Thesis for the degree

Doctor of Philosophy

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

עבודת גמר (תזה) לתואר

דוקטור לפילוסופיה

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

By
Yoach Rais

מאת
יואח רייס

תכנות ישיר ודטרמיניסטי של תאים בוגרים חזרה לשלב הרב-תכליתי
באמצעות עיכוב הקומפלקס Mbd3/NuRD

Deterministic Reprogramming of Somatic Cells to Pluripotency via Regulation of Mbd3/NuRD Activity

Advisor
Jacob (Yaqub) Hanna
MD PhD

מנחה
יעקוב חנא

April 2016

ניסן תשע"ו

Table of Contents

Abbreviations

2i	two inhibitors of the ERK1/2 (PD0325901) and GSK3 (CHIR99021)
2i/Lif	serum-free medium with addition of Lif, Erki and Gsk3βi (2i)
4OHT	4-hydroxytamoxifen
CoIP	Co-immunoprecipitation
ChIP	Chromatin Immunoprecipitation
CRISPR	Clustered Regularly-Interspaced Short Palindromic Repeats
Dox	doxycycline
EB	Embryonic Bodies
EpiSC	Epiblast Stem Cell
ESC	Embryonic Stem Cell
FBS	Fetal Bovine Serum
FDR	False Discovery Rate
hESC	Human Embryonic Stem Cell
ICM	Inner Cell Mass
iPSC	Induced Pluripotent Stem Cell
KO	knockout
KSR	Knockout Serum Replacement
LIF	Leukemia Inhibitory Factor
MBD	methyl-CpG-binding domain
MEF	Mouse Embryonic Fibroblast
mESC	Mouse Embryonic Stem Cell
NuRD	Nucleosome Remodeling and Deacetylase complex
OSKM	Oct4, Sox2, Klf4 and c-Myc
PSC	Pluripotent Stem Cell
RPKM	Reads Per Kilobase per Million reads
RQ	Relative Quantity
rtTA	Reverse tetracycline controlled transactivator
SCNT	Somatic Cell Nuclear Transfer
WT	Wild-type

Abstract

The discovery of cellular reprogramming and the generation of induced pluripotent stem cells (iPSCs) from somatic cells have revolutionized stem cell research. However, in most conventional iPSC reprogramming approaches, only a small fraction of the somatic cell population becomes pluripotent, thus making this process hard to investigate at sufficient molecular depth. In this work, we identify the Mbd3/NuRD complex as a major roadblock for reprogramming. Under naïve pluripotency promoting conditions, optimized depletion (50-80%) of Mbd3 at early stages of reprogramming led to a near-deterministic and synchronized reprogramming, as almost all the initial cell population became pluripotent simultaneously after only 8 days.

Previously, rapidly arrested proliferation of somatic upon complete ablation of Mbd3 has limited the ability to manipulate this pathway to achieve robust reprogramming. Therefore, by dissecting the role of Mbd3/NuRD complex during early stages of reprogramming, we aimed to identify ways to block Mbd3-dependent NuRD activity without the negative effect on somatic cell proliferation and viability. We identify the Mbd3/Gatad2a/Chd4 axis as a functional and biochemical barrier for re-establishment of pluripotency. Furthermore, we show that complete ablation of Gatad2a, a NuRD-specific subunit, disrupts Mbd3/NuRD repressive activity on the pluripotency circuit without compromising somatic cell proliferation, thereby yielding near-deterministic reprogramming.

Finally, we used the Mbd3- and Gatad2a-depleted platforms to conduct the first high-resolution mapping of authentic transcriptional and epigenetic dynamics during synchronized and deterministic reprogramming of somatic cell in mice by RNA sequencing, ChIP-seq, ATAC-seq, and bisulfite sequencing (RRBS and WGBS). These finding lead us to unravel and discuss novel insights regarding reprogramming factor-mediated regulation of chromatin changes.

תקציר

תהליך התכנות מחדש של תאים בוגרים לתאי גזע עובריים הביא למהפכה במחקר של תאי גזע. למרות השפעתה העצומה, טכנולוגיה זו לתכנות תאים בצורה ישירה מוגבלת ביעילותה, שכן רק אחוז קטן מהתאים הופכים לתאי גזע מושרים (iPSC). עובדה זו מגבילה את יכולתנו לחקור את תהליך התכנות מחדש ולהבין את המנגנונים התוך-תאיים אשר מעורבים בתהליך זה. במהלך לימודי, גיליתי ואפיינתי את הקומפלקס האפיגנטי Mbd3/NuRD כחסם לתהליך התכנות מחדש. עיכוב אופטימלי (50-80%) של החלבון Mbd3 גורם להורדת חסמים אפיגנטיים ולתהליך בעל מאפיינים דטרמיניסטיים, כאשר התהליך נעשה בסביבה שמעודדת את המצב הנאיבי של הפוטנציאל הרב תכליתי. בדרך זו, כמעט כל התאים מגיעים לפוטנציאל הרב תכליתי האותנטי בתוך 8 ימים בו-זמנית.

עיכוב החלבון Mbd3 עלול להגביל את השימוש במניפולציה זו כדי לתכנת תאים למצב רב-תכליתי, כיוון שהורדה הרמטית של החלבון גורמת להפסקת חלוקה של תאים סומטיים. לכן, בהמשך עבודתי התמקדנו בהבנת תפקידו של Mbd3 בשלבים הראשוניים של התכנות מחדש, על מנת למצוא דרכים אלטרנטיביות לעכב את תפקידו כחסם לתהליך בלי לפגוע בפוטנציאל החלוקה של התאים. גילינו שהחלבון פועל דרך ציר בקומפלקס ה- NuRD אשר כולל את Mbd3/Gatad2a/Chd4, ופעיל ביוכימית ופונקציונלית בעיכוב תהליך התכנות מחדש. בהמשך הראינו שבניגוד ל- Mbd3, הורדה הרמטית של החלבון Gatad2a אינו פוגע בחלוקת התאים ועדיין משמר את התכונות הדטרמיניסטיות של תהליך התכנות מחדש.

לבסוף, אנו משתמשים בפלטפורמות של התכנות הדטרמיניסטי בעזרת הורדה חלקית של החלבון Mbd3 והורדה מלאה של החלבון Gatad2a כדי לחקור לעומק את השינויים האפיגנטיים שמתרחשים בתאים אשר עוברים תכנות מחדש. השתמשנו במגוון שיטות למעקב אחר ביטוי גנים ומולקולות RNA קצרות אשר משתנות בתאים בזמן התהליך, כגון RNA-seq, small RNA-seq, Poly A RNA-seq. בהמשך בדקנו את השינויים אשר נעשים ברמת הכרומטין ע"י ChIP-seq ו- ATAC-seq ועקבנו אחרי השינויים האפיגנטיים ברמת ה- DNA ע"י RRBS ו- WGBS. בעזרת ניתוח מעמיק של מידע שאספנו כל 24 שעות עד השלמת התהליך אחרי 8 ימים, אנו מציגים בפעם הראשונה מעקב ברזולוציה גבוהה אחרי התהליכים והשינויים האפיגנטיים שמתרחשים בתאים אשר עוברים תכנות מחדש ומגיעים למצב האותנטי של תאי גזע מושרים בעלי פוטנציאל זהה לתאי גזע מוקדמים ביותר של עכבר.

Introduction

Pluripotency and ES cell developmental potential

The pluripotent state pertains to cells that can give rise to all known cell types from each of the three embryonic germ layers. In the 1960s and 1970s, studies of embryo aggregation and blastocyst chimerism by Mintz and colleagues have solidified the idea that the inner cell mass (ICM) cells of the mouse blastocyst were pluripotent[1-3].

The ability to isolate and propagate ES cells from the ICM *in vitro* has advanced the understanding of principles of early mammalian development, tissue formation and differentiation. Further, ES cells offer enhanced opportunities to model diseases, discover disease mechanisms and may ultimately be used as a platform for patient-specific cell therapy[4]. Yet, although ES cells are an excellent source of differentiated cells, the use of embryo-derived stem cells does not solve tissue-matching problems that arise during transplantation therapies.

The successful generation of cloned ES cells and animals by somatic cell nuclear transfer (SCNT), where the nucleus from donor somatic cell is transplanted into an a nucleated oocyte, has paved the way for realizing the ideal of generating 'customized' patient-specific ES cells[5]. Nuclear transfer-generated ES cell lines would capture a patient's complete genome in a cell that could be induced into any type, thus allowing differentiation into disease-relevant cell types for analysis or cell replacement therapy. However, ethical controversies regarding human cloning triggered opposition to human SCNT research. Further, nuclear transfer is extremely inefficient in rodents, rendering the approach obsolete. The more recently introduced cell fusion approach involves fusion of somatic cells with ES cells, resulting in reprogramming of the epigenetic state of the former. However, the generation of tetraploid fused stem cells produces genetically abnormal cells, which inhibits the clinical utilization of the technique. Although SCNT and cell fusion may not be applicative for generating customized ES cells[6], those strategies have shown the potential and ability to reprogram the epigenetic state of somatic cells to the pluripotent state.

Direct reprogramming of somatic cells to pluripotency with defined factors

The breakthrough in overcoming issues of tissue matching has been the *in vitro* derivation of reprogrammed pluripotent cells from differentiated somatic cells, termed induced pluripotent stem cells (iPSCs), by the ectopic expression of the four transcription factors Oct4, Sox2, Klf4 and c-Myc (OSKM) in mouse and human somatic cells [6,7]. The *in vitro* induction of pluripotency has had a dramatic effect on stem cell research. This approach enables generation of 'custom-made' genetically identical iPSC, Making it highly suitable for applications in the fields of regenerative medicine, disease modelling and drug discovery.

Reprogramming as a stochastic process

Direct reprogramming is a relatively inefficient process. Only a minority of somatic donor cells become reprogrammed over an extended period of time, with the first iPSCs (up to 5%) appearing only after 10-14 days of OSKM transcription factor expression. So far this has applied to both mouse and human fibroblasts, regardless of the combination of reprogramming factors used. Formation of iPSC colonies requires at least 2-3 weeks of continuous culturing and proliferation[8,9]. Reprogramming of terminally differentiated cell to iPS cells[10] suggests that all cells in a population have the potential to be reprogrammed and not only an 'elite' somatic cell or 'stem-like' cells. Studies of direct cell reprogramming showed that it is a stochastic process. The stochastic model predicts that most, if not all, cells within a donor population will generate iPS cells, although with varying latency[8,10-13]. The prolonged latency of the reprogramming process involves intermediate stages that are almost impossible to define molecularly and epigenetically, because of the heterogeneity of the obtained cells[14]. Therefore, developing a more efficient platform for studying this process is of high importance.

The iPSC reprogramming process can be accelerated in a cell division-dependent manner, e.g. by blocking the p53 pathway, or by increasing the intrinsic rate of conversion to iPSCs. An example for the latter was given by over-expression of the Nanog transcription factor, a regulator of pluripotency, together with OSKM during reprogramming. Notably, however, the process remained stochastic and asynchronous.

This suggests that to facilitate complete, stochastic and synchronized reprogramming, other yet unknown rate-limiting event(s) need to be bypassed. Such events could be related to epigenetic factors.

Chromatin regulators can alter reprogramming dynamics

From an epigenetic viewpoint, somatic cell reprogramming can be described as the reversal of differentiation. Whereas differentiation involves a shift from 'open' (euchromatin) to 'closed' (condensed/heterochromatin) chromatin conformation, reprogramming reverses the epigenetic architecture. Therefore, finding a way to enhance chromatin dynamic towards an 'open' conformation would affect the reprogramming process. Several studies reported minor positive effects of using such approaches. At the histone level, it was shown that using histone deacetylase (HDAC) inhibitors as valproic acid (VPN) or repression of histone methyl transferases SUV39H1[15] and Dot1L[16] as well as YY1[17] enhanced reprogramming. On the other hand, inhibition of UTX, a specific demethylase for K27Me3, results in a dramatic reduction in reprogramming[18], as is inhibition of the polycomb repressive complexes (PRC) 1 and 2. At the DNA level, demethylation agent as 5'-azacytidine has been used to enhance reprogramming[19], and repression of AID, a cytidine deaminase, inhibits reprogramming[20,21]. These approaches raise the questions of which chromatin regulators involved in early development could affect reprogramming, and whether it will be possible to transform conventional direct reprogramming from a stochastic to a deterministic process.

Mbd3/NuRD complex

One of the key compounds that has been shown to inhibit reprogramming is Mbd3/NuRD[22]. Mbd3 (methyl-CpG-binding domain 3) is a structural protein in the Nucleosome Remodeling and Deacetylation (NuRD) complex. Mbd2 and Mbd3 assemble into mutually exclusive distinct NuRD complexes[23], which mediate gene repression through histone deacetylation via Hdac1 and 2, and chromatin remodeling ATPase activity through its Chd3 (Mi2a) and Chd4 (Mi2b) subunits[24,25]. Mbd3/NuRD

preferentially binds and represses actively transcribed genes, and some of its components (e.g. Chd4) have also been implicated in transcriptional activation at certain loci[26,27]

During early development, the NuRD complex plays a critical role in the epigenetic regulation of loss of pluripotency. It was shown that Mbd3, a canonical NuRD subunit, is a major factor in this function[28-32].

Mbd3 belongs to the Methyl CpG Binding Domain (MBD) protein family[33,34] and, together with Mbd 1, 2 and 4, was originally characterized as a protein containing a region with high homology to the MBD of MeCP2 [26,37,38] MBD family members were shown to bind methylated DNA; however, Mbd3 lacks the ability to directly bind it[34]. While *Mbd3*[-/-] mice die during early embryogenesis[28], it was found that ES cells from these mice can conserve their self-renewal function and are severely deficient in their differentiation to embryonic lineages[29]. ChIP-seq analysis of Mbd3 gene showed that it occupies promoters of pluripotency regulators like Oct4, Tbx3, Klf4 and Nanog[32]. Mbd3 is a component of NuRD co-repressor complex and is essential for the assembly and proper activity of the NuRD complex during those stages. Therefore, it was suggested that Mbd3 and NuRD are crucial factors regulating differentiation through proper silencing of key naïve pluripotency genes [29,30,32,35]. Therefore, we have focused our work to evaluated Mbd3/NuRD function during reprogramming and how its ablation can affect conventional direct reprogramming.

Results

Part 1: Mbd3 acts as a roadblock during the iPSC reprogramming process

Most of the work described in this chapter was published in: Rais et al., Deterministic direct reprogramming of somatic cells to pluripotency, *Nature* 2013.

Neutralizing Mbd3 expression facilitates access to ground state pluripotency

Recent studies have pointed out the importance of chromatin derepression in converting somatic cells into iPSCs[18,36-38]. To determine how chromatin remodelers affect direct reprogramming, we first used a screening platform of primed-to-naïve reversion[18]. The naïve ground state of pluripotency involves an open chromatin configuration with reduced levels of repressive chromatin marks[39-41]. The naïve-to-primed transition is characterized by a massive regulation mechanism that starts to shape cellular differentiation[42,43]. This mechanism includes chromatin regulators, which play a major role in repressing the pluripotent circuit and activating various differentiation pathways. Therefore, with aim to improve the efficiency of reprogramming, we conducted a loss-of-function screen for selected epigenetic repressor factors. We initially focused on reverting primed epiblast stem cells (EpiSCs)[44,45]. In the absence of exogenous transcription factor, these cells can stochastically convert within 5 days into naïve pluripotent state in 2i/LIF (2i, PD0325901 ERK1/2 inhibitor and GSK3β inhibitor CHIR99021) growth conditions[46]. We utilized a primed EpiSC line carrying a Nanog-GFP knockin reporter that is active only in the naïve state[18,47], and applied siRNA screen to identify boosters of EpiSC reversion into Nanog-GFP[+] naïve pluripotent cells (**Fig. 1a**). Results showed that inhibition of DNA methylation or H3K27me3 by knockdown of Dnmt1 or Eed/Suz12, respectively, increased epigenetic reversion efficiency. Still, only a minority of the donor cells turned on the Nanog-GFP reporter (**Fig. 1a**). Remarkably, we noted that Mbd3 inhibition dramatically increased the EpiSC reversion efficiency. Up to 80% of the transfected cells turned on Nanog-GFP in 2i/LIF conditions, as compared to <10% in control EpiSCs (**Fig. 1a**). Such a dramatic effect was not observed with any of the other factors tested, including other Mbd family members (**Fig. 1a**).

Figure 1. primed-to-naïve reversion screen. (a) A siRNA screen for factors that can boost epigenetic reversion of primed EpiSCs into naïve ESCs. Nanog-GFP mouse EpiSCs were used for screening and Nanog-GFP reactivation was used as a specific marker for naïve pluripotency after expansion in 2i/LIF conditions and applying siRNA for the indicated factors. Percentage of GFP[+] cells detected by flow cytometry is indicated. Data from 3 independent experiments are presented as mean±sd. Asterisks indicate significant differences (*P* <0.01) in comparison to siScramble treatment.
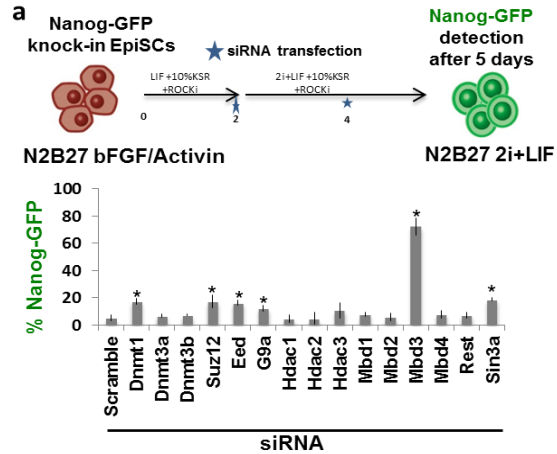
To validate the siRNA screening results, we used Mbd3$^{flox/-}$ ESCs and introduced Rosa26-CreER and Nanog-GFP knockin alleles by gene targeting (**Fig. 2a and b**). After validation of Mbd3 levels before and after 4OHT treatment, we showed that Mbd3$^{flox/-}$ cells exhibit reduced Mbd3 protein levels by ~40-20% relative to Mbd3$^{+/+}$ cells and that full knockout was achieved by using 4OHT (**Fig. 2c**). Next, EpiSCs were established from the latter engineered lines and from ESCs carrying a Nanog-GFP knockin reporter and Mbd3$^{+/+}$, following microinjection into E3.5 blastocysts and re-derivation at E6.5 from post-implantation epiblast (**Fig 2d**). In comparison to Mbd3$^{+/+}$ EpiSCs, Mbd3$^{flox/-}$ primed cells reverted with increased efficiency in 2i/LIF + 4OHT conditions (**Fig. 2e**), displaying homogenous Nanog-GFP reactivation consistent with reversion to ground state pluripotency[18,47] (**Fig. 2f**).

Single-cell clonal analysis for epigenetic reversion of EpiSCs, constitutively labeled with mCherry by viral infection as a control for plating efficiency, demonstrated 95% Nanog-GFP[+] single-cell reversion efficiency in Mbd3-depleted cells (**Fig. 2d**). Importantly, Mbd3$^{flox/-}$ EpiSCs, which retained hypomorphic (~20-40%) Mbd3 protein expression (**Fig. 2c**)[29], also yielded reverted ESCs with >93% efficiencies (**Fig. 2d**). Both reverted Mbd3$^{-/-}$, which underwent transgenic insertion of Mbd3 to rescue their differentiation deficiency[29,30], and Mbd3$^{flox/-}$ cells contributed to adult chimera formation (**Fig. 2g and h**). Reconstitution of Mbd3 expression in Mbd3$^{-/-}$ and Mbd3$^{flox/-}$ EpiSCs inhibited reversion efficiencies in 2i/LIF down to <20%, as typically observed in wild-type (WT) cells (**Fig 2d**). These results directly demonstrate that reduction of Mbd3 protein levels induces nearly complete reversion of murine EpiSCs to ground state pluripotency.
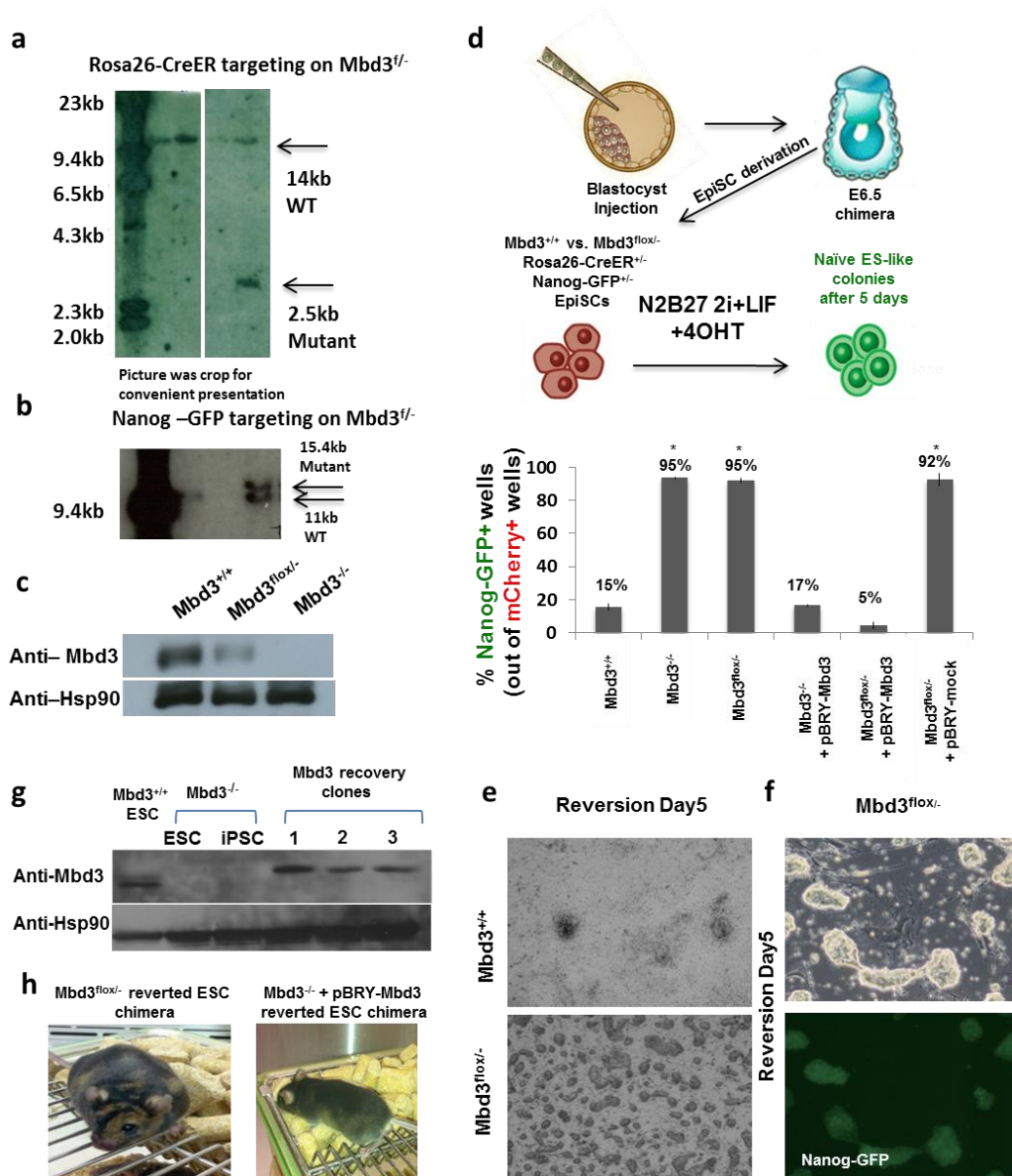
Figure 2. Boosting primed-to-naïve pluripotent stem cell epigenetic reversion.
(a) Southern blot analysis showing correct gene targeting of the Rosa26 locus with Cre-ER knock in construct introduced into Mbd3$^{flox/-}$ ESCs. (b) Southern blot analysis showing correct gene targeting of the Nanog locus with GFP knockin construct introduced into Mbd3$^{flox/-}$ ESCs. (c) Western blot indicates Mbd3 expression levels in different mutant stem cell lines. (d) Mbd3$^{+/+}$ and Mbd3$^{flox/-}$ EpiSCs carrying Nanog-GFP knockin reporter as indicated in the scheme, were tested for reversion into naïve cells in 2i/LIF + 4OHT. Single-cell reprogramming efficiency and quantification for EpiSC reprogramming from different mutant lines. pBRY-Mbd3 rescue construct, stably expressed in the indicated lines, reduced reprogramming efficiency back to those observed in Mbd3$^{+/+}$ WT cells. Asterisks indicate differences (P <0.01) in comparison to Mbd3$^{+/+}$ samples. (e) Note the dramatically increased reprogramming efficiency in Mbd3-depleted cells and the homogenous reactivation of Nanog-GFP marker (f). (g) Western blot analysis indicating Mbd3 expression in Mbd3$^{-/-}$ ESCs/iPSCs, with or without addition of pBRY-Mbd3 rescue transgene (recovery clones). (h) Agouti colored chimeras obtained from reverted ESCs following Mbd3 depletion.

13

Pre-implantation *in vivo* reprogramming and development are accompanied by depletion of Mbd3 expression

The finding that lack of Mbd3 promotes reversion to pluripotency is seemingly contradictory to previous *in vivo* studies, which had suggested that Mbd3 is essential for the establishment of the ground state of pluripotency after fertilization[38,43]. This conclusion was based on the fact that while Oct4[+] cells can be observed in the inner cell mass of Mbd3[-/-] E3.5 embryos, Mbd3[-/-] ESCs could not be derived *in vitro* upon explantation in serum/LIF derivation conditions[29,30]. However, established ES cells, which can tolerate loss of Mbd3 by gene targeting, show a propensity for trophoblast differentiation in serum-containing conditions (**Fig. 3a**)[31]. This could account for the previous technical inability to derived Mbd3[-/-] ESC in these conditions. Thus, we revisited ESC derivation from Mbd3[-/-] E3.5 embryos in serum-free 2i/LIF conditions[48] and, indeed, we successfully isolated Mbd3[-/-] ESCs at the expected Mendelian ratio. These cells expressed all pluripotency markers tested (**Fig. 3b-d**). These results indicate that Mbd3 is dispensable for establishing the ground state of pluripotency and mouse ESC derivation.
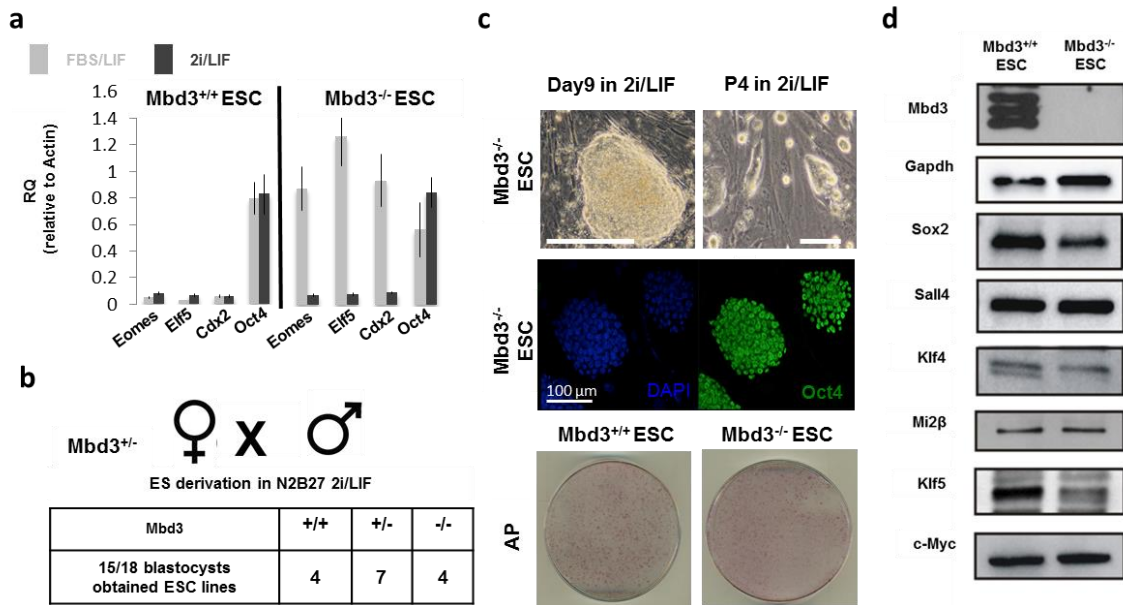
Figure 3. Derivation of Mbd3 KO ESCs from Mbd3[-/-] blastocysts. (a) RT-PCR analysis for Oct4 and trophoblast marker expression in Mbd3[+/+] and Mbd3[-/-] ESCs expanded either in FBS/LIF or 2i/LIF conditions. Only in Mbd3[-/-] ESCs in serum conditions trophoblast differentiation markers were upregulated. In stringent serum-free 2i/LIF conditions, Mbd3[-/-] ESCs are indistinguishable from Mbd3[+/+]

14

ESCs. Error bars indicate standard deviation of biological triplicate samples. (b) Mbd3[+/-] heterozygous mice were mated and ESCs were derived from blastocysts in naïve defined 2i/LIF conditions. Mbd3[-/-] ESCs were obtained at the expected ratio. (c) Top: ESC that was derived from blastocysts in naïve defined 2i/LIF conditions at day 9 and after 4 passage, middle: immunostaining for Oct4 in Mbd3[-/-] ESC line, bottom: AP normal staining of Mbd3[+/+] and Mbd3[-/-] ESCs.   (d) Western blot analysis also indicated that the derived Mbd3[-/-] ESC lines adequately expressed all tested pluripotency markers.

Consistently, nuclear expression of Mbd3 protein was reduced after fertilization throughout pre-implantation development, and became readily detected only at the late blastocyst stage and post-implantation epiblast (**Fig. 4a, b**). Nuclear Mbd3 protein expression was preserved after *in vitro* derivation in both naïve and primed pluripotent cells (**Fig. 4c**). These results indicate that pre-implantation *in vivo* reprogramming and development are accompanied by reduced expression of Mbd3, which is re-expressed once pluripotency is established, consistent with a critical role for Mbd3 in allowing differentiation and exit from naïve pluripotency[32]. Collectively, these data show that neutralizing Mbd3 expression facilitates access to ground state pluripotency from early embryonic Oct4-expressing cells.
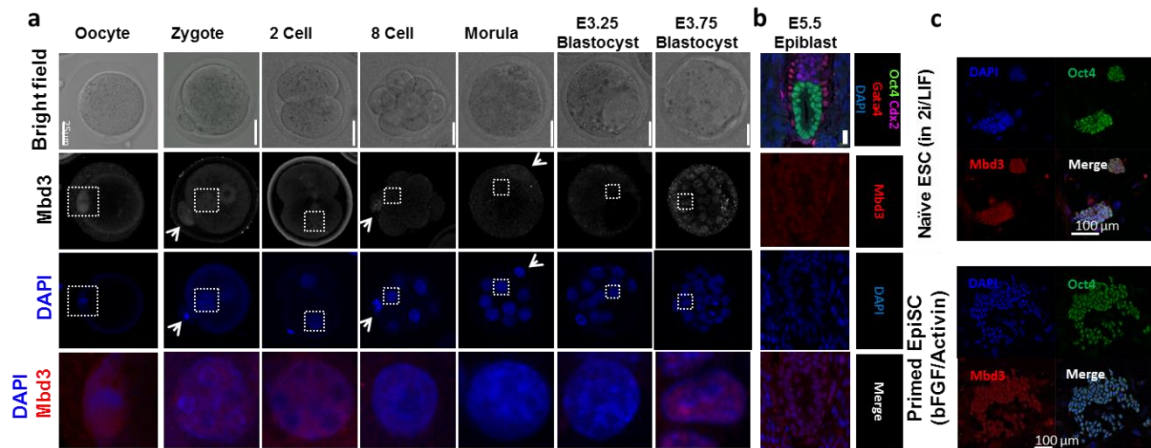
Figure 4. The dynamics of *in vivo* Mbd3 expression. (a) Confocal immunostaining images for Mbd3 expression in developing mouse embryos. Arrows indicate polar body. Note the reduction in Mbd3 expression after fertilization and re-expression at the late blastocyst stage. Immunostaining for Mbd3 in mouse oocytes indicates maternal inheritance of Mbd3. Scale bars: 25 μm. (b) Immunostaining for Mbd3 in E5.5 post-implantation epiblast, indicating prominent expression as seen in late-blastocyst stage. Scale bar: 25 μm. (c) Immunostaining analysis for Mbd3, showing prominent nuclear expression in pluripotent cells expanded in defined naïve and primed growth conditions. Scale bars: 100 μm.

Alleviating Mbd3 inhibition facilitates iPSC reprograming

We proceeded to test whether Mbd3 inhibition in somatic cells, which lack endogenous expression of pluripotency markers like Oct4 and are more developmentally restricted in comparison to EpiSCs, facilitates their high efficiency conversion to ground state pluripotency. To accurately evaluate reprogramming kinetics and efficiencies, we established a "secondary reprogrammable" platform[49,50]. We used Mbd3[+/+] and Mbd3[flox/-] transgenic cell lines harboring a Dox inducible OKSM polycistronic cassette[51], a constitutive nuclear mCherry marker, to allow tracking of individual cells and control for plating efficiency, and an Oct4-GFP specific reporter[52] (**Fig 5a**). The latter cells were injected into host blastocyst and secondary reprogrammable MEFs were derived and utilized for reprogramming quantifications[49]. Single-cell sorting of mCherry+ Mbd3[flox/-] MEFs and subsequent reprogramming in 2i-LIF+Dox and 5% O_2 conditions reproducibly yielded 100% iPSC derivation efficiency by day 8. In WT cells reprogrammed under identical conditions, no more than 20% of clones reactivated Oct4-GFP, whereas the majority of mCherry+ secondary fibroblasts neither reactivated Oct4-GFP marker nor acquired ES-like morphology (**Fig. 5b** and data not shown). Mbd3-depleted somatic cells had similar growth rates at early stages of reprogramming (after Dox induction), thus excluding cell proliferation as a potential cause for the observed differences in derivation efficiencies (**Fig. 5c**). Radically high single-cell reprogramming efficiency rates were obtained upon establishment of multiple Mbd3[flox/-] secondary lines harboring different OKSM integrations (**Fig. 5b**). All randomly tested clones stained positive for alkaline phosphatase (AP), Oct4 and Nanog pluripotency markers (**Fig. 5d**), and had high chimera contribution (**Fig. 5e**).

Next, we analyzed the reprogramming dynamics of "secondary" Mbd3[flox/-] and control Mbd3[+/+] fibroblasts by microscopic live imaging. Time-lapse measurements showed a dramatic increase in ES-like colony formation in Mbd3[flox/-] fibroblasts as compared to Mbd3[+/+] (**Fig. 6a and Movie 1[i]**). An in-house developed algorithm that allows segmentation of single mCherry colonies and tracking of Oct4-GFP reactivation dynamics during reprogramming on clonal populations was applied. By day 6 following Dox induction, >98% of Mbd3[flox/-] clonal populations reactivated the Oct4-GFP

---

[i] At: **http://www.nature.com/nature/journal/v502/n7469/fig_tab/nature12587_SV1.html**.

pluripotency marker, while only up to 15% efficiency was detected in control samples reprogrammed under identical conditions **(Fig. 6b and c).**
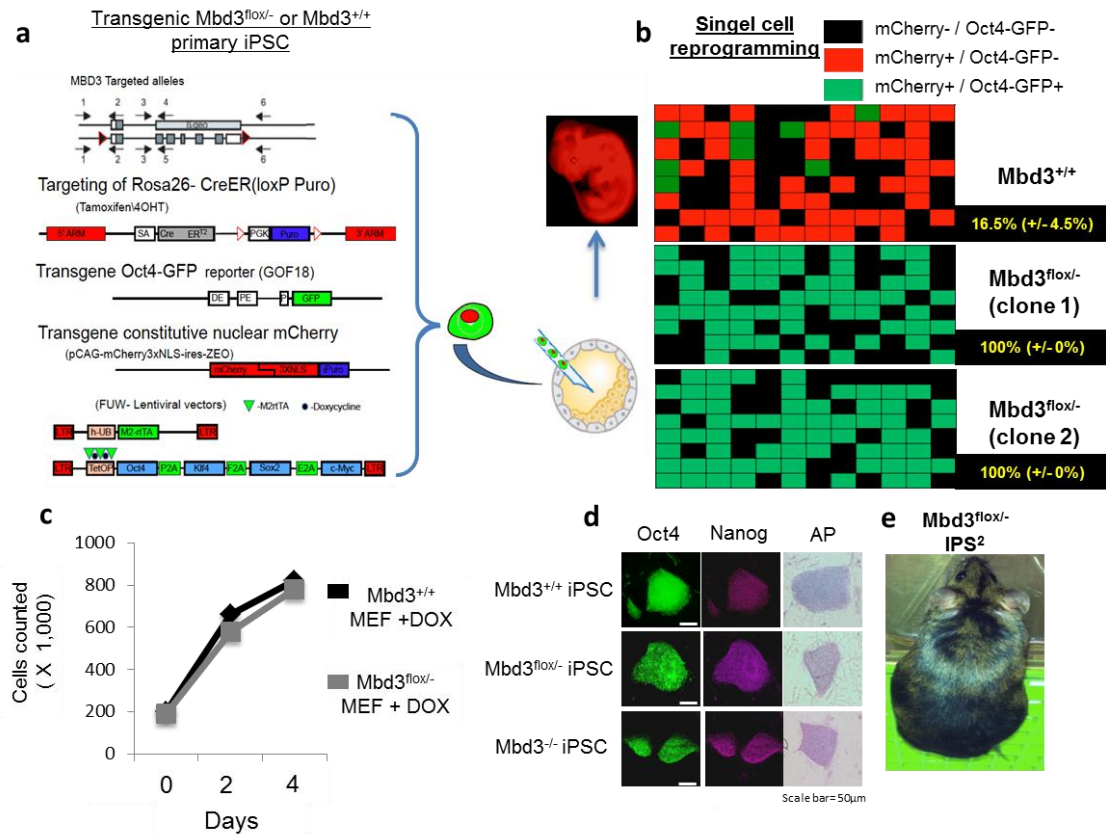
Figure 5. Radically efficient progression towards pluripotency in Mbd3-depleted somatic cells. (a)We established a reprogrammable mouse Mbd3[+/+] and Mbd3[flox/-] iPSC lines carrying *(1)* an Oct4-GFP reporter, *(2)* nuclear mCherry constitutively expressed marker, *(3)* m2RtTa transgene and (4) a TetO inducible 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 analyzed following Dox induction. This transgenic system allows non-restricted derivation of homogenous somatic cells for iPSC reprogramming studies. (b) Secondary reprogrammable fibroblasts, carrying an Oct4-GFP reporter and a mCherry constitutively expressed marker, were single-cell sorted and subjected to Dox reprogramming. Reprogramming efficiency at day 8 was calculated by dividing the number of Oct4-GFP+ wells by mCherry+ wells (mCherry was used to normalize for plating efficiency). Three independent experiments were conducted from each clone and data are presented as mean ± sd. Two independent Mbd3[flox/-] secondary iPSC lines are shown. (c) Similar growth kinetics was observed in Mbd3[+/+] and Mbd3[flox/-] MEFs upon Dox-mediated transgene induction. One representative experiment is shown out of two performed. Results exclude changes in transgene induction or proliferation as predominant causes for enhanced reprogramming in Mbd3-depleted somatic cells. (d) Immunostaining of representative iPSC clones for pluripotency markers. (e) Agouti coat colored chimera from Mbd3-depleted iPSCs.

By day 6, approximately 85% of cells in each individual Mbd3[flox/-] clonal population expressed GFP, whereas <2% of cells within successfully reprogrammed Mbd3[+/+] clones

17

turned on the Oct4-GFP marker (**Movies 2[ii], 3[iii] and 4[iv]).** The unbiased quantitative analysis demonstrated a dramatic intra- and interclonal synchronized reactivation of Oct4-GFP in Mbd3$^{flox/-}$ populations during a narrow time window at days 4.5-5.5 (**Fig. 6c**), highlighting a dramatic increase in reprogramming synchrony and efficiency following Mbd3 depletion in OKSM-transduced somatic cells. Detection of Oct4-GFP by flow cytometry on polyclonal populations demonstrated similar kinetics for iPSC reprogramming (**Fig. 6d).**
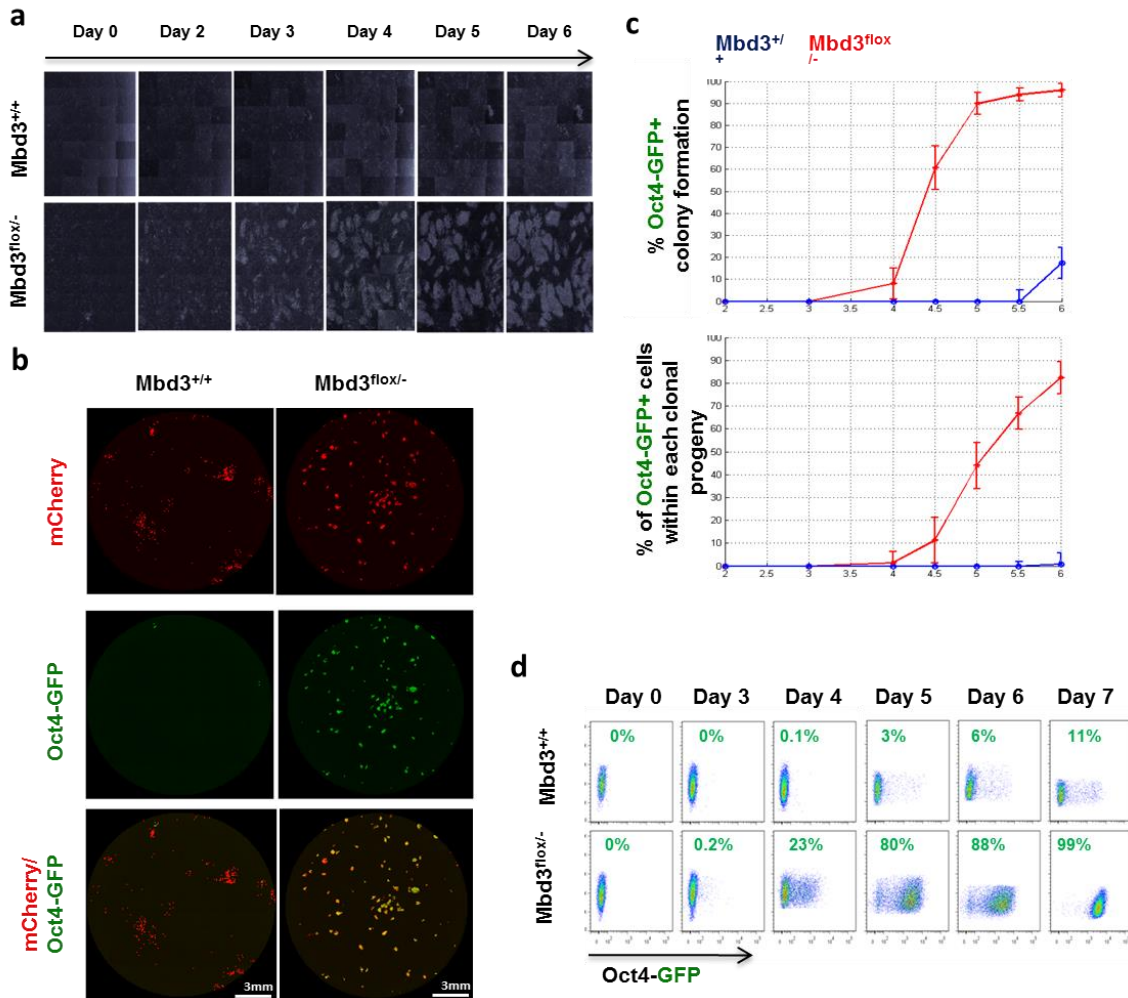
Figure 6. Synchronized and deterministic reprogramming of somatic cells to pluripotency. (a) Time-lapse imaging of reprogramming in equivalent regions and phase contrast. Note the dramatic increase in ES-like colony formation in Mbd3$^{flox/-}$ cells. (b) Full well mosaic images of mCherry, Oct4-GFP and combined channels, shown for Mbd3$^{flox/-}$ and Mbd3$^{+/+}$ at day 6, after starting with

150 cells plated per well. Abundant ES-like mCherry+ colonies are co-localize with Oct4-GFP$^+$ in Mbd3$^{flox/-}$ cells, but not in WT cells. (c) Upper graph indicates cumulative Oct4-GFP$^+$ colonies for Mbd3$^{flox/-}$ (red plot) and Mbd3$^{+/+}$ (blue plot) based on live imaging follow-up. Statistics of Oct4-GFP activation were calculated from all segmented colonies. Note the narrow window of synchronized Oct4-GFP activation at days 4-5. Lower graph indicates the average fraction of Oct4-GFP$^+$ cells within single colonies measured by live imaging follow-up. Approximately 85% of cells within individual Mbd3$^{flox/-}$ clonal population became Oct4-GFP$^+$ cells by day 6. Graph values indicate the mean and error bars indicate standard deviation calculated over four replicates (wells) in each sample and time point. (d) Flow cytometry measurement of Oct4-GFP reactivation dynamics in 2i/LIF following Dox (OSKM) induction. Mbd3$^{flox/-}$ secondary cells synchronously and rapidly reactivate Oct4-GFP by day 7 in the entire donor cell population. Importantly, wells at the indicated time points were harvested for analysis without prior passaging and splitting during the reprogramming course. One out of three independent experiments is shown (FSC, forward scatter). Note the dramatic reactivation of Oct4-GFP during the narrow time window at days 4-5, as also seen in microscopic time-lapse live imaging measurements.

Finally, re-infection with lentiviruses encoding Mbd3, but not Mbd2, before day 5 of reprogramming had a profound inhibitory effect on iPSC generation from Mbd3$^{flox/-}$ MEFs, whereas reinfection after day 5 had a diminished effect (**Fig. 7a**). The above kinetic analysis suggests that Mbd3 can inhibit reprogramming when introduced before the final stages of reprogramming, coinciding with endogenous Oct4/Nanog reactivation[47]. However, once pluripotency is reestablished, Mbd3 presence is tolerated and does not compromise the maintenance of pluripotency. To conclude, we show that Mbd3 acts as a barrier during the reprogramming and that optimized and controlled depletion of Mbd3 promotes synchrony and efficiency of reprogramming.

Figure 7. The effect of Mbd3 expression reconstitution during deterministic reprogramming. Scheme demonstrates experimental strategy for defining the ability of Mbd3 to inhibit iPS formation during reprogramming. Secondary OSKM reprogrammable Mbd3$^{flox/-}$ MEFs were tested for amenability to reprogramming following over-expression of Mbd3, Mbd2 or empty FUW lentiviruses at different time points during reprogramming. Mbd2 or mock-vector transfection did not decrease reprogramming efficiency. Introduction of Mbd3 before day 5 drastically reduced iPSC formation. One out of two representative experiments is shown. Average of duplicates is shown per condition, error bars indicate sd.

19
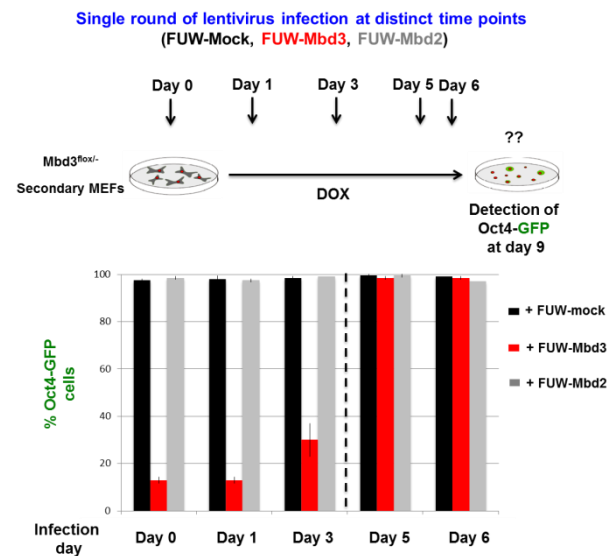
Part 2: The mechanisms of Mbd3-mediated inhibition of iPSC reprogramming

A part of the work described in this chapter was published in: Rais et al., Deterministic direct reprogramming of somatic cells to pluripotency, *Nature* 2013.

Direct OKSM-Mbd3 interactions are involved in inhibition of iPSC reprogramming

Optimized Mbd3 inhibition was a key contributor to the radically increased efficiency of reprogramming reported in the previous chapter. Thus, we aimed to decipher the mechanisms of Mbd3 inhibition that affect iPSC reprogramming. Inhibiting Mbd3 expression is not sufficient to induce iPSC formation in somatic cells, even in NPCs that can be reprogrammed with Oct4 expression alone[53]. Further, Mbd3 inhibitory effect on pluripotency cannot be predominantly attributed to its ability to repress endogenous Nanog, either directly or indirectly[11,22,54]. These observations raise the hypothesis that Mbd3 may act more globally in reprogramming regulation by directly interacting with other critical pluripotency promoting factors.

To test that hypothesis, we first established that Flag-Tagged Oct4, Klf4, Sox2 and c-Myc were specifically co-immunoprecipitated with Mbd3 following exogenous overexpression in HEK293 cells **(Fig 8a)**. Reciprocal experiments showed that Flag-tagged Mbd3 specifically co-immunoprecipitated with Oct4, Sox2, Klf4 and c-Myc, but not with Nanog **(Fig. 8b)**. Moreover, OKSM specifically co-immunoprecipitated with Mbd3/NuRD components in MEF cells undergoing reprogramming following 3 days of OKSM induction **(Fig. 8c)**. These interactions were abrogated upon the deletion of the MBD domain and co-immunoprecipitation with OKSM factors **(Fig. 8d)**. The MBD domains of Mbd3 and of Mbd2, which harbors two point mutations[55], interacted with OKSM factors (data not shown). Yet, as validated by co-immunoprecipitation experiments, Mbd2 and Mbd3 did not co-localize to the same NuRD complexes **(Fig. 8e)**, consistent with previous reports in a variety of somatic and cancer cell lines[23]. The above explains, at least in part, the limited influence of perturbing Mbd2 expression on reprogramming **(Fig. 7a)**.

Figure 8. MBD domain of Mbd3 is critical for direct interaction with OSKM reprogramming factors. (a) Constructs encoding Flag-tagged OCT4, c-MYC, KLF4, SOX2, NANOG or HDAC1 (used as a positive control) 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). The expression levels in whole-cell lysates or IP extract were determined by IB with anti-Mbd3. The analysis demonstrates direct interaction of Mbd3 with OSKM pluripotency factors, but not with Nanog. Hdac1 was used as a positive control. (b) Overexpression of Flag-tagged Mbd3 simultaneously with Oct4, Sox2, Klf4, c-Myc or Nanog in HEK293 cells was followed by co-immunoprecipitation (co-IP) assay.
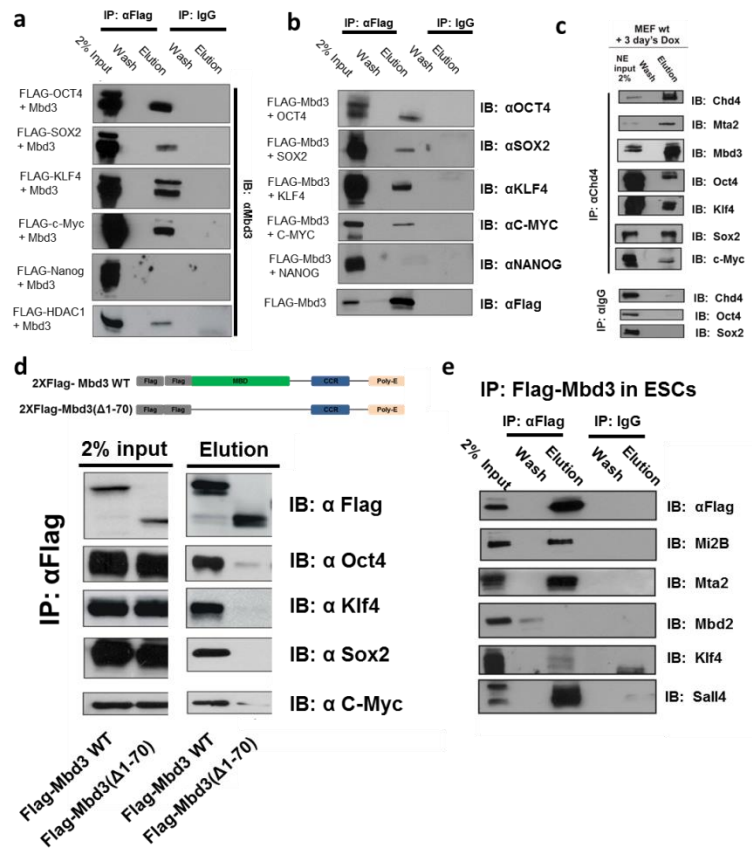
Immunoblot analysis (IB) using antibodies against Oct4, Sox2, Klf4, c-Myc and Nanog showed specific binding between Mbd3 and the pluripotent factors except Nanog. (c) Co-IP assay of Chd4 (Mi2b), the core subunit of the NuRD complex, in secondary Mbd3$^{+/+}$ fibroblasts 3 days after Dox induction under 2i/LIF conditions. Co-IP for NuRD component Chd4 followed by IB analysis indicated specific pull-down of other Mbd3/NuRD components (Mbd3 and Mta2) and OSKM reprogramming factors. (d) To find the binding region of Mbd3, flag-tagged WT and mutant constructs were co-transfected with Oct4, Sox2, Klf4 and c-Myc in HEK 293T cells for 48 hours, followed by co-immunoprecipitation with anti-flag beads and immunoblot against Oct4, Sox2, Klf4 and c-Myc. The analysis showed loss of binding and interaction between Mbd3 and OKSM when MBD domain was deleted. (e) We established an ESC line carrying recovery of flag-tagged Mbd3 transgene inserted in Mbd3$^{-/-}$ ESC line. Co-immunoprecipitation using flag beads followed by western blot shows that Mbd3 strongly binds the Mi2b and Mta2 members of the NuRD complex, but not Mbd2. Moreover, Mbd3 interacts with klf4 and Sall4 pluripotency reprogramming factor in ESCs.

Consistent with the direct protein interactions for Mbd3/NuRD complex with OKSM during reprogramming, genome-wide ChIP-seq analysis of Mbd3 binding in Dox-induced WT MEFs identified a global increase in Mbd3 recruitment and binding following OKSM induction (1177 binding regions in MEF, as compared to 8657 following OKSM induction, see online **Table 1**[v]). We found that only after Dox induction, Mbd3-bound genes were enriched for targets of Klf4, Oct4, Sox2 and Esrrb ($P<10^{-22}$), as well for genes with H3K4me3 active chromatin mark in ES cells ($P<10^{-38}$)

---

[v] At: **http://www.nature.com/nature/journal/v502/n7469/extref/nature12587-s2.xlsx**.

(Fig. 9a). Motif search analysis in Mbd3 binding locations identified 30 motifs ($P<0.001$), of which the motifs of Klf4, Sox2, and Oct4 were among the top six ($P<10^{-30}$) **(Fig. 9b)**. Importantly, in somatic MEFs prior to OKSM induction, Mbd3 was not localized to pluripotency factor target genes **(Fig. 9a).** Transcription level of Mbd3 target genes following 4 days of Dox was significantly upregulated in Mbd3-depleted samples **(Fig. 9c)**, consistent with predominant function of Mbd3/NuRD complex as a repressor of pluripotency gene network during differentiation and reprogramming[22,32,54]. Chromatin of Mbd3 direct targets was significantly more active and open in Mbd3-depleted samples during reprogramming, including statistically significant higher levels of H3K4me3 and H3K27ac, and reduced H3K27me3 repressive chromatin mark (**Fig. 9d**). Further, Mbd3 depletion allowed enhanced exogenous Oct4 binding to targets of Mbd3 at day 4 following Dox ($P<10^{-16}$, **Fig. 9d**). Finally, Mbd3 mutants with compromised ability to interact directly with OKSM were deficient in reducing reprogramming efficiency of Mbd3[flox/-] somatic cells (**Fig. 9e**), supporting the notion that direct OKSM-Mbd3 interactions are important for inhibiting iPSC formation.

Figure 9. Direct OKSM-Mbd3 interactions are important for inhibiting iPSC formation. (a) Functional enrichment of Mbd3 direct targets as measured by ChIP-seq in somatic MEFs before and after OSKM induction. Color levels indicate enrichment P-values above the FDR threshold of 0.0001; white indicates values below the threshold. (b) Motif enrichment analysis for Mbd3 binding regions following OSKM induction during reprogramming. Shown are the sequence logos of abundant motifs along with associated factors, Z-score and P values. (c) Distribution of gene expression fold change relative to MEF of Mbd3+/+ samples (blue) and Mbd3flox/- samples (red) throughout reprogramming (day 0, 4, 8 and 11 and iPSC/ESC). Box centers indicate the median value, and box edges indicate the 25th and 75th percentiles. All identified binding targets of Mbd3 (1400 genes) from

22
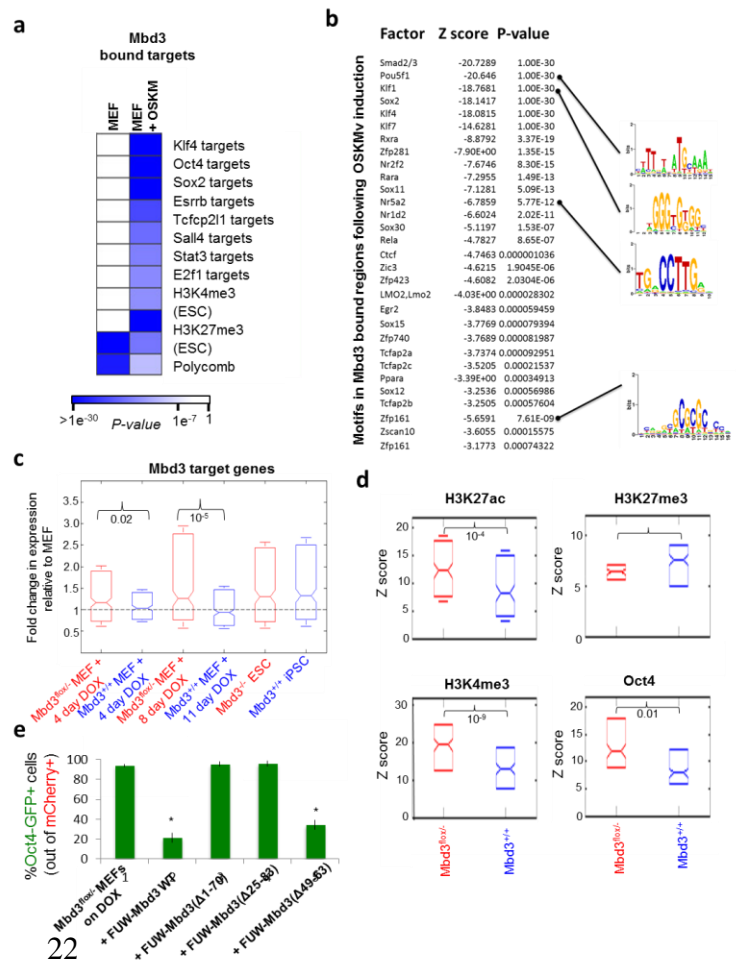
Mbd3+/+ cells following OSKM induction were analyzed. Results show general activation of Mbd3 targets throughout the reprogramming process and accelerated activation in Mbd3flox/- samples. (d) Distribution of histone marks and Oct4 binding levels in z-score values at day 4 after OSKM induction. All identified binding targets of Mbd3 from Mbd3+/+ cells were analyzed. Box centers indicate median and edges indicate the 25th and 75th percentiles. Results show a significant induction of H3K27ac and H3K4me3 and reduction in H3K27me3 in Mbd3flox/- relative to Mbd3+/+ cells, as well as induction of Oct4 binding in Mbd3flox/- samples. (e) Reprogramming efficiency of Mbd3flox/- MEFs after infection with lentiviruses encoding WT and mutant Mbd3 inserts, as indicated. Error bars indicate sd of biological triplicates; asterisks Indicate significant differences (P<0.001) from uninfected control sample.

Mbd3 acts through the Gatad2a/Chd4 within the NuRD complex to inhibit iPSC reprogramming

To better understand the inhibitory mechanism and to find whether Mbd3 effect was mediated by the NuRD complex, we used transient silencing assay with siRNAs to screen for other NuRD component during the reprogramming process. We used a transgenic "secondary" MEF cell line, which already contains heterozygote tetO-inducible OKSM under the Col1a loci and heterozygote rtTA (transactivator) cassettes under the Rosa26 loci (**Fig. 10a**). These cell lines are able to generate iPSCs following administration of doxycycline and without any other manipulation, thus enabling more reliable and accurate monitoring and quantification of the reprogramming process. The NuRD complex components can dramatically vary between pathological and physiological states[32,56,57]. Therefore, we examined several core complex members, namely Mbd3, the mutually exclusive Mbd2, Chd4, Chd3, Mta2, Hdac2, Gatad2B and Gatad2A (also known as p66β and p66α, respectively). Secondary MEF cells were seeded over feeders and treated with mESC medium and Dox. siRNA transfections took place 48 hours and 96 hours after Dox administration (**Fig. 10a**). After 9 days, reprogramming rate was determined by counting positive alkaline-phosphatase colonies. Treating the cells with siRNA for Mbd3 has significantly improved the reprogramming rate as expected (**Fig. 10a**). In addition, a significant improvement in the reprogramming rate was seen in cells treated with siRNA for Chd4 and Gatad2A. The depletion of the other NuRD components had no effect on the reprogramming rate.

Figure 10. siRNA treatment for Mbd3, Gatad2a and Chd4 improves reprogramming. (a) R26-RtTA homozygous mice were mated with Col1a1-tetO-OKSM; MEF were harvested at E12.5 and grown in MEF medium. 1000 cells/well were seeded on feeder coated 24 well plates and reprogramming was initiated by DOX addition. 48 and 96 hours post dox induction siRNA transfection for different NuRD components were applied. Cells were treated as seen in the scheme. At day 9, cells were fixated and stained for AP. One of 3 representative experiments is shown. Average of duplicates is shown per condition. Error bars indicate sd, asterisks indicate significant differences (P<0.01) in comparison to siControl treatment, showing the dramatic increment in colony formation in the siMbd3, siChd4 and siGatad2a treatments.
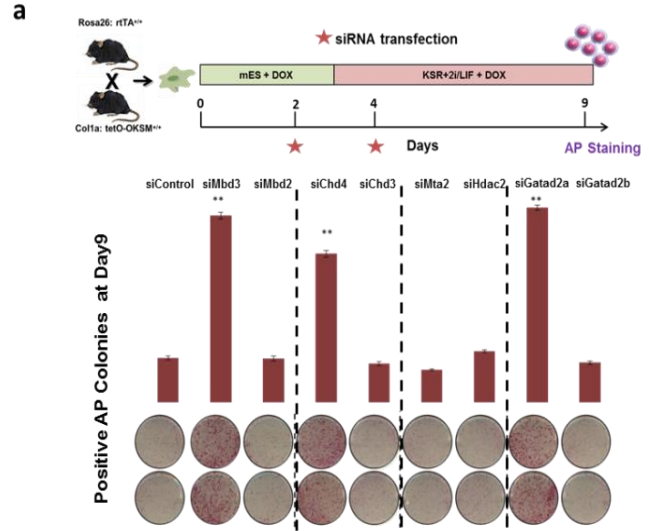
Previous reports on the NuRD complex structure and function show that the recruitment of Chd4 to the Mbd2-containing NuRD is mediated through Gatad2A. Mbd2 and Gatad2A can form a stable heterodimer complex through their coiled-coil regions. Thus, we first examined the similarity of the coiled-coil region between Mbd2 and Mbd3. Results showed that these regions are highly conserved across organisms and related proteins, including Mbd3 (**Fig. 11a**)[58,59]. To examine the binding of Gatad2A to Mbd3, we created a 2xflag Mbd3 expression vector and a mutant version of Mbd3, which lacks the coiled-coil region **(Fig. 11b).** The elimination of this region prevented Gatad2A binding to Mbd3 and, thereby, the recruitment of Chd4, while the mutant maintained its ability to bind other core members of the NuRD complex, such as Mta2. Moreover, examining the interaction between Mbd3 and Chd4 at day4 during reprogramming showed that the Chd4-Mbd3 interaction is disrupted by using siRNA for Gatad2a (**Fig.11c**). Together with the remarkable change in reprogramming efficiency upon Gatad2A knockdown (**Fig 10a**), this notion emphasizes the importance of Chd4 recruitment to the Mbd3-NuRD complex for its involvement in the process.
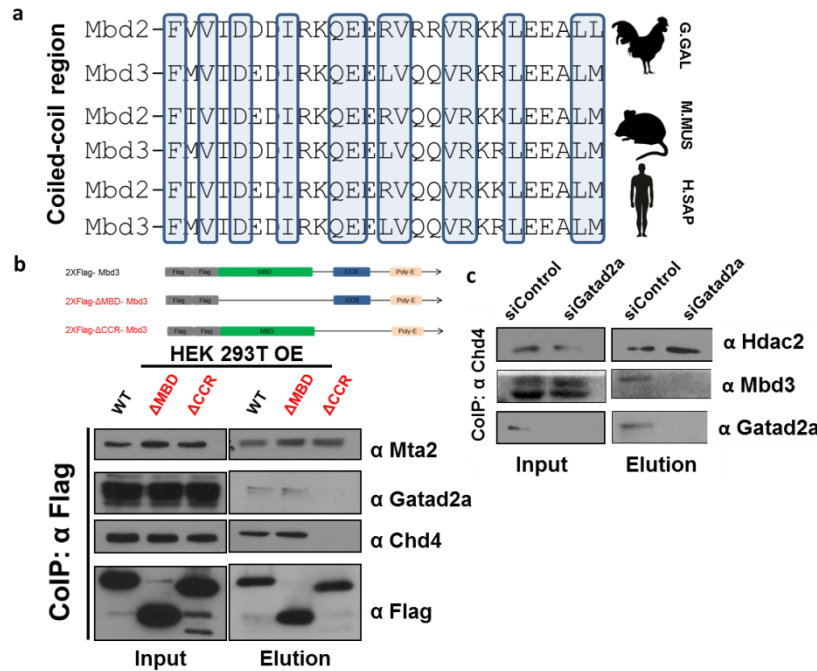
Figure 11. Gatad2a mediates Mbd3-Chd4 interaction through Mbd3 coiled-coil domain. (a) Alignment of the coiled-coil regions of Mbd2 and Mbd3 showing a high homology in this domain between the two proteins in *H. Sapiens*, *Mus musculus* and *Gallus gallus*. (b) Mbd3-WT was marked with a 2XFlag Tag. Two different plasmids were derived from the WT plasmid, expressing mutant versions of Mbd3: the first one lacks the methyl binding domain (ΔMBD) and the second lacks the coiled-coil region, which is necessary for the binding of Gatad2a (ΔCCR). The WT and two Mbd3 mutants were overexpressed in 293T cells. After 28 hours, the cells were harvested and Mbd3 interactions were examined using a Flag-CoIP. Whereas the elimination of the MBD did not diminish Mbd3 binding to different NuRD components, the elimination of the CCR selectively altered the NuRD structure, as the mutant maintained the ability to bind to Mta2, but lost the connection to both Gatad2a and Chd4. (c) Secondary MEFs were treated with Dox to initiate reprogramming and after 48 hours, some of the cells were transfected with siGatad2a. The cells were harvested after 96 hours, and Chd4 connections were examined by Chd4-CoIP. Cells that had been treated with siGatad2a continued to bind Hdac2, but could not bind to Mbd3. The dramatic reduction in Gatad2a protein levels can be observed in the input control.

We further examined the effect of Gatad2a abolishment on reprogramming by establishing Gatad2a KO secondary reprogrammable lines. For that, we used primary iPSC that were generated from MEF harboring R26-M2rtTA[+/-] STEMCCA Dox inducible OKSM polycistronic cassette[51], a constitutive nuclear mCherry marker, and an Oct4-GFP specific reporter[52]. The latter cells were injected into host blastocyst and secondary reprogrammable MEFs were derived and utilized for reprogramming quantifications[49] (**Fig. 12a**). As previously, single-cell sorting of mCherry+ Gatad2a KO

MEFs and subsequent reprogramming yielded 100% iPSC derivation efficiency by day 8. In the isogenic WT cells reprogrammed under identical conditions, no more than 6% of clones reactivated Oct4-GFP. Moreover, when 5 cells were seeded per well, no more than 8% of the wells reactivated the reporter, while the majority of mCherry secondary fibroblasts cells had neither reactivated Oct4-GFP marker nor acquired ES-like morphology (**Fig. 12b** and data not shown). Radically high reprogramming efficiency rates were obtained with multiple Gatad2A KO secondary lines established with different induction platforms, e.g. using knockin Col1a tetO-OKSM cassette or viral induction of the STEMCCA OKSM (**Fig. 12c**). All randomly tested clones stained positive for alkaline phosphatase (AP), SSEA1 and Nanog pluripotency markers (**Fig. 12d**), and contributed to three germ layers when injected to NSG mice to form teratomas (**Fig. 12e**). We also evaluated the Gatad2A KO reprogramming kinetics using FACS analysis. The reactivation of the Oct4-GFP reporter started to appear after 3 days of OKSM induction, and its rate kept increasing until reaching >93% after 8 days of reprogramming in total cell population. In the isogenic WT cells reprogrammed under identical conditions, no more than 20% of cell population was positive for that reporter (**Fig. 12f**). Finally, re-infection with lentiviruses encoding for Gatad2A resulted in resumption of Mbd3 interaction with Chd4 and led to a major reduction in iPSC reprogramming rates (**Fig. 12g, h**).

To conclude, we show that the inhibitory effect of Mbd3 is mediated by interaction with OKSM factors through the MBD domain to repress their targets during reprogramming. The inhibitory function of Mbd3 is NuRD-dependent and mediated through the recruitment of Chd4 and Gatad2A by Mbd3 coiled-coil domain. Finally, by using secondary reprogramming platform harboring Gatad2A KO, we show the same deterministic reprogramming effect observed with the hypomorphic Mbd3$^{flox/-}$ platform.
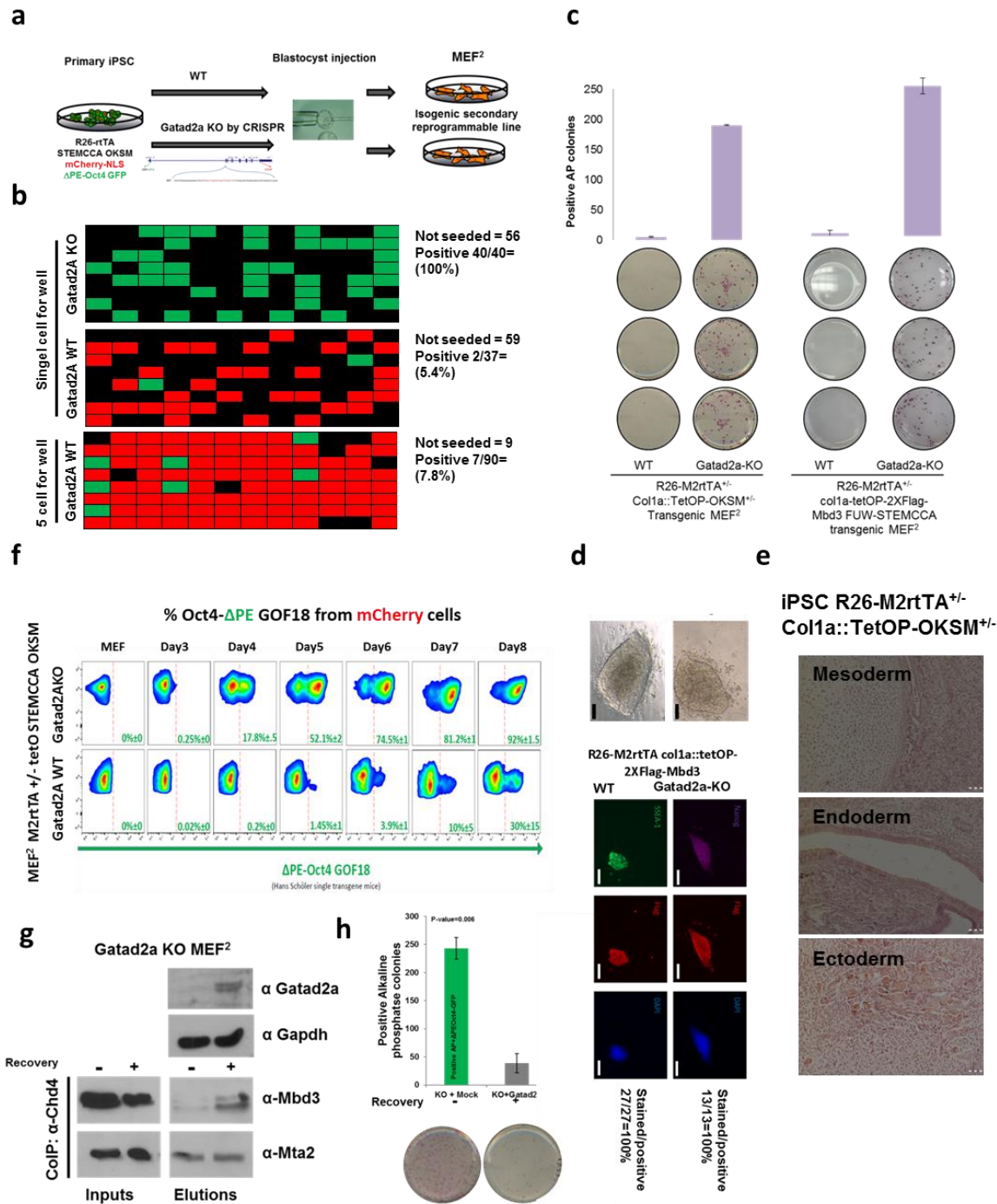
Figure 12. Gatad2a KO cell lines display deterministic reprogramming features. (a) Secondary isogenic cell lines were established using iPSC harboring R26-rtTA, FUW STEMCCA tetO OKSM polycistronic cassette, constitutive mCherry-NLS and ∆PE-Oct4-GFP naïve pluripotency specific marker. Gatad2a was targeted using CRISPR-Cas9, resulting in a KO line identical to the WT, besides lack of Gatad2a expression. Both the WT and Gatad2a-KO iPSC were then injected to a blastocyst and reprogrammable MEFs were derived from both cell lines. (b) Secondary MEFs were seeded as a 0.5 cells to each 96-wells plate and treated with Dox in order to initiate reprogramming. Reprogramming was estimated after 8 days by Oct4-GFP presence and the acquisition of an ES-like

27

morphology. The Gatad2a-KO cells yielded 100% iPSC derivation efficiency, as compared to 6% in WT cells treated by the same protocol. Seeding 5 cells per well improved reprogramming rate only to 8%. (c) Two other secondary cell lines were established in the same methodology as previously. Both show very high reprogramming efficiency compared to their isogenic control. (d) Arbitrary wells from single-cell experiment at day 8 using secondary reprogrammable line harboring R26:M2rtTA, Col1A:tetO Flag-Mbd3WT, FUW STEMCCA tetO OKSM cassette were selected. Two representative bright filed images of iPSC colony morphology are shown. Outer colonies were stained for Flag (seeding efficiency) and SSEA1 (27 out of 27 wells that tested) or Nanog (13/13 wells) for pluripotency markers showed 100% reprogramming success. (e) iPSC from the latter line were injected to NSG mice to test for teratoma formation. After 3 weeks, tumors were isolated and stained for H&E. All three germ layers were detected in all teratomas obtained. Representative images from each lineage are shown, confirming pluripotency and teratoma capability. (f) Flow cytometry measurement of ΔPE-Oct4-GFP reactivation dynamics in 2i/LIF conditions at 72 hours following Dox (OSKM) induction. Gatad2a KO secondary cells synchronously and rapidly reactivate ΔOct4-GFP by day 8 in over 90% of donor cell population, whereas isogenic WT reached only 30%. One out of 3 independent experiments is shown. (g,h) Reprogramming efficiency of Gatad2a KO MEFs after infection with lentiviruses encoding WT and mock for Gatad2a insert, as indicated in the panel. Error bars indicate sd of biological triplicates ($P<0.006$ in comparison to mock-infected control sample). One of 3 representative wells stained for alkaline phosphatase is presented.

Part3: High resolution profiling of deterministic iPSC reprogramming following Mbd3 or Gatad2a depletion

Secondary reprogrammable Mbd3$^{flox/-}$ and Gatad2a KO platforms enabled us to study reprogramming for the first time in a robust and synchronized system, which transforms somatic cells into genuine, induced pluripotent stem cell in only 8 days. More importantly, we could conduct our experiment without cell passage and without sorting for subpopulations during the entire process. Our platforms are highly suitable for analysis of chromatin and transcriptional dynamics during the reprogramming process and of its epigenetic trajectory, offering the advantage of overcoming the inefficiency and stochasticity of other currently used methods and systems[38,60-63]. In this section, I will describe a part from this project, which summarizes the OKSM function in driving the iPSC reprogramming process. We show that Oct4 and Sox2 are the main drivers of the pluripotent state, whereas Klf4 and c-Myc shift cell suitability from the somatic cell state into the pluripotent one.

Our reprogramming protocols **(Fig. 13a)** were applied on secondary MEFs reprogrammable of Mbd3$^{flox/-}$, Gatad2a KO and WT platforms. During reprogramming, cells were harvested every 24 hours until they completed iPSC transition at day 8, and were then used for library preparation followed by sequencing. The same was done with MEFs and established iPSCs (after 3 passages or more) from each platform to be used as controls. We generated 18 libraries at 11 time points, resulting in overall 198 samples for the Mbd3$^{flox/-}$ platform and additional 40 samples from selected time points to compare with Gatad2a KO and WT platforms. The libraries spanned transcriptome (RNA-seq, small RNA-seq, DGE-seq), chromatin modifications (ChIP-seq for H3K27ac, H3K27me3, H3K4me1, H3K4me3, H3K36me3, H3K9me3), DNA methylation (reduced representation bisulfite sequencing (RRBS), whole-genome bisulfite sequencing (WGBS)), chromatin accessibility (ATAC-seq) and transcription factor binding (ChIP-seq for Oct4, Sox2, Klf4, c-Myc and PolII) (**Fig 13b**). We aligned the reads to mm10 genome, and employed global peak calling for chromatin accessibility, histone modifications and TF binding samples. The highly reproducible data revealed high correlation between constitutive samples. Data from iPSC and mESC samples were

compared to previously published data, and showed high correlation and significant overlap with the previously established peaks (data not shown).
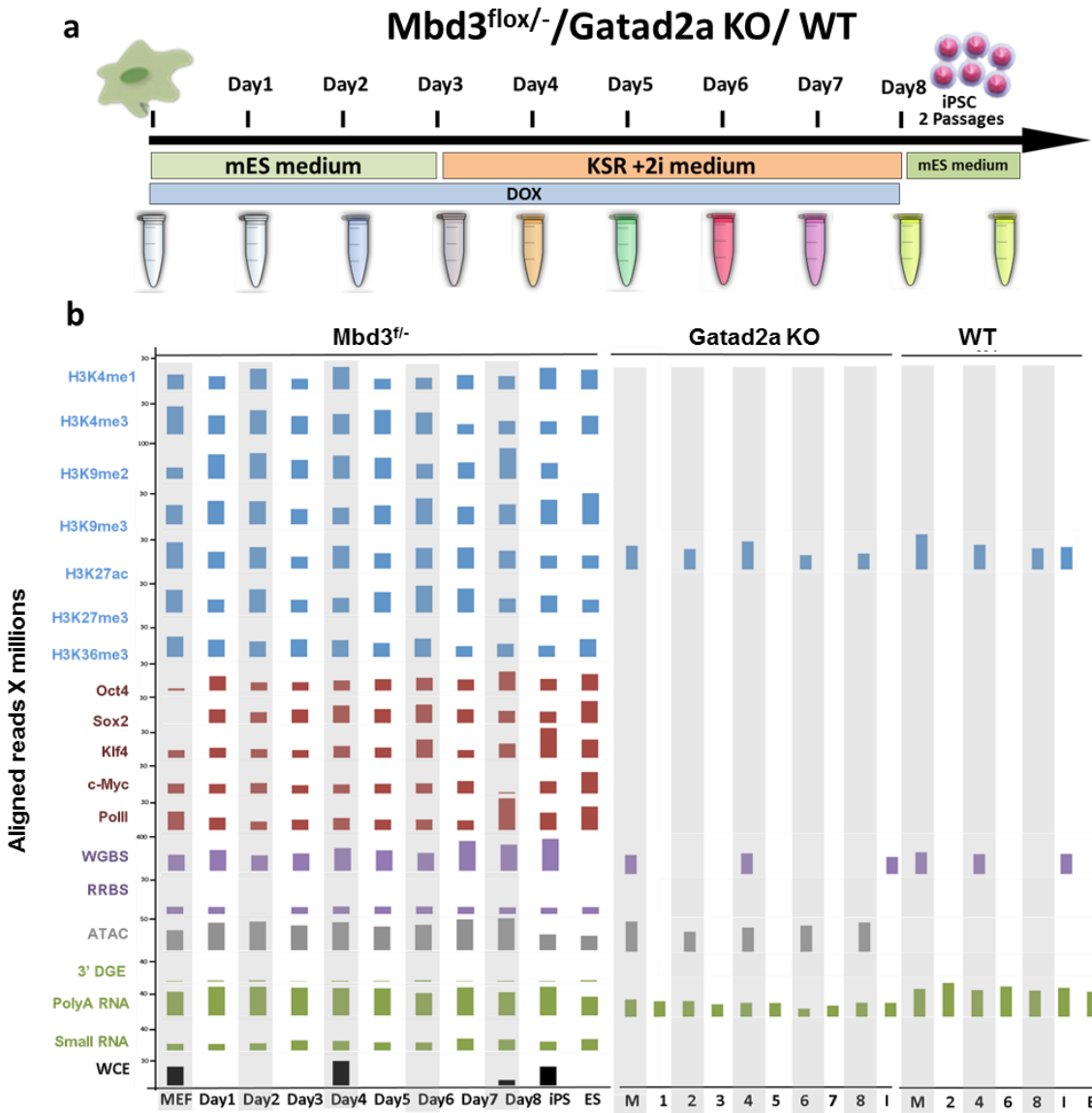
Figure 13. Summary of data collection during reprogramming and library preparation. (a) Experimental design for collecting samples for tracing of reprogramming. (b) The number of aligned reads mapped in each sample in the systems described here: Mbd3[f/-], Gatad2a KO and WT. Abbreviations: WGBS, whole-genome bi-sulfite; RRBS, reduced representation bi-sulfite; ATAC, assay for transposase-accessible chromatin; DGE, digital gene expression; WCE, whole-cell extract.

Transcriptional trajectory during reprogramming reveals hierarchical progression towards naïve pluripotency

To gain molecular insight into the transcriptional states of reprogrammed cells, we measured full transcriptomes using polyA RNA-seq from samples taken every 24 hours during reprogramming using the Mbd3$^{f/-}$ reprogrammable platform. We identified 8705 genes that were differentially expressed through the reprogramming process. These genes exhibited a sequential expression pattern and could be sorted according to expression peak time **(Fig 14a)**. Intriguingly, the pattern revealed a continuous and hierarchical transition from the somatic program to the pluripotent one. We next compared the transcriptional profile of Mbd3$^{f/-}$ system with that of the Gatad2a KO system and of two secondary WT reprogramming platforms. The clustered samples of Mbd3/NuRD-deficient systems showed the same hierarchical transition, whereas both WT samples until day8 showed high resemblance to early reprograming stages **(Fig 14a)**. Spearman correlation analysis of Mbd3$^{f/-}$ and Gatad2a KO revealed very high correlation between the two NuRD-deficient systems, suggesting that the two allow induction of fast and synchronized reprogramming by affecting the same pathway **(Fig 14b)**. Closer examination of the transcriptional profiles identified three major expression shifts during reprogramming. The first involves a large group of genes that are active in MEFs, which are downregulated as early as day 1. The second is a transient activation of genes between days 1 and 4, some of which are not expressed at the final iPS/ES cell state. Finally, there is a gradual establishment of iPS/ES signature starting at day 5 **(Fig 14a)**. Functional enrichment analysis at a single day resolution **(Figure 14d)** revealed that the downregulated MEF-associated genes were enriched for somatic program processes (e.g. developmental process, skeletal system, cytokine mediated signaling). Genes that were induced between days 1 and 6 were enriched for processes related to enhanced proliferation and DNA synthesis (purine biosynthesis processes, translation, ribosome biogenesis, DNA replication). These processes were followed by induction of genes enriched for epigenetic transformation and DNA repair processes (chromosome segregation, DNA repair, DNA recombination, response to DNA damage). Finally, at day 8 there was an induction of stem cell maintenance genes, including Nanog, Esrrb, Tbx3, Sall4, Prdm14 and others **(Fig 14a)**.
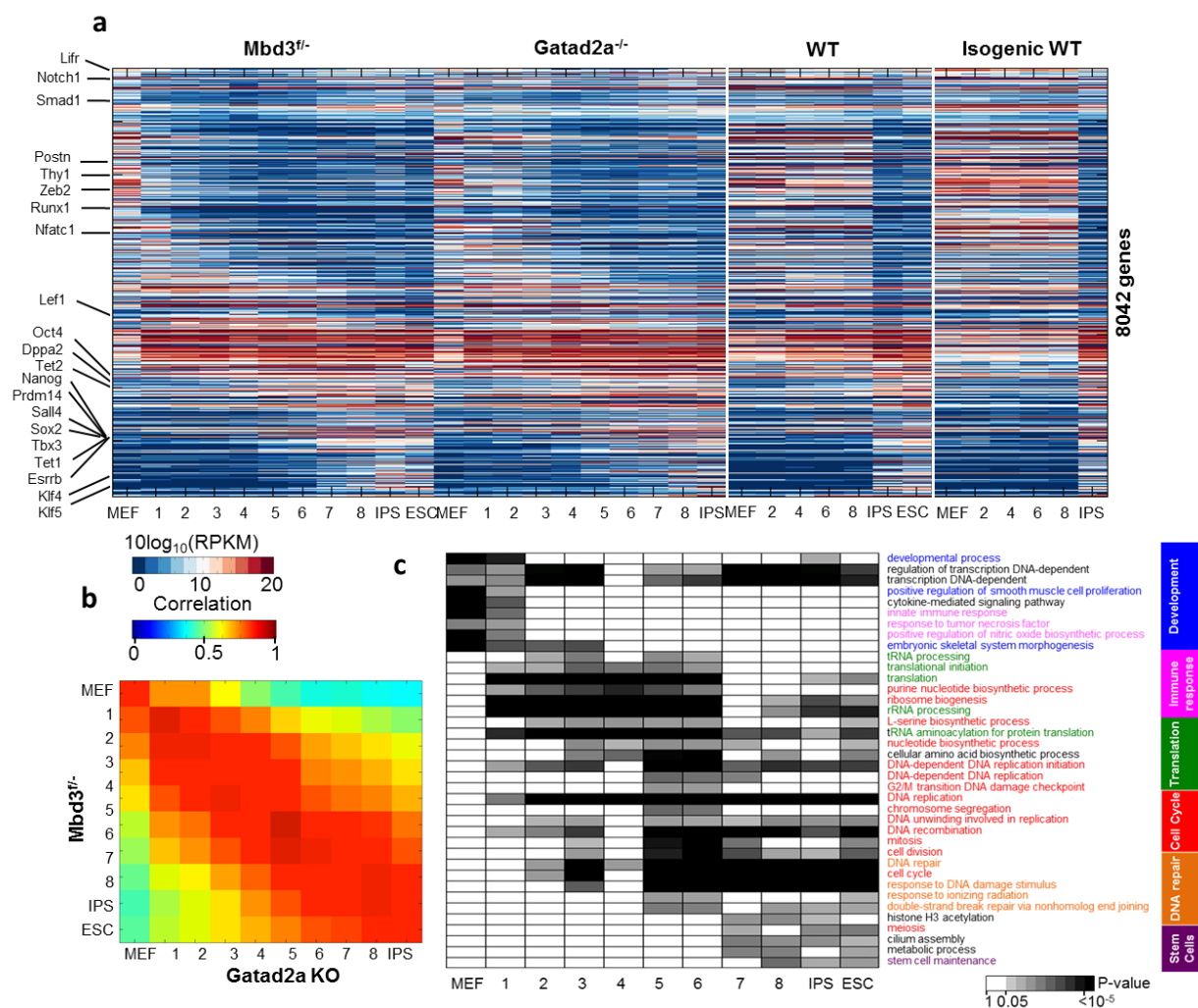
Figure 14. Global transcriptional changes in reprogrammed cells. (a) Global transcriptional pattern of 8705 differential genes, sorted by peak time, in Mbd3$^{f/-}$, Gatad2a KO, WT and isogeneic WT for Gatad2a KO cell lines. Heatmap shows unit-transformation of RPKM values, ranging from 0 (minimum) to 1. (b) Average Spearman correlation between Mbd3$^{f/-}$ and Gatad2a KO samples from the same time points was R=0.86. (c) Enriched GO categories in genes that are active at each stage on the Mbd3$^{f/-}$ platform. Categories where *P*<1E-6 at at least two stages are presented.

Two distinct functional groups play roles in cellular reprogramming

The reprogramming of cellular fate involves not only changes in transcription, but also dramatic epigenetic modifications, from the 'closed' heterochromatin in somatic cells to the 'open' euchromatin conformation in ESCs. Moreover, the epigenetic environment of pluripotency-promoting loci has to become accessible and to adopt the proper histone modification architecture and methylome state to be properly functional. Therefore, we next sought to detect epigenetic changes in the reprogramming system and

examine their regulatory role. For each differentially expressed gene with at least two chromatin marks in its promoter (n=7801), we calculated the correlation between its transcriptional pattern and chromatin mark patterns, measured around the TSS or TES. Analysis showed that chromatin marks were clustered into two groups **(Fig 15a)**. One cluster consisted of marks that were positively correlated with transcription, which are known to be associated with active transcription, such as H3K4me3, H327Ac and H3K36me3, and with accessible chromatin measured by ATAC-seq. The other cluster contained known repressive marks that were negatively correlated with transcription, such as H3K27me3 and H3K9me3. Interestingly, the genes were also clustered into two main groups. The first consists of **3593** genes that exhibited high correlation (either positive or negative) between transcription and chromatin marks, and the other consists of **4209** genes that did not show any correlation despite being differentially expressed **(Fig 15a)**.

To validate these results, we tested the actual transcriptional and epigenetic patterns for these two gene clusters, focusing on H3K27ac and H3K27me3 marks, which showed the highest positive and negative correlation to transcription. The genes in the first group showed a clear switch-like behavior between the marks **(Fig 15b),** correlated with the activation or repression of the gene. We therefore concluded that these are epigenetically regulated genes (ERGs). In the second group, the majority of the genes (n=3049, 72%) had differential transcription (fold change above 4), but with consistently high levels of H3K27ac and low levels of H3K27me3 (z-score<0.7, **Figure 15b**). The promoters of these genes displayed a constitutively active chromatin signature, suggesting that these genes are regulated at a different level. We termed this group CAPG, for constitutively active promoter genes. Interestingly, examination of the functional enrichment revealed distinct biological functions for the two groups. Whereas ERGs are enriched for cell fate determination, CAPGs are enriched for cellular adaptation processes, including proliferation, DNA repair and chromatin organization (**Fig 15c**). In addition, the two groups have different temporal behavior. ERGs activity gradually changes during reprogramming, whereas CAPG activity undergoes two major shifts, at day 1 and at the final stages of reprogramming **(Fig 15b).** Finally**,** the two defined groups were shown to be differently regulated by OSKM factors. Whereas ERGs

33

are enriched for targets of Sox2, Oct4 and Klf4 and not to the c-Myc, CAPGs are enriched for OSK and mostly for c-Myc (**Fig 15e**).

Figure 15. Epigenetically regulated genes (ERGs) and constitutively active promoter genes (CAPGs). (a) Spearman correlation analysis of differential genes that have at least two chromatin marks in their promoter (*n*=7801), with each of the indicated marks across all samples. Genes were clustered into two groups: genes with positive correlation to active marks and negative correlation to repressive marks (*n*=3593) and genes with no trend of correlation (*n*=4208). (b) Pattern of H3K27ac and H3K27me3 in ERGs (top) and in CAPGs (bottom). (c) Enriched GO categories among genes that are active at each stage clustered by ERPGs and CAPGs. (d) Correlation between differential ERPGs and CAPGs measured in different samples. (e) Enrichment of OSKM binding targets in promoters of ERGs compared to CAPGs. Fold changes in *P*-values are indicated.

Discussion

In the first part of my work, I show that the stochastic and asynchronized trajectory of classical direct reprograming by OSKM[12] can be coaxed to become radically more efficient and deterministic through modified reprogramming approaches. We highlight Mbd3/NuRD repressor complex, which is naturally depleted during normal pre-implantation development, as a predominant barrier preventing epigenetic reversion of EpiSCs and somatic cells to ground state pluripotency by defined signaling and transcriptional input. The ability of Mbd3/NuRD to block reprogramming of mouse somatic cells is supported by Luo at et.[22]. Here, we further emphasize the dramatic effect of Mbd3 elimination during reprogramming and provide the mechanism underlying this process.

The concept that reprogramming of somatic cell can be deterministic was already demonstrated by SCNT and cell fusion experiments[5,6]. Recent study by Buganim has indicated that successful direct reprogramming of somatic cells into iPSCs can be achieved at the later stages of the process, following an early stochastic phase[60]. Still, the rate-limiting event that would shift reprogramming from stochastic to deterministic was yet to be uncovered. We show that the stochastic phase can be bypassed to produce a single and dominant deterministic transition. We further show that an array of critical reprogramming factors directly interact with and recruit the Mbd3/NuRD complex, thereby forming a highly potent and "self-imposed" negative regulatory complex that restrains reactivation of pluripotency gene throughout the process. The inhibitory function of Mbd3/NuRD on pluripotency-related genes have been shown to be critical at early developmental stages, particularly during blastocyst implantation stage[29,30,32]. It was suggested that the repression of the pluripotent circuit by Mbd3/NuRD involves interaction with the promoters of active genes, which restrains their expression[32]. Finally, we show that inhibiting Mbd3 in human fibroblasts in pluripotency-promoting medium leads to radically fast and efficient reprogramming[64].

Despite the positive effect of Mbd3 reduction during reprogramming, complete ablation of Mbd3 rapidly yields a block in somatic cell proliferation. Cell fusion and SCNT experiment showed that proliferation-deficient cells cannot be reprogramed[65,66].

This side effect of Mbd3 abolishment explains the failure in reprogramming when it is done before preliminary transition occurs. This has limited the flexibility of manipulating the pathway to achieve robust reprogramming. Therefore, in the second part of my work I aimed to identify ways to block Mbd3 without the negative effect on somatic cell proliferation and viability. It was suggested that the mechanisms underlying Mbd3 inhibitory role in pluripotency repression is NuRD-dependent[32]. The NuRD is a highly dynamic epigenetic complex, whose composition varies according to its involvement in physiological and pathological states, including developmental processes, DNA damage response and cancer metastasis[32,56,57]. Moreover, many NuRD complex proteins play additional roles in the cell and take part in other complexes. For example, in addition to its role in cell cycle regulation, Chd4 participates in DNA damage response processes independently of NuRD[67]. Mbd3 and Hdac1, both NuRD components, have also been identified as cell cycle-related proteins, as they both co-localize with the aurora kinase in the centrosomes during mitosis[68]. Hdac1 and Hdac2, as well as RbAP46 and RbAP48, are not NuRD-specific and can be found in other epigenetic complexes such as Sin3[69]. Other proteins have also been suggested to be a part of the NuRD complex and to influence pluripotency or reprogramming. These include Doc-1[70], Zmynd8[71], a zinc finger protein, Lsd1[61,72] and Zfp281[54], which might be a part of the complex at specific context-dependent situations.

None of these possible interactions, however, explained the dramatic effect of Mbd3 depletion, suggesting a more fundamental role for Mbd3 in inhibiting this process. Indeed, knockdown of the canonical NuRD complex components Mbd3, Chd4 and Gatad2a led to highly efficient reprogramming. Moreover, we show that the three can form a sub-complex within the NuRD. Gatad2a can interact with Mbd3 through the coiled-coil domain and recruit Chd4 to the complex. Importantly, unlike Mbd3 or Chd4, the complete ablation of Gatad2a does not impair cell proliferation and viability. Mechanistically, we show that Mbd3- and Gatad2a-mediated reprogramming have the same trajectory and deterministic features of a fast and synchronized process. The Mbd3/Gatad2a/Chd4 axis within the NuRD complex have a specific and direct function related to protein-protein interaction during the dynamic states of differentiation and reprogramming (data not shown).

Mbd2 and Mbd3 can form a heterodimer and are considered mutually exclusive in the NuRD complex, as each combination presumably has a different function[23]. Our attempts to compensate for Mbd3 ablation using Mbd2 were unsuccessful. Indeed, in another study conducted in our lab we show that the Mbd3/NuRD complex in naïve or primed ESCs does not contain Mbd2 during differentiation or reprogramming (unpublished data). Similarly, the NuRD complex can contain Gatad2a or its homolog Gatad2b. Although we did not succeed to compensate for the lack of Gatad2a with Gatad2b, further analysis is required to determine whether Mbd3 can interact with Gatad2b and whether the latter has the same functional properties as its homolog.

To conclude, we characterize the Mbd3/Gatad2a/Chd4 axis within the NuRD complex as a key regulator of pluripotency and reprogramming. Due to the highly complex architecture and dynamic nature of NuRD, further studies are needed to fully decipher the function of this axis.

Mbd3/NuRD disruption may not be the only strategy for inducing high reprogramming rate. Lately, it was suggested that some chromatin remodelers are important to preserve genome architecture[73]. CAF-1 and Mbd3/NuRD inhibition may cause a unique cell state that can contribute to cell plasticity, facilitating the conversion between cell fates. The function of priming the cells for reprogramming was also shown to be performed by a short induction of C/EBPa, but only in B-cell reprogramming[61].

Having established two NuRD-deficient platforms with defined mechanistic features, we were ready to start asking questions regarding how reprogramming works. By using those independent rapid and synchronized reprogramming platforms with the appropriate controls we first validated that our MEF resemble transcriptionally and epigenetic-wise to the control MEF, then we started to follow direct reprogramming trajectory following OKSM administration day by day until the formation of authentic naïve iPSC after 8 days, without passaging the cells or sorting for specific population. Those powerful tools have showed us very high correlation and reproducible data between the two NuRD deficient platforms. The dynamic process in the total population was very robust and clean compared to other reprogramming data that was previously described[13,38,60-63]. Using this data enabled us to produce high detailed reprogramming maps, which include transcriptional dynamics of mRNA and small-RNA and the many

layers of the epigenetic dynamics, including histone modification, chromosome accessibility, DNA methylation and transcription factor binding sites. These maps enabled us to follow the causality of the process during reprogramming. I have showed a small example of our findings suggesting the mechanistic of the factors divided to two parallel paths: as one shifts cells adaptation (metabolism shift, codon usage, proliferation rate and etc.) to ESC state by c-Myc and its binding hardly changes during the reprogramming. The other path is Oct4 and Sox2 dependent, involving cell fate control, shifting epigenetic towards pluripotency, both have dynamic binding during reprogramming. Klf4 plays a role in both pathways. Both pathways contribute and support the other to promote reprogramming. For example: c-Myc induces key proteins from the PRC complexes help to shape together with Oct4 and Sox2 the chromatin structure. Moreover, c-Myc induces TET family enzymes that are involved in the removal of DNA methylation modification as a part of erasing of the epigenetic memory[74]. Further examination of c-Myc function, suggests that c-Myc contributes and co-operates with the OKS factors only at targets of the cell adaptation path. More importantly c-Myc is not involved in the cell fate path. Interestingly, c-Myc binding is restricted to promoters and in almost all cases its binding does not correlates to enhancers. On the other hand, during the earlier days, Oct4 and Sox2 massively bind enhancers and promoters of active genes that are repressed later during reprogramming, suggesting an inhibitory role for Oct4 and Sox2. Still, this insight has to be further validated, especially when at later days the binding is lost, suggesting that this is an unspecific function related to somatic state and does not necessarily promote the pluripotent state. Finally, we have observed a binding for Oct4 and Sox2 in "closed" chromatin validating that those factors act also as pioneer inducers during reprogramming. On the other hand we do not find different binding motifs in "closed" versus "open" binding as was suggested[75]. For future work we are plan to further characterize the function of c-Myc and the other OSK factors. We have very surprising preliminary result suggesting that c-Myc alone can prime the cell for the OSK function in the NuRD deficient reprogramming platform, thus suggesting that upon Mbd3/NuRD inhibition the cell is more "plastic" and prone to fate conversion as was showed with other chromatin remodellers[61,73].

In this work, I did not discuss an entire chapter that involves chromatin dynamics during reprogramming. The NuRD deficient reprogramming platforms are an excellent and very powerful tool, which can be used to study the causality of processes involving the activation and the repression of specific genes along cells conversion. Using those platforms makes it possible to trace global loci dynamics events, in multi epigenetic layers. It is a unique assay for global cell fate dynamics and we hope to find and prove some fundamentals concepts regarding mechanisms that govern cell fate decisions.

To conclude, this work has substantiated the concept of manipulating a cell endogenous regulatory pathway and to use it for our needs in controlling stem cells at will. Producing iPSCs with high reprogramming rate and efficiency is essential for understanding and uncovering the black box of reprogramming and the nature of this process. Hopefully, by using this basic knowledge and concepts will push the field closer to using patient specific stem cells routinely in regenerative medicine treatments.

Experimental methods and procedures

Mouse stem cell lines and cell culture

Reprogramming and maintenance of murine naïve pluripotent cells were conducted in serum-free chemically defined N2B27-based or KSR media. N2B27-based medium contained 500 ml KO-DMEM (Invitrogen), 5 ml N2 supplement (Invitrogen; 17502048), 5 ml B27 supplement (Invitrogen; 17504044), 15% knockout serum replacement (Invitrogen; 10828), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma), penicillin-streptomycin (Invitrogen), 5 mg/ml BSA (Sigma). KSR medium contained 500 ml DMEM (Invitrogen) 15% knockout serum replacement (Invitrogen; 10828), 1 mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1 mM β-mercaptoethanol (Sigma), penicillin-streptomycin (Invitrogen).

Naïve conditions for murine PSCs included 5 μg recombinant human LIF (Millipore; LIF1005). Where indicated, 2i was added 48 hours after OSKM induction: small-molecule inhibitors CHIR99021 (CH, 3 μM; Axon Medchem) and PD0325901 (PD, 1 μM; TOCRIS). Primed N2B27 medium for murine EpiSCs contained 8 ng/ml recombinant human bFGF (Peprotech Asia) and 20 ng/ml recombinant human activin (Peprotech Asia). Stem cell lines and mice deficient for Mbd3 (Mbd3$^{flox/flox}$, Mbd3$^{flox/-}$ and Mbd3$^{+/-}$) and their derived ES lines were obtained as previously described[38, 43].

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 lines, which were then subjected to selection with G418 (300 μg/ml) or puromycin (1 μg/ml). After 10 day of antibiotic selection, drug resistant clones were analyzed for correct targeting by PCR or southern blot analysis. Mbd3$^{+/-}$ male and female mice were mated and E3.5 blastocysts were harvested and explanted for ESC derivation in defined mouse 2i/LIF conditions on MEF-coated plates. NGFP1-Mbd3^{KD} was established by infection and subcloning of secondary NGFP1 iPSC line was performed with a ShRNA pLKO-Tet-On vector (Addgene) as previously described[12]. To produce rtTA-OKSM mice, R26-RtTA homozygous mice (Jackson 006965) were mated with Col1a1-tetO-OKSM (Jackson

01101) homozygotes. F1 mice were mated to create double homozygous offspring (rtTA[+\+]OKSM[+\+]). F2 offspring (double homozygous rtTA[+\+]OKSM[+\+]) were mated with TgOct4 homozygous reporter mice (Jackson 004654) to generate a double heterozygote with TgOct4 reporter offspring (rtTA[+\-]OKSM[+\-]TgOct4[+\-]).

Epigenetic reversion of primed murine epiblast cells

Male naïve V6.5 (Mbd3[+/+]) Nanog-GFP-positive ESCs[11] maintained in 2i/LIF conditions were injected into BDF2 blastocysts and epiblasts. 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 puromycin, allowing the isolation of Nanog-GFP+ EpiSCs. For epigenetic reversion of EpiSCs to naïve pluripotency, cells were passaged into N2B27 2i/LIF conditions on vitronectin- (1 µg/ml) and gelatin- (0.2%) coated plates, without overexpression of exogenous reprogramming factors. For single-cell plating, EpiSC growth medium was supplemented with ROCK inhibitor for 24 hours before trypsinization. siRNAs (ON-TARGETplus SMARTpool or Stealth (Invitrogene) or Silencer (Invitrogene)) and control siRNA (ON-TARGETplus Non-targeting pool D-001810-10-05 or Stealth Rnai Neg Ctl Med (Invitrogene) were purchased from Dharmacon. 75 nM siRNA or control was used for each transfection in 6-well plates with Lipofectamine RNAiMAX (Invitrogen).

Reprogramming of mouse somatic cells and cell infection

Supernatants containing the reprogramming viruses STEMCCA-OKSM polycistronic vector[18] and STEMCCA-OKS polycistronic vector (Dox-inducible and constitutively expressed), FUW-tetO-lox-hKLF4, 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-cMYC) were supplemented with the FUW-lox-M2rtTA virus (when necessary) and an equal volume of fresh culture medium for infection. Mouse fibroblasts and somatic cells of other types were isolated and single-cell sorted from secondary transgenic reprogrammable chimeras[18,65].

iPSCs reprogramming was performed using mouse naïve ESCs medium 2i/LIF and Dox (1 µg/ml; without 2i in the first 48 hours) under physiological 5% O_2 conditions. Mbd3[-/-] cells were reprogrammed by applying 4OHT (1 µg/ml) or tamoxifen to Mbd3[flox/-]

cells 48 hours after OSKM induction. Similarly, for acute knockdown of Mbd3 in somatic cells using Mbd3 siRNAs, transfection was conducted at least 48 hours after OSKM induction. Throughout the study, Mbd3$^{flox/-}$ or NGFP1-Mbd3^{KD} genetic backgrounds were preferably and predominantly used. Notably, in our reprogramming conditions, single cell plating of MEFs yielded approximately 70% survival efficiency (with or without Dox). Thus for live imaging, upon plating 150 MEFs per well we observed formation of 100-120 colonies that were tracked in Mbd3$^{flox/-}$ samples. Constitutive mCherry expression was used to assess survival after plating, providing accurate and unbiased quantification of reprogramming efficiency, defined as the number of Oct4- or Nanog-GFP-positive clones (cells) divided by the number of mCherry+ clones (%). Equivalent reprogramming efficiencies were obtained on mouse irradiated feeder cells or gelatin, matrigel or gelatin/vitronectin coated plates. Reprogramming on irradiated DR4 MEFs was preferably used for live imaging and single-cell reprogramming experiments in order to enhance cell survival and adherence. Mbd3 Stealth siRNA mix that included MSS-237238 ,MSS-275658 and MSS-275659 (Invitrogen) and Chd4 Stealth siRNA mix of MSS-200894, MSS-200895 and MSS-200896 (Invitrogen) were used for efficient knockdown in mouse cells. Transfections were conducted with RNAiMAX (Invitrogen) according to manufacturer's instructions.

Knockdown by siRNA transfection during reprogramming

siRNA transfection was performed with Lipofectamine RNAiMAX (Life Technologies, #13778075). Medium was changed after 24 hours and cells were harvested for protein samples after 48 hours. For reprogramming experiments, cells were re-transfected every 48 hours in order to maintain the knockdown. Knockdown was confirmed by western blot analysis. The siRNAs used are as follows:

Mouse-Mbd3	Invitrogen	MSS-237238
Mouse-Chd3	Invitrogen	MSS238382, MSS238383, MSS238384
Mouse-Chd4	Invitrogen	MSS- 200894, MSS-200895 , MSS-200896
Mouse-P66α	Invitrogen	MSS239240, MSS239241, MSS239242
Mouse- P66β	Invitrogen	MSS213785, MSS213786, MSS213787

DNA plasmids and CRISPR gene editing

The lentiviral and mammalian constitutive overexpression vectors pBRY-Mbd3-Ires-Zeocin were used in somatic and pluripotent cells. Lentiviruses FUW-Mbd2, FUW-Mbd3, FUW-GATAD2A and FUW-GATAD2B were generated by insert cloning into EcoRI sites of FUW vector to generate constitutive expression and stable integration in somatic or PSC lines following viral transduction. Flag-Mbd3 mutations and deletions were done by PCR with Q5 DNA polymerase (NEB). Three different ESC lines were genetically manipulated by CRISPR[41] to achieve P66α KO on different genetic backgrounds. The primers used for sgRNA were annealed and cloned into pX330. Cells were transfected (Xfect, clontech) with Cas9-sgRNA P66α (exon-4) and mCherry-NLS plasmids. mCherry-positive cells were sorted after 72 hours and seeded on feeders. After 8 days, colonies were picked and examined by HRM analysis (MeltDoctor HRM master mix, Life Technologies, #4415440). Candidate colonies were analyzed by western blot and confirmed by sequencing. The procedure was applied to three different lines: TgOct4 WT ES derivation, Kh2 Mbd3-flag and R26:M2rtTA$^{+\backslash-}$ OKSM$^{+\backslash-}$ TgOct4$^{+\backslash-}$ IPSCs. These IPSC lines were established by reprogramming. Cells were treated with Dox to induce the OKSM cassette and kept in MES medium for 14 days. Colonies were picked and kept in mES medium without Dox.

Immunofluorescence staining of pre- and post-implantation embryos

For pre-implantation staining, oocytes and one-cell embryos were collected from the oviducts of hormone-primed B6D2F1 mice and cultured in KSOM (Millipore) until reaching the desired stage. Immunostaining was performed as described previously with modifications[36]. Briefly, the zona pellucida was removed using acid Tyrode's solution (Sigma). Embryos were transferred to watch-glass dish (Genenet), fixed for 15 minutes in 4% PFA in phosphate buffer (PB), rinsed three times in PBS containing 3 mg/ml PVP, permeabilized in PBS/PVP with 1% triton X-100 for 30 minutes, and blocked in blocking solution (2% normal donkey serum, 0.05% BSA, 0.01% Tween in PBS) for 1 hour. Embryos were then incubated overnight at 4°C with primary antibodies diluted in blocking solution, washed three times in blocking solution for 15 minutes each, incubated with secondary antibodies for 1 hour at room temperature, counterstained with DAPI for

15 minutes, washed twice in PBS, and mounted on 96-well glass bottom plates for confocal imaging.

Post-implantation embryos were fixed and embedded in paraffin as described previously[69] with modification. Embryos in the maternal decidua were fixed in 4% PFA/PB overnight at 4°C, washed three times in PBS for 30 minutes each, dehydrated and embedded in paraffin using standard procedure. Embryonic paraffin sections (5-7 µm thick) were rehydrated, treated with antigen retrieval, rinsed in PBS, permeabilized in 0.1% Triton/PBS for 10 minutes, rinsed in PBT (0.02% Tween/PBS), and blocked in blocking solution (5% normal donkey serum, 0.05% BSA, in PBT) for 1 hour. Slides were then incubated with the appropriate primary and secondary antibodies diluted in blocking solution as described above, and processed as described previously[18]. 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 expression vector using jetPEI (Polyplus transfection) and were lysed 48 hours 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 from Addgene). 30 µl of anti-FlagM2 Magnetic beads (Sigma) were incubated for 6 hours 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 hours. Both the anti-flag and anti-IgG fractions 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 X3Flag peptide buffer (Sigma) for the anti-flagM2 beads, or by boiling with sample buffer, and analyzed by SDS–polyacrylamide gel electrophoresis and immunoblotting. 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 (ESCs or iPSCs) were injected into BDF2 diploid blastocysts. A flat-tip pipette was used for microinjection into blastocysts placed in M16 medium under mineral oil. A controlled number of 10-12 cells were injected into the blastocyst cavity. Then, blastocysts were returned to KSOM media (Invitrogen) and placed at $37^{o}C$ until transfer to recipient females. Ten to 15 injected blastocysts were transferred to each uterine horn of 2.5 days post coitum pseudo-pregnant females. 4n tetraploid complementation assay was performed by fusing BDF2 embryos at 2-cell stage. Embryos were allowed to develop until the blastocyst stage at day 3.5 and were then utilized for PSC microinjection. Embryos were either analyzed at different time points during development, or allowed to develop to full term. Germ line transmission was determined by mating chimeric animals with C57B/6 females and continuous checking for agouti colored pups. For teratoma formation and analysis, naïve hESCs and hiPSCs were harvested by trypsinization before injection. Cells were injected subcutaneously into NSG mice (Jackson laboratories). Animals were sacrificed before tumor size exceeded 1.5 cm in diameter. All animal studies were conducted according to the guidelines of and following approval by the Weizmann Institute IACUC (approval #00960212-3).

Southern blot analysis

Genomic DNA was extracted from each sub-clone targeted colony. 10-15 µg of genomic DNA was digested with restriction enzyme for 5 hours and separated by gel electrophoresis. The DNA was transferred to a nitrocellulose membrane, which was then hybridized with a radioactive-labeled probe and developed using ECL (Thermo Scientific).

Teratoma assay

ESC or iPSCs from indicated cell lines were expanded and injected subcutaneously to the flanks of immune-deficient NSG mice. After 4-6 weeks, all injected mice were sacrificed and the tumor mass extracted and fixed in 4% paraformaldehyde overnight. Slides were prepared from the paraffin-embedded fixed

tissue, which were next Hematoxylin & Eosin stained and inspected for representation of all three germ layers.

Alkaline phosphatase staining

Cells were grown on 6-, 12- or 24-well tissue culture plates and were washed three times with PBS and fixed with 4% paraformaldehyde for 3 minutes at room temperature. Alkaline phosphatase staining was performed according to the manufacturer's protocol (Millipore SCR004).

Western blot analysis

Following cell harvesting, whole-cell protein was extracted by a lysis buffer containing 150 mM NaCl, 150 mM Tris-Hcl (pH 7.4), 0.5% NP40, 1.5 mM MgCl2, and 10% glycerol. Protein concentration was determined by BCA Kit (Thermo). Blots were incubated with the following primary antibodies (diluted in 5% BSA in PBST):

Mbd3	1:1000	Bethyl A302-529A
Mbd2	1:1000	Bethyl A301-633A
P66α (H-162)	1:1000	Santa Cruz Sc-134712
P66β	1:1000	Bethyl A301-281A
Hdac2 (c-8)	1:1000	Santa Cruz sc-9959
Mta2 (c-20)	1:500	Santa Cruz sc-9447
Chd4	1:1000	Abcam Ab70469
Oct4 (H-134)	1:1000	Santa Cruz sc-9081
Klf4 (H-180)	1:1000	Santa Cruz Sc20691
Nanog	1:1000	Behtyl, A300-398A
Flag-M2	1:1000	Sigma F3165
Ha.11 16B12	1:1000	Covance MMS-101R
Gapdh	1:5000	Epitomics 2251-1
Hsp90β	1:1000	Epitomics 1492-1

Secondary antibodies used: Peroxide-conjugated AffiniPure goat anti-rabbit (1:10,000, 111-035-003; Jackson ImmunoResearch). Blots were developed using SuperSignal West Pico Chemiluminescent substrate (Thermo, #34080).

Immunostaining

Before staining, cell lines were cultured on glass cover slips (13 mm 1.5H; Marienfeld, 0117530), washed three times with PBS and fixed with 4% paraformaldehyde for 10 minutes at room temperature. Cells were then permeabilized and blocked in 0.1% Triton, 0.1% Tween, and 5% FBS in PBS for 15 min at room temperature. Primary antibodies were incubated for two hours at room temperature and then washed with 0.1% Tween and 1% FBS in PBS three times. Next, cells were incubated with secondary antibody for one hour at room temperature, washed and counterstained with DAPI, mounted with Shandon Immu-Mount (Thermo Scientific) and imaged. All secondary antibodies were diluted 1:200.

For staining of different cell lines, all cells were fixed and stained in the tissue culture wells or on the cover slips. The following antibodies were used: rabbit polyclonal Sox2 antibody (Millipore AB5603, 1:200), mouse monoclonal Oct4 antibody clone C10 (Santa Cruz SC5279, 1:200), rabbit polyclonal Nanog antibody (Bethyl A300-397A, 1:200), mouse monoclonal IgM SSEA1antibody (Hybridoma MC-480 clone, 1:20).

RT-PCR analysis

Total RNA was isolated using the RNeasy Kit (Qiagen). To remove potential contamination of genomic DNA, 3 μg of total RNA was treated with DNase I 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 re-suspended in 100 μl of water. Quantitative PCR analysis was performed in triplicates using 1/50 of the reverse transcription reaction on Viia7 platform (Applied Biosystems). For single-cell RT-PCR analysis, cells from different samples were sorted and Ambion® Single Cell-to-CT™ Kit was used for sample processing according to the manufacturer's instructions. TaqMan probe-based chemistry and TaqMan Real-Time PCR master mix were used on Viia7 platform for gene expression detection. The following TaqMan probes (Invitrogen) were used: Sall4 Mm00453037_s1, Esrrb Mm00442411_m1, Utf1 Mm00447703_g1, Lin28a Mm00524077_m1, Sox2 (endogenous) Mm03053810_s1, Nanog Mm02384862_g1, Gapdh Mm99999915_g1. C_T cutoff of 39 cycles was used as threshold for defining transcript detection.

Microscopy image acquisition and analysis

Secondary OKSM inducible Mbd3[+/+] and Mbd3[flox/-] MEFs carrying Oct4-GFP pluripotency reporter and constitutively expressed nuclear mCherry marker were plated in 12-well plates at low densities (150 cells per well) and imaged using AxioObserver Z1 (Zeiss) at 5% O_2, 5% CO_2, 37°C controlled conditions. Plates were taken out at day 3-4 for medium replacement (without passaging/splitting) and put back for the automated live imaging stage. Full-well mosaic images were taken every 12 hours for 6 days at 5x magnification, including phase contrast and two fluorescent wavelength images. An automated segmentation protocol was developed in-house and implemented in Matlab to analyze time-lapse measurements of full-well mosaics with fluorescent mCherry and Oct4-GFP markers. For fast segmentation of an unknown number of colonies in 10^8 pixels mosaic image, the following protocol was applied:

Adaptive Detection: For each time point and fluorescent wavelength, the plate margins were erased using a circular filter. Detection threshold was defined using median with offset (10% of the dynamic range), and a binary image of detected pixels was created.

Complexity Reduction: For this task, a morphological filter was applied to isolate mCherry+ colonies using median sliding filter $(60_{um}*60_{um})^{74}$. This filter retains only dense colonies, erasing noise and single isolated cells, which substantially reduces the extent of clustering.

Colony Segmentation: The segmentation was done using moving average filter (low-pass filter, $60_{um}*60_{um})^{74}$ to merge adjacent colony fragments into large colonies and connected components were clustered, labeling objects using 8-connected neighborhood.

Colony Feature Extraction: The features of each mCherry+ colony, including area, bounding box and centroid, were extracted. By overlaying mCherry colony segmentation on the GFP binary image (detected pixels), we extracted for each colony the GFP[+] indicator (0/1) and the fraction of GFP[+] and mCherry[+] pixels out of all mCherry[+] pixels. This protocol was run over time-lapse mosaics to collect information on colony formation dynamics, GFP expression dynamics and ratios of offspring Oct4-GFP[+] cells. Colony and reprogramming dynamics features were then statistically analyzed using Matlab, including estimation of the cumulative distribution, density function and box-plot graphical interpretation. The distribution of intra-colony Oct4-GFP reactivation was

analyzed for all segmented colonies, representing the distribution of offspring iPS cells within segmented colonies. Movies documenting the process dynamics were produced using customized Matlab program. The above program was validated by artificial input matrix and by collected ES mosaic images. In addition, the robustness of detection threshold and filter sizes was measured with varying parameters (data not shown).

Chromatin immuno-precipitation and sequencing library preparation

Chromatin immuno-precipitation followed by deep sequencing (ChIP-Seq) was measured for the following proteins at indicated time points. Approximately $40*10^6$ cells were cross-linked in formaldehyde (1% final concentration, 10 min at room temperature), and then quenched with glycine for 5 minutes. 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 x g for 10 minutes 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 ug of each antibody was pre-bound by incubating with Protein-G Dynabeads (Invitrogen100-07D) in blocking buffer (PBS supplemented with 0.5% Tween and 0.5% BSA) for 2 hours at room temperature. Washed beads were added to the chromatin lysate, which was then incubated overnight. Samples were washed 5 times with RIPA buffer, twice with RIPA buffer supplemented with 500 mM NaCl, twice with LiCl buffer (10 mM TE, 250 mM LiCl, 0.5% NP-40, 0.5% DOC), once with TE (10 Mm 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 hours and then treated sequentially with RNaseA (Roche, 11119915001) for 30 minutes and Proteinase K (NEB, P8102S) for two hours. 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[75]. 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 and sequenced on Illumina Nextseq500 sequencer according to standard Illumina protocols. The

following antibodies were used for ChIP experiments: Control IgG (ChIP grade, ab46540, Abcam), Anti-Histone H3 tri methyl K4 (ChIP grade, ab8580, Abcam), Anti-Histone H3 mono methyl K4 (ChIP grade, ab8895, Abcam), Anti-Histone H3 acetyl K27 (ChIP grade, ab4729, Abcam), anti-Histone H3 tri methyl K27 (ChIP grade, 07-449, Millipore), anti-Oct4 (sc5729 (C-10), Santa Cruz), anti-Sox2 (AB5603, Millipore), anti-Klf4 (AF3158, Millipore), anti-c-Myc(sc764, SantaCruz), anti-Pol2(sc899(N20), Santa Cruz), anti-Chd4 (ChIP Grade, ab70469, Abcam). For Mbd3 ChIP, 1:1 antibody mix was used: anti-Mbd3 (Bethyl laboratories A302-528A) and anti-Mbd3 (ab16057, Abcam).

Alignment and peak detection

We used Bowtie software[76] (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. Enriched intervals of H3K4me3, H3K27me3, H3K27ac, Mbd3 and Oct4 were identified using MACS version 1.4.1[77]. We used sequencing of whole-cell extract as a control to define a background model. Duplicate reads aligned to the exact same location were excluded by MACS default configuration. Enriched intervals were mapped to genes if they overlapped a single Kb symmetric interval around their transcription start sites (taken from RefSeq known gene table in UCSC genome browser). ChIP-seq data on WT samples were highly reproducible in comparison to previous publications[21,22] (data not shown).

Motif detection

Motifs that were enriched in Mbd3 binding regions were detected using SeqPos tool in Cistrome package (http://cistrome.org/ap/). Mbd3 peaks in MEF, MEF+Dox and iPSC were run against Cistrome curated motif database with a *P*-value cutoff of 0.001.

Histone mark profiles

Histone profiles were calculated using in-house script, which generates a matrix of read densities in given genomic intervals. The profiles of all 29,952 Entrez genes (mm9, taken from UCSC known gene tables) were calculated between 1kb upstream to TSS and TES. These read densities were then converted to z-score by normalizing each position with

the mean and standard deviation of the sample noise: ($\hat{X}_j = \dfrac{X_j - \mu_{Noise}}{\sigma_{Noise}}$). Noise parameters

were estimated for each sample from $6*10^7$ random bp 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 max value within that interval.

For distribution analysis and for the correlation and clustering of histone marks, each gene and each histone mark was represented with the maximal z-score measured in the profile of that gene. Clustering of histone marks was carried out on concatenated vectors that included all marks for every gene in tandem.

Annotation enrichment analysis

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[13,78-83]. Enrichment *P*-values were calculated using Fisher's exact test[76] and corrected for multiple hypotheses using false discovery rate (FDR) threshold of 0.0001%.

Poly-A RNA sequencing

RNA was extracted from Trizol pellets and utilized for RNA-seq by TruSeq RNA Sample Preparation Kit v2 (Illumina) according to manufacturer's instruction. To avoid DNA contamination, all samples were treated with DNase. RNA integrity was evaluated on Tapestation, requiring a minimal RNA integrity number (RIN) of 8.5. Libraries were prepared according to Illumina's instructions. Sequencing was carried out on Illumina Nextseq 500 according to the manufacturer's instructions.

RNA-seq analysis

Poly-A RNA sequencing was measured in mRNA extracted from the indicated samples. The paired-end reads were aligned to mouse genome version mm10 with TopHat2 aligner (v2.0.8b) using default input parameters. Transcriptional profiles were visualized using IGV v2.3. FPKM levels (fragments per kilobase per million reads) were estimated using Cufflinks package with "-p 3 –u" parameters and GTF file downloaded from Ensemble (version GRCm38.74).

Total RNA was isolated from indicated cell lines. The concentration of RNA was quantified and subjected to quality control on Agilent Bioanalyzer/ tapestation. 250 ng of RNA was simultaneously processed from each sample. cDNA was fragmented, labeled,

and hybridized to Affymetrix Mouse Gene 1.0 ST GeneChip (Affymetrix, Santa Clara, CA), which contain 35,557 probes. Transcripts levels were processed from image files using RMA method[84] , 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 no. GSE45352 (attached file includes access to confidential datasets on GEO website).

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$^{f/-}$, Mbd3$^{-/-}$) and ES samples (ES V6.5, Mbd3$^{-/-}$ ES ,Mbd3$^{f/-}$ IPS and Mbd3$^{+/+}$ iPS) were characterize using two-sample *t*-test and corrected for multiple hypotheses using false discovery rate (FDR)[77]. Differentially expressed gene signatures included genes below FDR threshold of 5% or fold change above 4, resulting in 1,323 genes. Sample clustering was performed either by hierarchical clustering using Spearman's rank correlation coefficient as a distance metric and average linkage, or by principle component analysis (PCA) to detect the components with the largest variation. Single gene progression (**Supplementary Fig. 17a-c**) was quantified using the following transformation:

$$\hat{X}_j(t) = \max\left(\frac{X_j(t) - X_j(MEF\_Mbd3^{+/+})}{\overline{X}_j(IPS) - \overline{X}_j(MEF)}, 0 \right)$$

where Xj(t) denotes gene j expression value at time t (e.g. Xj(4d) or Xj(MEF)) and $\overline{X}_j(IPS), \overline{X}_j(MEF)$ denotes the averaged expression value for IPS and MEF samples, respectively. The following transformation represent a distance from MEF expression values (set to 0) towards iPS values (set to 1); genes whose expression level changes towards their iPS value show $\hat{X}_j(t) > 0$. Distribution of gene expression fold change relative to MEF was calculated by paired samples *t*-test.

Preparation and analysis of RRBS and WGBS libraries

RRBS libraries were generated as described previously with slight modifications[30]. 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). For WGBS, we used illumina TruSeq DNA Methylation kit (EGMK91324, illumine). Libraries were prepared according to Illumina's instructions. Plasmids or DNA were treated with sodium bisulfite using the EZ DNA Methylation-Gold kit (Zymo) and the product was PCR-amplified using GoTaq Hot Start DNA polymerase (Promega). PCR products were A-taild 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 Nextseq 500 system. 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 five or more distinct sequencing reads across all libraries. The CpG content "experienced" by each CpG site was defined as the number of CpG dinucleotides found within a 500 bp window surrounding the site divided by the window size.

Transposase-accessible chromatin with high-throughput sequencing (ATAC-seq)

For ATAC sequencing, 50,000 cells were centrifuged at 500 g for 3 minutes, wash with 50 μL of cold PBS and centrifuged again at 500 g for 3 minutes. Cells were lysed using cold lysis buffer (10 mM Tris-HCl (pH 7.4), 10 mM NaCl, 3 mM $MgCl_2$ and 0.1% IGEPAL CA-630). Immediately after lysis, nuclei were spun at 500 g for 10 minutes using a refrigerated centrifuge. Next, the pellet was re-suspended in the transposase reaction mix (25 μL 2× TD buffer, 2.5 μL transposase (Illumina) and 22.5 μL nuclease-free water). The transposition reaction was carried out for 30 minutes at 37 °C and immediately put on ice. Immediately afterwards, the sample was purified using a MinElute kit (Qiagen). Then, library fragments were amplified using custom Nextera PCR primers for 12 cycles and libraries were purified using MinElute Kit.

ATAC-seq data analysis

Reads were aligned to mm10 mouse genome using Bowtie2 with the parameter X2000, allowing fragments up to 2 kb to align. Duplicated aligned reads were removed using Picard MarkDuplicates tool with the command REMOVE_DUPLICATES=true. To identify chromatin accessibility signal, we considered only short reads ($\leq$ 100bp) that corresponded to nucleosome-free region[21]. To detect and separate accessible loci in each sample, we used MACS version 1.4.2-1 with --call-subpeaks flag (PeakSplitter version 1.0). Next, summits in previously annotated spurious regions were filtered out using a custom blacklist targeted at mitochondrial homologues. For the blacklist, we generated 10,000,000 synthetic 34mer reads derived from the mitochondrial genome. After mapping and peak calling of the synthetic reads, we found 28 high-signal peaks for the mm10 genome. In all subsequent analyses, we discarded peaks falling within these regions.

Each peak in each sample was represented by a 300 bp region around the summit center. The peaks from all samples were unified and merged (using bedtools unionbedg and merge commands) to create a list of accessible loci. Accessibility signal and peaks, alongside previously published Oct4 binding signal, were visualized using IGV version 2.3.26 and to quantify the change in accessibility between samples, we estimated read coverage in all accessible loci using bedtools coverageBed command (version 2.16.2). Read coverage was normalized by peak length in Kbp and by million aligned reads per sample to give RPKM values. Further analysis was done using Matlab version R2011b. Differential peaks were defined by 4-fold change difference between MEF and ES samples. Correlation matrix was calculated using Spearman correlation and hierarchical clustering was performed using Spearman correlation as a distance metric and average linkage.

References

1   Gardner, R. L. Mouse chimeras obtained by the injection of cells into the blastocyst. *Nature* **220**, 596-597 (1968).

2   Brinster, R. L. The effect of cells transferred into the mouse blastocyst on subsequent development. *J Exp Med* **140**, 1049-1056 (1974).

3   Dewey, M. J., Martin, D. W., Jr., Martin, G. R. & Mintz, B. Mosaic mice with teratocarcinoma-derived mutant cells deficient in hypoxanthine phosphoribosyltransferase. *Proc Natl Acad Sci U S A* **74**, 5564-5568 (1977).

4   Yamanaka, S. A fresh look at iPS cells. *Cell* **137**, 13-17, doi:S0092-8674(09)00333-X [pii]
10.1016/j.cell.2009.03.034 (2009).

5   Gurdon, J. B. & Melton, D. A. Nuclear reprogramming in cells. *Science* **322**, 1811-1815, doi:322/5909/1811 [pii]
10.1126/science.1160810 (2008).

6   Yamanaka, S. Pluripotency and nuclear reprogramming. *Philos Trans R Soc Lond B Biol Sci* **363**, 2079-2087, doi:81R627X1V0591680 [pii]
10.1098/rstb.2008.2261 (2008).

7   Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861-872, doi:S0092-8674(07)01471-7 [pii]
10.1016/j.cell.2007.11.019 (2007).

8   Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663-676, doi:S0092-8674(06)00976-7 [pii]
10.1016/j.cell.2006.07.024 (2006).

9   Wernig, M. *et al.* In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318-324, doi:nature05944 [pii]
10.1038/nature05944 (2007).

10   Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250-264, doi:S0092-8674(08)00447-9 [pii]
10.1016/j.cell.2008.03.028 (2008).

11   Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462**, 595-601, doi:nature08592 [pii]
10.1038/nature08592 (2009).

12   Hanna, J. H., Saha, K. & Jaenisch, R. Pluripotency and cellular reprogramming: facts, hypotheses, unresolved issues. *Cell* **143**, 508-525, doi:S0092-8674(10)01144-X [pii]
10.1016/j.cell.2010.10.008 (2010).

13   Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**, 364-377, doi:S0092-8674(09)00007-5 [pii]
10.1016/j.cell.2009.01.001 (2009).

14   Jaenisch, R. & Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* **132**, 567-582, doi:S0092-8674(08)00115-3 [pii]
10.1016/j.cell.2008.01.015 (2008).

15   Schotta, G., Ebert, A. & Reuter, G. SU(VAR)3-9 is a conserved key function in heterochromatic gene silencing. *Genetica* **117**, 149-158 (2003).

16   Jones, B. *et al.* The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. *PLoS Genet* **4**, e1000190, doi:10.1371/journal.pgen.1000190 (2008).

17   Shi, Y., Seto, E., Chang, L. S. & Shenk, T. Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* **67**, 377-388, doi:0092-8674(91)90189-6 [pii] (1991).

18   Mansour, A. A. *et al.* The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming. *Nature* **488**, 409-413, doi:10.1038/nature11272 nature11272 [pii] (2012).

19   Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49-55, doi:10.1038/nature07056 nature07056 [pii] (2008).

20   Bhutani, N. *et al.* Reprogramming towards pluripotency requires AID-dependent DNA demethylation. *Nature* **463**, 1042-1047, doi:10.1038/nature08752 nature08752 [pii] (2010).

21   Deng, W. AID in reprogramming: quick and efficient: identification of a key enzyme called AID, and its activity in DNA demethylation, may help to overcome a pivotal epigenetic barrier in reprogramming somatic cells toward pluripotency. *Bioessays* **32**, 385-387, doi:10.1002/bies.201000014 (2010).

22   Luo, M. *et al.* NuRD blocks reprogramming of mouse somatic cells into pluripotent stem cells. *Stem Cells* **31**, 1278-1286, doi:10.1002/stem.1374 (2013).

23   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, doi:26/3/843 [pii] 10.1128/MCB.26.3.843-851.2006 (2006).

24   Ramirez, J. & Hagman, J. The Mi-2/NuRD complex: a critical epigenetic regulator of hematopoietic development, differentiation and cancer. *Epigenetics* **4**, 532-536, doi:10108 [pii] (2009).

25   Dege, C. & Hagman, J. Mi-2/NuRD chromatin remodeling complexes regulate B and T-lymphocyte development and function. *Immunol Rev* **261**, 126-140, doi:10.1111/imr.12209 (2014).

26   Gunther, K. *et al.* Differential roles for MBD2 and MBD3 at methylated CpG islands, active promoters and binding to exon sequences. *Nucleic Acids Res* **41**, 3010-3021, doi:10.1093/nar/gkt035 gkt035 [pii] (2013).

27   Shimono, Y. *et al.* Mi-2 beta associates with BRG1 and RET finger protein at the distinct regions with transcriptional activating and repressing abilities. *J Biol Chem* **278**, 51638-51645, doi:10.1074/jbc.M309198200 M309198200 [pii] (2003).

28   Hendrich, B. & Bickmore, W. Human diseases with underlying defects in chromatin structure and modification. *Hum Mol Genet* **10**, 2233-2242 (2001).

29   Kaji, K. *et al.* The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. *Nat Cell Biol* **8**, 285-292, doi:ncb1372 [pii] 10.1038/ncb1372 (2006).

30   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, doi:dev.02802 [pii] 10.1242/dev.02802 (2007).

31   Latos, P. A. *et al.* NuRD-dependent DNA methylation prevents ES cells from accessing a trophectoderm fate. *Biol Open* **1**, 341-352, doi:10.1242/bio.2012513 BIO2012513 [pii] (2012).

32   Reynolds, N. *et al.* NuRD Suppresses Pluripotency Gene Expression to Promote Transcriptional Heterogeneity and Lineage Commitment. *Cell Stem Cell* **10**, 583-594, doi:S1934-5909(12)00113-0 [pii]

10.1016/j.stem.2012.02.020 (2012).

33   Nan, X., Campoy, F. J. & Bird, A. MeCP2 is a transcriptional repressor with abundant binding sites in genomic chromatin. *Cell* **88**, 471-481, doi:S0092-8674(00)81887-5 [pii] (1997).

34   Hendrich, B. & Bird, A. Identification and characterization of a family of mammalian methyl-CpG binding proteins. *Mol Cell Biol* **18**, 6538-6547 (1998).

35   Kalkan, T. & Smith, A. Mapping the route from naive pluripotency to lineage specification. *Philos Trans R Soc Lond B Biol Sci* **369**, doi:10.1098/rstb.2013.0540

20130540 [pii]

rstb.2013.0540 [pii] (2014).

36   Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nat Genet* **45**, 34-42, doi:10.1038/ng.2491

ng.2491 [pii] (2013).

37   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, doi:10.1101/gad.2039011

25/10/1035 [pii] (2011).

38   Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994-1004, doi:10.1016/j.cell.2012.09.045

S0092-8674(12)01298-6 [pii] (2012).

39   Leitch, H. G. *et al.* Naive pluripotency is associated with global DNA hypomethylation. *Nat Struct Mol Biol* **20**, 311-316, doi:10.1038/nsmb.2510

nsmb.2510 [pii] (2013).

40   Marks, H. *et al.* The transcriptional and epigenomic foundations of ground state pluripotency. *Cell* **149**, 590-604, doi:10.1016/j.cell.2012.03.026

S0092-8674(12)00409-6 [pii] (2012).

41   Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339-344, doi:10.1038/nature10960

nature10960 [pii] (2012).

42   Clark, A. T. DNA methylation remodeling in vitro and in vivo. *Curr Opin Genet Dev* **34**, 82-87, doi:10.1016/j.gde.2015.09.002

S0959-437X(15)00095-7 [pii] (2015).

43   Geula, S. *et al.* Stem cells. m6A mRNA methylation facilitates resolution of naive pluripotency toward differentiation. *Science* **347**, 1002-1006, doi:10.1126/science.1261417

science.1261417 [pii] (2015).

44   Brons, I. G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191-195, doi:nature05950 [pii]

10.1038/nature05950 (2007).

45   Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196-199, doi:nature05972 [pii]

10.1038/nature05972 (2007).

46   Bao, S. *et al.* Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. *Nature* **461**, 1292-1295, doi:10.1038/nature08534

nature08534 [pii] (2009).

47   Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138**, 722-737, doi:10.1016/j.cell.2009.07.039

S0092-8674(09)00969-6 [pii] (2009).

48   Ying, Q. L. *et al.* The ground state of embryonic stem cell self-renewal. *Nature* **453**, 519-523, doi:10.1038/nature06968

nature06968 [pii] (2008).

49   Wernig, M. *et al.* A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. *Nat Biotechnol* **26**, 916-924, doi:10.1038/nbt1483

nbt1483 [pii] (2008).

50   Woltjen, K. *et al.* piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. *Nature* **458**, 766-770, doi:10.1038/nature07863

nature07863 [pii] (2009).

51   Sommer, C. A. *et al.* Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* **27**, 543-549, doi:10.1634/stemcells.2008-1075

stemcells.2008-1075 [pii] (2009).

52   Kehler, J. *et al.* Oct4 is required for primordial germ cell survival. *EMBO Rep* **5**, 1078-1083, doi:7400279 [pii]

10.1038/sj.embor.7400279 (2004).

53   Kim, J. B. *et al.* Direct reprogramming of human neural stem cells by OCT4. *Nature* **461**, 649-643, doi:10.1038/nature08436

nature08436 [pii] (2009).

54   Fidalgo, M. *et al.* Zfp281 mediates Nanog autorepression through recruitment of the NuRD complex and inhibits somatic cell reprogramming. *Proc Natl Acad Sci U S A* **109**, 16202-16207, doi:10.1073/pnas.1208533109

1208533109 [pii] (2012).

55   Saito, M. & Ishikawa, F. The mCpG-binding domain of human MBD3 does not bind to mCpG but interacts with NuRD/Mi2 components HDAC1 and MTA2. *J Biol Chem* **277**, 35434-35439, doi:10.1074/jbc.M203455200

M203455200 [pii] (2002).

56   Bowen, N. J., Fujita, N., Kajita, M. & Wade, P. A. Mi-2/NuRD: multiple complexes for many purposes. *Biochim Biophys Acta* **1677**, 52-57, doi:10.1016/j.bbaexp.2003.10.010

S0167478103002768 [pii] (2004).

57   Denslow, S. A. & Wade, P. A. The human Mi-2/NuRD complex and gene regulation. *Oncogene* **26**, 5433-5438, doi:1210611 [pii]

10.1038/sj.onc.1210611 (2007).

58   Gnanapragasam, M. N. *et al.* p66Alpha-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. *Proc Natl Acad Sci U S A* **108**, 7487-7492, doi:10.1073/pnas.1015341108

1015341108 [pii] (2011).

59   Walavalkar, N. M., Gordon, N. & Williams, D. C., Jr. Unique features of the anti-parallel, heterodimeric coiled-coil interaction between methyl-cytosine binding domain 2 (MBD2) homologues and GATA zinc finger domain containing 2A (GATAD2A/p66alpha). *J Biol Chem* **288**, 3419-3427, doi:10.1074/jbc.M112.431346

M112.431346 [pii] (2013).

60   Buganim, Y. *et al.* Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell* **150**, 1209-1222, doi:10.1016/j.cell.2012.08.023

S0092-8674(12)01021-5 [pii] (2012).

61   Di Stefano, B. *et al.* C/EBPalpha creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. *Nat Cell Biol* **18**, 371-381, doi:10.1038/ncb3326

ncb3326 [pii] (2016).

62   Hussein, S. M. *et al.* Genome-wide characterization of the routes to pluripotency. *Nature* **516**, 198-206, doi:10.1038/nature14046

nature14046 [pii] (2014).

63   Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151**, 1617-1632, doi:10.1016/j.cell.2012.11.039

S0092-8674(12)01424-9 [pii] (2012).

64   Rais, Y. *et al.* Deterministic direct reprogramming of somatic cells to pluripotency. *Nature* **502**, 65-70, doi:10.1038/nature12587

nature12587 [pii] (2013).

65   Tsubouchi, T. *et al.* DNA synthesis is required for reprogramming mediated by stem cell fusion. *Cell* **152**, 873-883, doi:10.1016/j.cell.2013.01.012

S0092-8674(13)00018-4 [pii] (2013).

66   Wang, B., Pfeiffer, M. J., Schwarzer, C., Arauzo-Bravo, M. J. & Boiani, M. DNA replication is an integral part of the mouse oocyte's reprogramming machinery. *PLoS One* **9**, e97199, doi:10.1371/journal.pone.0097199

PONE-D-14-11607 [pii] (2014).

67   Larsen, D. H. *et al.* The chromatin-remodeling factor CHD4 coordinates signaling and repair after DNA damage. *J Cell Biol* **190**, 731-740, doi:10.1083/jcb.200912135

jcb.200912135 [pii] (2010).

68   Sakai, H. *et al.* MBD3 and HDAC1, two components of the NuRD complex, are localized at Aurora-A-positive centrosomes in M phase. *J Biol Chem* **277**, 48714-48723, doi:10.1074/jbc.M208461200

M208461200 [pii] (2002).

69   Basta, J. & Rauchman, M. The nucleosome remodeling and deacetylase complex in development and disease. *Transl Res* **165**, 36-47, doi:10.1016/j.trsl.2014.05.003

S1931-5244(14)00166-2 [pii] (2015).

70   Spruijt, C. G. *et al.* CDK2AP1/DOC-1 is a bona fide subunit of the Mi-2/NuRD complex. *Mol Biosyst* **6**, 1700-1706, doi:10.1039/c004108d (2010).

71   Smits, A. H., Jansen, P. W., Poser, I., Hyman, A. A. & Vermeulen, M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. *Nucleic Acids Res* **41**, e28, doi:10.1093/nar/gks941

gks941 [pii] (2013).

72   Wang, Y. *et al.* LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell* **138**, 660-672, doi:10.1016/j.cell.2009.05.050

S0092-8674(09)00710-7 [pii] (2009).

73   Sihem Cheloufi, U. E., Barbara Hopfgartner, Youngsook L. Jung, Jernej Murn,Maria Ninova,Maria Hubmann, Aimee I. Badeaux, Cheen Euong Ang, Danielle Tenen, Daniel J. Wesche, Nadezhda Abazova, Max Hogue, Nilgun Tasdemir, Justin Brumbaugh, Philipp Rathert, Julian Jude, Francesco Ferrari, Andres Blanco, Michaela Fellner, Daniel Wenzel, Marietta Zinner, Simon E. Vidal, Oliver Bell, Matthias Stadtfeld, Howard Y. Chang, Genevieve Almouzni, Scott W. Lowe, John Rinn, Marius Wernig,Alexei Aravin, Yang Shi, Peter J. Park, Josef M. Penninger,

Johannes Zuber & Konrad Hochedlinger. The histone chaperone CAF-1 safeguards somatic cell identity. *Nature* **528**, 218–224 (2015).

74   Hill, P. W., Amouroux, R. & Hajkova, P. DNA demethylation, Tet proteins and 5-hydroxymethylcytosine in epigenetic reprogramming: an emerging complex story. *Genomics* **104**, 324-333, doi:10.1016/j.ygeno.2014.08.012

S0888-7543(14)00154-2 [pii] (2014).

75   Soufi, A. *et al.* Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555-568, doi:10.1016/j.cell.2015.03.017

S0092-8674(15)00304-9 [pii] (2015).

76   Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:gb-2009-10-3-r25 [pii]

10.1186/gb-2009-10-3-r25 (2009).

77   Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137, doi:gb-2008-9-9-r137 [pii]

10.1186/gb-2008-9-9-r137 (2008).

78   Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947-956 (2005).

79   Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553-560, doi:nature06008 [pii]

10.1038/nature06008 (2007).

80   Loh, Y. H. *et al.* The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. *Nat Genet* **38**, 431-440, doi:ng1760 [pii]

10.1038/ng1760 (2006).

81   Kim, J., Chu, J., Shen, X., Wang, J. & Orkin, S. H. An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* **132**, 1049-1061, doi:S0092-8674(08)00328-0 [pii]

10.1016/j.cell.2008.02.039 (2008).

82   Bruce, A. W. *et al.* Genome-wide analysis of repressor element 1 silencing transcription factor/neuron-restrictive silencing factor (REST/NRSF) target genes. *Proc Natl Acad Sci U S A* **101**, 10458-10463, doi:10.1073/pnas.0401827101

0401827101 [pii] (2004).

83   Zhao, X. D. *et al.* Whole-genome mapping of histone H3 Lys4 and 27 trimethylations reveals distinct genomic compartments in human embryonic stem cells. *Cell Stem Cell* **1**, 286-298, doi:S1934-5909(07)00123-3 [pii]

10.1016/j.stem.2007.08.004 (2007).

84   Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249-264, doi:10.1093/biostatistics/4.2.249

4/2/249 [pii] (2003).

Student declaration

I hereby declare that the thesis presented summarizes my independent research work under the supervision of Dr. Jacob H. Hanna at the Department of Molecular Genetics, the Weismann Institute of Science. Mirie Zerbib and I conducted microinjections in mice blastocysts. Some of the NGS samples were prepped and sequenced by INCPM. Bioinformatics analysis of NGS data was performed by Asaf Zviran and Elad Chomski under the supervision of Dr. Noa Novershtern. Dr. Rada Massarwa assisted with microscopy live imaging. Dr. Noam Stern-Ginossar helped with the Ribo-seq preparation and analysis. Biochemistry assays were conducted with the help of Shai Geula and Nofar Mor.

The work "Characterizing cell fate transition regulation during deterministic reprogramming" was done in collaboration with Asaf Zviran.

The work "Mechanism of the Gatad2a-Chd4-Mbd3/NuRD axis facilitates deterministic direct induction of murine naïve pluripotency" was done in collaboration with Nofar Mor.

List of publication and conferences

A. Zviran*, **Y. Rais*,** N. Mor*, H. Gingold, J.D. Buenrostro, E.Chomsky L. Weinberger, Y. S. Manor, O.Gafni, S.Viukov, V. Krupalnik, M. Zerbib, D.Jaitin, D. Larastiaso, I. Ultisky, G. Rechavi, I. Maza, S. Hanna, N. Stern-Ginossar, I. Amit, W. J. Greenleaf, Y. Pilpel, J. H. Hanna and N. Novershtern, "**Characterizing Cell Fate Transition Regulation during Deterministic Reprogramming** " **(in preperation)**

* contributed equally to this work.

N.Mor*, **Y. Rais***, S. Geula1, S. Viukov, D. Levin, A. Zviran, J. D. Buenrostro, E. Chomsky, Y. S. Manor, O. Gafni1, V. Krupalnik, M Zerbib, A. Tanay, W. J. Greenleaf, R. Massarwa, Y. Levin, N. Novershtern1 and J. H. Hanna," **A Gatad2a-Chd4-Mbd3/NuRD Axis Facilitates Deterministic Direct Induction of Murine Naïve Pluripotency" (in preperation)**

* contributed equally to this work

I. Maza, I. Caspi, A. Zviran, E. Chomsky, **Y. Rais**, S. Viukov, S. Geula, J. D. Buenrostro, L. Weinberger, V. Krupalnik, S. Hanna, M. Zerbib, J. R. Dutton, W. J. Greenleaf, R. Massarwa, N. Novershtern, and J. H. Hanna, **"Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors.,"** *Nature Biotechnology*, vol. 33, no. 7, pp. 769–774, 2015.

S. Geula, S. Moshitch-Moshkovitz, D. Dominissini, A. A. Mansour, N. Kol, M. Salmon-Divon, V. Hershkovitz, E. Peer, N. Mor, Y. S. Manor, M. S. Ben-Haim, E. Eyal, S. Yunger, Y. Pinto, D. A. Jaitin, S. Viukov, **Y. Rais**, V. Krupalnik, E. Chomsky, M. Zerbib, I. Maza, Y. Rechavi, R. Massarwa, S. Hanna, I. Amit, E. Y. Levanon, N. Amariglio, N. Stern-Ginossar, N. Novershtern, G. Rechavi, and J. H. Hanna, **"m6A mRNA methylation facilitates resolution of naïve pluripotency toward differentiation.,"** *Science*, vol. 347, no. 6225, pp. 1002–6, 2015.

O. Gafni, L. Weinberger, A. A. Mansour, Y. S. Manor, E. Chomsky, D. Ben-Yosef, Y. Kalma, S. Viukov, I. Maza, A. Zviran, **Y. Rais**, Z. Shipony, Z. Mukamel, V. Krupalnik, M. Zerbib, S. Geula, I. Caspi, D. Schneir, T. Shwartz, S. Gilad, D. Amann-Zalcenstein, S. Benjamin, I. Amit, A. Tanay, R. Massarwa, N. Novershtern, and J. H. Hanna, **"Derivation of novel human ground state naïve pluripotent stem cells.,"** *Nature*, vol. 504, pp. 282–6, 2013.

Y. Rais*, A. Zviran*, S. Geula*, O. Gafni, E. Chomsky, S. Viukov, A. A. Mansour, I. Caspi, V. Krupalnik, M. Zerbib, I. Maza, N. Mor, D. Baran, L. Weinberger, D. a. Jaitin, D. Lara-Astiaso, R. Blecher-Gonen, Z. Shipony, Z. Mukamel, T. Hagai, S. Gilad, D. Amann-Zalcenstein, A. Tanay, I. Amit, N. Novershtern, and J. H. Hanna, **"Deterministic direct reprogramming of somatic cells to pluripotency.,"** _**Nature,**_ vol. 502, pp. 65–70, 2013. *(238 citations)*
Selected for Faculty of 1000 Biology

Nature Highlight
Science Highlight
News and Views, Nature 2013

A.A. Mansour, O. Gafni, L. Weinberger, A. Zviran, M. Ayyash, **Y. Rais**, V.Krupalnik, M. Zerbib, D. Amann-Zalcenstein, I. Maza, S. Geula, S. Viukov, L. Holtzman, A. Pribluda, E. Canaani, S. Horn-Saban, I. Amit, N. Novershtern, and J. H. Hanna, **"The H3K27 demethylase Utx regulates somatic and germ cell epigenetic reprogramming,"** _**Nature**_, vol. 488. pp. 409–413, 2012.

ISSCR, June 2015, Stockholm, Sweden. **Oral presentation: Y. Rais,** N. Mor, A Zviran, E. Chomski, N. Novershtern and J. H. Hanna. **"**Deterministic Reprogramming of Somatic Cells to Pluripotency via Regulation of Mbd3/NuRD Activity".

NATURIMMUN, March 2013, Jerusalem, Israel **Oral presentation: Y. Rais,** "Deterministic direct reprogramming of somatic cells to pluripotency".

Acknowledgements

Like any journey that comes to the end, I wish to thank all my friends and colleagues that have supported and helped along this path, during the good times and especially during the less good ones. I learned so much from each one.

I would like to thanks my advisor Dr. Jacob Hanna, for your guidance, for the patience and for the knowledge. I am sure that I will take some of it for my future career.

Last but not least, I would like to thank my family: my children Matan, Shir, Lior and Hila for allowing me to deal with things other than science. To the one and only my wife one of a kind for your love, support, encouragement enabled the conclusion of this step in my life. Ayelet, I will make it up to you, it is a promise.