Deterministic direct reprogramming of somatic cells to pluripotency


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 naïve 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). 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 latencies7,8. A variety of chromatin modifiers have been implicated in facilitating epigenetic changes leading to authentic iPS cell reprogramming8. Despite these advances, the reprogramming efficiency of somatic cells remains extremely low7. Furthermore, the outcome of challenging the somatic epigenome with the overexpression of OSKM reprogramming factors is stochastic7. 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 factors7,8,9. The identity of such stochastic rate-limiting element(s) remains to be defined. Here we show that the Mbd3/NuRD repressor complex is the predominant molecular block preventing deterministic induction of ground-state pluripotency.

Mbd3 and establishment of naïve 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 cells8,9. 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 naïve pluripotent state in 2i/LIF growth conditions (where 2i is ERK1/2 and GSK3β inhibitors, and LIF is leukaemia inhibitory factor)10. We used a primed EpiSC line carrying a Nanog–GFP knock-in reporter that can be reactivated in the naïve state11, and applied short interfering RNA (siRNA) screening to identify boosters of EpiSC reversion into Nanog–GFP+ naïve 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 cells12. Mbd2 and Mbd3 assemble into mutually exclusive distinct NuRD complexes12, which can mediate gene repression through histone deacetylation and chromatin remodelling activities. To validate the siRNA screening results, we used Mbd3−/− and Mbd3fl/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). Mbd3fl/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 deficiency10,12) and Mbd3fl/fl cells can contribute to chimera formation (Extended Data Fig. 1f). Reconstitution of Mbd3 expression in Mbd3−/− and Mbd3fl/fl EpiSCs inhibited reversion efficiencies (Fig. 1b). These results directly demonstrate that reduction of Mbd3 protein levels renders nearly complete reversion of EpiSCs to naïve pluripotency.

We revisited ES-cell derivation experiments from Mbd3−/− E3.5 embryos12, and were able to isolate Mbd3−/− ES cells in serum-free 2i/LIF
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)\textsuperscript{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\textsuperscript{11}.

Finally, we aimed to test the influence of reducing Mbd3 expression in deterministically reprogramming somatic cells from mouse embryonic fibroblasts (Fig. 2a). To evaluate reprogramming efficiency and quantification for EpiSC reprogramming from different mutant lines. 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\textsuperscript{15}, and a negative effect on human primed iPS cell induction\textsuperscript{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 Mbd3fl/– and Mbd3fl–/– ES cell lines were Oct4–GFP\textsuperscript{+} at day 10, whereas only levels up to 18% were observed in control Mbd3fl/+ fibroblasts (Fig. 2a). To evaluate reprogramming efficiencies accurately, we established ‘secondary reprogrammable’ Mbd3fl/+ and Mbd3fl–/– transgenic cell lines harbouring a doxycycline-inducible OKSM polycistronic cassette\textsuperscript{18}, a constitutive nuclear mCherry reporter (to track individual cells and control for plating efficiency), and an Oct4–GFP reporter (Extended Data Fig. 3a). Single cell sorting of secondary mCherry Mbd3fl/+ 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 Mbd3 expression facilitates access to ground-state pluripotency from early embryonic Oct4-expressing cells.

**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\textsuperscript{+} 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. Mbd3fl–/– 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. Mbd3fl/+ and Mbd3fl–/– 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 Mbd3fl–/–. All error bars indicate s.d. from average.
cells reprogrammed under identical conditions, no more than 20% of clones reactivated Oct4–GFP (Fig. 2b). Teratomas (not shown) and chimaeras were obtained from iPSC 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’ Mbd3fl/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 Mbd3fl/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 Mbd3fl/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. 2d). The latter unbiased quantitative analysis demonstrated a markedly intra- and interclonal synchronized reactivation of Oct4–GFP occurring during a narrow window in Mbd3fl/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 iPSC 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 iPSC cell generation from Mbd3fl/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 iPSC cell and ES cell lines. Notably, by day 8, Mbd3fl/fl donor cells were transcriptionally indistinguishable from multiple ES cell lines and subcloned established iPSC 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 Mbd3fl/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 iPSC cell/ES cell-like methylation pattern could be seen in the Mbd3fl/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 Mbd3fl/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 cells19,20 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 Mbd3fl/fl OSKM-transduced partially reprogrammed cells and attempted to complete their reprogramming by Mbd3 inhibition. Notably, by introducing Mbd3 siRNA, all clones 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 iPSC cell reprogramming, we generated MBD3mut 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 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 Mbd3fl/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 iPSC cell/ES cell-like methylation pattern could be seen in the Mbd3fl/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 Mbd3fl/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 cells19,20 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 Mbd3fl/fl OSKM-transduced partially reprogrammed cells and attempted to complete their reprogramming by Mbd3 inhibition. Notably, by introducing Mbd3 siRNA, all clones19 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 iPSC cell reprogramming, we generated MBD3mut 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 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 Mbd3fl/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 iPSC cell/ES cell-like methylation pattern could be seen in the Mbd3fl/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 Mbd3fl/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.
LIN28 mRNA transfection\textsuperscript{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)\textsuperscript{5}. A secondary OSKM transgenic NGFP1-iPS cell line\textsuperscript{2}, carrying a Nanog–GFP reporter, was rendered transgenic for a doxycycline-inducible Mbd3 shRNA construct (NGFP1-Mbd3\textsuperscript{22}). Indeed, NGFP1-Mbd3\textsuperscript{KD} derived monoclonal B-cell populations converted into Nanog–GFP\textsuperscript{+} 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 detection on polyclonal NGFP1-control and NGFP1-Mbd3\textsuperscript{KD} B-cell populations. We next tested whether Mbd3\textsuperscript{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\textsuperscript{2} > 0.9, chi-squared = 52) of Mbd3\textsuperscript{KD} cells, but not of control Mbd3\textsuperscript{+}/+ cells (R\textsuperscript{2} = 0.55, chi-squared = 405) (Fig. 4c). Despite the observed similarity to a deterministic behaviour, variability was still evident in our Mbd3\textsuperscript{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\textsuperscript{+}/+ and Mbd3\textsuperscript{KD} samples, and further compare it to the inherent measured cell-cycle variability\textsuperscript{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 variability (s.d./drift) by a Brownian motion (BM) model (Extended Data Fig. 7b). Graphs show maximum likelihood estimations. Error bars indicate 95% confidence intervals of maximal likelihood value (Mbd3\textsuperscript{KD} reprogramming n = 7; control reprogramming n = 13; cell cycle for control and Mbd3\textsuperscript{KD} n = 20). e, Goodness of fit plots for the fitting of cell-cycle time distribution to the observed reprogramming latency.
on Mbd3 depletion, we modelled the dynamics of the OSKM reprogramming process using a multi-step Markov chain model (phase-type (PH) modelling) (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 Mbd3KD 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. We obtained a profound fit ($R^2 = 0.999$) between the Mbd3KD 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, 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 Esrb ($P < 10^{-22}$) (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 M2δ (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 Mbd3Δt/− (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 Mbd3Δt− 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.
interest to explore whether direct reprogramming in the absence of potency gene reactivation throughout the process. It will be of great interest to interact with OSK reprogramming factors directly (Extended Data Fig. 9d) were deficient in reprogramming efficiency of Mbd3ATC− somatic cells, supporting the notion that direct OSKM−Mbd3 interactions are important for inhibiting iPS cell formation (Fig. 5c).

We noted that a minimum 5-day exogenous transgene (doxycycline) induction was similarly required to obtain iPS cells from Mbd3ATG−/+ and Mbd3ATC− 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 reprogramming by OSKM factors can be coaxed to become nearly synchronized and deterministic with modified reprogramming approaches. We highlight a NuRD repressor complex component, Mbd3, 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.


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. Montau; and grants from the Leona M. and Harry B. Helmsley Charitable Trust, ERC (SG-281906) grant, BIRAX initiative, ISraEL Science Foundation (Bikura, ICORE and Regular programs), IDFR, Fritz Thysen Stiftung, the Benovert Endowment fund, Alon Scholar Program, and the Clor research prize. I.A. is supported by the HFSP Career Development Award, an ISF-Bikura and the ERC (SG-309788). A.M. is supported by the 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 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). Naïve conditions for murine iPS cells, ES cells and EGs included 5 μg recombinant human LIF (Propeptech). 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, 1 μM; Axon Medchem). Primed N2B27 media for murine conditions (EpiSCs) contained 8 ng ml⁻¹ recombinant human bFGF (Propeptech) and 20 ng ml⁻¹ recombinant human activin (Propeptech). Stem-cell lines and mice deficient for Mbd3 and their derived ES lines were obtained as previously described41,42. For additional gene targeting of mouse pluripotent stem cell lines (Nanog-GFP reporter, pBRY-Mbd3 rescue constructs, Rosas26-CreERT), 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 to C57BL/6J females. Blastocysts were collected and electroporated for ES cell derivation in defined mouse 2i/LIF conditions on gelatin/MEF-coated plates. NGFP1-Mbd3KO was established by infection and sub-cloning of secondary NGFP1 iPS cell line with a shRNA pLKO.Tet-On vector (Addgene) as previously described. Mbd3 target sequences selected for the latter strategy were CTAAGTGGATTGAGTGCCTTT and GCCGATATGGTCTCACAAC. Mycoplasma detection tests are conducted weekly to exclude the presence of any contaminated cells.

Epigenetic reversion of mouse primed epiblast cells. Male naive V6.5 (Mbd3⁻/⁻) and Nanog-GFP ES cells²⁵ maintained in 2i/LIF conditions were injected into BDF2 blastocysts. Chimeric 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⁺ GEP EpiSCs. For epigenetic reversion of murine EpiSCs to naïve 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.

Genetic reversion assay involved single-cell plating. EpiSC growth medium was supplemented with ROCK inhibitor (Y-27632) for 24 h before trypsinization. Genetic reversion assay involved single-cell plating, EpiSC growth medium was supplemented with ROCK inhibitor (Y-27632) for 24 h before trypsinization. Genetic reversion assay involved single-cell plating, EpiSC growth medium was supplemented with ROCK inhibitor (Y-27632) for 24 h before trypsinization.

Reprogramming of human somatic cells and cell infection. Reprogramming was conducted at 5% PO 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 hHSI147580 and hHS147581 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 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 (without excision of PGK-neo-pA cassette, to maintain interference with MBD3 expression). For generating OCT4-GFP reporter subcloned cell lines, 10⁻⁶ MBD3⁵⁷ and MBD3mut cells were electroporated with 30 μg of previously described OCT4-GFP 2A-PURO knock-in donor plasmid (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). Resistance clones were isolated and genomic DNA was extracted for Southern blot and PCR analysis. The WIBR3-MBD3mut and WIBR3-MBD3transgenic reprogrammable chimaeras were generated and labelled with constitutively expressed mCherry in the transgenic line (Mbd3 knock-in allele).

Immunofluorescence staining of pre- and post-implantation embryos. Immunostaining was performed as described previously with modifications46. 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. Idential 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 then rinsed three times 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. 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, FUS-OCVT4, FUS-KLF4, FUS-SOX2, FUS-S-MYC, FUS-Nanog, pCAGgs-Flag-Mbd3, pMSCV-Flag-OCVT4, pMSCV-Flag-SOX2, pMSCV-Flag-KLF4, pCAGgs-S-MYC, pCAGgs-Flag-Nanog, pDNA3.1-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 μl 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 of ×3 Flag 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, F316, Sigma), anti-Mbd3 (A300-258A, Bethyl), anti-Nanog (A300-397A, Bethyl), anti-OCVT4 (sc-9081, H34, Santa Cruz), anti-KLF4 (sc-20691, H180, Santa Cruz), anti-SOX2 (2748S, Cell Signalling) 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 pregnant recipient mice. Embryos were recovered for analysis at different time points throughout development or allowed to develop to full term. Determining germline transmission was performed by mating chimeraic 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 follow-up approval by the Weizmann Institute IACUC (approval 0096221-1-2). We did not exclude animals from our analysis, and did not apply randomization by blinding.

In the process of reprogramming, the proteins H3K4me3, H3K27me3, H3K27ac and Mbd3 in four different time points throughout reprogramming: 0 (MEF), 4 days, 8 days, subcloned iPS or ES cells. OCT4–GFP expression was also observed in ES cells (OCT4–GFP). The expression of OCT4–GFP was observed at subcloned iPS or ES cells. OCT4–GFP expression was also observed in ES cells (OCT4–GFP). The expression of OCT4–GFP was observed at subcloned iPS or ES cells. OCT4–GFP expression was also observed in ES cells (OCT4–GFP).

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; ab16287), rat anti-SSEA3 (1:50, Abcam; MGC631 ab16286), rabbit anti-Nanog (1:400, Bethyl; A300-397A), rabbit anti-Oct3/4 (1:400, Santa Cruz; H134 SC0981), 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 OCKS inductible Mbd3+/- and Mbd3fl/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 AxiocamObserver Z1 (Zeiss) in 5% O2, 5% CO2, 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 microscopy 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 104 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 μm×60 μm)11. This filter retains only dense colonies, erasing noise and single isolated cells (a single nucleus is approximately 6 μm×6 μm), 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 μm×60 μm)12 to merge adjacent colony fragments into large connected colonies and then apply connected components clustering, labelling connected objects using 8-connected neighbourhood. Colony features extraction of each individual colony was done by 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 dynamic 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 (MEP), 4 days, 8 days, subcloned iPS or ES lines. The binding of each protein was measured in both Mbd3+/- and Mbd3fl/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 for each reaction (1% formaldehyde, 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, 0469319001), centrifuged at 950g 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-45D) at −4°C to size ranges between 200 and 800 bp, and precipitated by centrifugation. 10 μg of each antibody was prebound 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, then washed once with lysis buffer and 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 (10M min 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 subsequently with RNaseA (Roche, 1119195001) 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 GenomeTools protocol and adapter 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, ab6540, Abcam), anti-histone H3 trimethyl K4 (ChIP grade, ab5880, 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 best single 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 publications19,20.

**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

\[ X_j = \frac{X_j - \mu_{here}}{\sigma_{here}}, \]

where, \(X_j\) corresponds to read density, and \(\mu_{here}\) and \(\sigma_{here}\) are the estimated noise mean and standard deviation, respectively. Noise parameters were estimated for each sample of 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 publications15,16. Enrichment P values were calculated using Fisher exact test26 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 RNA methylation software, 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 (2^10), resulting in 16,620 gene IDs. For gene expression analysis, we used Matlab version R2011b. Gene signatures differentially expressed between MEF samples (MbD3^+^, MbD3^-^, Mbd3^+^, Mbd3^-^, MEF samples) and ES samples (ES V6.5, MbD3^-^ ES, MbD3^+^ ES and MbD3^+^ /IP) were characterized using a two-sample t-test and corrected for multiple hypotheses using false discovery rate (FDR)20. 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

\[ \hat{X_j}(t) = \max \left( \frac{X_j(t)-X_j(MEF\ MBD3^+)}{X_j(IPS)-X_j(MEF)} \right), \]

where, \(X_j(t)\) denotes gene j expression value at time t (for example, Xj(4 days) or Xj(MEF)) and \(\hat{X_j}(IPS)\) and \(\hat{X_j}(MEF)\) denote the averaged expression value for iPS and MEF samples, respectively. The above transformation was chosen by the distance metric of 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 modifications24. 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 (Qagen). 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 CGs 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 published25. NGFFP1-Mbd325 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^+^ 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.

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<sup>−/−</sup> blastocysts. 

**a**, RT–PCR analysis for Oct4 and trophoblast marker expression of Mbd3<sup>+/−</sup> and Mbd3<sup>−/−</sup> ES cells expanded either in FBS/LIF or 2i/LIF conditions. Only Mbd3<sup>−/−</sup> ES cells, and only in serum conditions, upregulate trophoblast differentiation markers. Error bars indicate s.d. from average (n = 3).

**b**, Mbd3<sup>+/−</sup> heterozygous mice were mated, and Mbd3<sup>−/−</sup> ES cells were derived from blastocysts in naive defined 2i/LIF conditions. Western blot for pluripotency marker expression also indicated that the derived Mbd3<sup>−/−</sup> 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<sup>+/−</sup> and Mbd3<sup>−/−</sup> iPSC 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 moloney retroviruses encoding individual factors. c, Reprogramming efficiency after infection with indicated MEF lines with polycistronic OKSM encoding lentivirus. d, Mbd3<sup>−/−</sup> 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<sup>+/−</sup>, Mbd3<sup>−/−</sup> and Mbd3<sup>−/−</sup> MEFs, adult tail-tip-derived fibroblast (TTF) and neural precursor cells (NPC) were tested for iPSC cell formation in 2i/LIF with or without OKSM lentiviral transduction. Our analysis indicates that OKSM is essential for iPSC 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 OKSM combinations. Asterisk indicates t-test P value <0.01 relative to Mbd3<sup>+/−</sup> 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 $\text{Mbd3}^{\text{floxed}}$ 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<sup>+/+</sup> 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<sup>+/+</sup> and Mbd3<sup>−/−</sup> 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 MBD3WT and MBD3mut 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 iP 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 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 Mbd3/1/1 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).
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 for Mbd3<sup>+/+</sup> and Mbd3<sup>fl/fl</sup> 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).

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<sup>+/+</sup> and Mbd3<sup>fl/fl</sup> cells. Asterisks indicate t-test P value <0.01 in comparison to control siRNA sample. Error bars indicate s.d. from average (n = 3).

Representative data from one out of three biological replicates conducted.