*, error bars indicate mean ± s.d. ($n = 3$ independent biological experiments). **f**, Quantification of Sox2, H3K27ac and H3K27me3 levels at sites that are gained or lost in K36M versus WT samples and the effect of DMOG on these enrichment patterns (K36M only). Centre line indicates median ($n = 2$ biologically independent replicates were integrated for this analysis). **g**, Representative gene tracks (H3K36me2, H3K27ac, H3K27me3, Sox2 and whole-genome bisulfite sequencing (WGBS)) showing the mesenchymal gene *Prrx1* and the epithelial gene *Cdh1* for WT, K36M, K36M + DMOG conditions on day 4 of reprogramming ($n = 2$ biologically independent replicates). The data in square brackets represent the range of normalized read densities for CUT&Tag samples and the percentage of DNA methylation for WGBS. **h**, Schematic of DMOG washout experiment (top), and quantification of Oct4–GFP$^+$ cells on day 8 of reprogramming (bottom). K36M cells were either left untreated (control) or treated with DMOG for 8 days (DMOG day 0 to day 8) or 4 days, (DMOG day 0 to day 4). *P* values were determined by two-sided unpaired Student's *t*-test, error bars indicate mean ± s.d. ($n = 3$ biologically independent experiments). **i**, Summary highlighting the dual role of K36M-mediated H3K36me2 depletion on active mesenchymal versus repressed epithelial and pluripotency genes. NS, not significant.

https://doi.org/10.1038/s41556-023-01191-z

uncouples K36M-dependent ES cell enhancer activation, which is Tet dependent, from MEF enhancer decommissioning, which is Tet independent.

Finally, we determined whether the restoration of Tet function by withdrawal of DMOG from OKSM/K36M reprogramming cultures could restore their ability to acquire pluripotency. Remarkably, removal of DMOG from OKSM/K36M cultures that had been treated with DOX and DMOG for 4 days progressed from an Epcam$^-$/Oct4–GFP$^-$ state to an Epcam$^+$/Oct4–GFP$^+$ state in virtually every cell after as little as 4 days of additional DOX treatment (Fig. 6h), thus catching up to OKSM/K36M cultures exposed to DOX alone for the entire time period. These data imply that demethylation of epithelial/pluripotency genes is the key limiting factor for acquiring an epithelial and subsequently a pluripotent state in OKSM/K36M cells.



Discussion

Here we provide evidence that H3K36 methylation effectively reinforces cell identity via two complementary mechanisms that differentially impact gene expression (Fig. 6i). First, H3K36 methylation sustains the expression of cell-type-specific genes by keeping associated enhancers and gene bodies in a euchromatic, transcriptionally active state. This mechanism relies on the known ability of H3K36me2/3 to antagonize the repressive H3K27me3 mark and is consistent with previous observations in cancer cell lines[88]. Second, H3K36 methylation aids in the silencing of alternative lineage genes by maintaining associated enhancers in a methylated and closed chromatin state that opposes the binding of lineage-specific transcription factors. During reprogramming, this dual mechanism manifests itself as an enhanced decommissioning of mesenchymal enhancers leading to reduced ectopic OKSM binding, as well as an increased activation of pluripotency enhancers due to elevated OKSM binding to demethylated ES cell-specific targets. Mechanistically, our data suggest that DNA methylation of regulatory elements uncouples MEF enhancers that are decommissioned and transcriptionally silenced (methylation independent) from pluripotency enhancers that are commissioned and transcriptionally activated (methylation sensitive). We propose that the ultimate outcome of perturbed H3K36 methylation on cell fate will be context specific and dependent on the pre-existing chromatin and DNA methylation state as well as the availability of fate-instructive transcription factors.

In addition to illuminating the molecular basis by which K36M impacts cell identity, we establish a functional relationship between H3K36 methylation and TGFβ signalling. As K36M blocks signal integration independently of TGFβ and Smad phosphorylation, we propose that H3K36 methylation acts as an epigenetic relay mechanism for TGFβ signalling by modulating the ability of Smad proteins to access relevant chromatin targets such as *Zeb1* (ref. 89). Our finding that H3K36 methylation maintains a mesenchymal identity and resists acquisition of an epithelial identity is also supported by the recent identification of H3K36me2/3-modifying enzymes as modulators of epithelial plasticity in pancreatic cancer cell lines, suggesting a conserved mechanism across both physiological and pathological contexts[88].

Our data have implications for regenerative medicine and the treatment of cancer. For example, *NSD1* and *H3K36M* mutations were recently detected in a subset of head and neck squamous cell carcinomas associated with genomic hypomethylation and impaired differentiation[90]. Supporting this notion, we show that squamous stem cells expressing K36M indeed maintain expression of basal stem cell markers when exposed to differentiation cues. Furthermore, our data suggest that modulation of H3K36 methylation via K36M could be a simple approach to generate patient-specific human iPS cells or induced neurons at high efficiency. Beyond K36M, a recent study identified over one hundred additional mutants covering all histone genes and diverse types of cancer[25]. These mutations, like K36M, are thought to dominantly block methylation, and we speculate that their manipulation in experimental model systems should facilitate mechanistic insights into cell identity maintenance that have so far remained obscured by the study of the respective histone-modifying enzymes alone.

Online content

Any methods, additional references, Nature Portfolio reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41556-023-01191-z.

References

1. Tsankov, A. M. et al. Transcription factor binding dynamics during human ESC differentiation. *Nature* **518**, 344–349 (2015).
2. Ho, L. & Crabtree, G. R. Chromatin remodelling during development. *Nature* **463**, 474–484 (2010).
3. Mittnenzweig, M. et al. A single-embryo, single-cell time-resolved model for mouse gastrulation. *Cell* **184**, 2825–2842.e22 (2021).
4. Grosswendt, S. et al. Epigenetic regulator function through mouse gastrulation. *Nature* **584**, 102–108 (2020).
5. Wagner, D. E. et al. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science* **360**, 981–987 (2018).
6. Aguilera-Castrejon, A. et al. Ex utero mouse embryogenesis from pre-gastrulation to late organogenesis. *Nature* **593**, 119–124 (2021).
7. Takaoka, K. & Hamada, H. Cell fate decisions and axis determination in the early mouse embryo. *Development* **139**, 3–14 (2011).
8. Jambhekar, A., Dhall, A. & Shi, Y. Roles and regulation of histone methylation in animal development. *Nat. Rev. Mol. Cell Biol.* **20**, 625–641 (2019).
9. Suvà, M. L., Riggi, N. & Bernstein, B. E. Epigenetic reprogramming in cancer. *Science* **339**, 1567–1570 (2013).
10. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
11. Brumbaugh, J., Stefano, B. D. & Hochedlinger, K. Reprogramming: identifying the mechanisms that safeguard cell identity. *Development* **146**, dev182170 (2019).
12. Apostolou, E. & Hochedlinger, K. Chromatin dynamics during cellular reprogramming. *Nature* **502**, 462–471 (2013).
13. Onder, T. T. et al. Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483**, 598–602 (2012).
14. Borkent, M. et al. A serial shRNA screen for roadblocks to reprogramming identifies the protein modifier SUMO2. *Stem Cell Rep.* **6**, 704–716 (2016).
15. Yang, C.-S., Chang, K.-Y. & Rana, T. M. Genome-wide functional analysis reveals factors needed at the transition steps of induced reprogramming. *Cell Rep.* **8**, 327–337 (2014).
16. Qin, H. et al. Systematic identification of barriers to human iPSC generation. *Cell* **158**, 449–461 (2014).
17. Cheloufi, S. et al. The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528**, 218–224 (2015).
18. Brumbaugh, J. et al. Nudt21 controls cell fate by connecting alternative polyadenylation to chromatin signaling. *Cell* **172**, 106–120.e21 (2018).
19. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Exp. Mol. Med.* **49**, e324 (2017).
20. Kim, E. et al. Phosphorylation of EZH2 activates STAT3 signaling via STAT3 methylation and promotes tumorigenicity of glioblastoma stem-like cells. *Cancer Cell* **23**, 839–852 (2013).
21. Xu, K. et al. EZH2 oncogenic activity in castration-resistant prostate cancer cells is polycomb-independent. *Science* **338**, 1465–1469 (2012).
22. Shpargel, K. B., Sengoku, T., Yokoyama, S. & Magnuson, T. UTX and UTY demonstrate histone demethylase-independent function in mouse embryonic development. *PLoS Genet.* **8**, e1002964 (2012).
23. Miller, S. A., Mohn, S. E. & Weinmann, A. S. Jmjd3 and UTX play a demethylase-independent role in chromatin remodeling to regulate T-box family member-dependent gene expression. *Mol. Cell* **40**, 594–605 (2010).
24. Herz, H.-M. et al. Histone H3 lysine-to-methionine mutants as a paradigm to study chromatin signaling. *Science* **345**, 1065–1070 (2014).
25. Nacev, B. A. et al. The expanding landscape of 'oncohistone' mutations in human cancers. *Nature* **567**, 473–478 (2019).

26. Lewis, P. W. et al. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. *Science* **340**, 857–861 (2013).
27. Chan, K.-M. et al. The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. *Gene Dev.* **27**, 985–990 (2013).
28. Fang, D. et al. The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. *Science* **352**, 1344–1348 (2016).
29. Lu, C. et al. Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science* **352**, 844–849 (2016).
30. Schwartzentruber, J. et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* **482**, 226–231 (2012).
31. Mohammad, F. et al. EZH2 is a potential therapeutic target for H3K27M-mutant pediatric gliomas. *Nat. Med.* **23**, 483–492 (2017).
32. Brumbaugh, J. et al. Inducible histone K-to-M mutations are dynamic tools to probe the physiological role of site-specific histone methylation in vitro and in vivo. *Nat. Cell Biol.* **21**, 1449–1461 (2019).
33. Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nat. Methods* **7**, 53–55 (2010).
34. Chen, J. et al. H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat. Genet.* **45**, 34–42 (2013).
35. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
36. Sridharan, R. et al. Proteomic and genomic approaches reveal critical functions of H3K9 methylation and Heterochromatin Protein-1γ in reprogramming to pluripotency. *Nat. Cell Biol.* **15**, 872–882 (2013).
37. Fragola, G. et al. Cell reprogramming requires silencing of a core subset of polycomb targets. *PLoS Genet.* **9**, e1003292 (2013).
38. Ding, X. et al. The polycomb protein Ezh2 impacts on induced pluripotent stem cell generation. *Stem Cells Dev.* **23**, 931–940 (2014).
39. Wang, C. et al. Enhancer priming by H3K4 methyltransferase MLL4 controls cell fate transition. *Proc. Natl Acad. Sci. USA* **113**, 11871–11876 (2016).
40. Liang, G., He, J. & Zhang, Y. Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. *Nat. Cell Biol.* **14**, 457–466 (2012).
41. Wang, T. et al. The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell* **9**, 575–587 (2011).
42. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2**, 230–240 (2008).
43. Polo, J. M. et al. A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617–1632 (2012).
44. Brambrink, T. et al. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2**, 151–159 (2008).
45. Schwarz, B. A. et al. Prospective isolation of poised iPSC intermediates reveals principles of cellular reprogramming. *Cell Stem Cell* **23**, 289–305.e5 (2018).
46. Lengner, C. J. et al. Oct4 expression is not required for mouse somatic stem cell self-renewal. *Cell Stem Cell* **1**, 403–415 (2007).
47. Schiebinger, G. et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928–943.e22 (2019).
48. Li, R. et al. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7**, 51–63 (2010).
49. Samavarchi-Tehrani, P. et al. Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010).
50. Massagué, J. TGFβ signalling in context. *Nat. Rev. Mol. Cell Biol.* **13**, 616–630 (2012).
51. Maherali, N. & Hochedlinger, K. Tgfβ signal inhibition cooperates in the induction of iPSCs and replaces Sox2 and cMyc. *Curr. Biol.* **19**, 1718–1723 (2009).
52. Gregory, P. A. et al. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat. Cell Biol.* **10**, 593–601 (2008).
53. Park, S.-M., Gaur, A. B., Lengyel, E. & Peter, M. E. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Gene Dev.* **22**, 894–907 (2008).
54. Subramanyam, D. et al. Multiple targets of miR-302 and miR-372 promote reprogramming of human fibroblasts to induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 443–448 (2011).
55. Judson, R. L., Babiarz, J. E., Venere, M. & Blelloch, R. Embryonic stem cell–specific microRNAs promote induced pluripotency. *Nat. Biotechnol.* **27**, 459–461 (2009).
56. Vierbuchen, T. et al. Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**, 1035–1041 (2010).
57. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**, 987–1000 (1987).
58. Yagi, M. et al. Dissecting dual roles of MyoD during lineage conversion to mature myocytes and myogenic stem cells. *Gene Dev.* **35**, 1209–1228 (2021).
59. Bar-Nur, O. et al. Direct reprogramming of mouse fibroblasts into functional skeletal muscle progenitors. *Stem Cell Rep.* **10**, 1505–1521 (2018).
60. Beccari, L. et al. Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* **562**, 272–276 (2018).
61. van den Brink, S. C. et al. Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Development* **141**, 4231–4242 (2014).
62. Baillie-Johnson, P., Brink, S. C. van den, Balayo, T., Turner, D. A. & Arias, A. M. Generation of aggregates of mouse embryonic stem cells that show symmetry breaking, polarization and emergent collective behaviour in vitro. *J. Vis. Exp. JoVE* https://doi.org/10.3791/53252 (2015).
63. Diaz-Cuadros, M. et al. In vitro characterization of the human segmentation clock. *Nature* **580**, 113–118 (2020).
64. Kahata, K., Dadras, M. S. & Moustakas, A. TGF-β family signaling in epithelial differentiation and epithelial–mesenchymal transition. *Cold Spring Harbor Perspect. Biol.* **10**, a022194 (2018).
65. Mou, H. et al. Dual SMAD signaling inhibition enables long-term expansion of diverse epithelial basal cells. *Cell Stem Cell* **19**, 217–231 (2016).
66. Hillringhaus, L. et al. Structural and evolutionary basis for the dual substrate selectivity of human KDM4 histone demethylase family. *J. Biol. Chem.* **286**, 41616–41625 (2011).
67. Krogan, N. J. et al. Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* **23**, 4207–4218 (2003).
68. Streubel, G. et al. The H3K36me2 methyltransferase Nsd1 demarcates PRC2-mediated H3K27me2 and H3K27me3domains in embryonic stem cells. *Mol. Cell* **70**, 371–379.e5 (2018).

69. Wang, G. G., Cai, L., Pasillas, M. P. & Kamps, M. P. NUP98–NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat. Cell Biol.* **9**, 804–812 (2007).
70. Yuan, W. et al. H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* **286**, 7983–7989 (2011).
71. Rajagopalan, K. N. et al. Depletion of H3K36me2 recapitulates epigenomic and phenotypic changes induced by the H3.3K36M oncohistone mutation. *Proc. Natl Acad. Sci. USA* **118**, e2021795118 (2021).
72. Fang, Y. et al. The H3K36me2 methyltransferase NSD1 modulates H3K27ac at active enhancers to safeguard gene expression. *Nucleic Acids Res.* **49**, 6281–6295 (2021).
73. Chaouch, A. et al. Histone H3.3 K27M and K36M mutations de-repress transposable elements through perturbation of antagonistic chromatin marks. *Mol. Cell* **81**, 4876–4890.e7 (2021).
74. Barral, A. et al. SETDB1/NSD-dependent H3K9me3/H3K36me3 dual heterochromatin maintains gene expression profiles by bookmarking poised enhancers. *Mol. Cell* **82**, 816–832.e12 (2022).
75. Chronis, C. et al. Cooperative binding of transcription factors orchestrates reprogramming. *Cell* **168**, 442–459.e20 (2017).
76. Yin, Y. et al. Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* **356**, eaaj2239 (2017).
77. Kaluscha, S. et al. Evidence that direct inhibition of transcription factor binding is the prevailing mode of gene and repeat repression by DNA methylation. *Nat. Genet.* **54**, 1895–1906 (2022).
78. Weinberg, D. N. et al. The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281–286 (2019).
79. Baubec, T. et al. Genomic profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. *Nature* **520**, 243–247 (2015).
80. Yagi, M. et al. Identification of distinct loci for de novo DNA methylation by DNMT3A and DNMT3B during mammalian development. *Nat. Commun.* **11**, 3199 (2020).
81. von Meyenn, F. et al. Impairment of DNA methylation maintenance is the main cause of global demethylation in naive embryonic stem cells. *Mol. Cell* **62**, 848–861 (2016).
82. Charlton, J. et al. TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat. Genet.* **52**, 819–827 (2020).
83. Hu, X. et al. Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. *Cell Stem Cell* **14**, 512–522 (2014).
84. Gao, Y. et al. Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell* **12**, 453–469 (2013).
85. Zviran, A. et al. Deterministic somatic cell reprogramming involves continuous transcriptional changes governed by Myc and epigenetic-driven modules. *Cell Stem Cell* **24**, 328–341.e9 (2019).
86. Amouroux, R. et al. De novo DNA methylation drives 5hmC accumulation in mouse zygotes. *Nat. Cell Biol.* **18**, 225–233 (2016).
87. Sardina, J. L. et al. Transcription factors drive Tet2-mediated enhancer demethylation to reprogram cell fate. *Cell Stem Cell* **23**, 727–741.e9 (2018).
88. Yuan, S. et al. Global regulation of the histone mark H3K36me2 underlies epithelial plasticity and metastatic progression. *Cancer Discov.* **10**, 854–871 (2020).
89. Brabletz, S. & Brabletz, T. The ZEB/miR-200 feedback loop—a motor of cellular plasticity in development and cancer? *EMBO Rep.* **11**, 670–677 (2010).
90. Papillon-Cavanagh, S. et al. Impaired H3K36 methylation defines a subset of head and neck squamous cell carcinomas. *Nat. Genet.* **49**, 180–185 (2017).

Methods

Experimental procedures

Derivation of mouse embryonic fibroblasts. After timed mating, embryos were dissected from pregnant females at embryonic day (E)14.5. Head, limbs and internal organs were removed; the remaining tissue minced and incubated in trypsin–EDTA, cells were then expanded after quenching with MEF medium (Dulbecco's modified Eagle medium (DMEM), 10% fetal bovine serum (FBS), Minimum Essential Medium Non-Essential Amino Acids solution (MEM/NEAA), GlutaMAX and penicillin–streptomycin). Eight- to 10-week-old male and female mice of the following strains were used for breeding set-ups: Col1a1::tetO-OKSM with an EGFP reporter in the 3′ untranslated region (UTR) of the Pou5f1 gene, Col1a1::tetO-H3.3, Col1a1::tetO-K36M, Rosa26::M2-rtTA.

For reprogramming experiments, MEFs were derived from crosses leading to M2-rtTA at the Rosa26 locus (heterozygous), an EGFP reporter in the 3′ UTR of Pou5f1, an inducible OKSM cassette in one Col1a1 locus, and inducible H3.3 WT or K36M in the other Col1a1 locus. For additional experiments, MEFs were generated with the same set-up, but without a polycistronic OKSM cassette. Mice used in this study were housed and bred in specific-pathogen-free rooms located in the Association for Assessment and Accreditation of Laboratory Animal Care International-accredited Center for Comparative Medicine vivarium at Massachusetts General Hospital (MGH). Mice were housed in ventilated cages on a standard 12 h:12 h light cycle. All procedures involving mice adhered to the guidelines of the approved MGH Institutional Animal Care and Use Committee protocol no. 2006N000104.

Vectors. Lentiviral vectors for the DOX-inducible expression of WT H3.3, K4M, K9M, K27M, K36M, Zeb1, MyoD, Kdm2a, Kdm2b, Kdm4a, Kdm4c, Nsd2, Dnmt3a, Dnmt3b and an empty vector were purchased from VectorBuilder. Lentiviral vectors for the expression of short hairpin RNAs targeting Tet enzymes were also purchased from Vectorbuilder. FUW-TetO-Ascl1, FUW-TetO-Brn2, and FUW-Myt1l were gifts from Marius Wernig (addgene 27150, 27151 and 27152) (ref. 56). pHAGE2-TetOminiCMV-SKM, pHAGE2-TetOminiCMV-OSM and pHAGE2-TetOminiCMV-OKM were gifts from Hans Schöler (addgene 136551, 136555 and 136554) (ref. 91). pHAGE-STEMCCA vectors were used for constitutive and DOX-inducible expression of OKSM as previously described[92]. pLV-EF1a-IRES-Blast was a gift from Tobias Meyer (addgene 85133). H3.3 WT and K36M were introduced using Gibson Assembly (New England Biolabs).

Induction of pluripotency. MEFs and GMPs with DOX-inducible expression of OKSM and H3.3 WT or K36M were cultured as described previously[93]. FBS/LIF medium consisted of KO-DMEM, 15% de-activated FBS, GlutaMAX, MEM/NEAA, 1,000 U ml^{-1} leukaemia inhibitory factor (LIF) and 50 μM β-mercaptoethanol. Combinations of DOX (2 μg ml^{-1}, Sigma-Aldrich), ascorbic acid (50 μg ml^{-1}, Sigma-Aldrich) and GSK3-inhibitor CHIR99021 (3 μM, Axon Medchem) were supplemented (AGi medium). For reprogramming of GMPs, stem cell factor (SCF, 20 ng ml^{-1}, Peprotech), IL-3 (10 ng ml^{-1}, Peprotech) and IL-6 (10 ng ml^{-1}, Peprotech) were added for the first 72 h. For keratinocyte reprogramming, cells with DOX-inducible expression of OKSM were lentivirally transduced with DOX-inducible H3.3 WT or K36M vectors and cultured in SAGM medium (Lonza) supplemented with CHIR99021 (Tocris), Y27632 (Tocris) and A-8301 (Tocris)[65]. After 2 days of reprogramming, medium was changed to a 1:1 split with FBS/LIF. After two more days, cells were cultured in only FBS/LIF medium. Alkaline phosphatase (AP) staining was performed using a Vector Red kit (Vector Labs) according to the manufacturer's recommendations.

Human fibroblasts (BJ, CRL-2522 ATCC) cultured in DMEM/10% FBS were lentivirally transduced for the constitutive expression of H3.3 WT or K36M. STEMCCA lentivirus was introduced for the constitutive expression of OKSM, and cells were seeded on Matrigel-coated plates. Four days after transduction, medium was switched to ReproTeSR (Stem Cell Technologies). AP staining was performed as described above.

Teratoma formation assay. Approximately 2×10^6 cells were injected subcutaneously into 8-week-old female athymic nude mice (NU/J, Jackson Laboratories). Teratomas were monitored and removed when tumour size reached 1 cm or tumours ulcerated. Tumours were then processed for haematoxylin/eosin staining.

Blastocyst injections. To generate coat-colour chimaeras from iPS cells derived with transient expression of K36M, iPS cells were injected into albino B6 blastocysts at E3.5 as previously described[94,95].

MEF-to-neuron conversion. MEFs with DOX-inducible expression of H3.3 WT or K36M were infected with DOX-inducible lentiviral vectors (FUW-TetO-Ascl1, FUW-TetO-Brn2 and FUW-TetO-Myt1l) in MEF growth medium and plated on Matrigel. After 2 days, medium was changed to neuron induction medium supplemented with DOX[56]. After 3 days of half-medium changes, cells were fixed and stained for Tubb3 (BioLegend 801201).

MEF-to-myotube conversion, MEF-to-iMPC de-differentiation. MEFs with DOX-inducible expression of WT H3.3 or K36M were transduced with a DOX-inducible lentiviral vector for expression of MyoD. Cells were then cultured as previously described[58,59].

Gastruloid generation. Gastruloids were generated as described previously[61,62]. Briefly, KH2 ES cells inducibly expressing H3.3 WT or K36M (FBS/LIF) were plated in ultralow-attachment round-bottom 96-well plates (300 cells per well) in N2B27 medium and pulsed with 3 μM CHIR99021 for 24 h on day 2 after plating. Elongation was assessed at day 5.

Pre-somitic mesoderm differentiation. Pre-somitic mesoderm differentiation was induced as described previously[63]. In brief, KH2 ES cells inducibly expressing H3.3 WT or K36M (2i/LIF) were plated in N2B27 medium supplemented with 25 ng ml^{-1} Activin A (Peprotech) and 12 ng ml^{-1} basic fibroblast growth factor (Peprotech). After 48 h, medium was changed to high-glucose DMEM supplemented with GlutaMAX, sodium pyruvate, MEM/NEAA, 15% FBS, 50 μM β-mercaptoethanol, 3 μM CHIR99021 (Axon Medchem) and 0.5 μM LDN193189 (Tocris). Forty-eight hours later, cells were fixed for immunofluorescence or RNA was collected.

Embryoid body generation. Mouse iPS cells were resuspended in MEF medium, then centrifuged in anti-adherence microwell plates (Stem Cell Technologies). After 24 h, cell aggregates were transferred into low-adherence plates (Stem Cell Technologies) and cultured for four more days.

Keratinocyte culture. Keratinocytes were derived from a Rosa26-rtTA mouse, lentivirally transduced with vectors for the DOX-inducible expression of WT H3.3 or K36M, and cultured as previously described[65]. To assess the effect of K36M expression on TGFβ signalling in this system, A-83-01 was removed and replaced with recombinant TGFβ-1 (Peprotech) at 10 ng ml^{-1} after 2 days of DOX treatment. Cells were fixed and stained after four more days of culture.

qRT-PCR. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) reactions were set up in triplicate using Brilliant III SYBR Master Mix (Agilent) and run on a LightCycler 480 PCR machine (Roche) with 40 cycles of 30 s at 95 °C, 30 s at 60 °C and 30 s at 72 °C. Fold change was calculated using the $\Delta\Delta C_t$ method, and data were normalized to *Gapdh*. Kicqstart primers were purchased from Sigma-Aldrich

for mouse *Epcam*, *Cdh1* and *Vim*, and for human *EPCAM*, *CDH1*, *LIN28A*, *NANOG* and *GAPDH*.

For miRNA analysis, complementary DNA was generated using a TaqMan MicroRNA Reverse Transcription Kit according to the manufacturer's recommendations, and reactions were set up in triplicate with the TaqMan Universal PCR Master Mix (no AmpErase UNG, Applied Biosystems) and miRNA-specific TaqMan MicroRNA assays (TaqMan MicroRNA mmu-miR-290 assay ID: 000187, TaqMan MicroRNA hsa-miR-200b assay ID: 002251, TaqMan MicroRNA hsa-miR-205 assay ID: 000509, TaqMan Array MicroRNA U6 snRNA, assay ID: 001973, all Life Technologies). Reactions were run on a LightCycler 480 PCR machine (Roche) with 40 cycles of 15 s at 95 °C and 1 min at 60 °C. Fold change was calculated using the $\Delta\Delta C_t$ method, data were normalized to U6.

Flow cytometry of tissue cultures. Surface marker staining was performed using dye-conjugated antibodies against Thy1 (anti-Mouse Thy1.2 eFluor 450, Thermo Fisher Scientific), SSEA-1 (anti-Human/Mouse SSEA-1 eFluor 660, Thermo Fisher Scientific) and EpCAM (anti-Mouse EpCAM PE, Thermo Fisher Scientific). Intracellular staining was performed using a Fix & Perm Kit (Thermo Fisher Scientific) and an antibody targeting Nanog (anti-Nanog antibody ab80892, Abcam). CellTrace-Violet (Thermo Fisher Scientific) staining and Annexin-V/PI (Thermo Fisher Scientific) staining were performed according to the manufacturer's recommendations. Data were acquired on an LSRII flow cytometer (BD Biosciences) using Diva v.6.1.2 (BD Biosciences).

Bone marrow preparation and flow cytometry. Bone marrow from was incubated in RBC lysis buffer (BioLegend) for 8 min on ice. A total of 40–50 million cells were applied for lineage depletion according to the manufacturer's recommendations (Miltenyi Biotech). The cells were stained as follows to sort GMPs.

The following lineage markers were used: Ter119 (PE/Cy5; Thermo Fisher, catalog no. 15-5921-83), CD3e (PE/Cy5; BioLegend, catalog no. 100310), Gr1 (PE/Cy5; Thermo Fisher, catalog no. 15-5931-83); B220/CD45R (PE/Cy5; BioLegend, catalog no. 103210); TCRb (PE/Cy5; Thermo Fisher, catalog no. 15-5961-83); CD4 (PE/Cy5; Thermo Fisher, catalog no. 15-0041-83); CD8a (PE/Cy5; BioLegend, catalog no. 100710); Sca1 (PE/Cy7; Thermo Fisher, catalog no. 25-5981-82); c-Kit (APC-eFluor 780; Thermo Fisher, catalog no. 47-1171-82); CD34 (FITC; Thermo Fisher, catalog no. 11-0341-85); CD16/32 (FC gamma receptor) (PE; BD Biosciences, catalog no. 553145); 4′,6-diamidino-2-phenylindole (BD Biosciences, catalog no. 564907). Antibodies were diluted to 1 µl per million cells.

Immunofluorescence assays. The following primary antibodies were used: anti-H3K36M (anti-Histone H3 K36M Rabbit Monoclonal Antibody, Clone RM193, RevMab), anti-H3K36me3 (abcam 9050, Thermo Fisher MA5-24687), anti-H3K36me2 (active motif 39256), anti-Sox2 (Cell Signaling 23064), anti-Nanog (abcam 80892) anti-Epcam (Life Technologies 14-5791-81), anti-Vim (Cell Signaling 5741), anti-Tubb3 (BioLegend 801201), anti-Myh1 (Developmental Studies Hybridoma Bank, Iowa MF20), anti-p63 (Santa Cruz sc-8431) and anti-Cdh2 (Cell Signaling 13116).

Western blot assay. Immunoblotting was performed using the following antibodies: GAPDH (Cell Signaling 8884), phospho-Smad2 Ser465/467 (Cell Signaling 3108) and Smad2 (Cell Signaling 5339). Images were taken using a ChemiDoc Imaging System with enhanced chemiluminescence detection.

Mass spectrometry of histone post-translational modifications. Bulk histones were acid-extracted from cell pellets, propionylated and subjected to trypsin digestion as described previously[96]. Samples were analysed on a triple quadrupole (QqQ) mass spectrometer (Thermo Fisher TSQ Quantiva) directly coupled with an UltiMate 3000 Dionex nano-liquid chromatography system. Targeted analysis of unmodified and various modified histone peptides was performed. The process was repeated three separate times for each sample.

Dot blot assay. DNA samples were sonicated to generate fragments of 200–500 bp length. After denaturizing, samples were spotted on nitrocellulose membranes in a Bio-Dot apparatus (Bio-Rad) according to manufacturer recommendations. After washing, ultraviolet-crosslinking and blocking, membranes were incubated with anti-5hmC antibody (Active Motif 39769) overnight at 4 °C. After incubation with horseradish peroxidase-conjugated secondary antibody (Invitrogen G21234), images were taken using a ChemiDoc Imaging System with enhanced chemiluminescence detection. Quantification was performed using ImageJ.

Bisulfite sequencing. A total of 200 ng of DNA was bisulfite converted using an EZ DNA Methylation-Gold Kit (Zymo Research) according to the manufacturer's recommendations. PCR was performed with GoTaq (Promega), before cloning into a pCR4-TOPO vector (Invitrogen). Subcloned colonies were sequenced with M13 reverse primer. The primers used for amplification of the Cdh1 enhancer were AATAGGATAGAATAGATAAGGGAAAAAG (forward) and AAACATTTATTTCTAATCTTACAAACCA (reverse).

RNA-seq and ATAC-seq. For bulk RNA-seq, samples were collected at days 0, 2, 4, 6 and 8 of reprogramming for both genotypes. Specimens of iPS cells of the same backgrounds were passaged for seven passages in FBS/LIF on feeders before pre-plating and RNA extraction. RNA-seq libraries were constructed using polyA selection followed by NEBNext UltraDirectional kit protocol (New England Biolabs) and sequenced on the Illumina HiSeq2500 instrument, resulting in ~30 million reads per sample on average. Small RNA-seq libraries were generated using the NEB small RNA library kit E7330L (New England Biolabs), and sequencing was run on the NextSeq 2000 as PE50 on a P3 type flowcell to ~30 million reads per sample.

Assay for transposase-accessible chromatin using sequencing (ATAC-seq) was performed as previously described[97]. Briefly, nuclei of 60,000 cells per sample were resuspended in transposition buffer with Tn5 transposase. DNA was isolated using a MinElute kit (Qiagen), and libraries were amplified by PCR for 13 cycles using barcoded, Illumina-compatible primers. After amplification, DNA was size-selected for fragments between 100 bp and 1,000 bp using AMPure XP beads (Beckman Coulter Life Sciences). Libraries for RNA- and ATAC-seq underwent TapeStation 2200 (Agilent) analysis before sequencing. ATAC-seq libraries were sequenced in paired-end 50 bp mode on the Illumina HiSeq2500 instrument, resulting in ~40 million reads per sample.

Single-cell RNA-seq. Single-cell RNA-seq was performed for MEFs before induction and for reprogramming intermediates of both genotypes on days 2, 4, 6 and 8. An iPS cell sample was added after seven passages in FBS/LIF on feeders and pre-plating. Approximately 5,000 cells per sample were encapsulated using the 10x Chromium pipeline according to manufacturer guidelines using Single Cell 3′ chemistry v3 (10x Genomics). After mapping and quality control (QC), cells were embedded using the monocle framework and DPT (diffusion pseudotime)[98,99].

CUT&Tag assay. CUT&Tag was performed as previously described[100]. In brief, 100,000 bead-bound cells per sample were permeabilized and incubated with primary antibodies targeting anti-H3K36me3 (Thermo Fisher MA5-24687), anti-H3K36me2 (active motif 39256), anti-H3K4me3 (active motif 39060), anti-H3K27ac (Millipore MABE647), H3K27me3 (Cell Signaling 9733) and anti-Sox2 (Cell Signaling 23064). After incubation with secondary antibody (Guinea Pig

anti-Rabbit IgG, antibodies online ABIN101961), pAG-tethered transposase (pAG-Tn5, Epicypher) was bound in situ at target loci. After tagmentation and DNA clean-up, libraries were prepared by PCR amplification with barcoded Illumina-compatible primers. TapeStation 2200 analysis using High Sensitivity D1000 ScreenTape (Agilent) confirmed successful library preparation before sequencing in paired-end 50 bp mode with Illumina HiSeq2500, resulting in approximately 15 million reads per sample on average.

RRBS assay. RRBS was performed on 100 ng of genomic DNA for each sample using the NuGEN Ovation RRBS methyl-seq system according to the manufacturer's recommendations. Bisulfite conversion of DNA was performed using the Qiagen EpiTect fast bisulfite conversion kit. Libraries were purified with Agencourt RNAclean XP beads, QC was performed using TapeStation analysis. Libraries were sequenced on the Illumina NovaSeq6000, generating 100 bp single-end reads.

WGBS assay. Whole-genome bisulfite sequencing (WGBS) libraries were produced with the TrueMethyl oxBS-Seq Module (Tecan) and Accel-NGS Methyl-Seq DNA Library Kit (Swift) according to the manufacturer's instructions with the omission of the oxBS steps in the TrueMethyl protocol. Briefly, 300 ng of genomic DNA (gDNA) was sheared to average fragment size of 200 bp using a Covaris S2 sonicator for 1 min 30 s with the following settings: duty cycle 10%, intensity 5, cycles per burst 200. Resulting sheared gDNA was concentrated using AMPure XP beads (Beckman Coulter) and then subjected to bisulfite conversion and desulfunation as outlined in the TrueMethyl oxBS-Seq Module. This bisulfite-converted gDNA was then used as input for the Accel-NGS Methyl-Seq DNA Library Kit and yielded sequenceable libraries.

Statistical analyses

RNA-seq data analysis. RNA-seq reads were mapped by STAR v.2.5.0 aligner[101] to mm9 reference genome using ENSEMBL annotation. Read counts for individual genes were produced using the unstranded count feature in HTSeq v.0.11.2 (ref. 102). Differential expression analysis was performed using the edgeR package[103] after normalizing read counts and including only those genes with count per million reads >1 for one or more samples[104]. Differentially expressed genes (DEGs) were defined on the basis of the criteria of >2-fold change in expression value. Gene Ontology analysis was performed using GeneOntology.org[105].

Single-cell RNA-seq data analysis. Sequenced reads were first aligned using CellRanger v.3.0.2, and the filtered reads assigned to cell barcodes were analysed with the R package Seurat, v.4.0.0 (ref. 106). A first round of QC (input: 45,751 cells) involved removing all cells positioned at least three median absolute deviations below the mean for library size (total number of reads) and the number of detected genes, as well as above the mean for the percentage of mitochondrial genes expressed from the entire transcriptome (40,018 cells remaining). The R package monocle3, v.1.2.9 was used for normalization, dimensionality reduction and visualization (uniform manifold approximation and projection, UMAP)[98]. Further, 372 single cells corresponding to three separate clusters with low QC metrics and indicative of contamination were removed (39,646 cells remaining), and the data was normalized and clustered again. A new cluster with low QC metric (low number of genes expressed) consisting of 903 single cells emerged and was removed. The final cleaned dataset consisted of 38,743 cells. The heterogeneity score within samples was generated by computing the average Euclidean distance among all cells in the same condition on the basis of monocle3-derived UMAP coordinates. The correlation scores for MEFs and ES cells (ESCsV6.5) were calculated using all available genes, and were compared with publicly available bulk profiles[107]. The data were further *z*-scored. To assess the robustness of sample separation in the visual UMAP representation, we pre-processed the data following the standard Seurat framework as well, including normalizing with SCTransform. The pseudotime plots were generated on the Seurat-pre-processed data using the R package destiny v.3.4.0 (refs. 98,99).

Small RNA-seq data analysis. Raw sequencing reads were trimmed using trimmomatic to exclude adapter contamination and poor-quality bases[108]. Trimmed reads were aligned to reference genome mm9 with star aligner[101]. Read counts for individual genes were produced using the unstranded count feature in HTSeq 0.9.1 (ref. 102). Differential expression analysis was performed using the edgeR package[103] after normalizing read counts and including only those genes with count per million reads greater than 1 for one or more samples[104]. DEGs were defined on the basis of the criteria of minimum 2-fold change in expression value and *P* value less than 0.01.

Mass spectrometry of histone post-translational modification data analysis. Raw mass spectrometry files were imported and analysed in Skyline with Savitzky–Golay smoothing[109]. All Skyline peak area assignments for monitored peptide transitions were manually confirmed. Multiple peptide transitions were quantified for each modification. For each monitored amino acid residue, each (un)modified form was quantified by calculating the sum of peak areas of corresponding peptide transitions; the sum of all modified forms was then calculated for each amino acid to represent the total pool of modifications for that residue. Finally, each modification is then represented as a percentage of the total pool of modifications. This process was carried out for each of the three separate mass spectrometry runs and averaged per sample.

ATAC-seq data analysis. Sequencing reads were mapped using bwa v.0.5.9-r16 aligner[110] followed by peak calling using Hotspot[111]. DiffBind R package[112] was used for the analysis of differential accessible regions, based on the cut-offs of >2-fold change in peak intensity and false discovery rate <0.01.

CUT&Tag data analysis. Sequencing reads were mapped to mm9 reference genome using bwa aligner[110]. Peak calling was performed using Homer[113]. Heat maps and average profiles of CUT&Tag read densities were generated using deepTools[114]. Bedtools was used to calculate read densities over given regions[115]. Gene Ontology analysis was performed with ClusterProfiler after annotation using ChipSeeker[116,117]. Plots were generated in R using the 'ggplot2' package.

RRBS data analysis. Demultiplexed reads were trimmed using cutadapt and the Nugene diversity adapter trimming python script trimRRBSdiversityAdaptCustomers.py. Alignment to the mouse mm9 genome was performed using BSMAP[118] with the following parameters: -v 0.1 -s 12 -q 20 -w 100 -S 1 -u -R -D C-CGG. Duplicate reads were removed using the Nugene deduplication script nudup.py. Methylation ratios were called with the MOABS mcall module[119]. Differentially methylated regions (DMRs) were calculated in R, using package DSS to detect differentially methylated loci[120]. Locations were selected on the basis of a delta of 0.1 and a *P* value threshold of 0.001 across two replicates. Using R, all RRBS samples were filtered to keep only CpGs with at least 5× coverage. Data were then combined to perform analysis on only matched CpGs across all samples (n = 2,688,886). For Sox2 binding site analysis, CpGs were intersected with region sets using bedtools, and heat maps were generated using package 'pheatmap' after calculating average CpG methylation percentage for each binding site in R (ref. 115).

WGBS data analysis. QC and correcting for adaptor content was performed using fastQC and cutadapt. After this, reads were aligned to the mouse mm10 reference genome using BSmap with flags -v 0.1 -s 16 -w 100 -S 1 –q 20 –u -R. The methylation level of all CpGs captured was calculated using the mcall module in the MOABS software suite with standard parameter settings and CpGs <10× coverage were discarded from the analysis.

Statistics and reproducibility. Statistical tests are described in the corresponding figure legends. Unless otherwise indicated, statistical analyses were carried out using R or GraphPad Prism. Data distribution was assumed to be normal, but this was not formally tested. No statistical method was used to pre-determine sample size, no data were excluded from the analyses and the investigators were not blinded to allocation during experiments and outcome assessment.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

RNA-seq, ATAC-seq and CUT&Tag data have been deposited in the Gene Expression Omnibus under accession code GSE203492. Single-cell RNA-seq data are deposited under GSE203536. RRBS and WGBS data are available under GSE203606. The publicly available datasets used in this study are GSE90893, GSE111172 and GSE77420. All other data supporting the findings of this study are available from the corresponding authors upon reasonable request. Source data are provided with this paper.

Code availability

The code used to analyse the single-cell RNA-seq data presented here, and to generate the corresponding plots, is uploaded to GitHub https://github.com/Michorlab/H3K36_methylation_scRNAseq. Code for additional analyses available upon request.

References

91. Velychko, S. et al. Excluding Oct4 from Yamanaka cocktail unleashes the developmental potential of iPSCs. *Cell Stem Cell* **25**, 737–753.e4 (2019).
92. Sommer, C. A. et al. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* **27**, 543–549 (2009).
93. Bar-Nur, O. et al. Small molecules facilitate rapid and synchronous iPSC generation. *Nat. Methods* **11**, 1170–1176 (2014).
94. Nagy, A., Rossant, J., Nagy, R., Abramow-Newerly, W. & Roder, J. C. Derivation of completely cell culture-derived mice from early-passage embryonic stem cells. *Proc. Natl Acad. Sci USA.* **90**, 8424–8428 (1993).
95. Eggan, K. et al. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl Acad. Sci. USA* **98**, 6209–6214 (2001).
96. Zheng, Y., Thomas, P. M. & Kelleher, N. L. Measurement of acetylation turnover at distinct lysines in human histones identifies long-lived acetylation sites. *Nat. Commun.* **4**, 2203 (2013).
97. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
98. Trapnell, C. et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* **32**, 381–386 (2014).
99. Haghverdi, L., Büttner, M., Wolf, F. A., Buettner, F. & Theis, F. J. Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* **13**, 845–848 (2016).
100. Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat. Protoc.* **15**, 3264–3283 (2020).
101. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
102. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).
103. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
104. Anders, S. et al. Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. *Nat. Protoc.* **8**, 1765–1786 (2013).
105. Ashburner, M. et al. Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).
106. Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* **184**, 3573–3587.e29 (2021).
107. Hutchins, A. P. et al. Models of global gene expression define major domains of cell type and tissue identity. *Nucleic Acids Res.* **45**, 2354–2367 (2017).
108. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
109. MacLean, B. et al. Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. *Bioinformatics* **26**, 966–968 (2010).
110. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
111. John, S. et al. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43**, 264–268 (2011).
112. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
113. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576–589 (2010).
114. Ramírez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* **44**, W160–W165 (2016).
115. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
116. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics J. Integr. Biol.* **16**, 284–287 (2012).
117. Yu, G., Wang, L.-G. & He, Q.-Y. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* **31**, 2382–2383 (2015).
118. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinf.* **10**, 232 (2009).
119. Sun, D. et al. MOABS: model based analysis of bisulfite sequencing data. *Genome Biol.* **15**, R38 (2014).
120. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res.* **42**, e69 (2014).

Acknowledgements

We thank members of the Hochedlinger laboratory for their suggestions, and members of the MGH CRM/HSCI Flow Core, the Harvard Genome Modification Facility and the MGH Next Generation Sequencing Core for their support. We also thank A. Collier for suggestions and discussions, and H. Kretzmer for her help with DNA methylation data. Support from the Dana-Farber Cancer Institute's Center for Cancer Evolution and Physical Sciences-Oncology Center (U54CA193461; to F.M.) is gratefully acknowledged. M.S.H. was supported by the German Cancer Aid (Deutsche Krebshilfe). K.P. is supported by the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA, the David Geffen School of Medicine, the National Institutes of Health (GM099134) and a Faculty Scholar grant from the Howard Hughes Medical Institute. K.H. was supported by

funds from the MGH, the National Institutes of Health (R01 HD058013 and P01 GM099134), the Milky Way Research Foundation and the Gerald and Darlene Jordan Chair in Regenerative Medicine.

Author contributions

M.S.H. and K.H. conceived the study and wrote the manuscript. M.S.H., M.Y., B.D.S., A.J.H. and W.D. performed experiments and analysed the data. S.C. and F.M. analysed single-cell RNA-seq data. A.M. and J.C. performed RRBS. A.M. and C.H. performed WGBS. A.M., J.C., C.H., M.S.H., J.L. and K.P. performed DNA methylation analysis. R.I.S., L.P.W., J.L., K.P. and M.S.H. performed additional bioinformatics analyses.

Competing interests

F.M. is a co-founder of and has equity in Harbinger Health, has equity in Zephyr AI, and serves as a consultant for Harbinger Health, Zephyr AI and Red Cell Partners. F.M. declares that none of these relationships are directly or indirectly related to the content of this manuscript. The remaining authors declare no competing interests.

Additional information

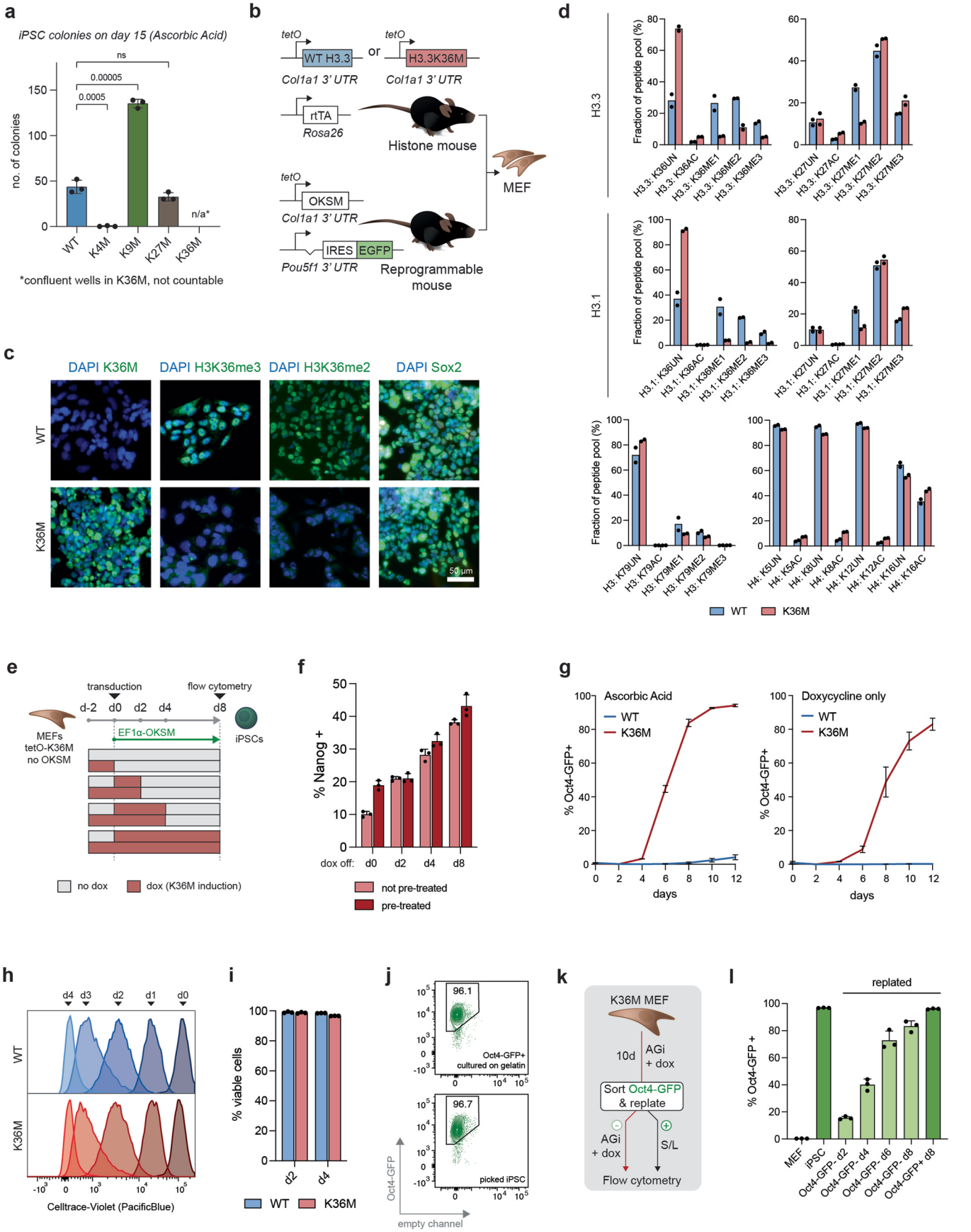
Extended data is available for this paper at https://doi.org/10.1038/s41556-023-01191-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41556-023-01191-z.

Correspondence and requests for materials should be addressed to Konrad Hochedlinger.

Peer review information *Nature Cell Biology* thanks Samantha Morris and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

https://doi.org/10.1038/s41556-023-01191-z

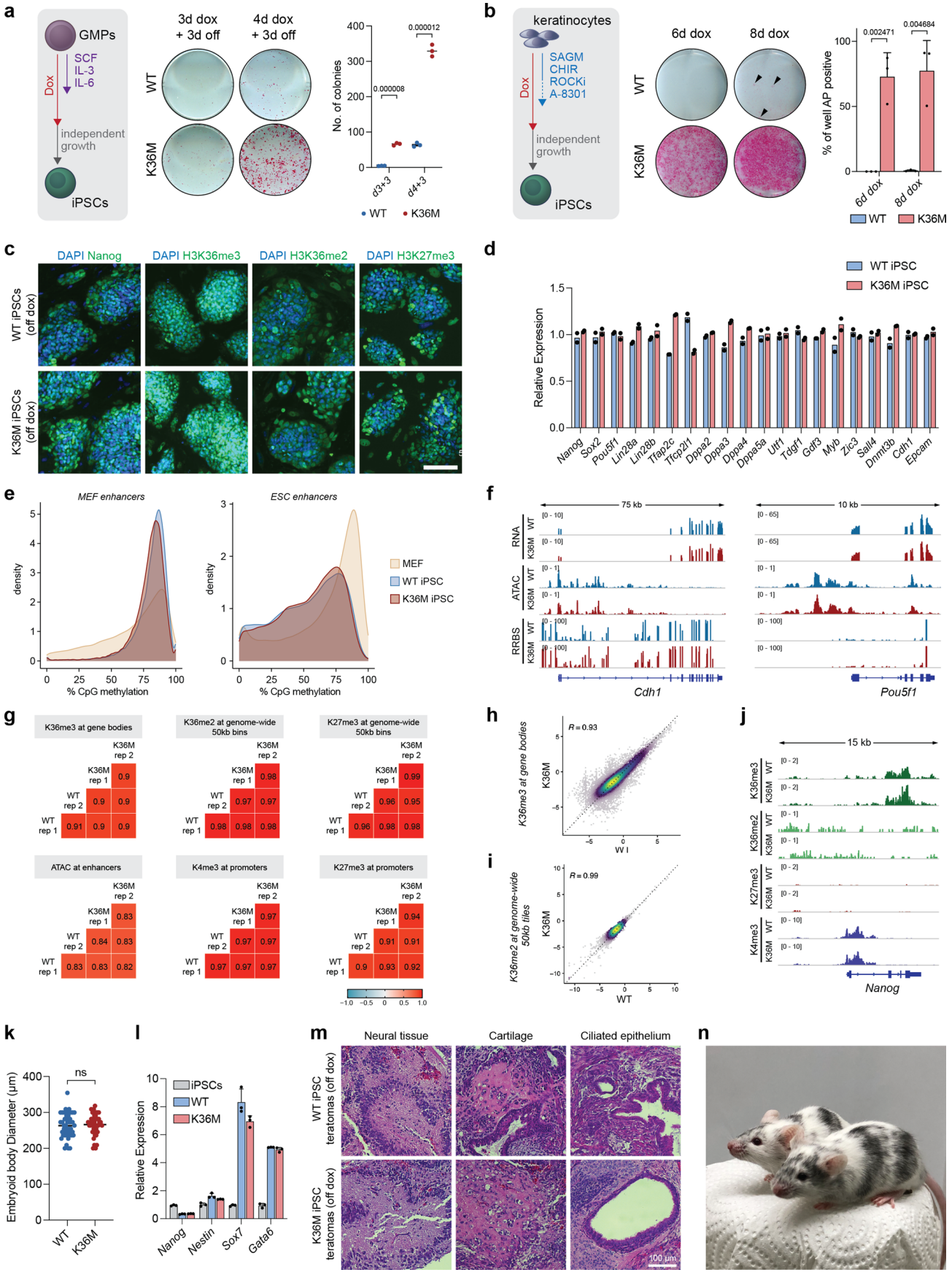


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s41556-023-01191-z

Extended Data Fig. 1 | Key role of H3K36-methylation in cell identity maintenance. (**a**) Colony counts for AP staining of reprogrammable MEFs transduced as indicated (Fig. 1b,c). K36M wells were confluent and could not be counted. P values were determined by two-sided unpaired Student's t test, error bars indicate mean ± SD (n = 3 biologically independent experiments). (**b**) Mice with dox-inducible alleles of WT H3.3 or K36M in the *Col1a1* locus were crossed with mice harboring a dox-inducible OKSM cassette in the same locus, and an EGFP reporter in the 3'UTR of *Pou5f1*. (**c**) Immunofluorescence of MEFs derived as in a). Scale bar 50 µm. (**d**) Mass spectrometry of histone modifications in day 4 reprogramming intermediates (n = 2 independent biological experiments). (**e**) MEFs without endogenous OKSM but with inducible H3.3 WT or K36M were doxycycline-treated for 2 days, then replated and transduced with constitutive OKSM virus. Doxycycline was added to the medium for the indicated intervals, intracellular flow cytometry for Nanog was performed on day 8. (**f**) Quantification of Nanog positivity by flow cytometry on day 8 in K36M cells treated with doxycycline for the indicated time. Cells were either not pre-treated or pre-treated with doxycycline 2 days prior to initiation of reprogramming (see k). Error bars indicate mean ± SD (n = 3 independent biological experiments). (**g**) Fraction of Oct4-GFP$^+$ cells during reprogramming in FBS/LIF medium supplemented with ascorbic acid (left) and without supplementation (right), error bars indicate mean ± SD (n = 3 independent biological experiments). (**h**) Membrane dye dilution assay for reprogramming cultures. (**i**) Percentage of viable cells as assessed by Annexin V/PI negativity on day 2 and 4 of reprogramming. Error bars indicate mean ± SD (n = 3 independent biological experiments). (**j**) Fraction of Oct4-GFP$^+$ cells after sorting of positive cells and expansion on gelatin (top), and in picked iPSCs passaged on feeders (bottom). (**k**) Day 10 K36M reprogramming cultures were sorted by Oct4-GFP reporter positivity. Positive cells were maintained in FBS/LIF, negative cells underwent continued reprogramming in AGi medium. (**l**) Quantification of Oct4-GFP$^+$ cells by flow cytometry in K36M cells sorted by Oct4-GFP reporter status (see i), error bars indicate mean ± SD (n = 3 independent biological experiments).

https://doi.org/10.1038/s41556-023-01191-z

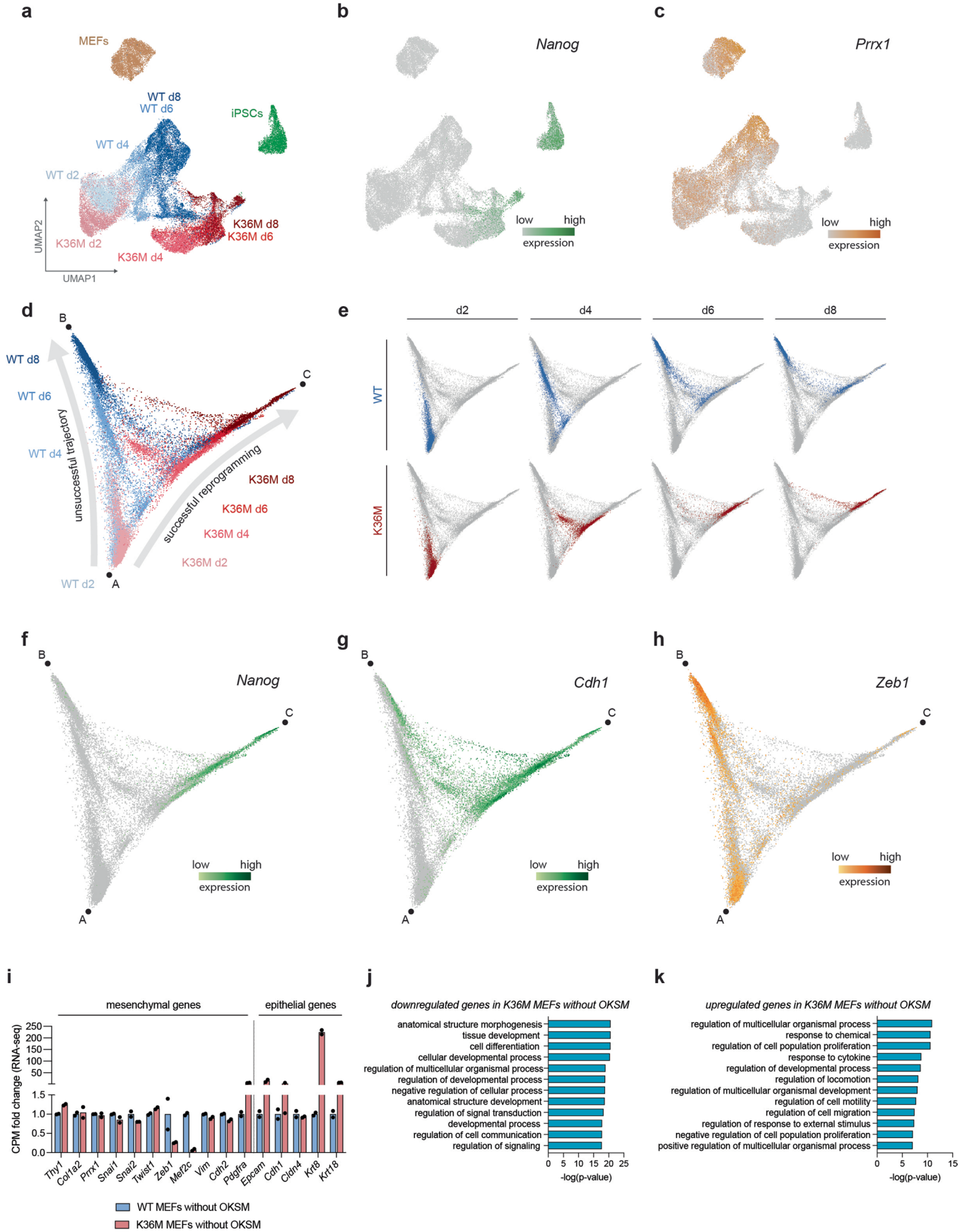


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s41556-023-01191-z

Extended Data Fig. 2 | K36M enhances the reprogramming of different cell types and generates iPSCs highly similar to control iPSCs. (**a**) Reprogramming of GMPs to iPSCs. Alkaline phosphatase staining of iPSC colonies at the indicated timepoints. Quantification of colony counts. P values were determined by two-sided unpaired Student's t test, n = 3 biologically independent experiments. (**b**) Reprogramming of keratinocytes to iPSCs. Alkaline phosphatase staining on day 13 of cells cultured for the indicated timeframes. Area percentage of well that is AP positive. P values were determined by two-sided unpaired Student's t test, error bars indicate mean ± SD (n = 3 independent biological experiments). (**c**) Immunofluorescence for Nanog, H3K36me3, H3K36me2, and H3K27me3 of passaged iPSC cultures from WT and K36M backgrounds cultured without doxycycline on irradiated feeders. (**d**) Relative expression (RNA-seq) of key pluripotency genes in passaged iPSCs of both backgrounds, n = 2 biologically independent experiments. (**e**) DNA methylation at MEF (n = 63,696) and ESC enhancers (n = 72,638) for MEFs and passaged iPSCs of both backgrounds (RRBS, n = 2 biologically independent uninduced MEF replicates were integrated, one replicate each for uninduced iPSCs of each genotype). (**f**) Representative gene tracks showing RNA-seq, ATAC-seq, and RRBS data for *Cdh1* and *Pou5f1* in passaged iPSCs of both backgrounds. (**g**) Correlation matrices for key histone modifications and chromatin accessibility in passaged WT and K36M iPSCs (CUT&Tag and ATAC-seq, n = 2 biologically independent replicates were integrated for this analysis). (**h**) Correlation of K36M vs. WT derived iPSCs for H3K36me3 over gene bodies (CUT&Tag, n = 2 biologically independent replicates were integrated for this analysis). (**i**) Correlation of K36M vs. WT derived iPSCs for H3K36me2 in genome-wide 50 kb bins (CUT&Tag, n = 2 biologically independent replicates were integrated for this analysis). (**j**) Representative gene tracks showing H3K36me3, H3K36me2, H3K27me3, and H3K4me3 at pluripotency gene *Nanog* (n = 2 biologically independent replicates). (**k**) Quantification of the diameter of embryoid bodies from passaged iPSCs of both backgrounds (without doxycycline). P value determined by two-sided unpaired Student's t test, n = 49 for WT, n = 44 for K36M. (**l**) qPCR for *Nanog*, *Nestin*, *Sox7*, and *Gata6* in embryoid bodies derived from passaged iPSCs of both backgrounds (without doxycycline), error bars indicate mean ± SD (n = 3 independent biological experiments). (**m**) H&E sections of teratomas generated with iPSCs of both backgrounds (without doxycycline). 4/4 WT and 6/6 K36M iPSC lines produced well-differentiated teratomas. Images depict tissue-like structures of all germ layers. (**n**) Coat-color chimeras generated by blastocyst injection of K36M iPSCs.

https://doi.org/10.1038/s41556-023-01191-z

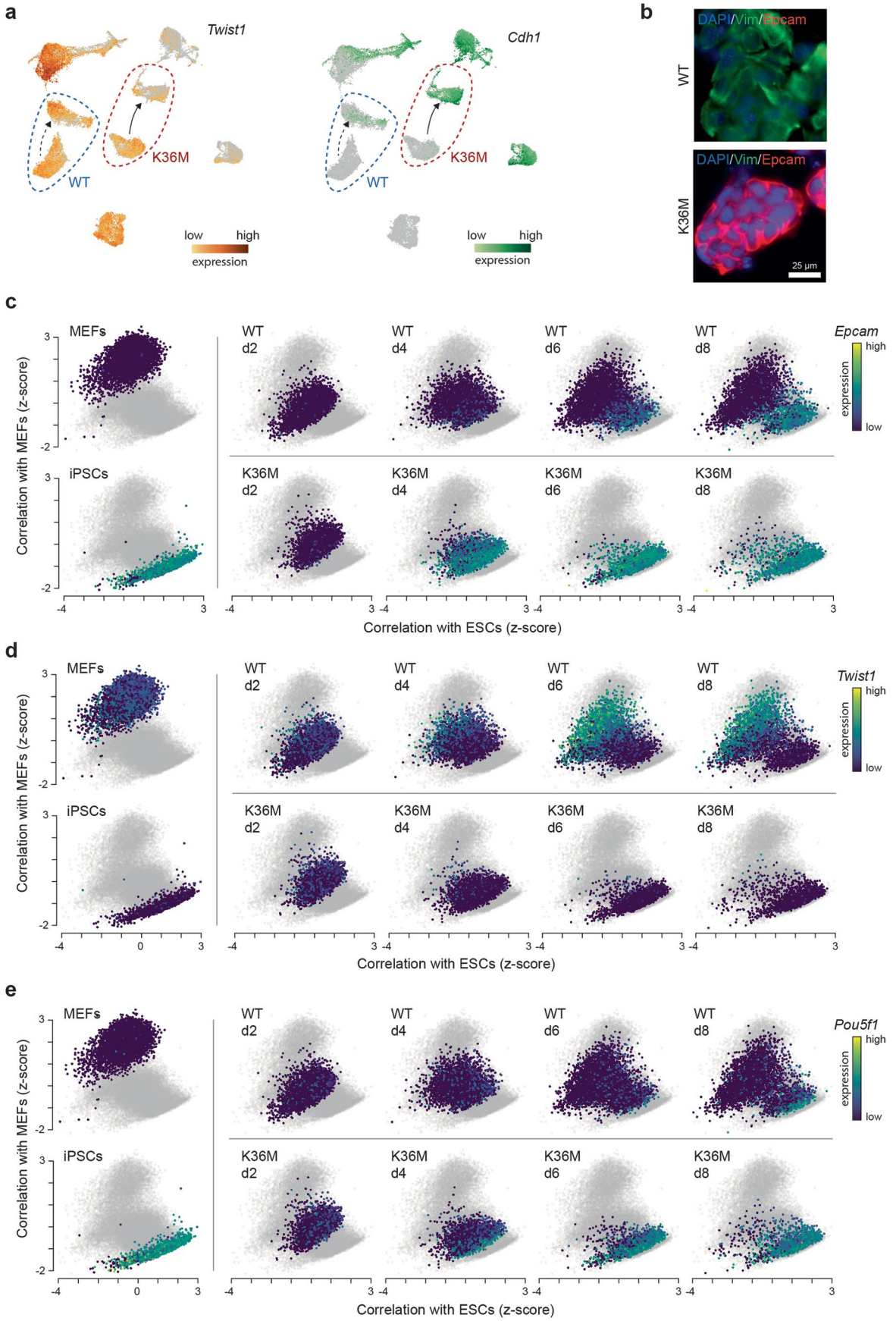


Extended Data Fig. 3 | See next page for caption.

https://doi.org/10.1038/s41556-023-01191-z

Extended Data Fig. 3 | Single-cell RNA-seq reveals main trajectories of WT and K36M reprogramming intermediates. (**a**) UMAP embedding of scRNA-seq data (Seurat framework) using MEFs, reprogramming intermediates on days 2, 4, 6, 8 for WT and K36M, as well as passaged iPSCs. For each indicated group, one sample was encapsulated leading to n = 38,743 cells total. (**b**) Expression of pluripotency gene *Nanog* projected on the same UMAP embedding as in (a). (**c**) Expression of mesenchymal gene *Prrx1* projected on the same UMAP embedding as in (a). (**d,e**) Diffusion pseudotime mapping of day 2 to day 8 intermediates undergoing reprogramming. WT cells are colored in blue, K36M cells in red. (**f**) Expression of pluripotency gene *Nanog* projected on the same pseudotime embedding as in (**d**). (**g**) Expression of epithelial gene *Cdh1* projected on the same pseudotime embedding as in (d). (**h**) Expression of mesenchymal gene *Zeb1* projected on the same pseudotime embedding as in (d). (**i**) Relative expression (RNA-seq) of mesenchymal and epithelial genes in MEFs expressing H3.3 WT or K36M, but not OKSM, n = 2 biologically independent experiments. (**j**) Gene ontology terms of genes downregulated in K36M MEFs without OKSM. Analysis and p values from geneontology.org. (**k**) Gene ontology terms of genes upregulated in K36M MEFs without OKSM. Analysis and p values from geneontology.org.

https://doi.org/10.1038/s41556-023-01191-z

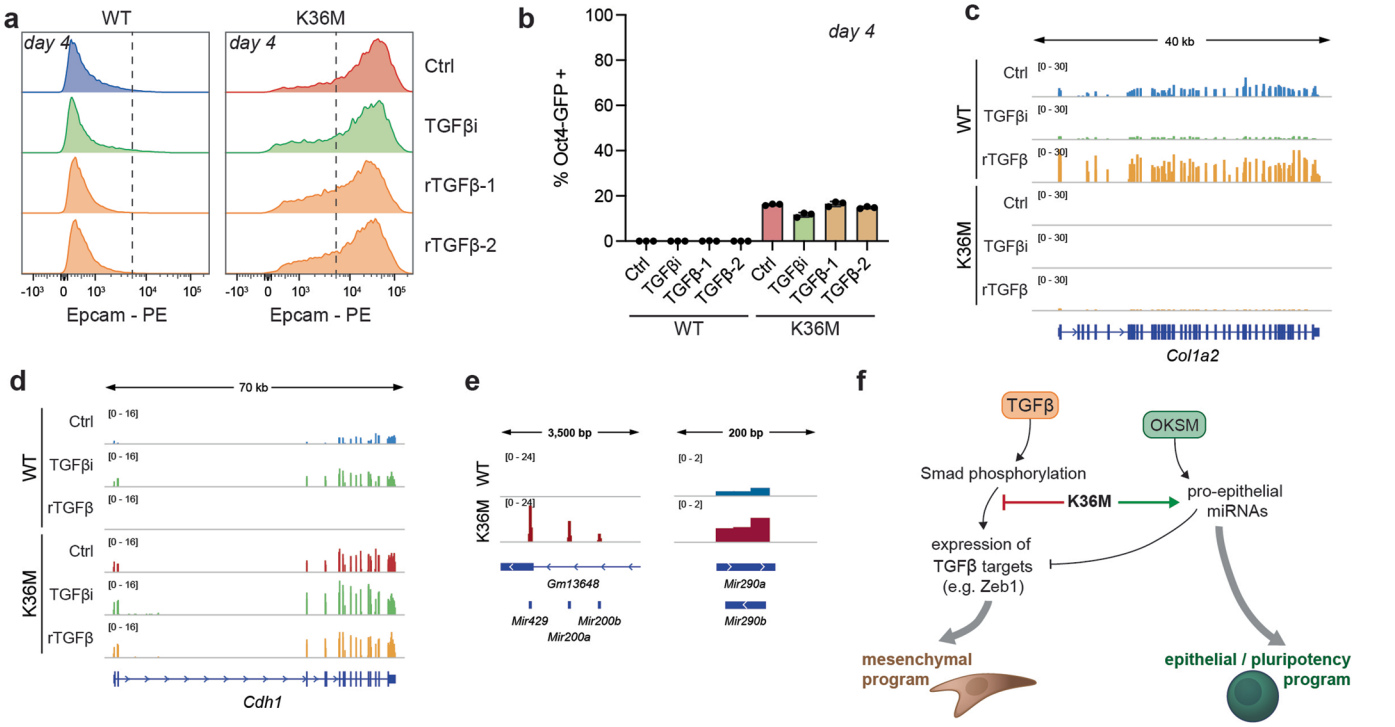


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s41556-023-01191-z

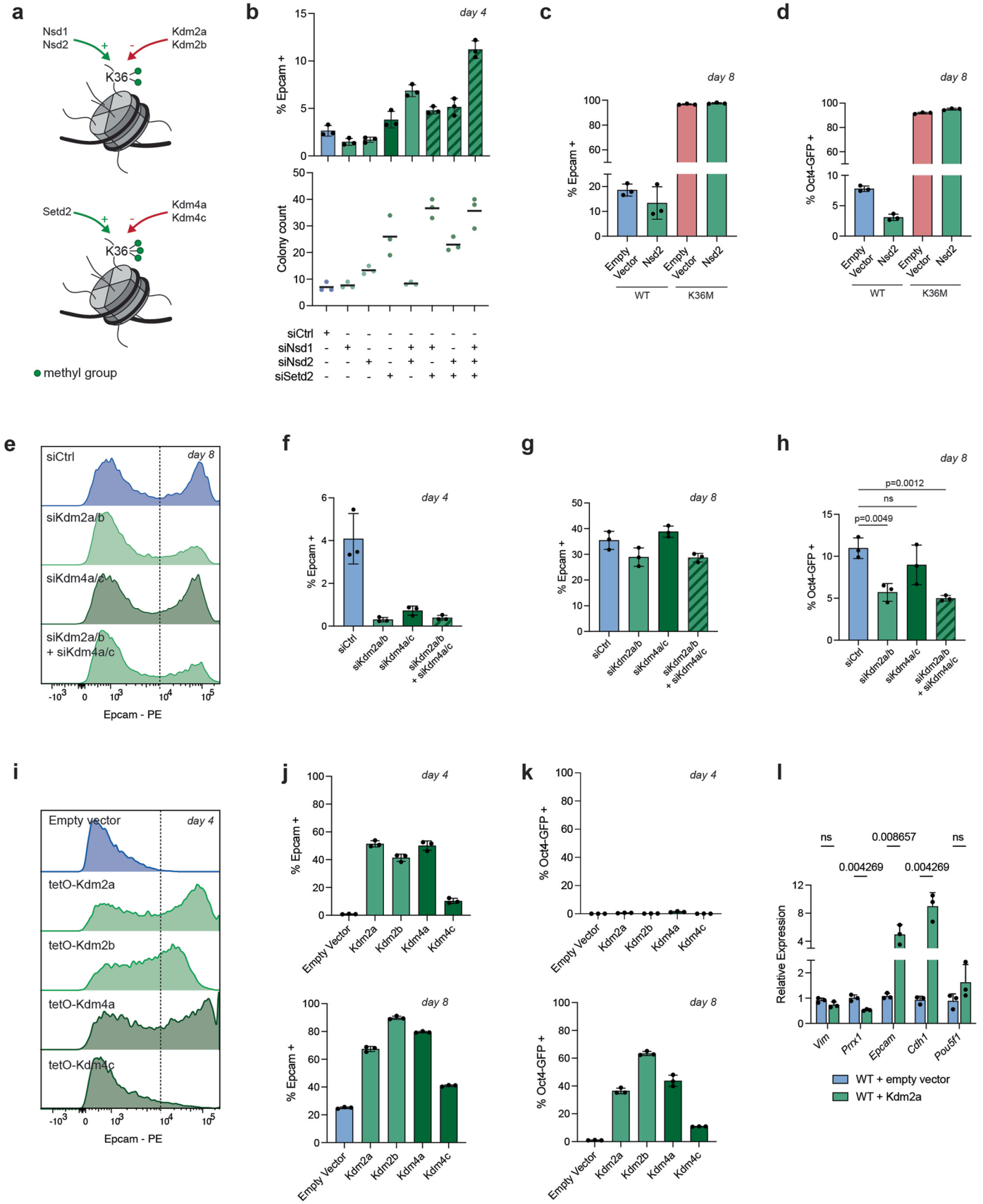
Extended Data Fig. 4 | K36M confers epithelial plasticity on cells undergoing reprogramming. (**a**) Gene expression of *Twist1* and *Cdh1* on the same UMAP embedding as used in Fig. 2d. Dashed circles encompass day-2 and day-4 samples for WT (blue circle) and K36M (red circle). Solid arrow indicates switch-like MET in K36M samples, dashed arrow indicates heterogeneous maintenance/activation of mesenchymal/epithelial programs in WT cells. For each group, one sample was encapsulated leading to n = 38,743 cells total. (**b**) Immunofluorescence for Vimentin and Epcam in WT and K36M cells on day 4 of reprogramming. Scale bar = 25 µm. Three independent biological experiments with similar results. (**c**) Correlation plots of single-cell RNA-seq data comparing transcriptional programs within each cell to MEFs (y-axis) and ESCs (x-axis)[107]. For each sample, the corresponding cells are colored according to their *Epcam* expression levels, whereas other cells are greyed out. (**d**) Correlation plots as in (c), expression data of *Twist1* is superimposed. (**e**) Correlation plots as in (c), expression data of *Pou5f1* is superimposed.

https://doi.org/10.1038/s41556-023-01191-z



Extended Data Fig. 5 | K36M disrupts TGFβ signaling and modulates epithelial plasticity in diverse contexts. (**a**) Flow cytometry histograms displaying Epcam expression in day-4 reprogramming intermediates for WT and K36M samples. Untreated control cells are compared to cells treated with 250 nM Repsox (TGFβi) or 2.5 ng/ml recombinant TGFβ-1 or -2 (rTGFβ-1, rTGFβ-2). (**b**) Fraction of Oct4-GFP$^+$ cells treated with TGFβi or rTGFβ in day 4 reprogramming intermediates. Error bars indicate mean ± SD (n = 3). (**c**) Representative tracks for expression of mesenchymal gene *Col1a2* on day 8 of reprogramming, WT or K36M cells were treated as indicated. (**d**) Representative tracks for expression of epithelial gene *Cdh1* on day 8 of reprogramming, WT or K36M cells were treated as indicated. (**e**) Representative tracks for *miR-200a* and *miR-290*. (**f**) Schematic of K36M's effect on TGFβ signaling and miRNA expression during reprogramming. (**g**) De-differentiation of MEFs to induced myogenic progenitor cells (iMPCs). qRT-PCR for myotube marker *Myh1* and iMPC marker *Pax7*, P values were determined by two-sided unpaired Student's t test, error bars indicate mean ± SD (n = 3 biologically independent experiments). Flow cytometry for Pax7-GFP reporter positive cells. (**h,i**) Immunofluorescence of Nanog and K36M in ES cells of both backgrounds, cultured in S/L (g) or 2iL (h) conditions. Result is representative of three independent biological experiments. (**j**) Differentiation of ESC aggregates to elongated gastruloids. Representative brightfield images (scale bar = 500 μm) and quantification of long axis diameter (line = mean). P value was determined by two-sided unpaired Student's t test, n = 19 for WT, n = 20 for K36M. (**k**) Differentiation of ESCs to pre-somitic mesoderm. Representative immunofluorescence for Cdh2 (scale bar = 50 μm). qRT-PCR for mesodermal transcription factors *Tbx6* and *Msgn1*. P values were determined by two-sided unpaired Student's t test, error bars indicate mean ± SD (n = 3 independent biological experiments).

https://doi.org/10.1038/s41556-023-01191-z

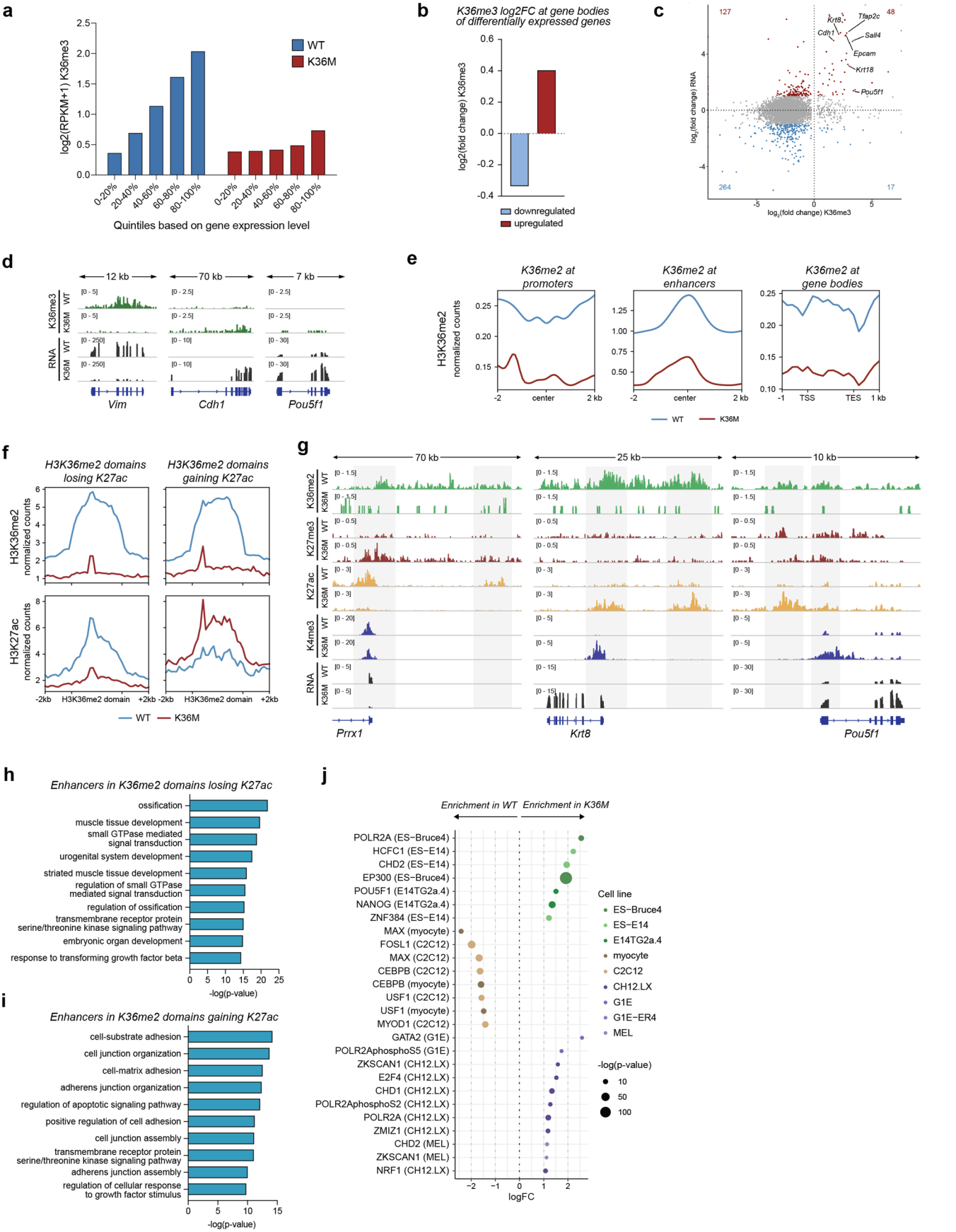


Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s41556-023-01191-z

Extended Data Fig. 6 | H3K36me2 and H3K36me3 cooperatively safeguard cell identity. (**a**) Histone methyltransferases and demethylases implicated in the regulation of H3K36me2 (top) and H3K36me3 (bottom). (**b**) Fraction of Epcam$^+$ cells on day 4 of reprogramming (top) in WT cells with knockdown of indicated histone methyltransferases. Colony counts after 6 days of doxycycline followed by 4 days of independent growth (bottom), error bars indicate mean ± SD (n = 3 biologically independent experiments). (**c,d**) Fraction of Epcam$^+$ (c) or Oct4-GFP$^+$ (d) cells on day 8 of reprogramming in WT and K36M cells transduced with either empty vector or dox-inducible *Nsd2*, error bars indicate mean ± SD (n = 3 biologically independent experiments). (**e**) Flow cytometry for Epcam on day 8 of reprogramming in WT cells with knockdown of the indicated histone demethylases. (**f,g**) Fraction of Epcam$^+$ cells on day 4 (f) and day 8 (g) of reprogramming in WT cells with knockdown of the indicated histone demethylases, error bars indicate mean ± SD (n = 3 biologically independent experiments). (**h**) Fraction of Oct4-GFP$^+$ cells on day 8 of reprogramming in WT cells with knockdown of the indicated histone demethylases. P values were determined by unpaired Student's t test, error bars indicate mean ± SD (n = 3 biologically independent experiments). (**i,j**) Flow cytometry for Epcam during reprogramming in WT cells with overexpression of the indicated histone demethylases, error bars indicate mean ± SD (n = 3 biologically independent experiments). (**k**) Fraction of Oct4-GFP$^+$ cells on day 4 and day 8 of reprogramming in WT cells with overexpression of the indicated histone demethylases, error bars indicate mean ± SD (n = 3 biologically independent experiments). (**l**) qRT-PCR for mesenchymal genes *Vim* and *Prrx1*, epithelial genes *Epcam* and *Cdh1*, and pluripotency gene *Pou5f1* on day 4 of reprogramming in WT cells overexpressing *Kdm2a* vs. empty vector control. P values were determined by unpaired Student's t test, error bars indicate mean ± SD (n = 3 independent biological experiments).

https://doi.org/10.1038/s41556-023-01191-z

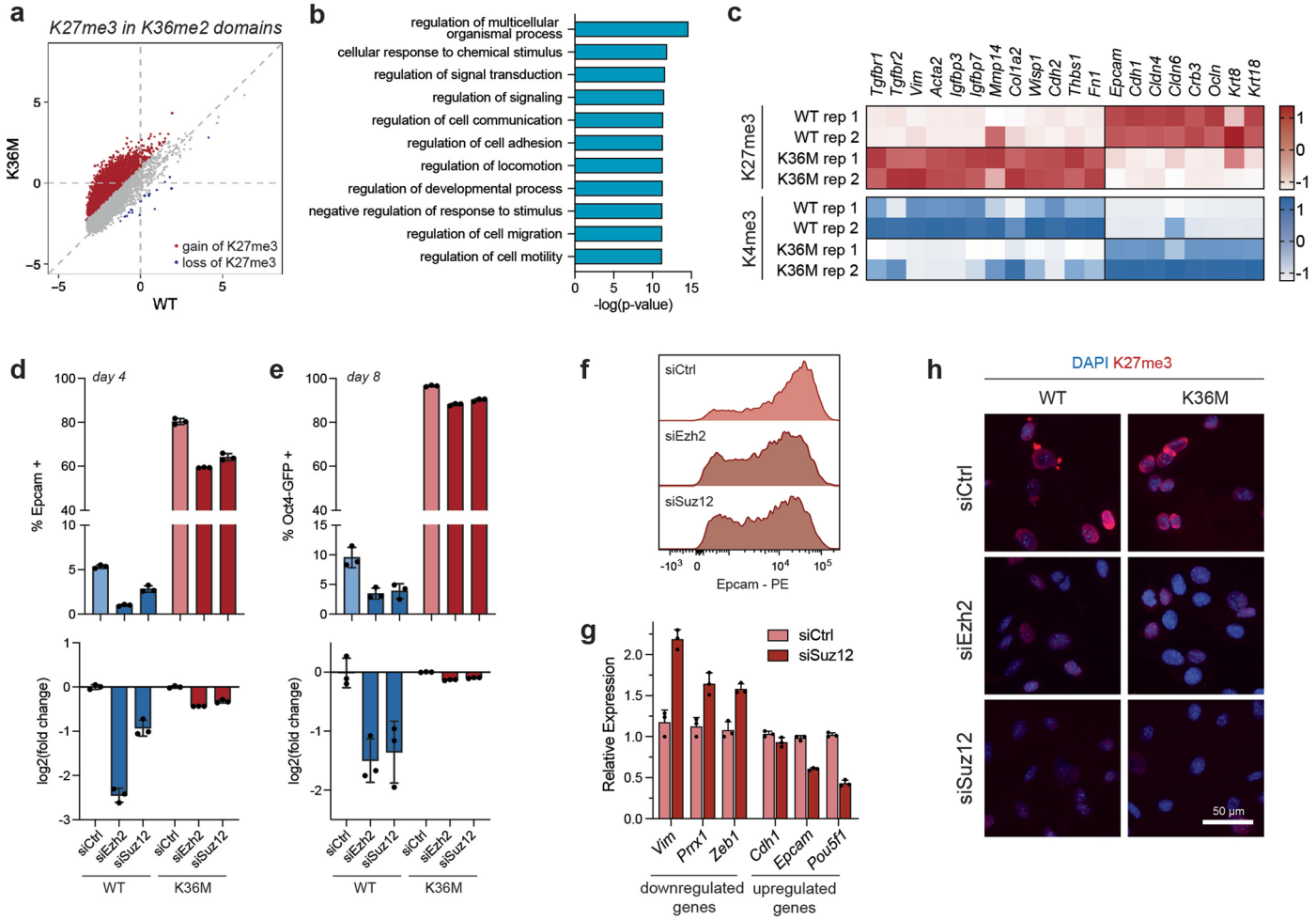


Extended Data Fig. 7 | See next page for caption.

https://doi.org/10.1038/s41556-023-01191-z

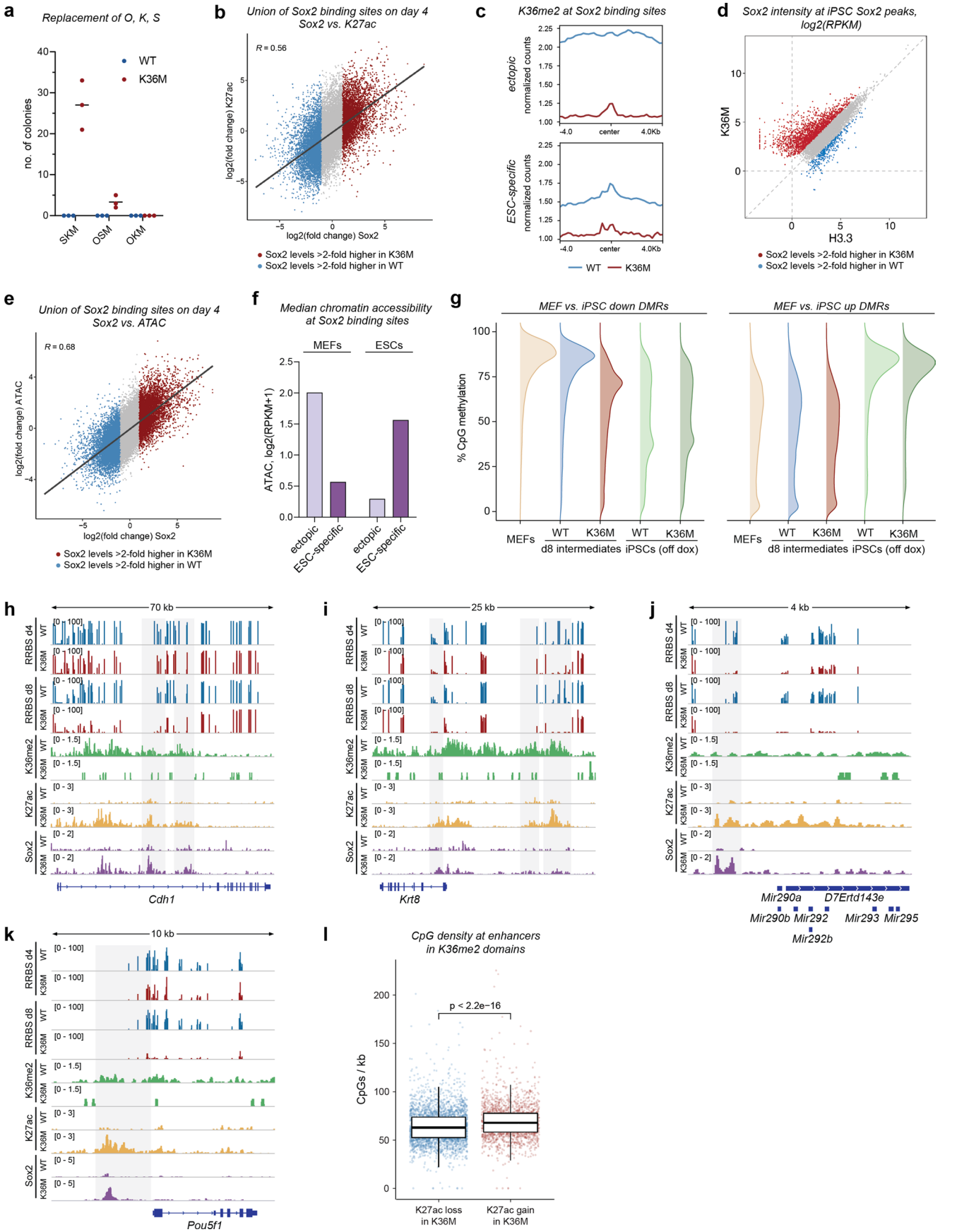
Extended Data Fig. 7 | K36M alters H3K36me2/3 deposition and gene expression. (**a**) Median H3K36me3 deposition at expressed genes (RPKM > 0.1, n = 10,251 for WT, n = 10,496 for K36M) of indicated expression quintiles for WT (left) and K36M (right) samples on day 4 of reprogramming (n = 2 biologically independent replicates were integrated for this analysis). (**b**) Median fold change of H3K36me3 density over gene bodies of differentially expressed genes (upregulated = red, downregulated = blue, n = 1,872) on day 4 (n = 2 biologically independent replicates were integrated for this analysis). (**c**) Fold change of gene expression (y axis) vs. fold change of H3K36me3 (x axis) between K36M and WT samples on day 4 of reprogramming (n = 2 biologically independent replicates were integrated for this analysis). (**d**) Representative gene tracks for H3K36me3 and RNA levels at mesenchymal gene *Vim*, epithelial gene *Cdh1*, and pluripotency gene *Pou5f1* (n = 2 biologically independent replicates). (**e**) Profile plots of H3K36me2 at promoters, enhancers, and gene bodies (n = 2 biologically independent replicates were integrated for this analysis). (**f**) Profile plots of H3K36me2 and H3K27ac at H3K36me2 domains containing down- or upregulated enhancers (n = 2 biologically independent replicates were integrated for this analysis). (**g**) Representative tracks for *Prrx1*, a mesenchymal gene downregulated in K36M samples on day 4, *Krt8*, and *Pou5f1*, epithelial/pluripotency genes upregulated in K36M samples on day 4 (n = 2 biologically independent replicates). Putative regulatory elements highlighted in grey. (**h**) Gene ontology terms of genes closest to H3K36me2 domain-embedded enhancers that are differentially downregulated in K36M cells. Analysis and p values from geneontology.org. (**i**) Gene ontology terms of genes closest to H3K36me2 domain-embedded enhancers that are differentially upregulated in K36M cells. Analysis and p values from geneontology.org. (**j**) Dot plot representing enrichment of ENCODE data for differentially active enhancers within H3K36me2 domains. P values were determined by Fisher's exact test.

https://doi.org/10.1038/s41556-023-01191-z



Extended Data Fig. 8 | PRC2 contributes to the K36M-dependent silencing of the somatic program. (**a**) H3K27me3 deposition within H3K36me2 domains in K36M vs. WT samples in day-4 reprogramming intermediates. Domains gaining H3K27me3 are colored in red, domains losing H3K27me3 are colored in blue. (**b**) Ontology terms for genes with promoters overlapping H3K36me2 domains and gaining H3K27me3. Analysis and p values from geneontology.org. (**c**) Heatmaps for H3K27me3 and H3K4me3 at promoters mesenchymal and epithelial genes in WT and K36M samples. (**d,e**) Fraction of Epcam$^+$ and Oct4-GFP$^+$ cells in WT (blue) and K36M (red) samples with knockdown of indicated PRC2 components (top). Log2(fold change) of fraction normalized to control siRNA (bottom). Error bars indicate mean ± SD (n = 3 independent biological experiments). (**f**) Representative histograms of flow cytometry for Epcam in K36M cells with control siRNA and knockdown of *Ezh2* or *Suz12*. (**g**) qRT-PCR for mesenchymal (*Vim*, *Prrx1*, *Zeb1*), epithelial (*Cdh1*, *Epcam*), and pluripotency (*Pou5f1*) marker genes, error bars indicate mean ± SD (n = 3 biologically independent experiments). (**h**) Immunofluorescence for H3K27me3 in WT and K36M cells transfected with control siRNA or knockdown of *Ezh2* or *Suz12*. Representative result from three independent biological experiments.

https://doi.org/10.1038/s41556-023-01191-z

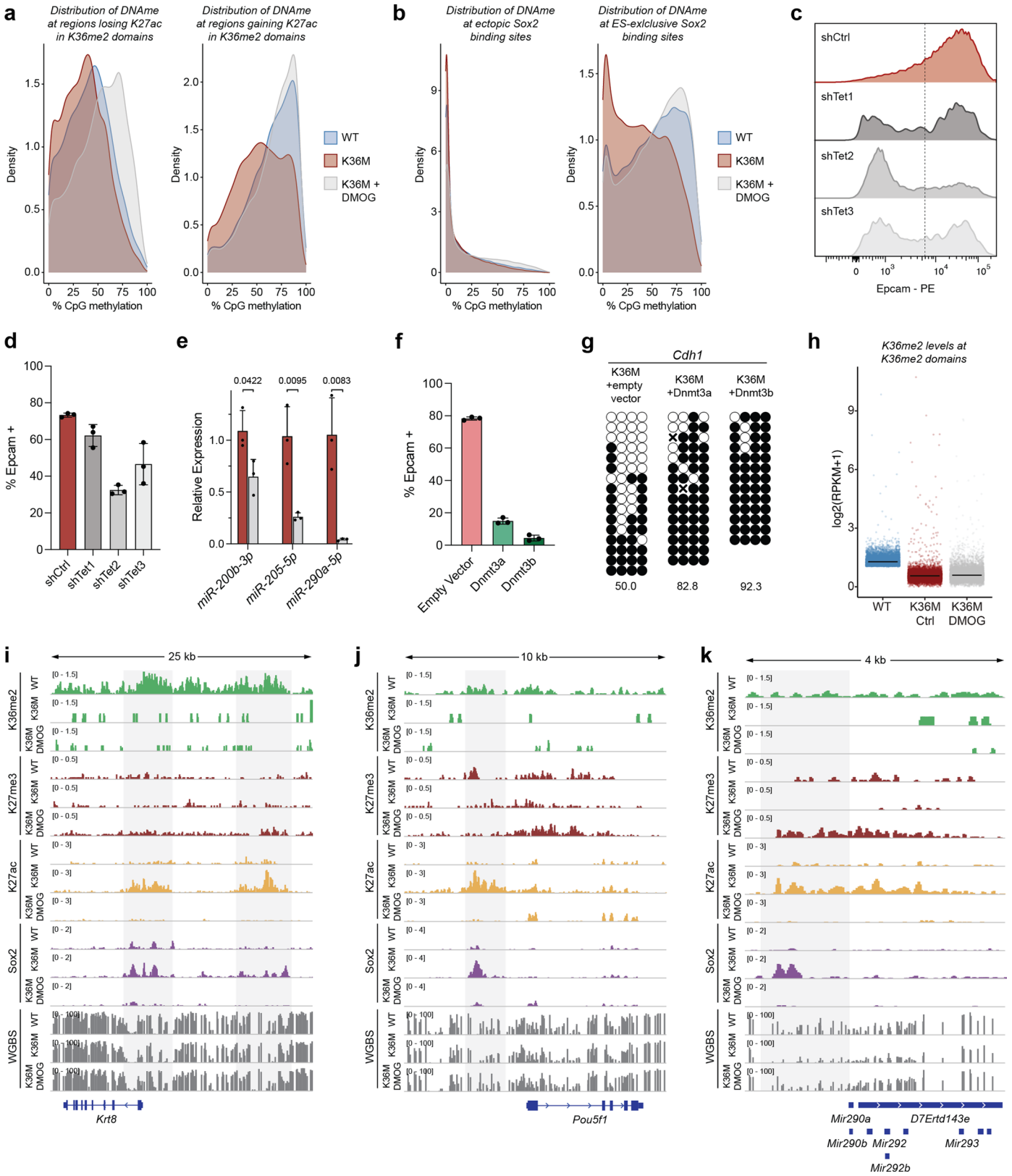


Extended Data Fig. 9 | See next page for caption.

https://doi.org/10.1038/s41556-023-01191-z

Extended Data Fig. 9 | K36M rewires DNA methylation patterns. (**a**) Number of colonies following alkaline phosphatase staining of WT and K36M cell cultures transduced with non-selectable, dox-inducible lentiviruses for the expression of SKM, OSM, or OKM. Cultures were induced for 12 days and stained on day 15, n = 3 independent biological experiments. (**b**) Correlation plot of log2(fold-change) differences (K36M vs. WT) at Sox2 peaks called in WT and K36M samples. Differences of Sox2 enrichment are correlated with differences in H3K27ac abundance (n = 2 biologically independent replicates were integrated for this analysis). Pearson's R = 0.56. (**c**) Profile plots showing H3K36me2 abundance at ectopic and ESC-specific Sox2 sides in WT and K36M cells on day 4 of reprogramming. (**d**) Sox2 enrichment at Sox2 binding sites as defined in iPSCs, log2(RPKM). (**e**) Correlation plot of log2(fold-change) differences (K36M vs. WT) at Sox2 peaks called in WT and K36M samples. Differences of Sox2 enrichment are correlated with differences in chromatin accessibility (as measured by ATAC-seq, n = 2 biologically independent replicates were integrated for this analysis). Pearson's R = 0.68. (**f**) Median chromatin accessibility (as measured by ATAC-seq, n = 2 biologically independent replicates were integrated for this analysis) at ectopic and ESC-specific Sox2 binding sites in MEFs and passaged iPSCs. (**g**) Density ridge plots of DNA methylation at differentially methylated regions losing (left, n = 30,294) or gaining (right, n = 28,060) methylation in iPSCs vs. MEFs (n = 2 biologically independent replicates for uninduced MEFs, for d4 and d8 samples were integrated, one sample for each genotype in uninduced iPSCs). (**h**-**k**) Representative gene tracks of *Cdh1*, *Krt8*, the *miR-290* cluster, and *Pou5f1* (n = 2 biologically independent replicates). Putative regulatory elements affected by DNA demethylation are highlighted in grey. (**l**) CpG density at differentially active enhancers in H3K36me2 domains (n = 4,939). P value determined by two-sided Wilcoxon rank sum test. Center line indicates median; lower/upper hinges indicate 25^{th}/75^{th} percentiles; whiskers extend to 1.5x IQR. P value determined by two-sided Wilcoxon rank sum test.

https://doi.org/10.1038/s41556-023-01191-z



Extended Data Fig. 10 | See next page for caption.

https://doi.org/10.1038/s41556-023-01191-z

Extended Data Fig. 10 | DNA demethylation is limiting for K36M-dependent enhancer rewiring. (**a**) Distribution of DNA methylation (WGBS, n = 2 biologically independent replicates were integrated for this analysis) at differentially active enhancers in H3K36me2 domains (n = 4,939) in day-4 reprogramming intermediates of WT, K36M and K36M + DMOG samples. (**b**) Distribution of DNA methylation (WGBS, n = 2 biologically independent replicates were integrated for this analysis) at ectopic (n = 45,095) and ESC-exclusive (n = 27,708) Sox2 binding sites in day-4 reprogramming intermediates of WT, K36M and K36M + DMOG samples. (**c**) Representative histogram plots from flow cytometric analysis for Epcam of K36M cells with knockdown of the indicated *Tet* demethylases. (**d**) Fraction of Epcam$^+$ cells in K36M cells with *Tet* knockdown on day 4 of reprogramming, error bars indicate mean ± SD (n = 3 biologically independent experiments). (**e**) qPCR of miRNAs *miR-200b-3p*, *miR-205-5p*, and *miR-290a-5p* in untreated (K36M Ctrl) and DMOG-treated K36M cells (K36M DMOG). P values were determined by unpaired Student's t test, error bars indicate mean ± SD (n = 3 independent biological experiments). (**f**) Fraction of Epcam$^+$ cells in K36M cultures transduced with either an empty vector or dox-inducible overexpression vectors for *Dnmt3a* and *Dnmt3b*, error bars indicate mean ± SD (n = 3 independent biological experiments). (**g**) Bisulfite-seq of a *Cdh1* enhancer in K36M cells transduced with either empty vector (left), or overexpression of *Dnmt3a* (middle) or *Dnmt3b* (right). (**h**) H3K36me2 levels within H3K36me2 domains (n = 7,610) on day 4 of reprogramming in WT, untreated K36M cells (K36M Ctrl), and DMOG-treated K36M cells (K36M DMOG). Crossbars indicate median (n = 2 biologically independent replicates were integrated for this analysis). (**i-k**) Representative gene tracks of *Krt8*, *Pou5f1*, and the *miR-290* cluster (n = 2 biologically independent replicates). Putative regulatory elements highlighted in grey.

nature portfolio

Corresponding author(s): Konrad Hochedlinger

Last updated by author(s): Feb 27, 2023

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	For flow cytometry, cells were analyzed on an LSRII flow cytometer (BD Biosciences) using Diva v6.1.2 (BD Biosciences). Quantitative PCR was performed on a Lightcycler 480 system (Roche) using Brilliant III Sybr-Green (Agilent) and cDNA from each sample in triplicate. The efficiencies of each primer set used was found to be >90%, and fold change was calculated using the ΔΔCt method. RNA-seq libraries were constructed using polyA selection followed by NEBNext UltraDirectional kit protocol (New England Biolabs) and sequenced on the Illumina HiSeq2500 instrument, resulting in approximately 30 million reads per sample on average. For single-cell RNA-sequencing, approx. 5000 cells per sample were encapsulated using the 10x Chromium pipeline according to manufacturer guidelines using Single Cell 3' chemistry v3 for embedding, reverse transcription and library preparation (10x Genomics). ATAC-seq libraries were sequenced in paired-end 50 bp mode on the Illumina HiSeq2500 instrument, resulting in approximately 40 million reads per sample on average. CUT&Tag libraries were sequenced in paired-end 50 bp mode on the Illumina HiSeq2500 instrument, resulting in approximately 15 million reads per sample on average. RRBS libraries were sequenced on the Illumina NovaSeq6000, generating 100-bp single-end reads.
Data analysis	For flow cytometry, data analysis was performed using FlowJo software v10.2 (FlowJo, LLC). RNA-seq reads were mapped by STAR v.2.5.0 aligner to mm9 reference genome using ENSEMBL annotation. Read counts for individual genes were produced using the unstranded count feature in HTSeq version 0.11.2.. Differential expression analysis was performed using the edgeR package after normalizing read counts and including only those genes with count per million reads (CPM) > 1 for one or more samples. Differentially expressed genes were defined based on the criteria of >2-fold change in expression value. Sequenced reads were first aligned using CellRanger version 3.0.2, and the filtered reads assigned to cell barcodes were analyzed with the R package Seurat, version 4.0.0. A first round of quality control (input: 45,751 cells) involved all cells positioned at least 3 median absolute deviations below the mean for library size (total number of reads) and the number of detected genes, above the mean for the percentage of mitochondrial genes expressed from the entire transcriptome (40,018 cells remaining). The R package monocle was used for dimensionality reduction (UMAP). Further, 372 single cells corresponding to 3 separate clusters with low QC metrics and indicative of contamination were removed (39,646 cells remaining), and the data was normalized and clustered again. A new low quality cluster consisting of 903 single cells emerged and was removed. The final cleaned dataset consisted of 38,743 cells. The average distance boxplot was generated by computing the average Euclidean distance among all cells in the same condition based on UMAP coordinates. Outliers larger than 7.5 were removed for visual purposes. The correlation scores for MEFs and ESCs (ESCsV6.5) were calculated

using all available genes, correlated with publicly-accessible bulk profiles downloaded from an existing publication. The data was further z-scored. The pseudotime plots were generated on the Seurat-preprocessed data using the R package destiny version 3.4.0. The code used to analyze the scRNAseq data presented here, and to generate the corresponding plots, is uploaded to GitHub https://github.com/Michorlab/H3K36_methylation_scRNAseq. For ATAC-seq, sequencing reads were mapped using bwa v.0.5.9-r16 aligner followed by peak calling using Hotspot. DiffBind R package was used for the analysis of differential accessible regions (DAR), based on the cutoffs of > 2-fold change in peak intensity and false discovery rate (FDR) < 0.01. For CUT&Tag, sequencing reads were mapped to mm9 reference genome using bwa aligner. Peak calling was performed using Homer. Heatmaps and average profiles of CUT&Tag read densities were generated using deepTools v3.5.1. Bedtools v2.30.0 was used to calculate read densities over given regions. Gene ontology analysis was performed using geneontology.org after annotation using ChipSeeker v1.34.1. For RRBS analysis, demultiplexed reads were trimmed using cutadapt and the Nugene diversity adapter trimming python script trimRRBSdiversityAdaptCustomers.py. Alignment to the mouse mm9 genome was performed using BSMAP v2.90 with the following parameters: -v 0.1 -s 12 -q 20 -w 100 -S 1 -u -R -D C-CGG. Duplicate reads were removed using the Nugene deduplication script nudup.py. Methylation ratios were then called with the MOABS mcall module. DMRs were calculated in R software, using package DSS to detect differentially methylated loci. Locations were selected based on a delta of 0.1, and a p-value threshold of 0.001, across two replicates. Using R software, all RRBS samples were filtered to keep only CpGs with at least 5 reads coverage. Data was then combined to perform analysis on only matched CpGs across all samples (n = 2,688,886). For Sox2 binding site analysis, CpGs were intersected with region sets using bedtools, and heatmaps were generated using package 'pheatmap' after calculating average CpG methylation percentage for each binding site in R. Statistical analysis was carried out using R or GraphPad Prism.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

RNA-seq, ATAC-seq, and CUT&Tag data have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE203492. Single-cell RNA-seq data are deposited under GSE203536. RRBS and WGBS data are available under GSE203606. The publicly available datasets used in this study are: GSE90893, GSE111172, GSE77420. All other data supporting the findings of this study are available from the corresponding authors upon reasonable request. mm9 reference genome available under https://www.ncbi.nlm.nih.gov/assembly/GCF_000001635.18/.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical method was used to determine sample size. Sample sizes were chosen similar to previous publications (Bar-Nur O, 2014; Brumbaugh J, 2018; Brumbaugh J, 2019).
Data exclusions	No data were excluded from this manuscript.
Replication	All attempts at reproducing experiments reported here were successful. Reproducibility statements are included in the figure legends or methods section for each figure.
Randomization	Samples were not randomized for this study.
Blinding	No blinding was performed for this study.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☐	☒ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used

DAPI solution (BD Biosciences 564907), 1:1000 IF, 1:1000 FACS
anti-Thy1.2 eFluor 450, 53-2.1 (Thermo Fisher Scientific Cat# 48-0902-82), 1:200 FACS
anti-SSEA-1 eFluor 660, eBioMC-480 (Thermo Fisher Scientific Cat# 50-8813-42), 1:200 FACS
anti-Sca-1 PE-Cy7, 25-5981-82 (Thermo Fisher Scientific Cat# 25-5981-82), 1:200 FACS
anti-EpCAM PE, G8.8 (Thermo Fisher Scientific Cat# 12-5791-82), 1:200 FACS
anti-Histone H3K36M Monoclonal Antibody, RM193 (RevMab 31-1085-00-L), 1:500 IF
anti-H3K36me3 (abcam 9050), 1:500 IF
anti-H3K36me2, C75H12 (Cell Signaling 2901), 1:200 IF
anti-Sox2, D9B8N (Cell Signaling 23064), 1:400 IF, 1:50 CUT&Tag
anti-EpCAM, G8.8 (Life Technologies 14-5791-81), 1:200 IF
anti-Nanog (abcam 80892), 1:500 IF, 1:200 FACS
anti-Ter119 (PE/Cy5; Thermo fisher, Cat# 15-5921-83), 1µl/million cells FACS
anti-CD3e (PE/Cy5; Biolegend, Cat# 100310), 1µl/million cells FACS
anti-Gr1 (PE/Cy5; Thermo fisher, Cat# 15-5931-83), 1µl/million cells FACS
anti-B220/CD45R (PE/Cy5; Biolegend, Cat# 103210), 1µl/million cells FACS
anti-TCRb (PE/Cy5; Thermo fisher, Cat# 15-5961-83), 1µl/million cells FACS
anti-CD4 (PE/Cy5; Thermo fisher, Cat# 15-0041-83), 1µl/million cells FACS
anti-CD8a (PE/Cy5; Biolegend, Cat# 100710), 1µl/million cells FACS
anti-c-Kit (APC-eFluor 780; Thermo fisher, Cat# 47-1171-82), 1µl/million cells FACS
anti-CD34 (FITC; Thermo fisher, Cat# 11-0341-85), 1µl/million cells FACS
anti-CD16/32 (FC gamma receptor) (PE; BD Biosciences, Cat# 553145), 1µl/million cells FACS
anti-Vim, D21H3 (Cell Signaling 5741), 1:500 IF
anti-Tubb3 (Biolegend 801201), 1:400 IF
anti-Myh1, MF 20 (Developmental Studies Hybridoma Bank, Iowa, #MF20-c), 1:200 IF
anti-p63 (Santa Cruz sc-8431), 1:200 IF
HRP-conjugated anti-GAPDH, D16H11 (Cell Signaling 8884), 1:1000 WB
anti-phospho-Smad2 Ser465/467, 138D4 (Cell Signaling 3108), 1:1000 WB
anti-Smad2, D43B4 (Cell Signaling 5339), 1:1000 WB
anti-N-Cadherin, D4R1H (Cell Signaling 13116), 1:400 IF
anti-5hmC (Active Motif 39769), 1:5000 dot blot
HRP-conjugated secondary antibody, G-21234 (Invitrogen G21234), 1:5000 dot blot
anti-H3K36me3 (Thermo fisher MA5-24687), 1:100 CUT&Tag
anti-H3K36me2 (active motif 39256), 1:50 CUT&Tag
anti-H3K4me3 (active motif 39060), 1:100 CUT&Tag
anti-H3K27ac, RM172 (Millipore Sigma MABE647), 1:100 CUT&Tag
anti-H3K27me3, C36B11 (Cell Signaling 9733), 1:100 CUT&Tag
Guinea Pig anti-Rabbit IgG (antibodies online ABIN101961), 1:100 CUT&Tag

Validation

anti-Thy1.2 eFluor 450, 53-2.1 (Thermo Fisher Scientific),
https://www.thermofisher.com/antibody/product/CD90-2-Thy-1-2-Antibody-clone-53-2-1-Monoclonal/48-0902-82

anti-SSEA-1 eFluor 660, eBioMC-480 (Thermo Fisher Scientific),
https://www.thermofisher.com/antibody/product/SSEA1-Antibody-clone-eBioMC-480-MC-480-Monoclonal/50-8813-42

anti-Sca-1 PE-Cy7, 25-5981-82 (Thermo Fisher Scientific),
https://www.thermofisher.com/antibody/product/Ly-6A-E-Sca-1-Antibody-clone-D7-Monoclonal/25-5981-82

anti-EpCAM PE, G8.8 (Thermo Fisher Scientific),
https://www.thermofisher.com/antibody/product/CD326-EpCAM-Antibody-clone-G8-8-Monoclonal/12-5791-82

anti-Histone H3K36M Monoclonal Antibody, RM193 (RevMab),
https://www.revmab.com/index.php/product/anti-histone-h3-k36m-rabbit-monoclonal-antibody-clone-rm193-histone-h3-k36m-mutant/

anti-H3K36me3 (abcam 9050),
https://www.abcam.com/histone-h3-tri-methyl-k36-antibody-chip-grade-ab9050.html

anti-H3K36me2, C75H12 (Cell Signaling 2901)

https://www.cellsignal.com/products/primary-antibodies/di-methyl-histone-h3-lys36-c75h12-rabbit-mab/2901

anti-Sox2, D9B8N (Cell Signaling 23064),
https://www.cellsignal.com/products/primary-antibodies/sox2-d9b8n-rabbit-mab/23064

anti-EpCAM, G8.8 (Life Technologies 14-5791-81),
https://www.thermofisher.com/antibody/product/CD326-EpCAM-Antibody-clone-G8-8-Monoclonal/14-5791-81

anti-Nanog (abcam 80892),
https://www.abcam.com/nanog-antibody-ab80892.html

anti-Ter119 (PE/Cy5; Thermo fisher, Cat# 15-5921-83),
https://www.thermofisher.com/antibody/product/TER-119-Antibody-clone-TER-119-Monoclonal/15-5921-82

anti-CD3e (PE/Cy5; Biolegend, Cat# 100310),
https://www.biolegend.com/en-us/punchout/punchout-products/product-detail/pe-cyanine5-anti-mouse-cd3epsilon-antibody-26?GroupID=BLG6742

anti-Gr1 (PE/Cy5; Thermo fisher, Cat# 15-5931-83),
https://www.thermofisher.com/antibody/product/Ly-6G-Ly-6C-Antibody-clone-RB6-8C5-Monoclonal/15-5931-82

anti-B220/CD45R (PE/Cy5; Biolegend, Cat# 103210),
https://www.biolegend.com/en-ie/products/pe-cyanine5-anti-mouse-human-cd45r-b220-antibody-448?GroupID=GROUP658

anti-TCRb (PE/Cy5; Thermo fisher, Cat# 15-5961-83),
https://www.thermofisher.com/antibody/product/TCR-beta-Antibody-clone-H57-597-Monoclonal/15-5961-82

anti-CD4 (PE/Cy5; Thermo fisher, Cat# 15-0041-83),
https://www.thermofisher.com/antibody/product/CD4-Antibody-clone-GK1-5-Monoclonal/15-0041-82

anti-CD8a (PE/Cy5; Biolegend, Cat# 100710),
https://production.biolegend.com/en-us/products/pe-cyanine5-anti-mouse-cd8a-antibody-156

anti-c-Kit (APC-eFluor 780; Thermo fisher, Cat# 47-1171-82),
https://www.thermofisher.com/antibody/product/CD117-c-Kit-Antibody-clone-2B8-Monoclonal/47-1171-82

anti-CD34 (FITC; Thermo fisher, Cat# 11-0341-85),
https://www.thermofisher.com/antibody/product/CD34-Antibody-clone-RAM34-Monoclonal/11-0341-82

anti-CD16/32 (FC gamma receptor) (PE; BD Biosciences, Cat# 553145),
https://www.bdbiosciences.com/en-us/products/reagents/flow-cytometry-reagents/research-reagents/single-color-antibodies-ruo/pe-rat-anti-mouse-cd16-cd32.553145

anti-Vim, D21H3 (Cell Signaling 5741),
https://www.cellsignal.com/products/primary-antibodies/vimentin-d21h3-xp-rabbit-mab/5741

anti-Tubb3 (Biolegend 801201),
https://www.biolegend.com/ja-jp/products/purified-anti-tubulin-beta-3-tubb3-antibody-11580

anti-Myh1, MF 20 (Developmental Studies Hybridoma Bank, Iowa),
https://dshb.biology.uiowa.edu/MF-20

anti-p63 (Santa Cruz sc-8431),
https://datasheets.scbt.com/sc-8431.pdf

anti-GAPDH, D16H11, HRP conjugated (Cell Signaling 8884),
https://www.cellsignal.com/products/primary-antibodies/gapdh-d16h11-xp-rabbit-mab/5174

anti-phospho-Smad2 Ser465/467, 138D4 (Cell Signaling 3108),
https://www.cellsignal.com/products/primary-antibodies/phospho-smad2-ser465-467-138d4-rabbit-mab/3108

anti-Smad2, D43B4 (Cell Signaling 5339),
https://www.cellsignal.com/products/primary-antibodies/smad2-d43b4-xp-rabbit-mab/5339

anti-N-Cadherin, D4R1H (Cell Signaling 13116),
https://www.cellsignal.com/products/primary-antibodies/n-cadherin-d4r1h-xp-rabbit-mab/13116

anti-5hmC (Active Motif 39769),
https://www.activemotif.com/catalog/details/39769.html

HRP-conjugated secondary antibody, G-21234 (Invitrogen G21234),
https://www.thermofisher.com/antibody/product/Goat-anti-Rabbit-IgG-H-L-Cross-Adsorbed-Secondary-Antibody-Polyclonal/G-21234

anti-H3K36me3 (thermo fisher MA5-24687),
https://www.thermofisher.com/antibody/product/H3K36me3-Antibody-clone-RM491-Recombinant-Monoclonal/MA5-24687

anti-H3K36me2 (active motif 39256),
https://www.activemotif.com/catalog/details/39255

anti-H3K4me3 (active motif 39060),
https://www.activemotif.com/catalog/details/39159

anti-H3K27ac, RM172 (Millipore Sigma MABE647),
https://www.sigmaaldrich.com/US/en/product/mm/mabe647

H3K27me3, C36B11 (Cell Signaling 9733),
https://www.cellsignal.com/products/primary-antibodies/tri-methyl-histone-h3-lys27-c36b11-rabbit-mab/9733

Guinea Pig anti-Rabbit IgG, antibodies online ABIN101961,
https://www.antibodies-online.com/antibody/101961/Guinea+Pig+anti-Rabbit+IgG+Heavy++Light+Chain+antibody+-+Preadsorbed/

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Mouse embryonic fibroblasts were derived from crosses of histone mutant- and OKSM cassette- harboring mice that additionally carry an Oct4-EGFP reporter. KH2 ES cell lines were originally obtained from the Jaenisch lab and were previously used in prior studies by the lab to generate histone mutant mice (Brumbaugh et al.). BJ fibroblasts (CRL-2522) and HEK293T (CRL-2316) were obtained from ATCC.
Authentication	Primary derivation of cell lines as outlined in the text, KH2 ES cells were authenticated in prior publications (Beard et al., Brumbaugh et al.; PCR).
Mycoplasma contamination	Cell lines were tested weekly for mycoplasma contamination using the LONZA Mycoplasma Detection Kit. All cell lines tested negative for Mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals	Eight to ten week-old female Col1a1::tetO-OKSM mice with an EGFP reporter in the 3' UTR of the Pou5f1 gene were crossed with either male Col1a1::tetO-K36M mice or Col1a1::tetO-H3.3 mice to generate corresponding mouse embryonic fibroblasts. To generate mouse embryonic fibroblasts without OKSM cassette, female Rosa26::M2-rtTA mice were used. For teratoma assays, 8 week-old female NU/J mice were used. Mice used in this study were housed and bred in specific-pathogen-free rooms located in the AAALAC-accredited Center for Comparative Medicine vivarium at Massachusetts General Hospital. Mice were housed in ventilated cages on a standard 12:12 light cycle. All procedures involving mice adhered to the guidelines of the approved Massachusetts General Hospital Institutional Animal Care and Use Committee (IACUC) protocol no. 2006N000104.
Wild animals	This study did not involve wild animals.
Field-collected samples	This study did not involve animals collected from the field.
Ethics oversight	All procedures involving mice adhered to the guidelines of the approved Massachusetts General Hospital Institutional Animal Care and Use Committee (IACUC) protocol no. 2006N000104.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Cells were trypsinized for 5 minutes, then quenched with serum-containing medium. Cells were then centrifuged at 350x g

Sample preparation	for 5 minutes and washed with buffer containing 2% FBS in PBS, and filtered through 0.2 µm mesh. Staining was performed on ice for 30 min before additional washing. For GMP sorting, bone marrow from euthanized mice was collected by flushing both femurs and tibiae using a 27.5-gauge needle with 10 ml of FACS buffer (PBS+2% FBS (Hyclone)). The bone marrow was then dissociated via trituration with a 22.5-gauge needle. The resulting cells were centrifuged at 1,200 rpm for 5 min and resuspended in 5 ml of cold RBC lysis buffer (Biolegend) and incubated for 8 min on ice. The lysis was then quenched immediately with 30 ml of FACS buffer and again centrifuged at 1,200 rpm for 5 min. The resulting cells were resuspended in 10 ml of FACS buffer, filtered through 100µm cell strainer (Corning), and counted. 40-50 million cells were applied for lineage depletion kit (Miltenyi Biotech) according to the manufacturer's recommendations.
Instrument	Cells were analyzed on an LSRII flow cytometer (BD Biosciences) or collected on an Aria II flow cytometer (BD Biosciences).
Software	Data was collected using Diva v6.1.2 (BD Biosciences) and data analysis was performed using FlowJo software v10.2 (FlowJo, LLC).
Cell population abundance	For GMP sorting, we reanalyzed purified population to ensure greater than 90% purity.
Gating strategy	Dead cells were identified with DAPI and gated out. Cell debris was gated out using a FSC vs SSC gate. Cell aggregates were excluded using a FSC-W vs FSC-A approach. The GMP gating strategy is depicted as schematic within Extended Data figures.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.