doi:10.1038/nature12587

Deterministic direct reprogramming of somatic cells to pluripotency

Yoach Rais[1]*, Asaf Zviran[1]*, Shay Geula[1]*, Ohad Gafni[1], Elad Chomsky[1], Sergey Viukov[1], Abed AlFatah Mansour[1], Inbal Caspi[1], Vladislav Krupalnik[1], Mirie Zerbib[1], Itay Maza[1], Nofar Mor[1], Dror Baran[1], Leehee Weinberger[1], Diego A. Jaitin[2], David Lara-Astiaso[2], Ronnie Blecher-Gonen[2], Zohar Shipony[3,4], Zohar Mukamel[3,4], Tzachi Hagai[5], Shlomit Gilad[6], Daniela Amann-Zalcenstein[6], Amos Tanay[3,4], Ido Amit[2], Noa Novershtern[1] & Jacob H. Hanna[1]

Somatic cells can be inefficiently and stochastically reprogrammed into induced pluripotent stem (iPS) cells by exogenous expression of Oct4 (also called Pou5f1), Sox2, Klf4 and Myc (hereafter referred to as OSKM). The nature of the predominant rate-limiting barrier(s) preventing the majority of cells to successfully and synchronously reprogram remains to be defined. Here we show that depleting Mbd3, a core member of the Mbd3/NuRD (nucleosome remodelling and deacetylation) repressor complex, together with OSKM transduction and reprogramming in naive pluripotency promoting conditions, result in deterministic and synchronized iPS cell reprogramming (near 100% efficiency within seven days from mouse and human cells). Our findings uncover a dichotomous molecular function for the reprogramming factors, serving to reactivate endogenous pluripotency networks while simultaneously directly recruiting the Mbd3/NuRD repressor complex that potently restrains the reactivation of OSKM downstream target genes. Subsequently, the latter interactions, which are largely depleted during early pre-implantation development *in vivo*, lead to a stochastic and protracted reprogramming trajectory towards pluripotency *in vitro*. The deterministic reprogramming approach devised here offers a novel platform for the dissection of molecular dynamics leading to establishing pluripotency at unprecedented flexibility and resolution.

Induced pluripotent stem cells can be generated from somatic cells by ectopic expression of different transcription factors, originally Oct4, Sox2, Klf4 and Myc (OSKM)[1]. The reprogramming process requires initial cell proliferation, after which a fraction of the cell progeny successfully converts into an embryonic stem (ES)-like state with different time latencies[2,3]. A variety of chromatin modifiers have been implicated in facilitating epigenetic changes leading to authentic iPS cell reprogramming[4,5]. Despite these advances, the reprogramming efficiency of somatic cells remains extremely low[3]. Furthermore, the outcome of challenging the somatic epigenome with the overexpression of OSKM reprogramming factors is stochastic[2]. Experimental and theoretical modelling approaches for characterizing the nature of stochastic elements acting in iPS cell reprogramming have suggested that the existence of as few as one dominant rate-limiting element may adequately recapitulate the experimentally measured kinetics for clonal iPS cell formation by OSKM factors[2,6]. The identity of such stochastic rate-limiting element(s) remains to be defined. Here we show that the Mbd3/NuRD repressor complex[7] is the predominant molecular block preventing deterministic induction of ground-state pluripotency.

Mbd3 and establishment of naive pluripotency

We tested whether additional genetic manipulations may enable deterministic reprogramming towards ground-state pluripotency by OSKM factors, where all donor somatic cells and their progeny synchronously convert into iPS cells. Recent studies have pointed out the importance of chromatin derepression in converting somatic cells into iPS cells[5,8,9]. Therefore, we aimed to conduct a loss-of-function screen for selected epigenetic repressor factors in an attempt to markedly boost the efficiency of reprogramming to ground-state pluripotency. We initially focused on reverting murine primed pluripotent epiblast stem cells (EpiSCs) that can stochastically convert within 5 days into a naive pluripotent state in 2i/LIF growth conditions (where 2i is ERK1/2 and GSK3β inhibitors, and LIF is leukaemia inhibitory factor)[5]. We used a primed EpiSC line carrying a Nanog–GFP knock-in reporter that can be reactivated in the naive state[5], and applied short interfering RNA (siRNA) screening to identify boosters of EpiSC reversion into Nanog–GFP^{+} naive cells (Fig. 1a and Extended Data Fig. 1a). Notably, only Mbd3 inhibition markedly increased the EpiSC reversion efficiency, where up to 80% of the transfected cells turned on Nanog–GFP in 2i/LIF conditions (Fig. 1a).

Mbd3 is a key component in the NuRD complex, ubiquitously expressed in all somatic cells[10]. Mbd2 and Mbd3 assemble into mutually exclusive distinct NuRD complexes[11], which can mediate gene repression through histone deacetylation and chromatin remodelling activities. To validate the siRNA screening results, we used *Mbd3*$^{+/+}$ and *Mbd3*$^{fl/-}$ ES cells and introduced *Rosa26-creER* and *Nanog-GFP* knock-in alleles before converting them into EpiSCs (Fig. 1b and Extended Data Fig. 1b–e). Clonal analysis for epigenetic reversion of EpiSCs demonstrated 95% Nanog–GFP^{+} single-cell reversion efficiency in Mbd3-null cells (Fig. 1b). *Mbd3*$^{fl/-}$ EpiSCs, which retain hypomorphic (~20%) Mbd3 protein expression levels (Fig. 1c), also yielded reverted ES cells with >90% efficiency (Fig. 1b). Both reverted *Mbd3*$^{-/-}$ (after transgenic insertion of Mbd3 to rescue their differentiation deficiency[10,12]) and *Mbd3*$^{fl/-}$ cells can contribute to chimaera formation (Extended Data Fig. 1f). Reconstitution of Mbd3 expression in *Mbd3*$^{-/-}$ and *Mbd3*$^{fl/-}$ EpiSCs inhibited reversion efficiencies (Fig. 1b). These results directly demonstrate that reduction of Mbd3 protein levels renders nearly complete reversion of EpiSCs to naive pluripotency.

We revisited ES-cell derivation experiments from *Mbd3*$^{-/-}$ E3.5 embryos[10], and were able to isolate *Mbd3*$^{-/-}$ ES cells in serum-free 2i/LIF

[1]The Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel. [2]The Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel. [3]The Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel. [4]The Department of Computer Science, Weizmann Institute of Science, Rehovot 76100, Israel. [5]The Department of Structural Biology, Weizmann Institute of Science, Rehovot 76100, Israel. [6]The Israel National Center for Personalized Medicine, Weizmann Institute of Science, Rehovot 76100, Israel.
*These authors contributed equally to this work.

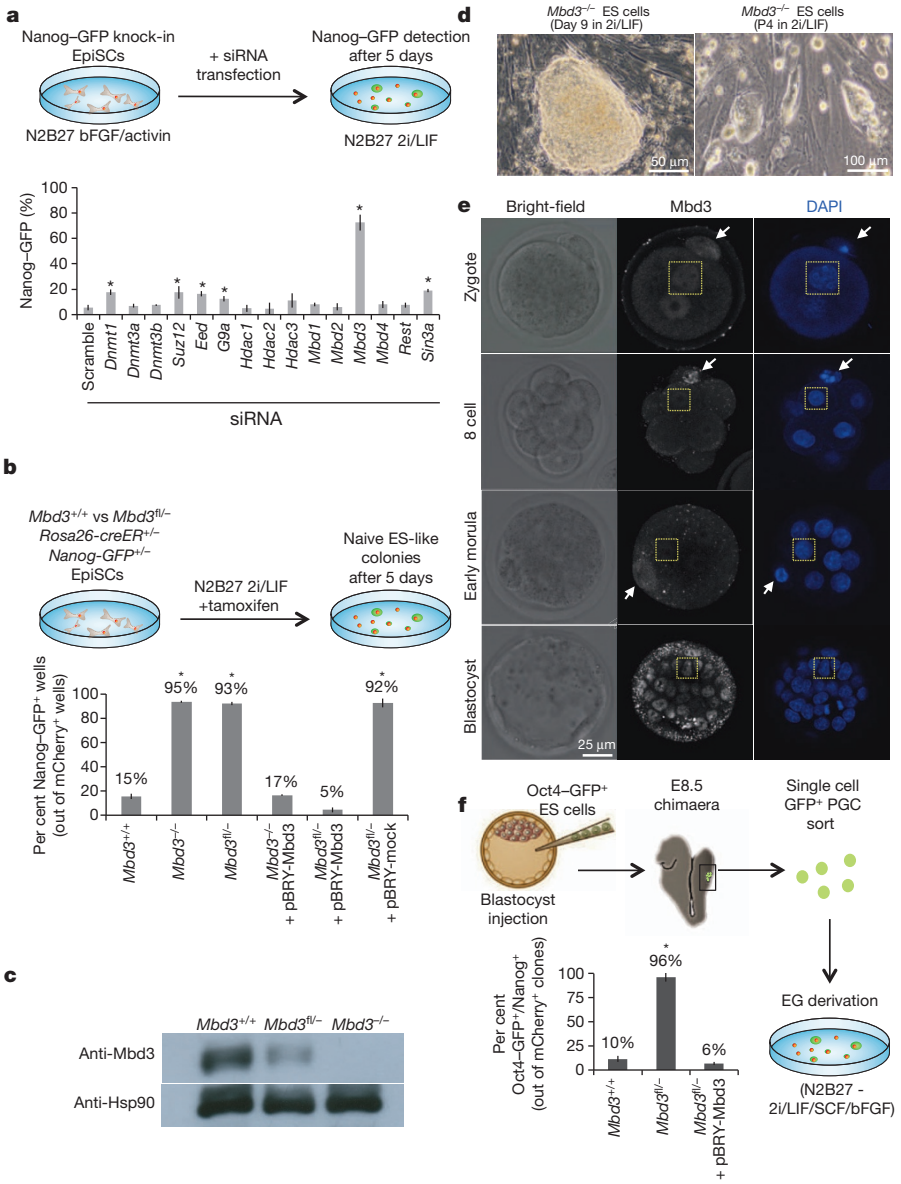


Figure 1 | Boosting primed to naive pluripotency reversion. **a**, An siRNA screen for factors that can boost epigenetic reversion of primed EpiSCs into naive ES cells. Percentage of naive Nanog–GFP$^+$ cells detected by flow cytometry is indicated ($n = 3$). **b**, Single-cell reprogramming efficiency and quantification for EpiSC reprogramming from different mutant lines. The pBRY-Mbd3 rescue construct was stably expressed in the indicated lines ($n = 4$). **c**, Western blot analysis for Mbd3 expression in ES cells. **d**, *Mbd3*$^{-/-}$ ES cell derivation from blastocysts in 2i/LIF. **e**, Representative confocal immunostaining images for temporal Mbd3 expression in developing mouse embryos. Arrows indicate polar body ($n = 15$ embryos stained per stage). Scale bar, 25 µm. **f**, *Mbd3*$^{fl/-}$ and *Mbd3*$^{+/+}$ ES cell lines (with or without pBRY-Mbd3 overexpression allele) were targeted with an Oct4–GFP reporter and a constitutively expressed mCherry reporter, and injected into host chimaeras. Embryonic day (E)8.5 primordial germ cells (PGC) were sorted into defined conditions and evaluated for efficiency to generate EG cells ($n = 6$). Asterisk indicates *t*-test *P* value <0.01 in comparison to *Mbd3*$^{+/+}$. All error bars indicate s.d. from average.

conditions (Fig. 1d and Extended Data Fig. 2a, b). This indicates that Mbd3 is dispensable for establishing the ground state of pluripotency and ES cell derivation. Consistent with an antagonistic role for Mbd3 in establishing pluripotency, Mbd3 is largely depleted after fertilization and throughout pre-implantation development (from 4-cell until early morula stages), and its nuclear expression becomes consolidated at the late morula, blastocyst and post-implantation epiblast (Fig. 1e and Extended Data Fig. 2c, d)[10]. These results indicate that early pre-implantation *in vivo* reprogramming and development are accompanied by depletion of Mbd3 expression, which gets re-expressed as pluripotency is consolidated in the inner cell mass (ICM). The latter dynamic Mbd3 expression pattern is also consistent with a critical role for Mbd3 in restricting aberrant trophoblast lineage specification of the ICM and facilitating adequate differentiation of the assembled pluripotent epiblast[13]. Finally, we aimed to test the influence of reducing Mbd3 expression in reprogramming Oct4$^+$ primordial germ cells (PGCs) into ES-like pluripotent embryonic germ (EG) cells[14]. Single cell isolated *Mbd3*$^{fl/-}$ Oct4–GFP$^+$ E8.5 PGCs from chimaeric mice were proficient in forming EG cell colonies and lines (>95% efficiency), whereas PGCs isolated from chimaeras that were generated by micro-injecting *Mbd3*$^{+/+}$ or *Mbd3*$^{fl/-}$ cells carrying an exogenous *Mbd3* transgene reprogrammed at less than 10% efficiency (Fig. 1f). Collectively, these findings show that neutralizing Mbd3 expression facilitates access to ground-state pluripotency from early embryonic Oct4-expressing cells.

Deterministic reprogramming of somatic cells

We next moved to test whether Mbd3 depletion in somatic cells facilitates their conversion to pluripotency at efficiencies nearing 100%. Recent studies have described a mild positive effect for Mbd3 short-hairpin RNA (shRNA)-mediated knockdown on mouse iPS cell formation[15], and a negative effect on human primed iPS cell induction[16]. We revisited these experiments while using optimized *Mbd3* genetic depletion, OSKM transgene delivery and 2i/LIF containing naive pluripotency conditions (Extended Data Fig. 3a–d). Notably, 95% of *Mbd3*$^{fl/-}$ and *Mbd3*$^{-/-}$ cells were Oct4–GFP$^+$ at day 10, whereas only levels up to 18% were observed in control *Mbd3*$^{+/+}$ fibroblasts (Fig. 2a). To evaluate reprogramming efficiencies accurately, we established 'secondary reprogrammable'[17] *Mbd3*$^{+/+}$ and *Mbd3*$^{fl/-}$ transgenic cell lines harbouring a doxycycline-inducible OKSM polycistronic cassette[18], a constitutive nuclear mCherry marker (to track individual cells and control for plating efficiency), and an Oct4–GFP reporter (Extended Data Fig. 3a). Single cell sorting of secondary mCherry$^+$ *Mbd3*$^{fl/-}$ mouse embryonic fibroblasts (MEFs) and subsequent reprogramming in 2i/LIF plus doxycycline conditions reproducibly yielded 100% iPS cell derivation efficiency by day 8. Wild-type

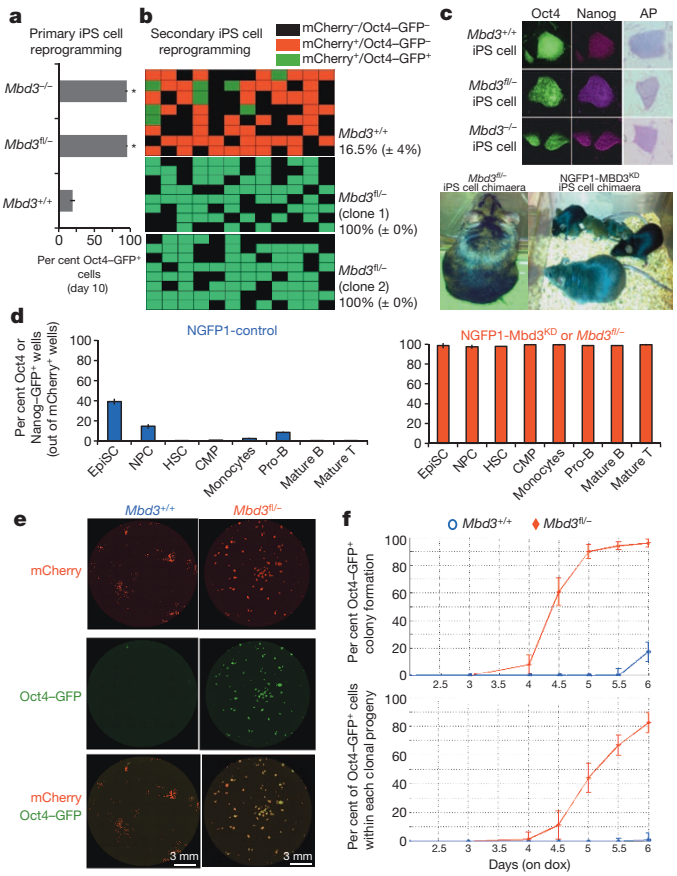


Figure 2 | Deterministic and synchronized iPS cell reprogramming. **a**, Mbd3 wild-type and depleted ($Mbd3^{fl/-}$ or $Mbd3^{-/-}$) MEFs were directly infected with lentiviruses expressing a polycistronic OKSM cassette. Reprogramming efficiency (Oct4–GFP) was measured by FACS at day 10. Asterisk indicates *t*-test *P* value <0.01 in comparison to $Mbd3^{+/+}$ samples. Error bars indicate s.d. from average ($n = 5$). **b**, Secondary reprogrammable fibroblasts carrying an Oct4–GFP reporter and an mCherry constitutively expressed marker were single-cell-seeded and subjected to doxycycline-induced reprogramming. Reprogramming efficiency at day 8 was calculated by dividing the number of Oct4–GFP$^+$ wells by mCherry$^+$ wells ($n = 3$ per clone, ± indicates s.d. from average). **c**, Top: immunostaining of representative iPS cell clones for pluripotency markers (original magnification, ×100). Bottom: agouti-coat-coloured chimaera and germline transmission from Mbd3-depleted iPS cells. **d**, The indicated somatic cell types from NGFP1-control or NGFP1-Mbd3^{KD} adult chimaeras were isolated and subjected to single-cell reprogramming and evaluation of Nanog–GFP expression after 8 days of doxycycline ($n = 4$). CMP, common myeloid progenitor; HSC, haematopoietic stem cell; NPC, neural precursor. **e**, Full-well mosaic images of mCherry, Oct4–GFP and combined channels, shown for $Mbd3^{fl/-}$ and $Mbd3^{+/+}$ at day 6. Scale bar, 3 mm. **f**, Characterization of Oct4–GFP$^+$ dynamic for $Mbd3^{fl/-}$ (red plot) and $Mbd3^{+/+}$ (blue plot) based on live imaging. Top graph indicates cumulative Oct4–GFP$^+$ colonies; bottom graph indicates fraction of Oct4–GFP$^+$ cells within colonies. Dox, doxycycline. Graphs show means and s.d. of all tracked colonies in three biological replicates (one out of four biological data sets is shown).

cells reprogrammed under identical conditions, no more than 20% of clones reactivated Oct4–GFP (Fig. 2b). Teratomas (not shown) and chimaeras were obtained from iPS cell clones (Fig. 2c). High single-cell reprogramming efficiency rates were obtained from a variety of adult progenitor and terminally differentiated cells (Fig. 2d and Extended Data Fig. 3e, f).

We analysed the reprogramming dynamics of 'secondary' $Mbd3^{fl/-}$ and control $Mbd3^{+/+}$ fibroblasts by applying microscopic live imaging and an algorithm that allows segmentation of single mCherry$^+$ colonies and tracking of Oct4–GFP reactivation dynamics during reprogramming (Supplementary Videos 1–4). By day 6 after doxycycline induction, >98% of $Mbd3^{fl/-}$ clonal populations reactivated the Oct4–GFP pluripotency marker, whereas only up to 20% efficiency was detected in control samples reprogrammed in identical growth conditions (Fig. 2e, f). By day 6, approximately 85% of cells within each individual $Mbd3^{fl/-}$ clonal population became Oct4–GFP$^+$ cells, whereas <2% of cells within successfully reprogrammed $Mbd3^{+/+}$ clones turned on the Oct4–GFP marker (bottom panel in Fig. 2f). The latter unbiased quantitative analysis demonstrated a markedly intra- and interclonal synchronized reactivation of Oct4–GFP occurring during a narrow window in $Mbd3^{fl/-}$ clonal populations at days 4.5–5.5 (Fig. 2f and Supplementary Video 2), and highlights a marked increase in reprogramming synchrony and efficiency after Mbd3 depletion in OSKM-transduced somatic cells. Detection of Oct4–GFP by flow cytometry on polyclonal populations demonstrated similar iPS cell reprogramming kinetics (Extended Data Fig. 4a). Re-infection with lentiviruses encoding Mbd3, but not Mbd2, before day 5 of reprogramming had a profound inhibitory effect on iPS cell generation from $Mbd3^{fl/-}$ MEFs, whereas re-infection after day 5 had a diminished effect (Extended Data Fig. 4b). The above kinetic analysis indicates that Mbd3 can inhibit reprogramming when introduced before the final stages of reprogramming. However, once pluripotency is re-established, Mbd3 does not compromise the maintenance of pluripotency.

We next conducted global gene expression analysis on donor MEFs at days 0, 4 and 8 after doxycycline induction without cell passaging, and compared them to iPS cell and ES cell lines. Notably, by day 8, $Mbd3^{fl/-}$ donor cells were transcriptionally indistinguishable from multiple ES cell lines and subcloned established iPS cell lines (Fig. 3a). Genome-wide chromatin mapping for H3K27me3, H3K4me3 and H3K27ac histone marks by chromatin immunoprecipitation followed by sequencing analysis (ChIP-seq) also confirmed that only $Mbd3^{fl/-}$ transduced MEFs had assumed an ES-like chromatin profile by day 8 (Extended Data Fig. 5a). Genome-wide DNA methylation mapping by reduced representation bisulphite sequencing (RRBS) confirmed that an iPS cell/ES cell-like methylation pattern could be seen in the $Mbd3^{fl/-}$ polyclonal population sample after 8 days of doxycycline treatment (Extended Data Fig. 5b, c). Single cell polymerase chain reaction with reverse transcription (RT–PCR) analysis confirmed that nearly 100% of single cells tested expressed key endogenous pluripotency markers only in $Mbd3^{fl/-}$ reprogrammed samples (Extended Data Fig. 5d). Collectively, the above results indicate that Mbd3 depletion after OSKM induction yields authentic molecular re-establishment of pluripotency in the entire population of donor somatic cells and their progeny.

After the depletion of Mbd3 expression, we were not able to isolate stable, partially reprogrammed cells[19,20] that did not reactivate Oct4–GFP or Nanog–GFP and could be stably expanded *in vitro*, as typically can be obtained from OSKM-transduced wild-type somatic cells (Fig. 3b). We next took $Mbd3^{+/+}$ OSKM-transduced partially reprogrammed cells and attempted to complete their reprogramming by Mbd3 inhibition. Notably, by introducing *Mbd3* siRNA, all clones[19] markedly turned on Oct4–GFP or Nanog–GFP pluripotency markers after continued OSKM expression in 2i/LIF (Fig. 3c). To functionally test a conserved inhibitory role for MBD3 in human iPS cell reprogramming, we generated MBD3^{mut} human ES cells by gene editing with TALE nuclease effectors (Fig. 3d), and validated hypomorphic MBD3 protein expression in a selected bi-allelically targeted clone (Fig. 3e). Single-cell reprogramming efficiency was tested after introducing an OCT4–GFP knock-in reporter and a constitutive mCherry in MBD3^{WT} and MBD3^{mut} iPS cells harbouring doxycycline-inducible OSKM factors. Only *in vitro* differentiated secondary fibroblasts from MBD3^{mut} cells reprogrammed at near 100% efficiency in optimized conditions (Fig. 3f and Extended Data Fig. 6a). Markedly increased human iPS cell formation was obtained by applying *MBD3* siRNA starting at 2 days after doxycycline induction in MBD3^{WT} OSKM transgenic secondary fibroblasts (Extended Data Fig. 6b). *MBD3* siRNA treatment allowed reproducible generation of human iPS cells from adult human patient-specific fibroblasts only after two rounds of OSKM with

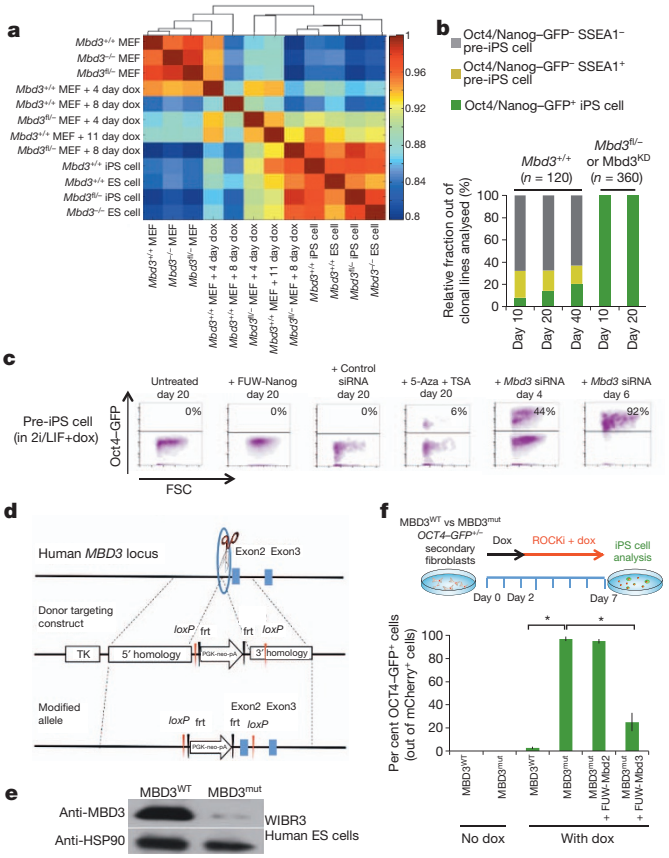


Figure 3 | Alleviating Mbd3 expression facilitates transition to pluripotency. **a**, Spearman correlation matrix between the indicated mouse samples, measured over gene expression levels of all 16,620 expressed genes. The matrix is clustered with hierarchical clustering. **b**, Monoclonal lines established from *Mbd3*$^{+/+}$ and *Mbd3*$^{fl/-}$ secondary cells and reprogrammed in 2i/LIF plus doxycycline. Fraction of pre-iPS clones that did not reactivate Oct4/Nanog GFP markers (either SSEA1 positive or negative) is shown. **c**, Representative partially reprogrammed cell line containing OSKM transgenes, and that did not reactivate GFP reporter, was subjected to the indicated manipulations and analysed for completion of reprogramming as assayed by FACS. FUW indicates lentiviral backbone vector used. **d**, Targeting strategy for human *MBD3* locus. **e**, Biallelically targeted MBD3^{mut} clone displayed ~90% reduction in MBD3 protein expression levels. **f**, MBD3^{WT} and MBD3^{mut} iPS cells carrying doxycycline-inducible OKSM transgenes were labelled with constitutively expressed mCherry and targeted with an *OCT4-GFP* knock-in allele. ROCKi, Rho-associated protein kinase inhibitor. *In vitro* differentiated fibroblasts from the latter lines were reprogrammed as indicated in the scheme. Error bars indicate s.d. of average ($n = 6$). Asterisks indicate *t*-test *P* value <0.01 in comparison to *MBD3*^{WT}.

LIN28 mRNA transfection[21] (Extended Data Fig. 6c). Taken together, these results demonstrate that inhibiting MBD3 alleviates predominant obstacles for human iPS cell reprogramming.

Numerical modelling of reprogramming

We next sought to characterize the reprogramming latency distribution for both wild-type and Mbd3-depleted samples quantitatively, as this may allow a comparison of reprogramming dynamics to known deterministic and stochastic dynamic models. We applied a previously described approach for monoclonal pre-B-cell weekly follow-up for reactivation of Nanog–GFP (Fig. 4a)[2]. A secondary OSKM transgenic NGFP1-iPS cell line[2], carrying a Nanog–GFP reporter, was rendered transgenic for a doxycycline-inducible *Mbd3* shRNA construct (NGFP1-Mbd3^{KD}). Indeed, NGFP1-Mbd3^{KD}-derived monoclonal B-cell populations converted into Nanog–GFP^{+} iPS cells at day 7 at near 100% efficiency (Fig. 4b and Extended Data Fig. 7a). Subsequently, during the first 10 days of reprogramming we conducted daily Nanog–GFP

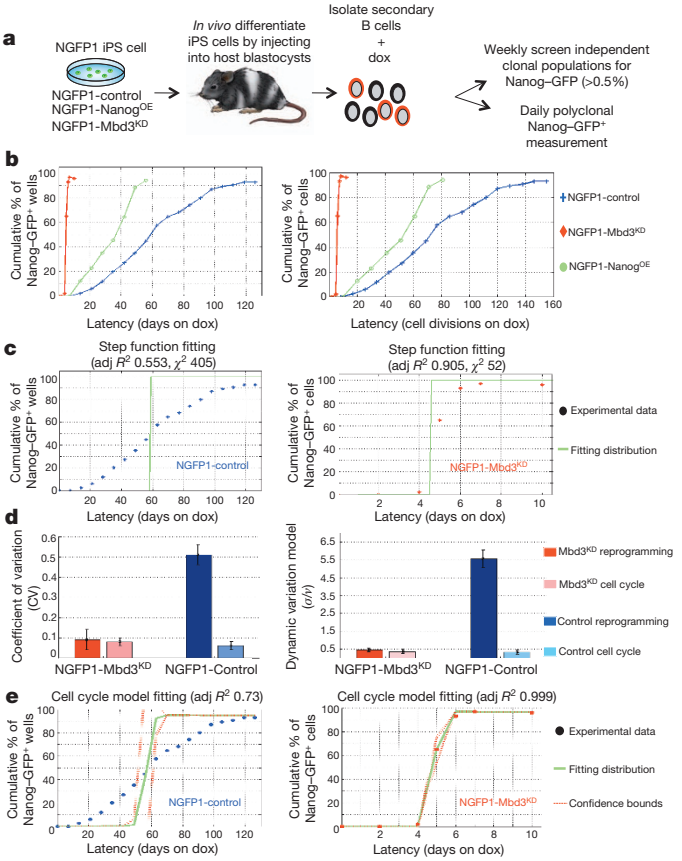


Figure 4 | Numerical description of reprogramming after Mbd3 depletion. **a**, Scheme demonstrating the monoclonal and polyclonal follow-ups for Nanog–GFP reactivation. KD, knockdown; OE, overexpression. **b**, Cumulative percentage of Nanog–GFP^{+} wells versus time on doxycycline, measured for various clonal B-cell-derived populations. **c**, Goodness of fit plots for the fitting of deterministic step function model to the observed reprogramming latency. **d**, Comparison of the calculated variation of *Mbd3*^{KD} and control (dark red and blue colours, respectively), and the cell cycle variation of each sample (light red and blue colours, respectively). Sample variation was calculated using coefficient of variation (left panel) and dynamic variation (right panel) (Extended Data Fig. 7b). Graphs show maximum likelihood estimations. Error bars indicate 95% confidence intervals of maximal likelihood value (Mbd3^{KD} reprogramming $n = 7$; control reprogramming $n = 13$; cell cycle for control and Mbd3^{KD} $n = 20$). **e**, Goodness of fit plots for the fitting of cell-cycle time distribution to the observed reprogramming latency.

detection on polyclonal NGFP1-control and NGFP1-Mbd3^{KD} B-cell populations. We next tested whether Mbd3^{KD} cell reprogramming behaves like a deterministic function. Such deterministic behaviour is well approximated by a step function with 0% reprogramming before a fixed deterministic time after doxycycline induction, and 100% iPS cell formation afterwards. Fitting the clonal cell reprogramming dynamics to such deterministic step-function revealed a tight fit ($R^2 > 0.9$, chi-squared = 52) of *Mbd3*^{KD} cells, but not of control *Mbd3*$^{+/+}$ cells ($R^2 = 0.55$, chi-squared = 405) (Fig. 4c). Despite the observed similarity to a deterministic behaviour, variability was still evident in our Mbd3^{KD} sample under the optimized conditions devised herein. Thus, we sought to quantify and compare the variability detected in the reprogramming latency measurements in both *Mbd3*$^{+/+}$ and *Mbd3*^{KD} samples, and further compare it to the inherent measured cell-cycle variability[2]. For this purpose we used two modelling schemes. By calculating the dynamic variability (s.d./drift) by a Brownian motion (BM) model (Extended Data Fig. 7b), and the coefficient of variation (s.d./mean) by Gaussian model fitting (Extended Data Fig. 7c), we show a consistent reduction in Mbd3^{KD} reprogramming variability and tight proximity to cell-cycle variability (Fig. 4d). To further support reduction in rate-limiting barrier(s)

on Mbd3 depletion, we modelled the dynamics of the OSKM reprogramming process using a multi-step Markov chain model (phase-type (PH) modelling[22]) (Extended Data Fig. 7d). Although this model does not directly argue for stochasticity, still there was a clear connection between the reduction in barriers and the reduction in process variability (Extended Data Fig. 7d). Finally, we proposed that the reduction in rate-limiting barriers in Mbd3^{KD} samples may reduce reprogramming variability to variability explained by cell cycle alone. For this purpose we fitted the observed reprogramming latency to a cell-cycle model that captures the required reprogramming duration[2,23]. We obtained a profound fit ($R^2 = 0.999$) between the Mbd3^{KD} dynamic and cell-cycle model, but not in the control *Mbd3*$^{+/+}$ dynamics ($R^2 = 0.73$) (Fig. 4e). Altogether, these results consistently show a reduction in OSKM reprogramming variability and increased proximity to deterministic dynamic behaviour upon Mbd3 depletion.

Deterministic reprogramming mechanism

We aimed to define the mechanisms for Mbd3 inhibition of iPS cell reprogramming. Inhibiting Mbd3 expression was not sufficient to induce iPS cell formation in the absence of exogenous OSKM overexpression in somatic cells (Extended Data Fig. 3e, f). Contrary to previous reports[15], Mbd3 depletion without OSKM expression does not independently lead to endogenous reactivation of bona fide pluripotency genes (Extended Data Fig. 8a). We established that Flag-tagged OCT4, KLF4, SOX2 and MYC specifically co-immunoprecipitated with Mbd3 after exogenous overexpression in HEK293 cells (Fig. 5a and Extended Data Fig. 9a). OSKM specifically co-immunoprecipitated with Mbd3/NuRD components in MEFs undergoing reprogramming (Extended Data Fig. 9b). These interactions were mediated via the MBD domain of Mbd3, as defined deletions introduced into the MBD domain abrogated co-immunoprecipitation of Mbd3 with OSKM (Extended Data Fig. 9c). Consistent with the direct protein interactions of the Mbd3/NuRD complex with OSKM reported above, genome-wide ChIP-seq analysis of Mbd3 binding in doxycycline-induced wild-type MEFs identified a global increase in Mbd3 recruitment and binding after OSKM induction (1,177 binding regions in MEFs compared to 8,657 after OSKM induction) (Supplementary Data set 1). Importantly, in somatic MEFs before OSKM induction, Mbd3 is not localized to pluripotency factor target genes (Fig. 5b). Only after doxycycline induction are Mbd3-bound genes enriched for targets of Klf4, Oct4, Sox2 and Esrrb ($P < 10^{-22}$) (Fig. 5b). The NuRD component Chd4 was similarly recruited to downstream targets only after doxycycline induction, indicating NuRD recruitment with Mbd3 (Fig. 5b). Chd4 knockdown in *Mbd3*$^{+/+}$ MEFs undergoing reprogramming enhanced iPS cell formation (Extended Data Fig. 8b). Transcription levels of Mbd3 target genes after 4 days of doxycycline were significantly upregulated in Mbd3-depleted samples (Fig. 5c), consistent with the predominant function of the Mbd3/NuRD complex as a repressor of pluripotency. Chromatin of Mbd3 and/or OSKM direct targets was significantly more active and open in Mbd3-depleted samples during reprogramming, including

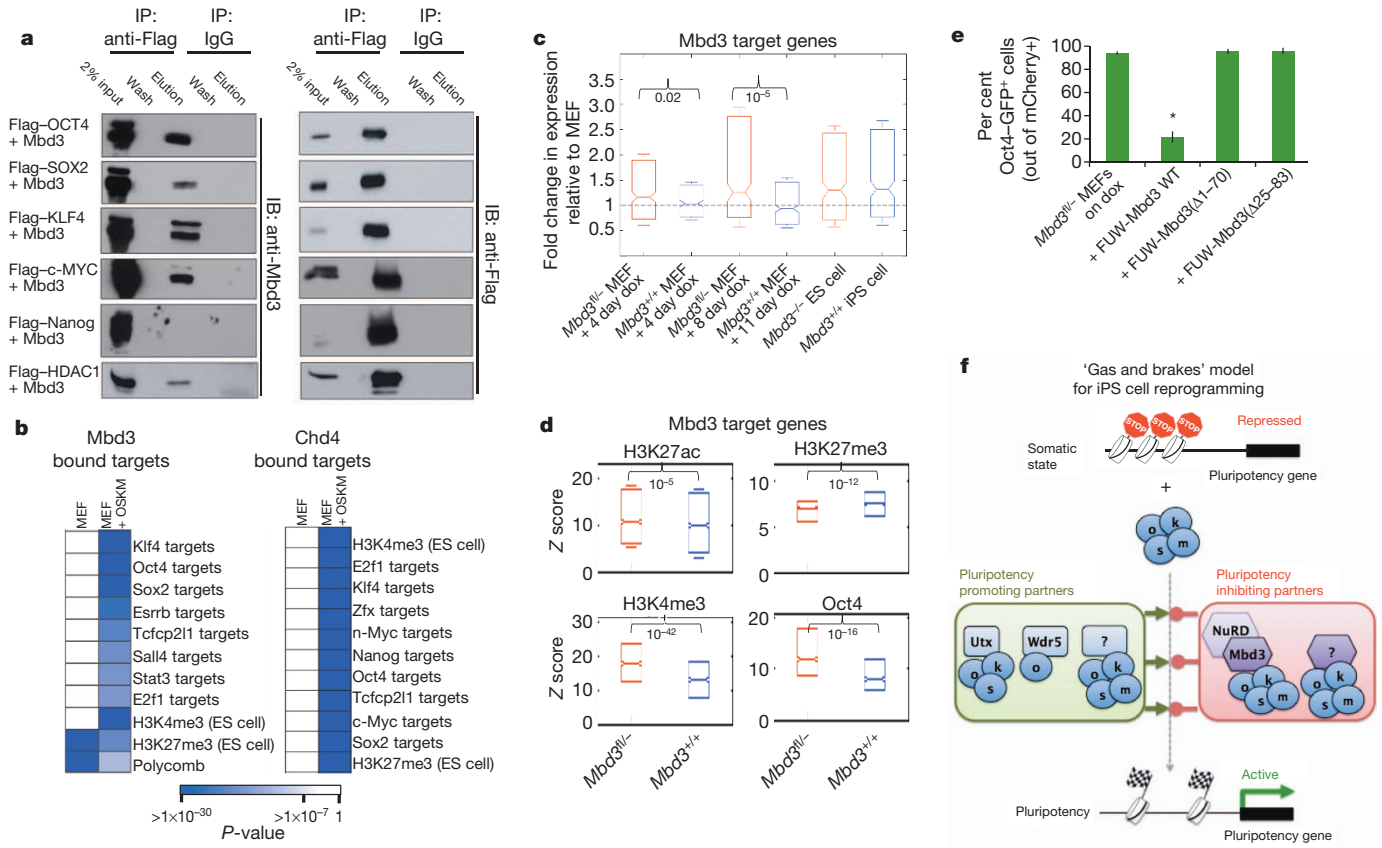


Figure 5 | Mechanisms for Mbd3 inhibitory effect on induced pluripotency. **a**, Constructs encoding Flag-tagged OCT4, SOX2, KLF4, MYC, Nanog or HDAC1 were transfected into HEK293T cells in combination with Mbd3. The cell lysates were immunoprecipitated (IP) with an anti-Flag antibody (or anti-IgG as control), followed by an immunoblot analysis (IB) ($n = 3$ biological replicates). **b**, Functional enrichment of Mbd3 and Mi2β (Chd4) direct targets, measured in MEFs before and after OSKM induction. Colour levels indicate enrichment *P* values (by Fisher's exact test) that pass the false discovery rate (FDR) threshold of 0.0001%. **c**, Distribution of gene expression fold change relative to MEFs of *Mbd3*$^{+/+}$ (blue) and *Mbd3*$^{fl/-}$ (red) samples throughout reprogramming. Graphs show box-plot medians and 25th/75th percentiles, and *P* values by paired sample *t*-test. **d**, Distribution of histone marks and Oct4 binding levels in *z*-score values at day 4 after OSKM (doxycycline) induction. Graphs show box-plot medians and 25th/75th percentiles, and *P*-values by paired sample *t*-test. **e**, Reprogramming efficiency of *Mbd3*$^{fl/-}$ MEFs after infection with lentiviruses encoding wild type and different mutant Mbd3 inserts. Error bars indicate s.d. from average ($n = 6$). Asterisk in **e** indicates *t*-test *P* value <0.01. **f**, Mechanistic model scheme.

statistically significant higher levels of Oct4 binding, H3K4me3 and H3K27ac, and reduced H3K27me3 repressive chromatin mark (Fig. 5d and Extended Data Fig. 8c–e). Mbd3 mutants with a compromised ability to interact with OSKM reprogramming factors directly (Extended Data Fig. 9d) were deficient in reducing reprogramming efficiency of *Mbd3*$^{fl/-}$ somatic cells, supporting the notion that direct OSKM–Mbd3 interactions are important for inhibiting iPS cell formation (Fig. 5e).

We noted that a minimum 5-day exogenous transgene (doxycycline) induction was similarly required to obtain iPS cells from *Mbd3*$^{+/+}$ and *Mbd3*$^{fl/-}$ cells (Extended Data Fig. 10a), and that the expression of Utx and Wdr5 (refs 4, 5)—OSK interacting partners that positively propel reprogramming to pluripotency—was also essential for iPS cell formation in Mbd3-depleted cells (Extended Data Fig. 10b, c). Collectively, these results establish a 'gas and brakes' paradigm. Whereas exogenous OSKM factors interact with multiple epigenetic complexes that de-repress pluripotency-promoting gene networks (such as Wdr5- or Utx-containing complexes), they also directly assemble with the Mbd3/NuRD repressor complex (Fig. 5f). As a result, Mbd3/NuRD is directly recruited to downstream OSKM target genes that are essential for propelling the reprogramming process, and potently counteracts their robust reactivation. In the absence of an Mbd3 inhibitory effect, OSKM interactions with pluripotency-promoting epigenetic regulators predominate functionally, and drive uninterrupted progression of direct reprogramming to pluripotency.

Concluding remarks

Here we show that the stochastic and asynchronized trajectory of direct reprograming by OSKM factors[3] can be coaxed to become nearly synchronized and deterministic with modified reprogramming approaches. We highlight a NuRD repressor complex component, Mbd3, which is normally depleted during early pre-implantation development, as a predominant barrier preventing epigenetic reversion of EpiSCs, primordial germ cells and somatic cells to ground-state pluripotency by defined signalling and transcriptional input. Several critical reprogramming factors directly interact and recruit Mbd3/NuRD complex, and thus form a highly potent negative regulatory complex that restrains pluripotency gene reactivation throughout the process. It will be of great interest to explore whether direct reprogramming in the absence of Mbd3 repression also improves the quality of reprogrammed cells and reduces the frequency of obtaining aberrantly reprogrammed iPS cells[24]. Finally, the deterministic reprogramming strategy reported here may allow the dissection of authentic molecular events accompanying a synchronized and non-saltatory progression pattern towards iPS cells, which is critical for deciphering the black box of reprogramming.

METHODS SUMMARY

Details of cell lines, plasmids, siRNAs and antibodies used, as well as descriptions of methods for reprogramming, immunofluorescence, immunoprecipitation, embryo micromanipulation, time-lapse microscopic imaging, bioinformatics, statistical analyses and mathematical modelling, are provided in Methods.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 April; accepted 23 August 2013.
Published online 18 September 2013.

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
2. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462,** 595–601 (2009).
3. Hanna, J. H., Saha, K. & Jaenisch, R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* **143,** 508–525 (2010).
4. Ang, Y.-S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145,** 183–197 (2011).
5. Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488,** 409–413 (2012).
6. Smith, Z. D., Nachman, I., Regev, A. & Meissner, A. Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. *Nature Biotechnol.* **28,** 521–526 (2010).
7. Hu, G. & Wade, P. A. NuRD and pluripotency: a complex balancing act. *Cell Stem Cell* **10,** 497–503 (2012).
8. Pawlak, M. & Jaenisch, R. *De novo* DNA methylation by Dnmt3a and Dnmt3b is dispensable for nuclear reprogramming of somatic cells to a pluripotent state. *Genes Dev.* **25,** 1035–1040 (2011).
9. 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).
10. Kaji, K., Nichols, J. & Hendrich, B. Mbd3, a component of the NuRD co-repressor complex, is required for development of pluripotent cells. *Development* **134,** 1123–1132 (2007).
11. Le Guezennec, X. *et al.* MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol. Cell. Biol.* **26,** 843–851 (2006).
12. Kaji, K. *et al.* The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nature Cell Biol.* **8,** 285–292 (2006).
13. Reynolds, N. *et al.* NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. *Cell Stem Cell* **10,** 583–594 (2012).
14. Leitch, H. G. *et al.* Embryonic germ cells from mice and rats exhibit properties consistent with a generic pluripotent ground state. *Development* **137,** 2279–2287 (2010).
15. Luo, M. *et al.* NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. *Stem Cells* **31,** 1278–1286 (2013).
16. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483,** 598–602 (2012).
17. Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133,** 250–264 (2008).
18. Sommer, C. A. *et al.* Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* **27,** 543–549 (2009).
19. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
20. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136,** 364–377 (2009).
21. Warren, L. *et al.* Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. *Cell Stem Cell* **7,** 618–630 (2010).
22. Bolch, G., Greiner, S., de Meer, H. & Trivedi, K. S. *Queueing Networks and Markov Chains: Modeling and Performance Evaluation with Computer Science Applications* (John Wiley, 2006).
23. Duffy, K. R. *et al.* Activation-induced B cell fates are selected by intracellular stochastic competition. *Science* **335,** 338–341 (2012).
24. Stadtfeld, M. *et al.* Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465,** 175–181 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements J.H.H. is supported by a generous gift from I. and P. Mantoux; and grants from the Leona M. and Harry B. Helmsley Charitable Trust, ERC (StG-281906) grant, BIRAX initiative, Israel Science Foundation (BIKURA, ICORE and Regular programs), ICRF, Fritz Thyssen Stiftung, The Benoziyo Endowment fund, Alon Scholar Program, and the Clore research prize. I.A is supported by the HFSP Career Development Award, an ISF-Bikura and the ERC (StG-309788). A.A.M. is supported by a Weizmann Dean fellowship. We thank N. Barkai and her group, K. Saha, B. Hendrich, J. Nichols and A. Surani, for reagents and advice. We thank the Weizmann Institute management for providing critical financial and infrastructural support.

Author Contributions Y.R., A.Z., S.Ge., N.N. and J.H.H. conceived the idea for this project, designed and conducted experiments and wrote the manuscript. S.Ge. conducted protein biochemical analysis. A.Z. conducted numerical modelling analysis. O.G., L.W. and N.M. assisted in chromatin immunoprecipitation experiments. N.N. and A.Z. conducted bioinformatics analysis. Y.R. and A.Z. conducted live imaging experiments and analysis. S.V. engineered human stem cell lines. I.A., D.A.J., D.L.-A., S.Gi., D.A.-Z. and R.B.-G. assisted with ChIP-seq experiments. E.C., Z.S., Z.M. and A.T. conducted RRBS analysis. Y.R. and M.Z. conducted microinjections. Y.R. and A.A.M. conduced embryo staining. Y.R., S.Ge. and J.H.H. conducted reprogramming experiments with help from I.C., I.M., V.K., T.H. and D.B.

Author Information Chromatin immunoprecipitation data are available at the National Center for Biotechnology Information Gene Expression Omnibus database under the series accession number GSE49766. Microarray data are available at the National Center for Biotechnology Information Gene Expression Omnibus database under the series accession number GSE45352. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H.H. (jacob.hanna@weizmann.ac.il) or N.N. (noa.novershtern@weizmann.ac.il).

METHODS

Mouse stem cell lines and cell culture. Reprogramming and maintenance of murine naive pluripotent cells were conducted in serum-free chemically defined N2B27-based media: 500 ml KO-DMEM (Invitrogen), 5 ml N2 supplement (Invitrogen; 17502048), 5 ml B27 supplement (Invitrogen; 17504044), 15–20% knockout serum replacement (Invitrogen; 10828), 1 mM glutamine (Invitrogen), 1% non-essential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma), 1% penicillin–streptomycin (Invitrogen), 5 mg ml^{-1} BSA (Sigma). Naive conditions for murine iPS cells, ES cells and EGs included 5 μg recombinant human LIF (Peprotech). Throughout the study 2i was applied in reprogramming 48 h after OSKM induction: small-molecule inhibitors CHIR99021 (CH, 3 μM; Axon Medchem) and PD0325901 (PD, 1μM; Axon Medchem). Primed N2B27 media for murine cells (EpiSCs) contained 8 ng ml^{-1} recombinant human bFGF (Peprotech) and 20 ng ml^{-1} recombinant human activin (Peprotech). Stem-cell lines and mice deficient for Mbd3 and their derived ES lines were obtained as previously described[10,12]. For additional gene targeting of mouse pluripotent stem cell lines (Nanog-GFP reporter, pBRY-Mbd3 rescue constructs, Rosa26-CreER), 50 μg DNA of the targeting construct was linearized and electroporated into the indicated pluripotent cell lines, which were then subjected to selection with puromycin (1 μg ml^{-1}). After 10 days of antibiotic selection, drug-resistant or GFP^+ clones were analysed for correct targeting by PCR or Southern blot analysis. *Mbd3*$^{+/-}$ male and female mice were mated and E3.5 blastocysts were collected and explanted for ES cell derivation in defined mouse 2i/LIF conditions on gelatin/MEF-coated plates. NGFP1-Mbd3^{KD} was established by infection and sub-cloning of secondary NGFP1 iPS cell line with a shRNA pLKO-Tet-On vector (Addgene) as previously described[2]. Mbd3 target sequences selected for the latter strategy were CTAAGTGGATTGAGTGCCTTT and GCGCTATGATTCTTCCAACCA. Mycoplasma detection tests are conducted weekly to ensure exclusion of any contaminated cells.

Epigenetic reversion of mouse primed epiblast cells. Male naive V6.5 (*Mbd3*$^{+/+}$) and Nanog–GFP ES cells[15] maintained in 2i/LIF conditions were injected into BDF2 blastocysts. Chimaeric embryos were dissected at day E6.5 and explanted on gelatin/vitronectin-coated plates in N2B27 bFGF/activin conditions supplemented with 1 μg ml^{-1} puromycin, allowing the isolation of Nanog–GFP EpiSCs. For epigenetic reversion of murine EpiSCs to naive pluripotency, cells were passaged into N2B27 2i/LIF conditions on vitronectin (1 μg ml^{-1}) and gelatin (0.2%)-coated plates (without overexpression of exogenous reprogramming factors). When epigenetic reversion assay involved single-cell plating, EpiSC growth medium was supplemented with ROCK inhibitor (Y-27632) for 24 h before trypsinization. siRNAs (ON-TARGETplus SMARTpool) and the control siRNA (ON-TARGETplus non-targeting pool D-001810-10-05) were obtained from Dharmacon. 10 nM siRNA or control was used for each transfection with RNAiMAX (Invitrogen). For EG derivation experiments, Oct4–GFP^+ primordial germ cells were sorted from E8.5 dissected chimaeric embryos and plated in N2B27 15% KSR, LIF (20 ng ml^{-1})/SCF (10 ng ml^{-1})/bFGF (8 ng ml^{-1}) medium and 2i (supplemented 48 h later). Nuclear mCherry labelling of EpiSCs and their derived primordial germ cells was used to allow calculating plating efficiency and reprogramming efficiency (reprogramming efficiency % = Oct4 or Nanog–GFP^+ clones/mCherry$^+$ clones).

Reprogramming of mouse somatic cells and cell infection. Virus-containing supernatants of the different reprogramming viruses: STEMCCA-OKSM polycistronic vector (doxycycline-inducible and constitutive expression)[5], STEMCCA-OKS polycistronic vector (doxycycline-inducible and constitutive expression), FUW-tetO-lox-KLF4, FUW-tetO-lox-OCT4 and FUW-tetO-lox-SOX2, FUW-tetO-Klf4, FUW-tetO-Oct4, FUW-tetO-Sox2, FUW-tetO-c-Myc, FUW-Oct4-2A-Sox2, FUW-Oct4-2A-Klf4, FUW-tetO-lox-SOX2, pMXs-OCT4, pMXs-SOX2, pMXs-KLF4, pMXs-MYC was supplemented with the FUW-M2rtTA virus (when necessary) and an equal volume of fresh culture medium for infection. Mouse fibroblast and other somatic cells types were isolated and single-cell sorted from secondary transgenic reprogrammable chimaeras[5,25]. iPS cells were reprogrammed using mouse naive ES cell medium 2i/LIF plus doxycycline (1 μg ml^{-1}) (without 2i in the first 48 h) (under physiological 5% pO_2 conditions for fibroblast cells). *Mbd3*$^{-/-}$ MEFs (but not pluripotent cells) experience accelerated senescence and proliferation capacity loss[26], and thus *Mbd3*$^{-/-}$ somatic cells were reprogrammed by applying tamoxifen on *Mbd3*$^{fl/-}$ cells only after 48 h of OSKM induction. Similarly, for acute knockdown of Mbd3 in somatic cells with *Mbd3* siRNAs in an attempt to boost reprogramming, transfections were conducted at least after 48 h of OSKM induction. Alternatively, somatic cells with hypomorphic expression (rather than complete ablation) of Mbd3 do not demonstrate proliferation defects or accelerated senescence due to the residual Mbd3 expression levels, yet they retain sufficiently reduced Mbd3 levels that allow deterministic synchronized reprogramming by OSKM (Fig. 2b). Thus, the following systems were preferably and predominantly used throughout this study: *Mbd3*$^{fl/-}$ or WIBR3-MBD3^{mut} genetic backgrounds. Notably, in our reprogramming conditions, single-cell plating of MEFs yielded approximately 70% survival efficiency (with or without doxycycline). Thus, for live imaging, upon plating 150 MEFs per well we observed formation of 100–120 colonies that were tracked in *Mbd3*$^{fl/-}$ samples. Notably, constitutive mCherry allowed one to control for survival after plating to obtain accurate and unbiased reprogramming efficiencies (reprogramming efficiency % = Oct4 or Nanog–GFP^+ clones (cells)/mCherry + clones (cells)). Equivalent reprogramming efficiencies were obtained on mouse irradiated feeder cells or gelatin-, Matrigel- or gelatin/vitronectin-coated plates (data not shown). Reprogramming on irradiated DR4 MEFs was preferably used for live imaging and single-cell reprogramming experiments to enhance cell survival and adherence. Mbd3 Stealth siRNA mix that includes MSS-237238, MSS-275658 and MSS-275659 components (Invitrogen), and Chd4 Stealth siRNA mix that includes MSS-200894, MSS-200895 and MSS-200896 (Invitrogen), were used for efficient knockdown in mouse cells. Transfections were conducted with RNAiMAX (Invitrogen) according to the manufacturer's instructions.

DNA plasmids and TALEN gene editing. The following lentiviral and mammalian constitutive overexpression vectors were used in somatic and pluripotent cells: pBRY-Mbd3-IRES-zeocin or pBRY-Mbd3-IRES-puromycin. Constitutively expressed lentiviruses FUW-Mbd2 and FUW-Mbd3 were generated by insert cloning into EcoRI sites of FUW lentiviral vector to generate constitutive expression following viral transduction and stable integration in somatic or iPS cell/ES cell lines. Flag–Mbd3 mutations and deletions were done by PCR with Q5 DNA polymerase (NEB). TALEN-expressing plasmids were designed and cloned using GoldenGate TALEN kit 2.0 purchased from Addgene[27] according to the published protocol. For targeting G we have used NN-type repeat. Donor construct was made with DNA fragments amplified from WIBR3 human ES cell genomic DNA. 10^7 ES cells were electroporated with 30 μg of donor plasmid and 10 μg each of the TALEN-expressing plasmids and grown in the presence of G418 (75 μg ml^{-1}) and ganciclovir (1 μM). Resistant clones were isolated and genomic DNA was extracted for Southern blot and PCR analysis. The WIBR3-MBD3^{mut} subcloned cell line was correctly targeted on both endogenous alleles as confirmed by 5′ and 3′ Southern blot strategies (not shown), and was subsequently used for generating a doxycycline-inducible secondary reprogramming system[28] (without excision of PGK-neo-pA cassette, to maintain interference with *MBD3* expression). For generating OCT4–GFP reporter subcloned cell lines, 10^7 MBD3^{WT} and MBD3^{mut} cells were electroporated with 30 μg of previously described OCT4-GFP-2A-PURO knock-in donor plasmid[29] (provided by R. Jaenisch through Addgene) and 10 μg each of the TALEN-expressing plasmids and grown in the presence of puromycin (0.4 μg ml^{-1}). Resistant clones were isolated and genomic DNA was extracted for Southern blot and/or PCR analysis.

Reprogramming of human somatic cells and cell infection. Reprogramming was conducted at 5% pO_2 in doxycycline (1–2 μg ml^{-1}) supplemented conditions. In the first 48 h, cells were incubated in conventional human ES medium (hESM; see below). Afterwards, cells were transferred until day 7–8 into modified hESM with ROCKi (Y-27632; 5 μM final concentration) and , more optimally, with 2i/LIF containing supplement. After 8 days doxycycline was withdrawn and cells were expanded as stable iPS cell lines. MBD3 Stealth siRNAs that include HSS147580 and HSS147581 components (catalogue number 1299003) were used for efficient MBD3 knockdown in human cells. Transfections were conducted with RNAiMAX (Invitrogen) according to the manufacturer's instructions. Conventional human ES conditions (hESM) include: 475 ml knockout DMEM (Invitrogen 10829), 15–20% KSR (Invitrogen), 8 ng ml^{-1} recombinant bFGF (Peprotech) and 1 ng ml^{-1} recombinant TGFβ1 (Peprotech), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma), 1% penicillin–streptomycin (Invitrogen). Mbd3^{mut} human ES cells were generated by genetic engineering with TALE nuclease effectors and differentiated *in vitro* into fibroblasts. Next, MBD3^{WT} and MBD3^{mut} iPS cells carrying doxycycline-inducible OKSM transgene were generated and labelled with constitutively expressed mCherry and targeted with an *OCT4-GFP* knock-in allele. *In vitro* differentiated fibroblasts from human ES cells/iPS cells were generated as previously described[28]. The Weizmann Institute ESCRO committee approved human cell line studies.

Immunofluorescence staining of pre- and post-implantation embryos. Immunostaining was performed as described previously with modifications[30]. Embryos were collected from the oviducts and uteri of hormone-primed B6D2F1 6-week-old females mated with C57BL/6 males. At least 15 embryos at each stage were analysed in total, and representative images were shown. Identical simultaneous staining and imaging analysis was conducted for all embryonic stages analysed. Briefly, embryos were transferred to a watch-glass dish (Genenet), fixed for 15 min in 4% PFA in phosphate buffer (PB), rinsed three times in PBS containing 3 mg ml^{-1} PVP, permeabilized in PBS/PVP with 0.5% Triton X-100 for 30 min, and blocked in blocking solution (2% normal donkey serum, 0.01% Tween in PBS) for 1 h. Embryos were then incubated overnight at 4 °C in primary antibodies diluted in blocking solution, washed three times in blocking solution for 15 min each, incubated with secondary antibodies for 1 h at room temperature, counterstained with DAPI for 15 min, washed twice in PBS, and placed in 96-well glass-bottom

plates for confocal imaging. Post-implantation embryos in the maternal decidua were fixed in 4% PFA/PB overnight at 4 °C, washed three times in PBS for 30 min each, dehydrated and embedded in paraffin using standard procedure. Embryonic paraffin sections (5–7 μm) were rehydrated, treated with antigen retrieval, rinsed in PBS, permeabilized in 0.1% Triton/PBS for 10 min, rinsed in PBT (0.02% Tween/PBS), and blocked in blocking solution (5% normal donkey serum, 0.05% BSA, in PBT) for 1 h. Slides were then incubated in the appropriate primary and secondary antibodies diluted in blocking solution as described above, and processed as described previously[5]. The following antibodies were used: mouse anti-Oct4 (1:100, C-10; Santa Cruz SC-5279), goat anti-Mbd3 (1:50, C-18; Santa Cruz SC-9402).

Immunoprecipitation and immunoblotting analyses. HEK293T cells were transfected with each cDNA clone in an expression vector using jetPEI (Polyplus transfection) and were lysed 48 h later in lysis buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 1% Triton, 0.1% NP40 and 1.5 mM EDTA). The following plasmids were used for transfections in different combinations: pCaggs-Mbd3, FUW-OCT4, FUW-KLF4, FUW-SOX2, FUW-c-MYC, FUW-Nanog, pCaggs-Flag-Mbd3, pMSCV-Flag-OCT4, pMSCV-Flag-SOX2, pMSCV-Flag-KLF4, pCaggs-Flag-c-MYC, pCaggs-Flag-Nanog, pcDNA3.1-Flag-HDAC1 (obtained through addgene). 30 μl of anti-FlagM2 Magnetic beads (Sigma) were incubated for 6 h in cell lysate fractions, for IgG control 6 μg of IgG and 50 μl of protein-G Dynabeads (Invitrogen) were added to the cell lysate for 6 h. Both fractions (the anti-Flag and anti-IgG) were loaded on Invitrogen magnetic separator and the beads were washed six times with lysis buffer. The binding proteins were eluted with 0.5 μg ml^{-1} of ×3Flag peptide (Sigma) buffer for the anti-FlagM2 beads or by boiling with sample buffer and analysed by SDS–polyacrylamide gel electrophoresis and immunoblotting. The immunoblot analyses were performed using the following primary antibodies: anti-Flag (clone M2, F3165, Sigma), anti-Mbd3 (A300-258A, Bethyl), anti-Nanog (A300-397A, Bethyl), anti-OCT4 (sc-9081, H134, Santa Cruz), anti-KLF4 (sc20691, H180, Santa Cruz), anti-SOX2 (2748s, Cell Signaling) and anti-c-Myc (9402s, Cell Signaling).

Mouse embryo micromanipulation and teratoma formation. Pluripotent stem cells (ES cells or iPS cells) were injected into BDF2 diploid blastocysts. Microinjection into blastocysts placed in M16 medium under mineral oil was done by a flat-tip microinjection pipette. A controlled number of 10–12 cells were injected into the blastocyst cavity. After injection, blastocysts were returned to KSOM media (Invitrogen) and placed at 37 °C until transferred to recipient females. Ten to fifteen injected blastocysts were transferred to each uterine horn of 2.5 days post coitum pseudo-pregnant females. Embryos were recovered for analysis at different time points during development or allowed to develop into full term. Determining germline transmission was performed by mating chimaeric animals with C57BL/6 females, and continuous checking for agouti-coloured pups. For teratoma formation and analysis, ES cells and iPS cells were collected by trypsinization before injection. Cells were injected subcutaneously into female 4–8-week-old NOD-SCID mice (Jackson laboratories). Tumours generally developed within 4–6 weeks and animals were killed before tumour size exceeded 1.5 cm in diameter. All animal studies were conducted according to the guideline and following approval by the Weizmann Institute IACUC (approval 00960212-3). We did not exclude animals from our analysis, and did not apply randomization by blinding.

RT–PCR analysis. Total RNA was isolated using the RNeasy kit (Qiagen). 3 μg of total RNA was treated with DNase I to remove potential contamination of genomic DNA using a DNA Free RNA kit (Zymo Research). 1 μg of DNase-I-treated RNA was reverse transcribed using a First Strand Synthesis kit (Invitrogen) and ultimately re-suspended in 100 μl of water. Quantitative PCR analysis was performed in triplicate using 1/50 of the reverse transcription reaction on Viia7 platform (Applied Biosystems). Error bars indicate standard deviation of technical triplicate for each measurement. For single-cell RT–PCR analysis, single cells from different samples were single cell plated, and Ambion Single Cell-to-CT kit was used for sample processing according to manufacturer instructions. TaqMan probe based chemistry and TaqMan Real-Time PCR master mix were used on Viia7 platform for gene expression detection. The following TaqMan (Invitrogen) probes were used: Sall4 Mm00453037_s1, Utf1 Mm00447703_g1; Sox2 (endogenous mouse allele specific) Mm03053810_s1; Nanog Mm02384862_g1; Gapdh Mm99999915_g1. Undetected expression indicates lack of amplification even after 50 amplification cycles (red boxes).

Immunocytochemistry and FACS analysis. Cells were fixed in 4% paraformaldehyde in PBS and immunostained according to standard protocols using the following primary antibodies: mouse anti-TRA-1-60 (1:500, Abcam; ab16288), mouse anti-TRA-1-81 (1:500, Abcam; ab16289), mouse anti-SSEA1 (1:100, Abcam; MC480 ab16285), mouse anti-SSEA4 (1:50, Abcam; MC813 ab16287), rat anti-SSEA3 (1:50, Abcam; MC631 ab16286), rabbit anti-Nanog (1:400, Bethyl; A300-397A), rabbit anti-Oct3/4 (1:400, Santa Cruz; H134 SC9081), mouse anti-Oct4 (1:200, Santa Cruz; C-10 SC5279), rabbit anti-Sox2 (1:500, Millipore; AB5603). Appropriate Alexa Fluor dye-conjugated secondary antibodies (1:200, Jackson ImmunoResearch) were used. FACS data were collected on BD FACS ARIA III and analysed with Flowjo software.

Microscopy image acquisition and analysis. Secondary OKSM inducible $Mbd3^{+/+}$ and $Mbd3^{fl/-}$ MEFs carrying the Oct4–GFP pluripotency reporter and constitutively expressed nuclear mCherry marker were plated in 24-well plates at low densities (~150 cells per well) and imaged using AxioObserver Z1 (Zeiss) in 5% O_2, 5% CO_2, 37 °C controlled conditions. Plates were taken out at day 3–4 for media replacement (but without passaging/splitting) and put back for the automated live imaging stage. Full-well mosaic images were taken every 12 h for 6 days at ×50 magnification, including phase contrast and two fluorescent wavelength images. In-house automated segmentation protocol was developed and implemented in Matlab to analyse time-lapse measurements of full-well mosaics with fluorescent mCherry and Oct4–GFP markers.

The challenge in this protocol was to implement fast segmentation of unknown number of colonies in 10^8 pixels mosaic image. The protocol includes the following main steps. Adaptive detection: erasing the plate margins with circular filter. Defining detection threshold using median with offset (10% of the dynamic range), and creating a binary image of detected pixels. These steps were carried out separately for each time point and each fluorescent wavelength. Complexity reduction: for this task we applied a morphological filter to isolate mCherry colonies using median sliding filter $(60_{[\mu m]} \times 60_{[\mu m]})$[31]. This filter retains only dense colonies, erasing noise and single isolated cells (a single nucleus is approximately $6_{[\mu m]} \times 6_{[\mu m]}$), this step is crucial for reducing the dimension of the clustering task. Colony segmentation: the segmentation was done using moving average filter (low-pass filter) $(60_{[\mu m]} \times 60_{[\mu m]})$[31] to merge adjacent colony fragments into large connected colonies and then apply connected components clustering, labelling connected objects using 8-connected neighbourhood. Colony feature extraction: extracting the features of each mCherry colony including area, bounding box and centroid. By overlaying mCherry colony segmentation on the GFP binary image (detected pixels) we extract for each colony the GFP^+ indicator (0/1) and the fraction of GFP^+ and $mCherry^+$ pixels out of all $mCherry^+$ pixels.

This segmentation protocol was run over time-lapse mosaics collecting information on colony formation dynamics, colony GFP^+ dynamics and ratios of offspring Oct4–GFP^+ cells. Colony and reprogramming dynamics features were then statistically analysed using Matlab program, including estimation of the cumulative distribution and density function. In addition, videos characterizing the process dynamics were produced using customized Matlab program. The above algorithm was validated by artificial input matrix and by ES mosaic image collection. In addition, robustness of detection threshold and filter size were measured with varying parameters (data not shown).

Chromatin immunoprecipitation and sequencing library preparation. Chromatin immunoprecipitation followed by deep sequencing (ChIP-Seq) was measured for the proteins H3K4me3, H3K27me3, H3K27ac and Mbd3 in four different time points throughout reprogramming: 0 (MEF), 4 days, 8 days, subcloned iPS or ES lines. The binding of each protein was measured in both $Mbd3^{+/+}$ and $Mbd3^{fl/-}$ conditions, as well as in $Mbd3^{-/-}$ ES cells. Oct4 was measured in all the above conditions, excluding 8 days. Approximately 40×10^6 cells were crosslinked in formaldehyde (1% final concentration, 10 min at room temperature), and then quenched with glycine (5 min at room temperature). Fixed cells were lysed in 50 mM HEPES KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 alternative, 0.25% Triton supplemented with protease inhibitor at 4 °C (Roche, 04693159001), centrifuged at 950*g* for 10 min and re-suspended in 0.2% SDS, 10 mM EDTA, 140 mM NaCl and 10 mM Tris-HCl. Cells were then fragmented with a Branson Sonifier (model S-450D) at −4 °C to size ranges between 200 and 800 bp, and precipitated by centrifugation. 10 μg of each antibody was pre-bound by incubating with Protein-G Dynabeads (Invitrogen 100-07D) in blocking buffer (PBS supplemented with 0.5% TWEEN and 0.5% BSA) for 2 h at room temperature. Washed beads were added to the chromatin lysate, and then incubated overnight. Samples were washed five times with RIPA buffer, twice with RIPA buffer supplemented with 500 mM NaCl, twice with LiCl buffer (10 mM TE, 250mM LiCl, 0.5% NP-40, 0.5% DOC), once with TE (10Mm Tris-HCl pH 8.0, 1mM EDTA), and then eluted in 0.5% SDS, 300 mM NaCl, 5 mM EDTA, 10 mM Tris HCl pH 8.0 at 65 °C. Eluate was incubated in 65 °C for 8 h, and then treated sequentially with RNaseA (Roche, 11119915001) for 30 min and proteinase K (NEB, P8102S) for 2 h. DNA was purified with The Agencourt AMPure XP system (Beckman Coulter Genomics, A63881). Libraries of cross-reversed ChIP DNA samples were prepared according to a modified version of the Illumina Genomic DNA protocol, as described previously[32]. Briefly, ChIP DNA was ligated to Illumina adaptors and subjected to 14 cycles of PCR amplification. Amplified products between 200 and 800 bp were purified on a 2% agarose gel. Roughly 5 pmol of DNA library was then applied to each lane of the flow cell and sequenced on Illumina Hiseq2000 sequencer according to standard Illumina protocols. The

following antibodies were used for chromatin immunoprecipitation experiments: control IgG (ChIP grade, ab46540, Abcam), anti-histone H3 trimethyl K4 (ChIP grade, ab8580, Abcam), anti-histone H3 acetyl K27 (ChIP grade, ab4729, Abcam), anti-histone H3 trimethyl K27 (ChIP grade, 07-449, Millipore), anti-Oct4 (sc5729 (C-10), Santa Cruz), anti-Chd4 (ChIP Grade, ab70469, Abcam). For Mbd3 chip 1:1 antibody mix was used: anti-Mbd3 (Bethyl laboratories A302-528\9A) and anti-Mbd3 (ab16057, Abcam). Chromatin immunoprecipitation data are available at the National Center for Biotechnology Information Gene Expression Omnibus database under the series accession number GSE49766.

Alignment and peak detection. We used bowtie software version 0.12.5 to align reads to mouse mm9 reference genome (UCSC, July 2007). We only considered reads that were uniquely aligned to the genome with up to a single mismatch, taking the single best match of each read. We identified enriched intervals of H3K4me3, H3K27me3, H3K27ac, Mbd3 and Oct4 using MACS version 1.4.1. We used sequencing of whole-cell extract as control to define a background model. Duplicate reads aligned to the exact same location are excluded by MACS default configuration. Enriched intervals were mapped to genes if they overlapped a single kilobase symmetric interval around their transcription start sites (TSS; taken from RefSeq known gene table in UCSC genome browser). ChIP-seq data on wild-type samples were highly compatible with those provided in previous publications[19,20].

Histone mark profiles. Histone mark profiles were calculated using in-house script. Briefly, this script generates a matrix of read densities in given genomic intervals. In this case, the profiles of all 29,952 Entrez genes (mm9, taken from UCSC known gene tables) were calculated between 1 kb upstream to TSS and TES (transcription end site). These read densities were then converted to z-score by normalizing each position by the mean and standard deviation of the sample noise ($\hat{X}_j = \frac{X_j - \mu_{\text{Noise}}}{\sigma_{\text{Noise}}}$), where X_j corresponds to read density, and μ_{noise} and σ_{noise} are the estimated noise mean and standard deviation, respectively . Noise parameters were estimated for each sample from 6×10^7 random base pairs across the genome. Finally, to present aligned profiles, the z-score profile of each gene was binned to 20 bins upstream to TSS and another 100 quantiles between TSS to TES. The value of each bin or quantile was selected to be the maximum value within that interval. In the histone mark distribution analysis and in the correlation and clustering of histone marks (Extended Data Fig. 5a), each gene and each histone mark is represented with the maximal z-score measured in the profile of that gene, where the profiles were calculated as described above. Clustering of histone marks was carried out on concatenated vectors that include all marks for every gene in tandem.

Annotation enrichment analysis. Mbd3 target genes were tested for enrichment of functional gene sets taken from Gene Ontology (GO, http://www.geneontology.org). Protein–DNA binding annotations were taken from various publications[33–36]. Enrichment P values were calculated using Fisher exact test[37] and corrected for multiple hypotheses using false discovery rate (FDR) threshold of 0.0001%.

Gene expression data acquisition. Total RNA was isolated from indicated cell lines. The concentration of RNA was quantified and subjected to quality control on Agilent Bioanalyzer. 250 ng of RNA was simultaneously processed from each sample. cDNA was fragmented, labelled, and hybridized to Affymetrix Mouse Gene 1.0 ST GeneChip (Affymetrix), which contain 35,557 probes. Transcript levels were processed from image files using RMA method[38], which corrects for non-biological sample variation using quantile normalization, implemented by the Affymetrix 'Expression Console' software. Microarray data are available at the National Center for Biotechnology Information Gene Expression Omnibus database under the series accession number GSE45352.

Gene expression analysis. Probes were mapped to Entrez Gene IDs and further filtered to include IDs that have at least one call higher than 32 ($=2^5$), resulting in 16,620 gene IDs. For gene expression analysis, we used Matlab version R2011b. Gene signatures differentially expressed between MEF samples (*Mbd3*$^{+/+}$, *Mbd3*$^{fl/-}$, *Mbd3*$^{-/-}$ MEF samples) and ES samples (ES V6.5, *Mbd3*$^{-/-}$ ES, Mbd3$^{fl/-}$ iPS and *Mbd3*$^{+/+}$ iPS) were characterized using a two-sample t-test and corrected for multiple hypotheses using false discovery rate (FDR)[39]. Differentially expressed gene signatures include genes that are under FDR threshold of 5%, as well as above fourfold change, resulting in 1,323 genes. Sample clustering with all 16,620 genes (Fig. 3a) was done with hierarchical clustering using Spearman correlation as a distance metric and average linkage. Single gene progression in reprogramming (Extended Data Fig. 8a) were quantified using the following transformation

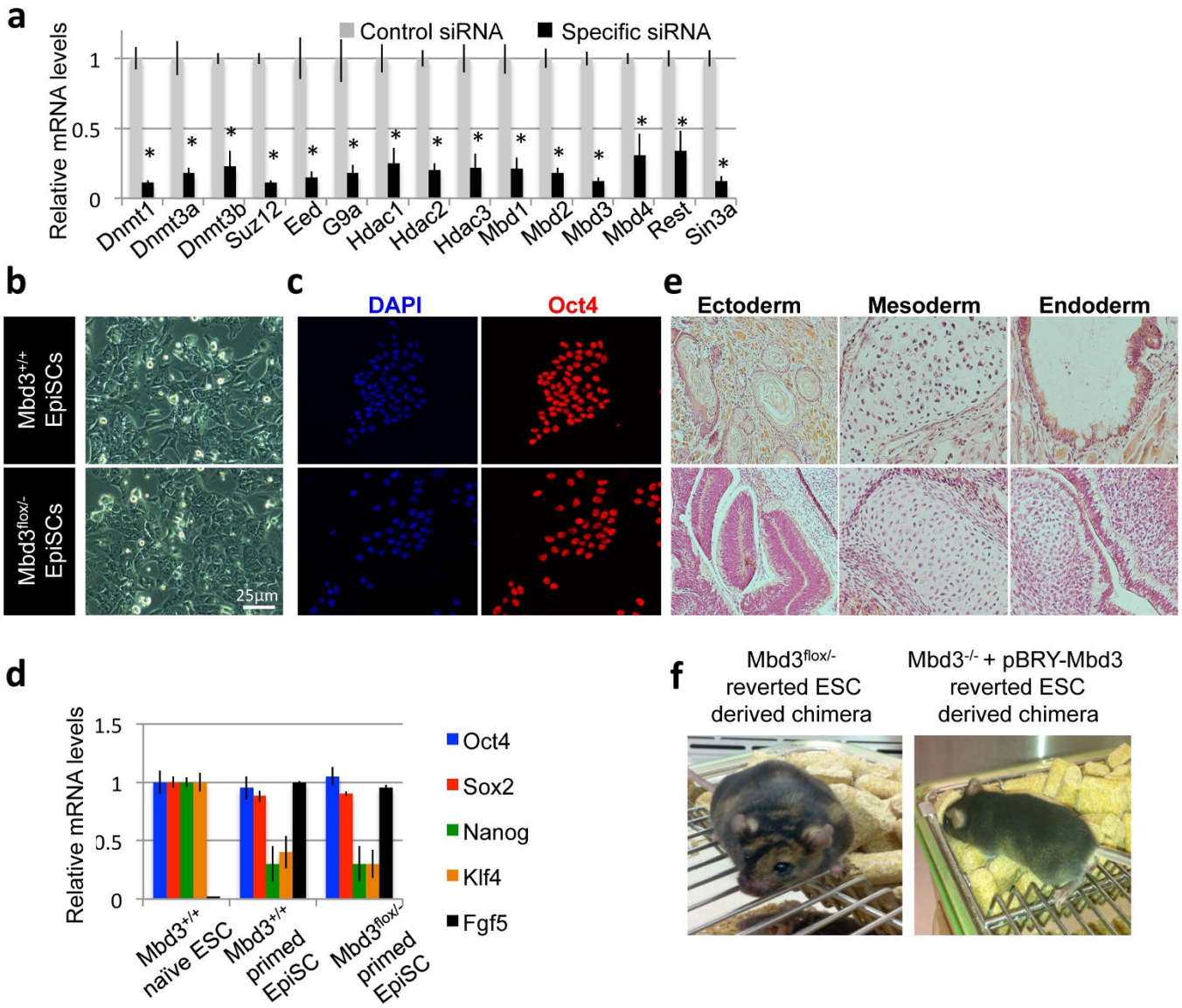
$$\hat{X}_j(t) = \max\left(\frac{X_j(t) - X_j(\text{MEF } Mbd3^{+/+})}{\bar{X}_j(\text{iPS}) - \bar{X}_j(\text{MEF})}, 0\right)$$

where, $X_j(t)$ denotes gene j expression value at time t (for example, X_j(4 days) or X_j(MEF)) and $\bar{X}_j$(iPS), $\bar{X}_j$(MEF) denote the averaged expression value for iPS and MEF samples, respectively. The above transformation represents a distance from MEF expression values (set to 0) towards iPS values (set to 1), where genes whose expression changes towards (up/down-regulating) their iPS value show $\hat{X}_j(t) > 0$. Distribution of gene expression fold-change, relative to MEF, is presented by box plots (Fig. 5c and Extended Data Fig. 8c). Distribution difference significance was calculated with a paired samples t-test.

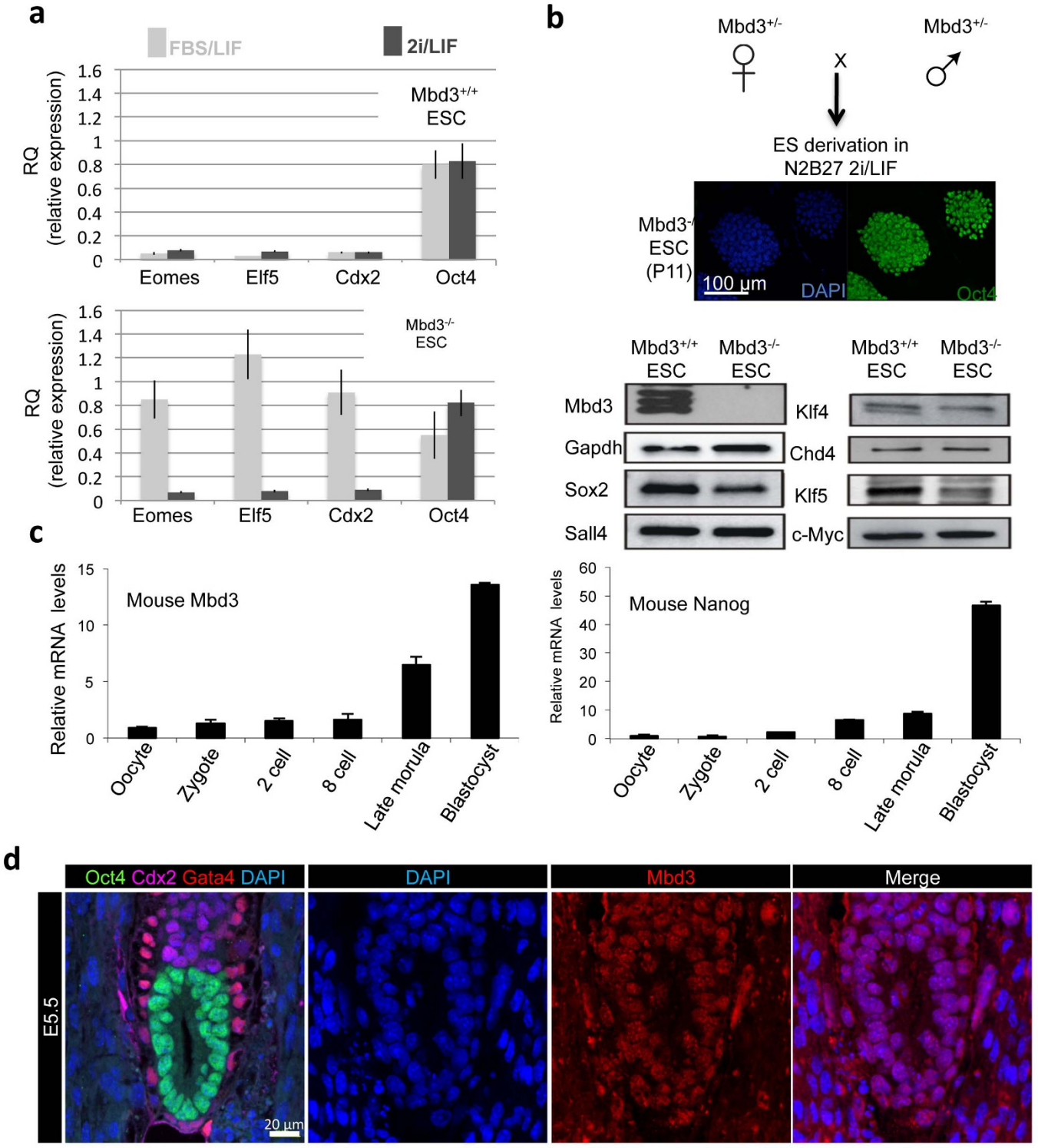
Preparation and analysis of reduced representation bisulphite sequencing libraries. RRBS libraries were generated as described previously with slight modifications[40]. Briefly, DNA was isolated from snap-frozen cell pellets using the Quick-gDNA mini prep kit (Zymo). Isolated DNA was then subjected to MspI digestion (NEB), followed by end repair using T4 PNK/T4 DNA polymerase mix (NEB), A-tailing using Klenow fragment (3′→5′ exo-) (NEB), size selection for fragments shorter than 500 bp using SPRI beads (Beckman Coulter) and ligation into a plasmid using quick T4 DNA ligase (NEB). Plasmids were treated with sodium bisulphite using the EZ DNA Methylation-Gold kit (Zymo) and the product was PCR amplified using GoTaq Hot Start DNA polymerase (Promega). The PCR products were A-tailed using Klenow fragment, ligated to indexed Illumina adapters using quick T4 DNA ligase and PCR amplified using GoTaq DNA polymerase. The libraries were then size-selected to 200–500 bp by extended gel electrophoresis using NuSieve 3:1 agarose (Lonza) and gel extraction (Qiagen). Libraries were pooled and sequenced on an Illumina HiSeq 2500 system. The sequencing reads were aligned to the Mouse Genome Build 37 (mm9) using Bismark. Methylation levels were calculated and averaged only for CpGs that were covered by 5 or more distinct sequencing reads across all libraries.

Numerical modelling analysis. Clonal reprogramming measurements and some data sets used were previously published[2]. NGFP1-*Mbd3*^{KD} and NGFP1-*Mbd3*$^{+/+}$ (control) reprogramming distributions were fitted to multiple modelling schemes, comparing reprogramming variability to known deterministic and stochastic dynamic models. This includes: (1) fitting to a deterministic step function, where the deterministic transition time was estimated by the fitting procedure. (2) Fitting to Gaussian distribution and calculating the mean, variance and coefficient of variation (CV = s.d./mean) for each sample. (3) Fitting to inverse Gaussian distribution, according to the 'first passage time of Brownian motion' model, and calculating the dynamic variability as the ratio of the Brownian motion standard deviation divided by the Brownian motion drift parameter. (4) Fitting to multi-step Markov chain (phase-type) model that infers possible structure for the reprogramming process. For this purpose, we constructed a nested fitting procedure for the fitting of *Mbd3*^{KD} and control *Mbd3*$^{+/+}$ dynamics to multiple models with 1 to 5 exponential transitions. We also compared reprogramming dynamics to a cell-cycle model, where we estimated the number of generations according to the reprogramming duration, and fit the cell cycle time distribution to the observed reprogramming latency. All model fittings were implemented by Matlab program performing nonlinear regression fitting with adjusted R^2 statistic and/or by using maximum likelihood estimator. For more detailed information see Supplementary Information numerical modelling analysis section.

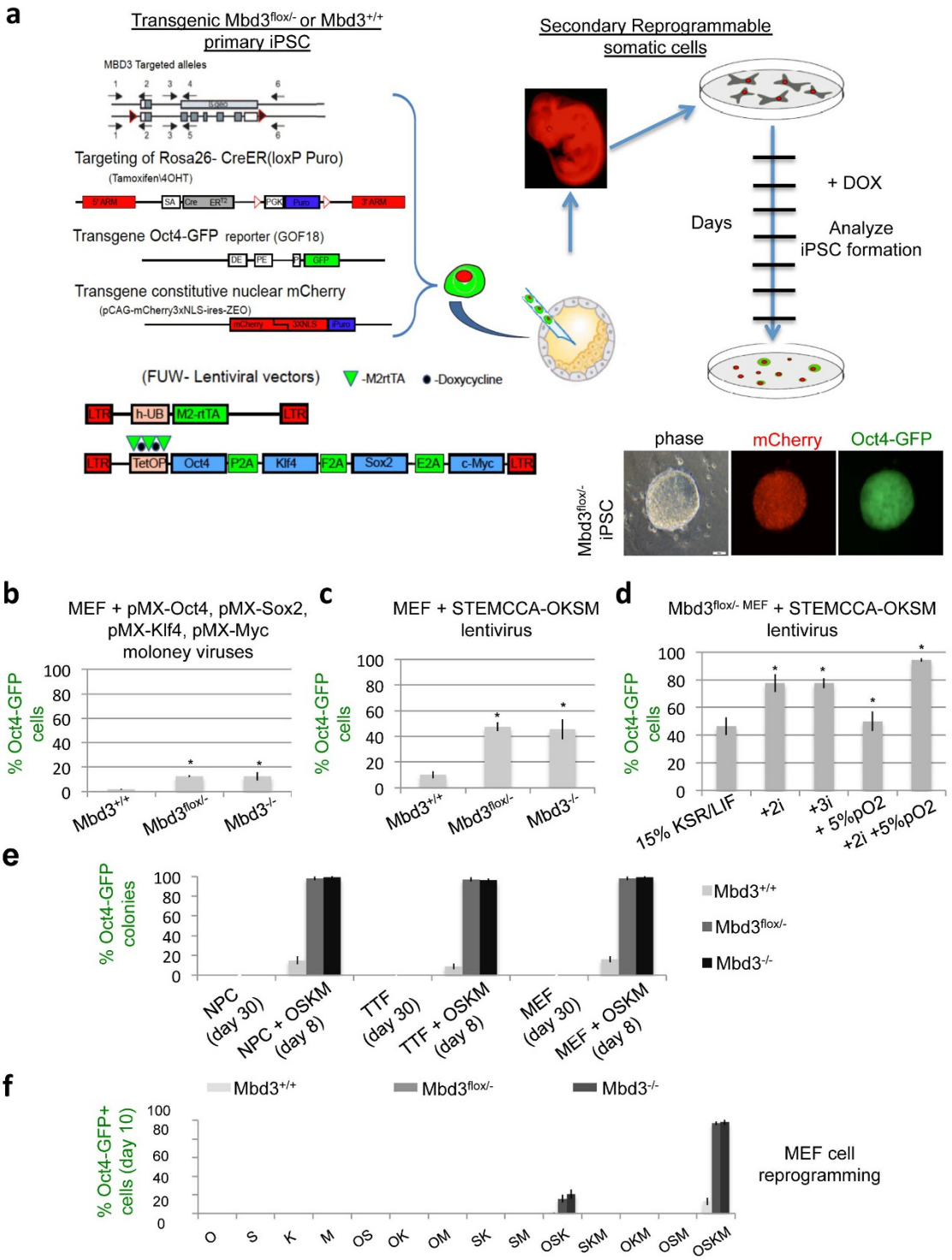
25. Hanna, J. *et al.* Metastable pluripotent states in NOD-mouse-derived ESCs. *Cell Stem Cell* **4,** 513–524 (2009).
26. Reynolds, N. *et al.* NuRD-mediated deacetylation of H3K27 facilitates recruitment of Polycomb Repressive Complex 2 to direct gene repression. *EMBO J.* **31,** 593–605 (2011).
27. Bedell, V. M., Wang, Y., Campbell, J. M. & Poshusta, T. L. *In vivo* genome editing using a high-efficiency TALEN system. *Nature* **491,** 114–118 (2012).
28. Hockemeyer, D. *et al.* A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. *Cell Stem Cell* **3,** 346–353 (2008).
29. Hockemeyer, D. *et al.* Genetic engineering of human pluripotent cells using TALE nucleases. *Nature Biotechnol.* **29,** 731–734 (2011).
30. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138,** 722–737 (2009).
31. Arce, G. R. *Nonlinear Signal Processing: A Statistical Approach* (Google Books, 2005).
32. Blecher-Gonen, R. *et al.* High-throughput chromatin immunoprecipitation for genome-wide mapping of *in vivo* protein-DNA interactions and epigenomic states. *Nature Protocols* **8,** 539 (2013).
33. Boyer, L. A., Gifford, D. K., Melton, D. A., Jaenisch, R. & Young, R. A. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122,** 947–956 (2005).
34. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448,** 553–560 (2007).
35. Kim, J. *et al.* An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132,** 1049–1061 (2008).
36. Loh, Y.-H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nature Genet.* **38,** 431–440 (2006).
37. Fisher, S., Genetiker, S., Fisher, R. A. & Genetician, S. *Statistical Methods for Research Workers* (Oliver and Boyd, 1970).
38. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4,** 249–264 (2003).
39. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* **57,** 289–300 (1995).
40. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484,** 339–344 (2012).



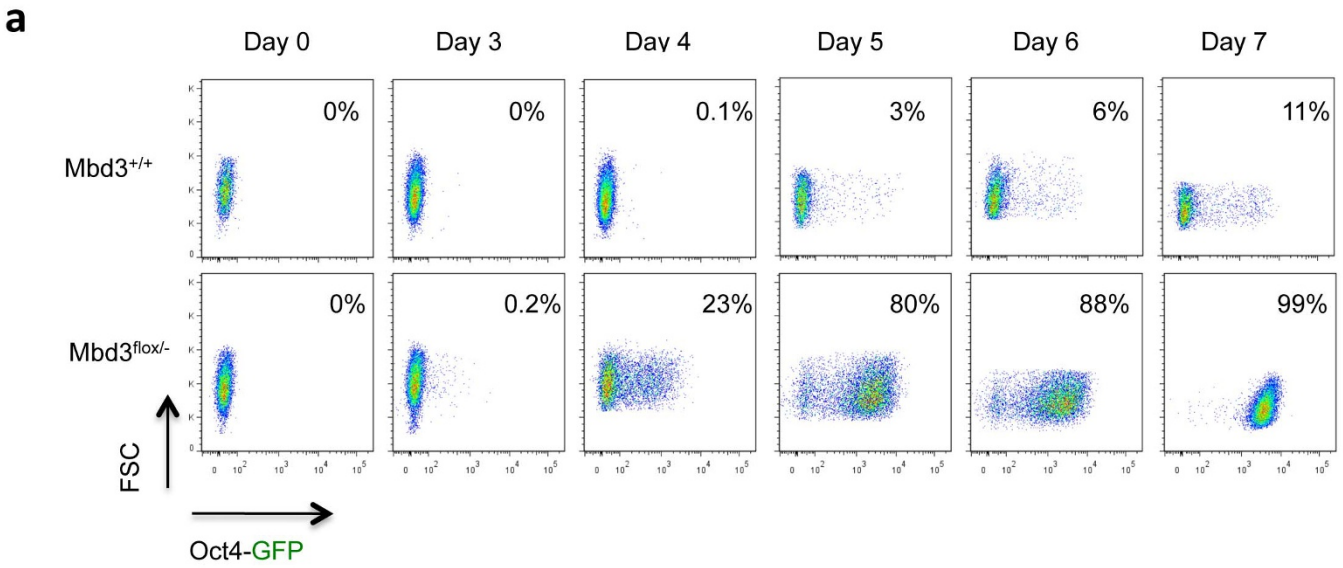
Extended Data Figure 1 | Knockdown screen for epigenetic repressors in EpiSCs. **a**, Knockdown efficiency of the indicated siRNA pools in EpiSCs measured by qRT–PCR. Expression values for each gene were normalized to those measured in control siRNA. Error bars indicated s.d. from average. Asterisks indicate *t*-test *P* value <0.05. **b**, Phase images of *Mbd3*$^{+/+}$ and *Mbd3*$^{fl/-}$ EpiSC lines in this study. **c**, Oct4 immunostaining on EpiSC lines. **d**, RT–PCR expression level validation for pluripotency genes in naive V6.5 ES cells and primed *Mbd3*$^{+/+}$ and *Mbd3*$^{fl/-}$ EpiSCs. In comparison to naive ES cells, primed EpiSCs downregulate naive pluripotency markers Nanog and Klf4, and upregulate FGF5 transcription ($n = 3$). **e**, EpiSC lines were pluripotent as evident by their ability to form mature differentiated teratomas. **f**, Representative agouti-coloured chimaeras obtained from reverted EpiSCs after Mbd3 depletion.

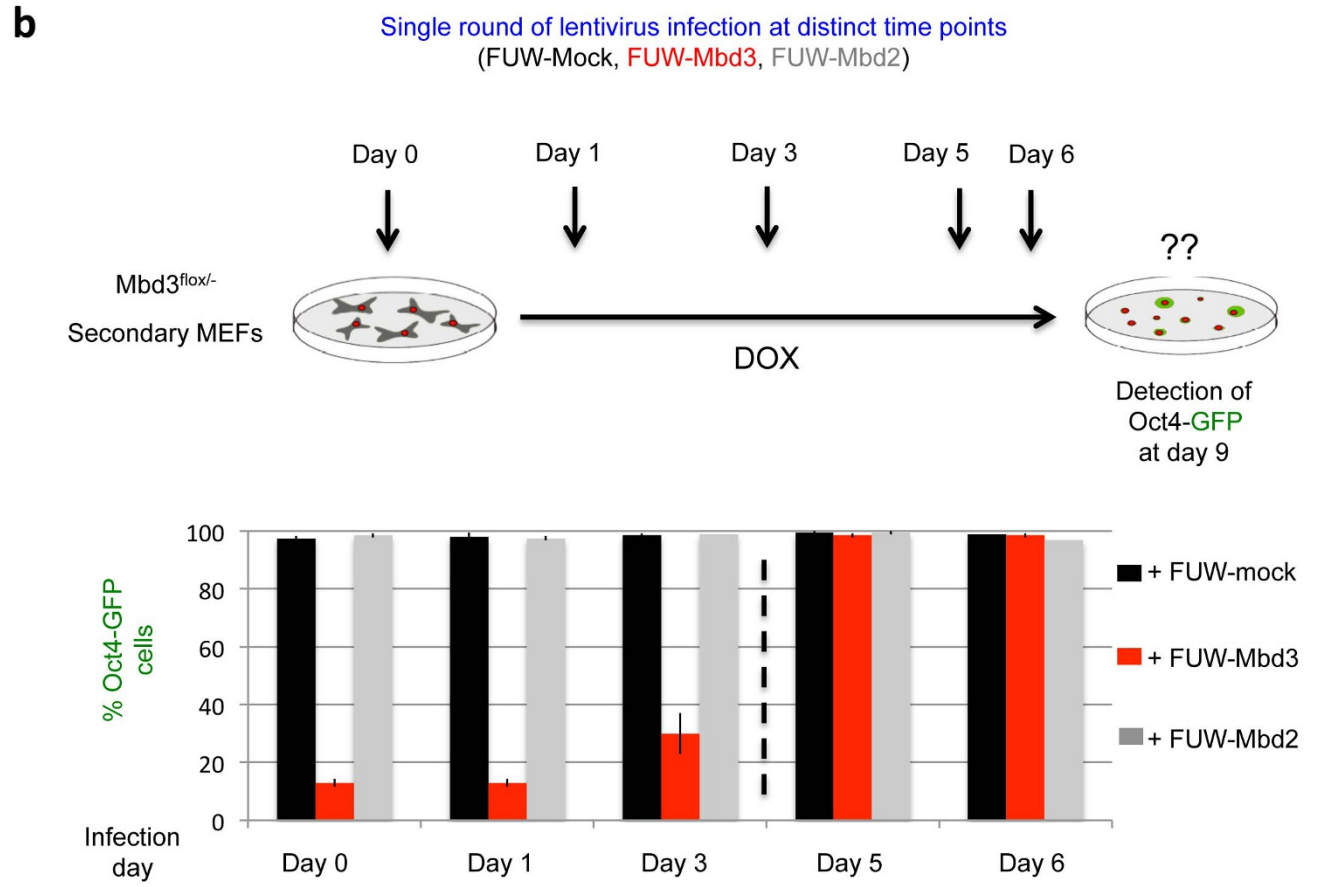


Extended Data Figure 2 | Derivation of ES cells from *Mbd3*$^{-/-}$ blastocysts. **a**, RT–PCR analysis for Oct4 and trophoblast marker expression of *Mbd3*$^{+/+}$ and *Mbd3*$^{-/-}$ ES cells expanded either in FBS/LIF or 2i/LIF conditions. Only *Mbd3*$^{-/-}$ ES cells, and only in serum conditions, upregulate trophoblast differentiation markers. Error bars indicate s.d. from average ($n = 3$). **b**, *Mbd3*$^{+/-}$ heterozygous mice were mated, and ES cells were derived from blastocysts in naive defined 2i/LIF conditions. Western blot for pluripotency marker expression also indicated that the derived *Mbd3*$^{-/-}$ ES cell lines adequately expressed all pluripotency factors tested. **c**, Transcriptional expression of Mbd3 and Nanog during pre-implantation development. RT–PCR analysis demonstrating the expression of Mbd3 during early mouse development, presented as a relative quantification column scheme. Error bars indicate s.d. from average ($n = 3$). *Mbd3* transcript is detected at low levels in oocytes whereas Mbd3 protein is weakly detected by immunostaining in oocytes and zygotes (Fig. 1e), consistent with maternal inheritance. *Mbd3* transcription becomes increased towards the end of pre-implantation development at the morula and blastocyst stages, consistent with strong re-expression of Mbd3 protein at the blastocyst stage (Fig. 1e). **d**, Immunostaining for Mbd3 and lineage markers in E5.5 post-implantation epiblast, indicating prominent expression ($n = 3$ embryos stained).

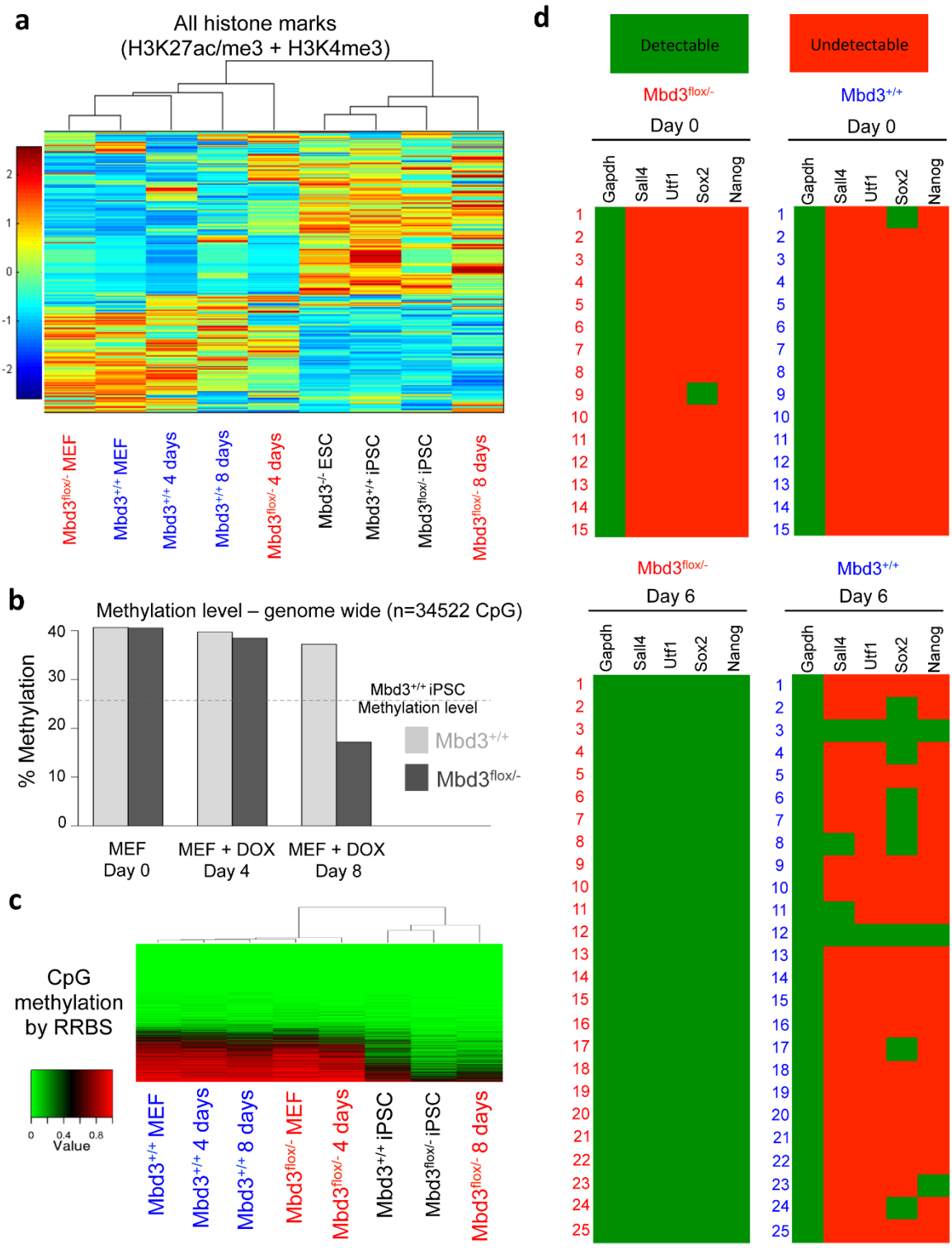


Extended Data Figure 3 | Genetically engineered systems for deterministic reprogramming in mouse cells. **a**, We established a reprogrammable mouse *Mbd3*$^{+/+}$ and *Mbd3*$^{fl/-}$ iPS cell lines carrying (1) an Oct4–GFP reporter, (2) nuclear mCherry constitutively expressed marker, (3) m2RtTa transgene and (4) a TetO inducible STEMCCA-OKSM polycistronic cassette. These lines were injected into host blastocysts, and their differentiated derivatives were re-isolated *in vitro*. Subsequently, reprogramming efficiency and progression were analysed after doxycycline induction. **b**, Reprogramming efficiency after infection with indicated MEF lines with moloney retroviruses encoding individual factors. **c**, Reprogramming efficiency after infection with indicated MEF lines with polycistronic OKSM encoding lentivirus. **d**, *Mbd3*$^{fl/-}$ MEFs were infected with polycistronic OKSM vector in LIF-containing ES medium with or without the indicated exogenous supplements. Reprogramming efficiency was evaluated by Oct4–GFP levels on day 9 after transduction without cell splitting during the process. **e**, *Mbd3*$^{+/+}$, *Mbd3*$^{-/-}$ and *Mbd3*$^{fl/-}$ MEFs, adult tail-tip-derived fibroblast (TTF) and neural precursor cells (NPC) were tested for iPS cell formation in 2i/LIF with or without OKSM lentiviral transduction. Our analysis indicates that OKSM is essential for iPS formation, and that Mbd3 depletion alone is not sufficient to reprogram any of these cells types to pluripotency (even after 30 days of follow up). **f**, Reprogramming efficiency of MEFs after transduction with the indicated combinations of reprogramming factors at day 10. Polycistronic lentiviral vectors were used for OSK and OSKM combinations. Asterisk indicates *t*-test *P* value <0.01 relative to *Mbd3*$^{+/+}$ control. Error bars indicate s.d. from average ($n = 4$).

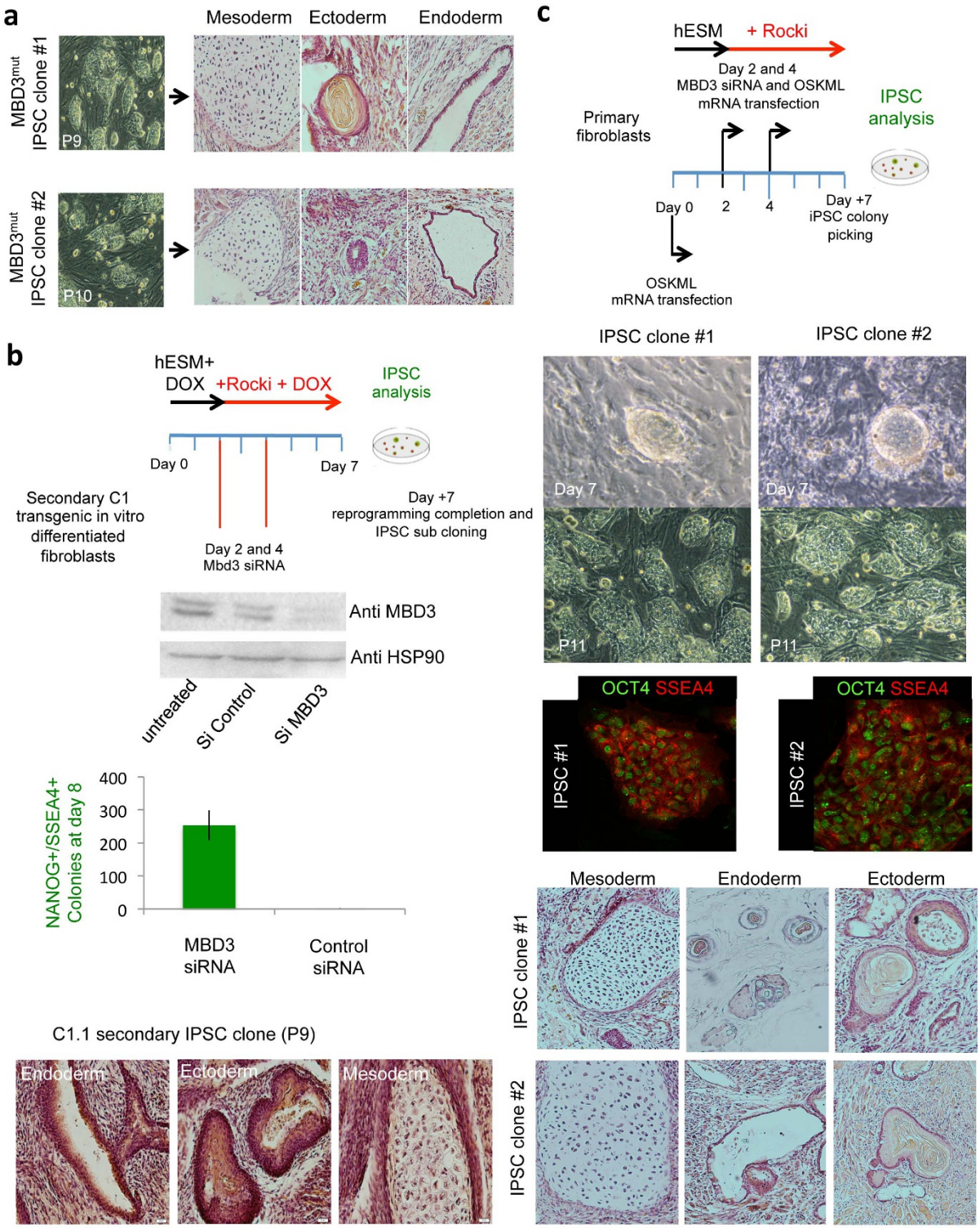




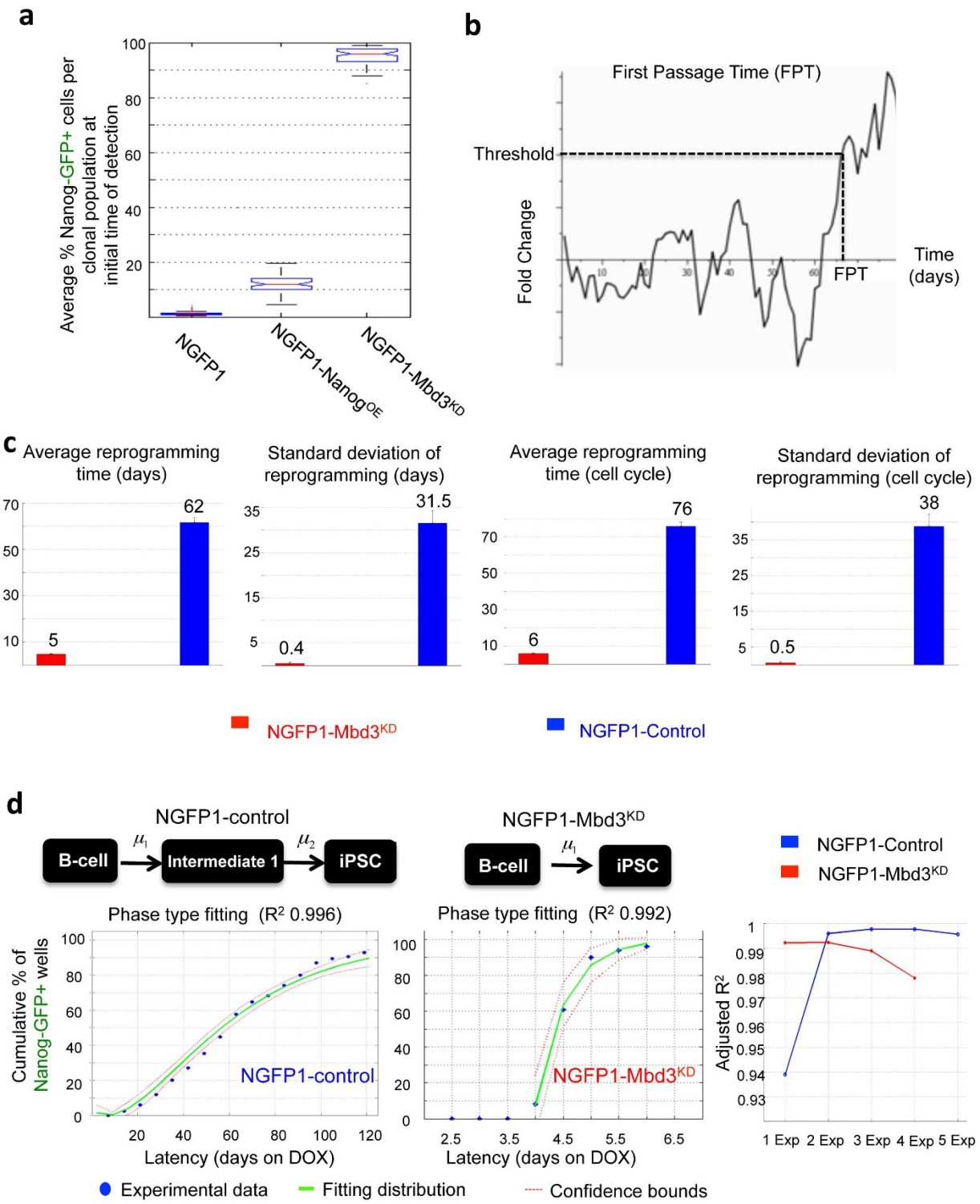
Extended Data Figure 4 | Reprogramming kinetics on perturbation of Mbd3 expression. a, Flow cytometry measurements of Oct4–GFP reactivation dynamics in 2i/LIF after doxycycline (OSKM) induction. Notably, wells at the indicated time points were collected for analysis without prior passaging and splitting during the reprogramming course. 1 out of 3 independent experiments is shown. FSC, forward scatter. **b**, Characterizing the effect for Mbd3 expression reconstitution during deterministic reprogramming of somatic cells to pluripotency. Scheme demonstrates experimental strategy for defining the temporal ability of Mbd3 during reprogramming to inhibit iPS formation. Secondary OSKM reprogrammable *Mbd3*$^{fl/-}$ MEFs were tested for their amenability to reprogramming after overexpression of Mbd3, Mbd2 or empty FUW lentiviruses at different time points during reprogramming. Mbd2 or mock-vector transfection did not result in a decrease in iPS cell reprogramming efficiency. Error bars indicate s.d. from average ($n = 3$). One out of two representative data sets is shown.



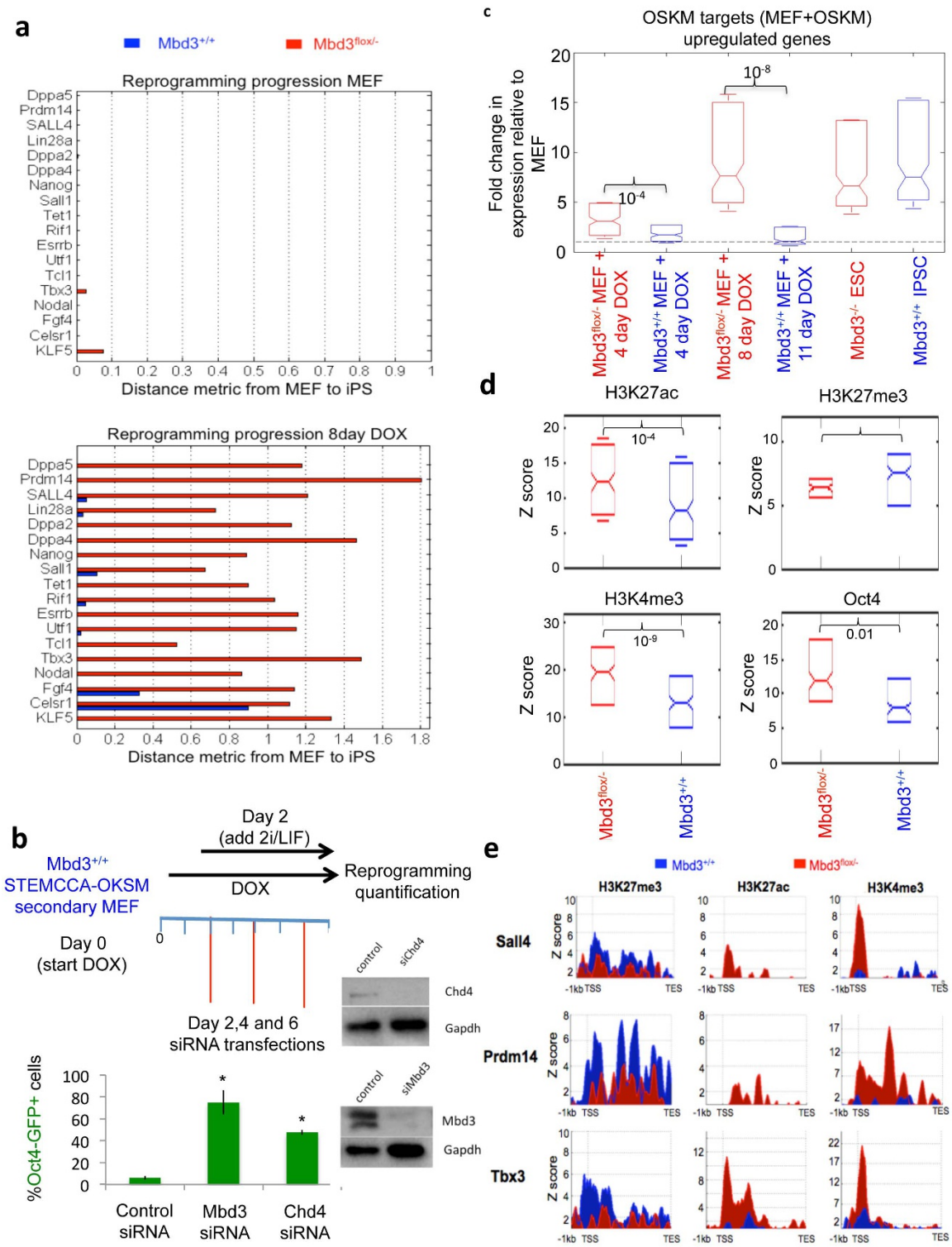
Extended Data Figure 5 | Genetic and epigenetic changes during iPS cell reprogramming after Mbd3 depletion. **a**, Hierarchical clustering was carried out on chromatin IP-seq measurements in fibroblasts before and after doxycycline induction. Clustering was calculated over concatenate vectors including *z*-scores of all histone marks (H3K4me3, H3K27me3 and H3K27ac) for each gene ($n = 1{,}323$ genes with differential gene expression between MEFs and ES cells). Spearman correlation was used as a distance metric and average linkage. **b**, Graph shows genome-wide methylation levels as measured by reduced representation bisulphite sequencing (RRBS). Results are averaged over all CpGs that were covered by five or more distinct sequencing reads (34,522 CpG sites in total). The average methylation level of low-passage $Mbd3^{+/+}$ iPS cells is provided as a dashed line for reference. **c**, Hierarchical clustering for CpG methylation was made using Ward's method and the Pearson correlation score as the similarity matrix. **d**, Single cell RT–PCR analysis for detection of pluripotency gene markers. Analysis was conducted on $Mbd3^{+/+}$ and $Mbd3^{fl/-}$ MEFs before and 6 days after doxycycline induction. Undetected expression (marked by red boxes) indicates lack of amplification even after 50 amplification cycles are marked in red. Expressed genes are marked by green boxes. One biological replicate is shown of two performed.



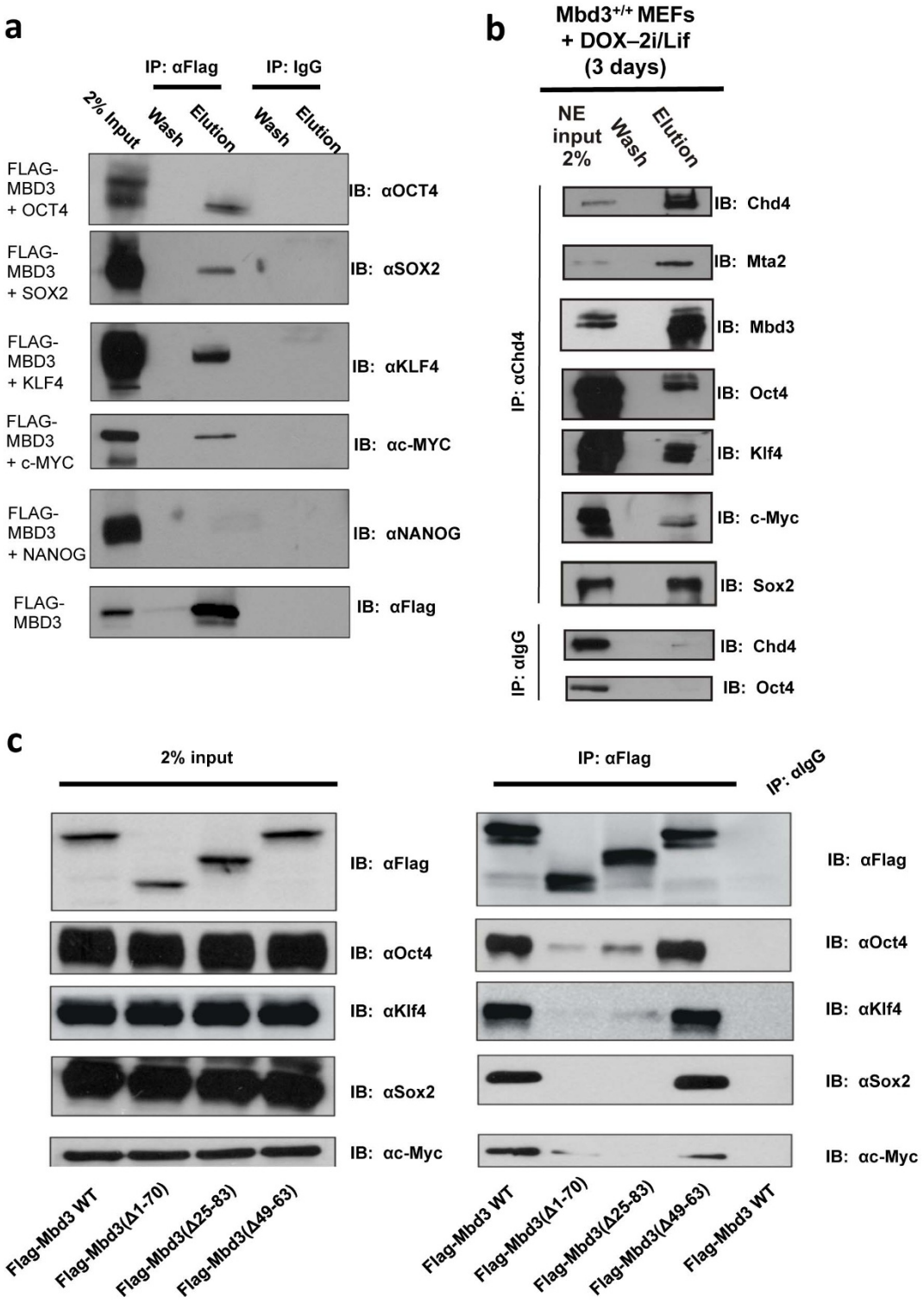
Extended Data Figure 6 | Depleting Mbd3 expression facilitates human iPS cell formation. **a**, *In vitro* differentiated fibroblasts from MBD3^{WT} and MBD3^{mut} iPS cells carrying the doxycycline-inducible OKSM transgenes, were reprogrammed as indicated in Fig. 3f. Pluripotency of randomly selected iPS cell clones is shown as evident by teratoma. **b**, Secondary human reprogrammable C1 fibroblasts carrying doxycycline-inducible OSKM transgenes were subjected to the depicted reprogramming protocol. Knockdown of Mbd3 at days 2 and 4, but not with scrambled control siRNA, markedly increased the reprogramming efficiency as evaluated by formation of NANOG/SSEA4^{+} colonies. Pluripotency of a randomly selected iPS cell clone expanded and validated by *in vivo* teratoma formation. Western blot confirmed specific and significant decrease in MBD3 protein expression after *MBD3* siRNA transfection. Error bars indicate s.d. from average ($n = 3$). One out of three representative experiments is shown. **c**, *MBD3* siRNA treatment of human primary fibroblasts allows generation of iPS cells by only two rounds of reprogramming with mRNA transfection with OSKM and LIN28 (OSKML) factors. Representative human iPS cell clones are shown at different time points and passages (P indicates passage number). Pluripotency of randomly selected clones is shown by specific staining for OCT4 and SSEA4 pluripotency markers and teratoma formation. These results indicate that inhibition of MBD3 expression and/or function promotes iPS cell formation by transient mRNA or other transient transfection protocols for iPS cell reprogramming.



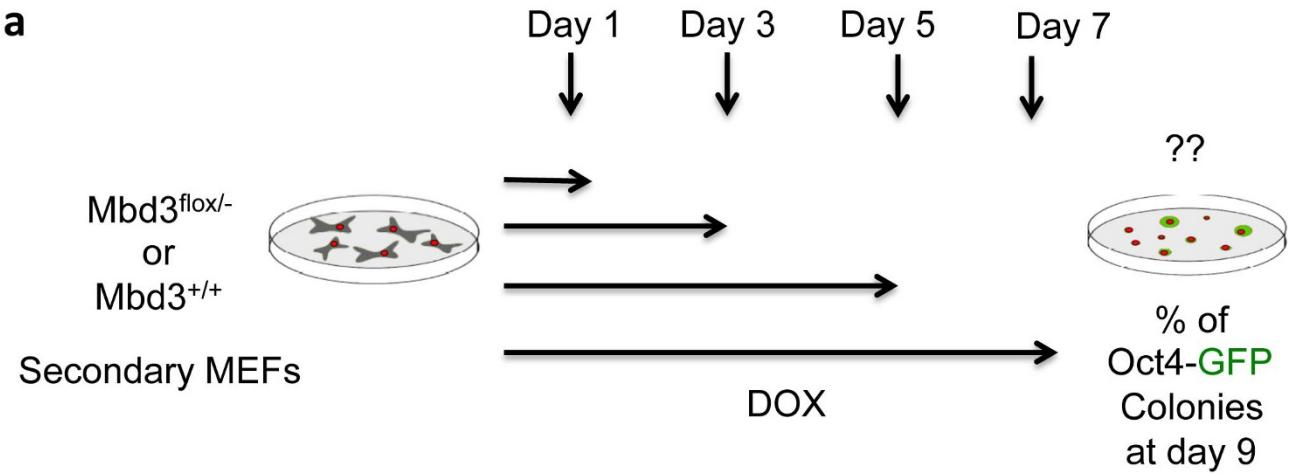
Extended Data Figure 7 | Statistical analysis of iPS cell reprogramming after Mbd3 depletion. **a**, Distribution of Nanog–GFP^{+} cells at initial time of detection, by quantifying the amount of Nanog–GFP^{+} cells detected above the 0.5% threshold. Graphs show box-plot medians and 25th/75th percentiles. **b**, Illustration of the first passage time model. In this model, we assume that reprogramming time depends on the first time in which some master regulator (that is, Nanog or Oct4) makes a transition from a low state to a high state of expression. **c**, *Mbd3*^{KD} and *Mbd3*$^{+/+}$ reprogramming dynamics were fit to Gaussian distribution. Figures show maximum likelihood estimates of mean and standard deviation, with 95% confidence intervals. **d**, *Mbd3*^{KD} and *Mbd3*$^{+/+}$ reprogramming dynamics were fit to multiple tandem rate-limiting step models, where convergence of adjusted R^2 indicates the best fit (right panel). Results show that *Mbd3*$^{+/+}$ (blue) fit best to a multi-phase process with one or two intermediate states, whereas *Mbd3*^{KD} (red) fit best to a single exponential transition with no intermediate states.



Extended Data Figure 8 | Effect of Mbd3 depletion on OSKM target genes. **a**, Normalized single gene expression for selected group of genes in MEF and 8 days after doxycycline induction. Expression values represent distance from MEF expression values (set to 0) towards iPS values (set to 1), indicating absence of transcription in MEF and fast activation after doxycycline induction in Mbd3 depleted samples. **b**, Reprograming efficiency of *Mbd3*$^{+/+}$ secondary MEFs after knockdown of Mbd3 or Chd4. Error bars indicate s.d. from average ($n = 3$). Asterisks indicate Student's *t*-test *P* value <0.01. Western blot indicating protein depletion efficiency on siRNA transfection of either Mbd3 or Chd4 targeting siRNA pools. **c**, Distribution of gene expression fold-change relative to MEF, calculated over 2,928 genes bound by at least one of the OSKM factors[20] and upregulated during reprogramming. Graphs show box-plot medians and 25th/75th percentiles, and *P* values by paired sample *t*-test. **d**, Distribution of histone marks and Oct4 binding levels in *z*-score values at day 4 after OSKM (doxycycline) induction, calculated over the same set of 2,928 genes described above. **e**, Histone mark *z*-score profiles for three representative OSKM target genes, calculated between 1 kb upstream to TSS and TES.

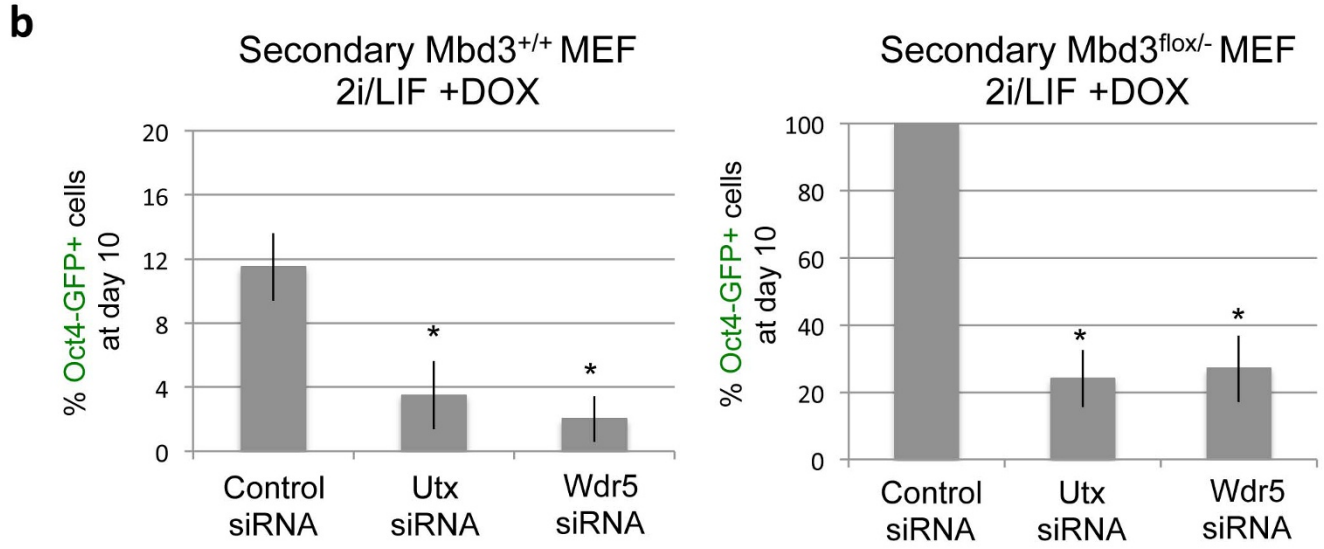


Extended Data Figure 9 | Direct interaction of Mbd3 with OSKM pluripotency factors during reprogramming. **a**, Overexpression of Flag-tagged Mbd3 simultaneously with OCT4, SOX2, KLF4, MYC or Nanog in HEK293 cells was followed by co-immunoprecipitation (co-IP) assay. Immunoblot analysis (IB) using antibodies against Oct4, Sox2, Klf4, Myc and Nanog showed specific binding between Mbd3 and the pluripotent factors except Nanog ($n = 2$). **b**, Co-immunoprecipitation assay of Chd4 (Mi2b), the core subunit of the NuRD complex, in secondary *Mbd3*$^{+/+}$ fibroblasts 3 days after doxycycline induction. Co-immunoprecipitation for NuRD component, Chd4, followed by immunoblot analysis indicated specific pull-down of other Mbd3/NuRD components (Mbd3 and Mta2) and OSKM reprogramming factors ($n = 3$). **c**, Deletion mutations in the MBD site of Mbd3 was planned to find the binding region of Mbd3. Flag-tagged mutation constructs were co-transfected with Oct4, Sox2, Klf4 and Myc in HEK293T cells for 48 h followed by co-immunoprecipitation with anti-Flag beads and immunoblotted against OSKM. This analysis shows loss of binding and interaction between OSKM and selected Mbd3 mutants ($n = 3$).



DOX stopped at:

	Day 1	Day 3	Day 5	Day 7
Mbd3flox/-	0%	4%	100%	100%
Mbd3+/+	0%	2%	6%	7%



Extended Data Figure 10 | Pluripotency-promoting epigenetic activators are essential for both deterministic and stochastic iPS cell formation. **a**, Requirement for doxycycline-mediated transgene induction during iPS cell reprogramming form *Mbd3*$^{+/+}$ and *Mbd3*$^{fl/-}$ secondary MEFs. Percentage of Oct4–GFP colonies was quantified at final set time point on day 9. Similar time frame for minimal doxycycline induction was required for iPS cell formation in both cell samples (irrespective of the total iPS formation efficiency obtained). Representative data from one out of three biological replicates conducted. **b**, **c**, Specific knockdown of Utx and Wdr5 epigenetic regulators that are required for iPS cell formation significantly inhibited iPS cell formation in both *Mbd3*$^{+/+}$ and *Mbd3*$^{fl/-}$ cells. Asterisks indicate *t*-test *P* value <0.01 in comparison to control siRNA sample. Error bars indicate s.d. from average ($n = 3$).