Published in final edited form as:
Cell Stem Cell. 2018 September 06; 23(3): 412–425.e10. doi:10.1016/j.stem.2018.07.004.

Neutralizing Gatad2a-Chd4-Mbd3/NuRD Complex Facilitates Deterministic Induction of Naïve Pluripotency

Nofar Mor[#1], **Yoach Rais**[#1], **Daoud Sheban**[1,2], **Shani Peles**[1], **Alejandro Aguilera-Castrejon**[1], **Asaf Zviran**[1,3], **Dalia Elinger**[4], **Sergey Viukov**[1], **Shay Geula**[1], **Vladislav Krupalnik**[1], **Mirie Zerbib**[1], **Elad Chomsky**[5], **Lior Lasman**[1], **Tom Shani**[1], **Jonathan Bayerl**[1], **Ohad Gafni**[1], **Suhair Hanna**[6], **Jason D. Buenrostro**[7,8], **Tzachi Hagai**[9], **Hagit Masika**[10], **Gintautas Vainorius**[11], **Yehudit Bergman**[10], **William J. Greenleaf**[12,13], **Miguel A. Esteban**[14], **Ulrich Elling**[11], **Yishai Levin**[4], **Rada Massarwa**[1], **Yifat Merbl**[2], **Noa Novershtern**[1,#,@], **Jacob H. Hanna**[1,#,%,@]

[1]The Department of Molecular Genetics, Weizmann Institute of Science, 234 Herzl, Rehovot 76100, Israel

[2]The Department of Immunology, Weizmann Institute of Science, Rehovot, Israel

[3]The New York Genome Center, NYC, NY, USA

[4]The Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science, Rehovot, Israel

[5]Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel

[6]Department of Pediatrics, Rambam Health Care Campus, Haifa, Israel

[7]Broad Institute of MIT and Harvard, Cambridge, MA, USA

[8]Harvard Society of Fellows, Harvard University, Cambridge, MA, USA

[9]EMBL-European Bioinformatics Institute, Wellcome Genome Campus, Cambridge, UK

[10]Department of Developmental Biology and Cancer Research, Hebrew University, Jerusalem, Israel

[11]Institute of Molecular Biotechnology (IMBA), Vienna, Austria

[12]Department of Applied Physics, Stanford University, Palo Alto, CA, USA

[@]Correspondence: N.N. (noa.novershtern@weizmann.ac.il), Y.R. (yoach.rais@weizmann.ac.il), J.H.H (jacob.hanna@weizmann.ac.il).
[#]Senior author
[%]Lead Contact

Author Contributions

N.M and Y.R. conceived the idea for this project, designed and conducted experiments, and wrote the manuscript. M.Z., R.M., A.A.C and Y.R. conducted micro-injections. A.Z. assisted in RNA-seq analysis. N.M., Y.L. and D. E. conducted MS analysis. W.J.G. and J.B. assisted in ATAC-seq. E.C. assisted in DNA methylation analysis. N.N. supervised all bioinformatics analysis and analyzed data. S.P., S.H., T.S., J.B., O.G, L.L and V.K. assisted in tissue culture, preparing DNA reagents and reprogramming. M.A.E, G.V. and U.E. confirmed reprogramming efficiencies in independent labs and assisted in generating Gatad2a mutants. H.M. and Y.B. conducted replication timing experiments. S.G. assisted in IP experiments. Y.R. and N.M engineered cell lines under S.V. supervision. T.H. conducted predictive domain analysis. D.S. and Y.M. conducted and supervised experiments related to SUMOylation. N.N., Y.R. and J.H.H. supervised executions of experiments, adequate analysis of data and presentation of conclusions made in this paper.

Declaration of Interests

J.H.H is an advisor to Accelta and Biological Industries Ltd. J.H.H, N.N. and Y.R. filed patents and issued commercial licenses related to some of the findings described herein.

Europe PMC Funders Author Manuscripts

[13]Chan Zuckerberg Biohub, San Francisco, CA, USA

[14]Key Laboratory of Regenerative Biology, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou, China

[#] These authors contributed equally to this work.

Abstract

Mbd3, a member of Nucleosome Remodeling and Deacetylase (NuRD) co-repressor complex, was previously identified as an inhibitor for deterministic iPSC reprogramming where up to 100% of donor cells successfully complete the process. NuRD can assume multiple mutually-exclusive conformations, and it remains unclear whether this deterministic phenotype can be attributed to a specific Mbd3/NuRD sub-complex. Moreover, since complete ablation of Mbd3 blocks somatic cell proliferation, we aimed to explore functionally relevant alternative ways to neutralize Mbd3-dependent NuRD activity. We identify Gatad2a, a NuRD-specific subunit, whose complete deletion specifically disrupts Mbd3/NuRD repressive activity on the pluripotency circuitry during iPSC differentiation and reprogramming without disrupting somatic cell proliferation. Ablation of Gatad2a facilitates deterministic murine iPSC reprogramming within 8-days. We validate a distinct molecular axis, Gatad2a-Chd4-Mbd3, within Mbd3/NuRD as being critical for blocking re-establishment of naïve pluripotency, and further highlight signaling-dependent and post-translational modifications of Mbd3/NuRD that influence its interactions and assembly.

Introduction

Somatic cell reprogramming is a technically simple process in which the induction of exogenous factors, classically Oct4, Klf4, Sox2 and Myc (OKSM), induces somatic cells to convert back into ESC-like cells, termed induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka, 2016). Conventional iPSC reprogramming is an inefficient and a-synchronized process, where less than 10% of the donor somatic cells undergo reprogramming over a period of 4 weeks (Hanna et al., 2009). Further, while donor somatic cells reprogram with different efficiencies, it is not possible to a priori definitively predict among individual identical donor somatic cells, which and when they will convert into iPSCs. The latter attributes have supported the conclusion that conventional iPSC formation is a stochastic process, but amenable to acceleration (Hanna et al., 2009).

Multiple studies have devised alternative reprogramming protocols where rapid and up to 100% reprogramming efficiency can be obtained within a relatively short period (Rais et al., 2013; Di Stefano et al., 2013). Such methods are termed “deterministic reprogramming.” Our group found that controlled and partial reduction of Mbd3, a key component of Mbd3/NuRD (Nucleosome Remodeling and Deacetylation) co-repressor complex, in concert with optimized OKSM delivery in naïve pluripotency conditions (2i/LIF, where 2i is applied 60 hours after initiation), can lead to highly efficient and rapid iPSC formation with up to 100% reprogramming efficiency within 8 days (Rais et al., 2013). These high efficiencies and identity of iPSCs generated have been validated using Cytof single cell analysis (Lujan et al., 2015). Grummt and colleagues similarly showed that Mbd3 depletion promotes reprogramming (Luo et al., 2013).

Essential to the promotion of reprogramming upon Mbd3 depletion is an optimal incomplete depletion of Mbd3 activity during a critical early reprogramming window (Rais et al., 2013). Our group and others showed that using $Mbd3^{-/-}$ somatic cells as starting material does not yield a boost in reprogramming (Rais et al., 2013; Dos Santos et al., 2014). In fact, as these cells can no longer proliferate *in vitro* after complete Mbd3 depletion, they cannot reprogram as cell proliferation is indispensable for iPSC formation (Hanna et al., 2009). Further, the decrease in cell proliferation occurs prior and irrespective to OKSM induction, further supporting the notion that the inhibition of reprogramming efficiency trivially results from hampering cell proliferation in the somatic state, rather than establishing an epigenetic blockade for reprogramming. As such, Rais et al. focused mostly on using on $Mbd3^{flox/-}$ cells and compared them to wild-type $Mbd3^{+/+}$ cells. It should be noted that several iPSC boosting strategies rely on incomplete depletion of epigenetic modulators. For instance, Caf1, NCoR/SMRT and Ubc9 partial, but not complete, depletion was critical for boosting iPSC efficiencies by OKSM (Cheloufi et al., 2015; Zhuang et al., 2018).

Recent findings suggest that NuRD co-repressive complex can assume multiple distinct complexes based on differential subunit composition. In addition to Mbd2 and Mbd3, which form two distinct mutually exclusive Mbd2/NuRD and Mbd3/NuRD complexes (Le Guezennec et al., 2006), other canonical subunits include: Chd3 or Chd4, RbAP48 and RbAP46, Hdac1 and Hdac2, which have deacetylation activity; Mta1, Mta2 and Mta3, Gatad2a or Gatad2b (also known as P66α and P66β, respectively) whose function remains to be fully defined (Alqarni et al., 2014). Quantitative mass spectrometry-based proteomics for Mbd3/NuRD complex has indicated that each NuRD unit contains six units of RbAP48 or RbAP46, three units of Mta1/2/3, two units of Gatad2a or Gatad2b, one unit of Chd3 or Chd4, and one unit of Hdac1 or Hdac2 (Smits et al., 2013). The latter is further complicated by the identification of non-canonical components which include Doc-1, Zmynd8, and Lsd1 (Spruijt et al., 2016). Moreover, some of the proteins that contribute to the NuRD complex, maintain additional roles in the cell and take part in other complexes. For example, Mbd3 has been reported to co-localize with aurora kinase at the mitotic spindle during mitosis and thus regulate cell cycle progression through mechanisms that may be NuRD independent (Sakai et al., 2002). These facts complicate assigning the outcome of perturbing components like Mbd3 or Chd4 to NuRD complex exclusively, since the observed functional changes might be stemming from perturbing only a certain sub complex of NuRD with a unique conformation or altering other different complexes that share some components with NuRD (e.g. Chd4 or Hdac1).

In this study, we set out to identify alternative strategies to block Mbd3 dependent NuRD activity while preserving somatic cell proliferation and viability. By dissecting the mechanisms and putative components of Mbd3/NuRD relevant during early stages of reprogramming, we have identified Gatad2a-Chd4-Mbd3 as a functional axis underlying potent inhibition of naïve pluripotency.

Results

Screening for alternative components to define Mbd3/NuRD subcomplexes relevant for iPSC reprogramming

Optimized incomplete Mbd3 depletion during reprogramming has been shown to significantly improve iPSC formation efficiency (Luo et al., 2013; Rais et al., 2013) (Fig. S1AE), but a complete reduction of this protein at the somatic stage results in cell cycle malfunctions (Rais et al., 2013) and subsequently loss of reprogramming ability. In order to better understand the mechanisms of Mbd3 inhibitory effect on reprogramming, we adapted a transient silencing assay with siRNAs for screening other known NuRD components, aiming to identify those whose inhibition might enhance iPSC reprogramming while having minimal effect on somatic cell proliferation, even if completely depleted at the protein level. We used a transgenic "secondary reprogramming" platform of MEF derived reprogrammable cell line, which already contains heterozygote TetO-inducible OKSM (Stadtfeld et al., 2010). We targeted canonical NuRD complex core members: Mbd3 (used as a positive control), the mutually exclusive Mbd2, Chd4, Chd3, Mta2, Hdac2, Gatad2a and Gatad2b components and the non-canonical co-factor Zmynd8 Fig. 1A and Fig. S1F-H). siRNA was applied 48 hours and 96 hours after DOX administration (reprogramming initiation). Consistent with our previous results, treating the cells with siRNA for Mbd3 and Chd4, but not Chd3 or Mbd2, has significantly improved the reprogramming rate as measured 9 days after DOX addition (Fig. 1A-B) (Cheloufi et al., 2015; Rais et al., 2013). Remarkably, a significant and equivalent improvement in the reprogramming rate was seen in cells treated with validated siGatad2a (Fig. 1A-B, Fig. S1G-H). However, while depleting the NuRD components Mbd3 and Chd4 compromised the cell growth rate (Fig. 1C), depleting Gatad2a had a minimal effect on changing proliferation rate of MEFs or apoptosis (Fig. 1C-F).

We next used modified regimens in order to further optimize the process, by repressing NuRD components at earlier time points during reprogramming. Chd4, Gatad2a, Chd3 and Mbd3 targeted siRNAs and siScramble were applied in two different transfection cycles: the first transfection cycle started in somatic cells before OKSM induction (pre-DOX), and the second one started 24 hours after reprogramming was initiated (post-DOX). siRNA mediated knockdown of Mbd3 and Chd4 achieved improved reprogramming only in the second regimen of transfection (post-DOX), while the siRNA transfection which started before OKSM induction (pre-DOX) had a negative effect on the reprogramming process (Fig. 1G-H). The latter is consistent with previous reports indicating that complete ablation of Mbd3 in the earlier somatic cell-like states hinders reprogramming due to higher propensity for blocking somatic cell proliferation (Rais et al., 2013) which is essential for iPSC reprogramming (Hanna et al., 2009). However, as Gatad2a knockdown did not profoundly alter cell proliferation or viability, comparable enhancement in iPSC reprogramming rate was obtained in both transfection regimens (Fig. 1G-H). iPSC lines acquired following siRNA treatment for Gatad2a were pluripotent as evident by uniform expression of pluripotency markers and teratoma formation (Fig. S1J-K). These findings underscore Gatad2a as a potential alternative and technically more flexible way to inhibit Mbd3/NuRD repressive activity (i) early in the reprogramming process and (ii) without profoundly blocking somatic cell proliferation when completely depleted.

Gatad2a ablation facilitates deterministic iPSC reprogramming

We next set out to examine the effect of abolishing Gatad2a during iPSC reprogramming by establishing sets of genetically modified Gatad2a knock out (Gatad2a-KO or Gatad2a$^{-/-}$) secondary reprogrammable lines that were generated from defined parental isogenic Gatad2a $^{+/+}$ wild-type (WT) lines. We first generated two different Gatad2a$^{+/+}$ secondary reprogramming cell lines (Fig. 2A, Fig. S2A): i) WT primary MEFs were reprogrammed with viruses encoding FUW-M2rtTA and FUW-TetO-STEMCCA-OKSM. A randomly picked iPSC clone was selected and rendered transgenic for a constitutively expressed mCherry allele (to allow tracking of viable somatic cells) and a stringent ΔPE-Oct4-GFP transgenic reporter for naïve pluripotency (Rais et al., 2013). An isolated clone was then used as a WT control and as the source for an isogenic matched Gatad2a$^{-/-}$ null cell line. Notably, CRISPR/Cas9 was used to generate Gatad2a$^{-/-}$iPSCs/ESCs (Fig. 2B-C). Gatad2a$^{-/-}$ clones were obtained with ~40% successful targeting efficiency (Fig. S2B-D). ii) Reprogrammable Gatad2a$^{+/+}$ line was established by deriving an ES line from mice carrying Rosa26-M2rtTA$^{+/-}$; m.Col1a-TetO-OKSM$^{+/-}$and heterozygous for a single copy OG2-ΔPE-Oct4-GFP transgenic allele (Boiani et al., 2004), and was then subjected to CRISPR/Cas9 Gatad2a KO strategy (Fig. S2). Isogenic WT and Gatad2a$^{-/-}$ transgenic reprogrammable iPSCs/ESCS were injected into host blastocysts, and secondary MEFs were isolated (Fig. 2D and Fig. S3A). In all experiments, isogenic sets were used side by side for comparative analysis (Fig. 2A). As previously optimized for Mbd3$^{flox/-}$ donor cells (Rais et al., 2013), iPSC reprogramming was done in mES medium supplemented with DOX and 5% O_2 conditions for the first three days, followed by changing to 20% O_2 and knockout serum replacement (KSR) based medium supplemented medium with 2i/LIF (Fig. 2A). For all different lines tested, a profound improvement in the reprogramming rate and efficiency was observed, as the Gatad2a-KO cells achieve high rates AP+ colonies in 6 days only (Fig. 2E). Reprogramming rates were also measured by FACS analysis by quantification of ΔPE-Oct4-GFP positive cells every 24 hours for 8 days without cell passaging to avoid any biases (Fig. 2F, Fig. S3B). Reminiscent of kinetics measured in Mbd3$^{flox/-}$ systems (Rais et al., 2013), ΔPE-Oct4-GFP starts to appear after 3 days of reprogramming, and its rates ascending prominently on day 4 and reach >95% after 8 days of OKSM induction (Fig. 2F, Fig. S3B).

We next utilized the fact that these lines harbor constitutive expression of mCherry marker that allows detection of viable somatic cells after single cell plating, and quantified reprogramming efficiency of MEFs at the single cell level by detecting Oct4-GFP. Gatad2a-KO MEF yielded up to 100% iPSC derivation efficiency by day 8 in independent isogenic systems tested, as determined by mCherry+/ΔPE-Oct4-GFP+ expression in iPSC colonies obtained (Fig. 2G). When isogenic Gatad2a WT cells reprogrammed under identical conditions, <10% of clones reactivated ΔPE-Oct4–GFP and the majority of clones were only positive for mCherry, but not Oct4-GFP (Fig. 2G). In order to validate expression of pluripotency network, 96 well plates were fixed and found positive for endogenous Nanog/SSEA-1 and Oct4/Esrrb pluripotency marker expression in all clones tested at day 8 (Fig. S3C). FACS analysis of ICAM-1/CD44 surface markers (O'Malley et al., 2013), indicated uniform iPSC-like pattern acquisition on day 8 in Gatad2a$^{-/-}$ cells (Fig. S3D). Similar reprogramming efficiencies were observed upon reprogramming of other cell types including macrophages, neural progenitor cell (NPCs) and Pro-B cells (Fig. 2H). Finally, we

utilized NGFP2 primary iPSC line carrying Dox inducible lentiviral vectors individually encoding O, S, K and M, and a knock-in Nanog-GFP reporter (Hanna et al., 2008) and generated isogenic Gatad2a reprogrammable system which also yielded similar high efficiencies (Fig. S3E).

We then applied microscopic live imaging of the reprogramming dynamics (Rais et al., 2013), by seeding somatic cells constitutively labeled with mCherry marker, and their reprogramming was evaluated by ΔPE-Oct4-GFP reactivation (Fig. S4, Video S1-S2). Time-lapse measurements showed a dramatic increase in ES-like colony formation in Gatad2a$^{-/-}$, 6 days after DOX induction: more than 98% of Gatad2a$^{-/-}$ clonal populations reactivated ΔPE-Oct4-GFP pluripotency marker, compared to 15% in isogenic WT control cells, reprogrammed in identical conditions. Notably, while the most efficient conditions of reprogramming Gatad2a$^{-/-}$ and Mbd3$^{flox/-}$ cells involve applying 2i after 3 days (up to 100% at day 8), Gatad2a$^{-/-}$ cells reprogram with dramatically higher efficiency at day 7-8 (40-60%) in all different conditions tested, compared to their isogenic WT controls (Fig. S3F-G). Following Gatad2a depletion, we were not able to isolate stable partially reprogrammed cells that did not reactivate ΔPE-Oct4-GFP and could be stably expanded *in vitro* (Fig. S3H). The residual <15% ΔPE-Oct4-GFP negative colonies seen at day 6 in Gatad2a$^{-/-}$ cells, rapidly become GFP+ with additional reprogramming (Fig. S3H).

Gatad2a$^{-/-}$ derived iPSC lines from each of the different systems were established independently of DOX. All 3 randomly tested clones retained normal karyotype, stained positive for pluripotency markers, formed mature teratomas and chimeras consistent with adequate reprogramming (Fig. S5A-C). We conducted gene expression analysis on bulk donor MEFs at days 1-8 following DOX induction without cell passaging or sorting and compared them to iPSC and ESC lines. Gatad2a$^{-/-}$ somatic cells and not Gatad2a$^{+/+}$ cells clustered separately from donor fibroblasts already at day 1 following DOX (Fig. 3A-B). Gatad2a$^{-/-}$ MEFs cluster together with WT MEFs, thus ruling out a possibility of dramatically different transcriptional state before DOX (Fig.3A-B, Fig. S5D, Excel Sheet S1). By day 8 Gatad2a$^{-/-}$ cells were transcriptionally indistinguishable from multiple ESC and subcloned established iPSC lines (Fig. 3A-B). Key regulators of mouse fibroblasts or pluripotency were similarly expressed in Gatad2a$^{-/-}$ and Gatad2a$^{+/+}$ lines (Fig. 3A-B, Fig. S5D). Genome wide chromatin mapping for H3K27acetyl (K27ac) by Chromatin Immunoprecipitation followed by sequencing analysis (ChIP-seq), also confirmed that by day 8, only Gatad2a$^{-/-}$ transduced MEFs had assumed an ESC-like chromatin profile (Fig. 3C). Genome wide DNA methylation mapping by whole genome bisulfite sequencing (WGBS) confirmed that the Gatad2a$^{-/-}$ polyclonal population of MEFs and iPSCs are positively correlated to their WT counterparts (Fig. 3C). Moreover, FISH experiments for asynchronous DNA replication show that cells undergoing reprogramming start from asynchronous replication (in their somatic state) but adopt synchronous replication pattern which suits ground state naïve pluripotent stem cells (Masika et al., 2017) as early as day 5 (Fig. S5E), coinciding with the reactivation of ΔPE-Oct4-GFP reporter in the majority of cells.

Inhibiting Gatad2a expression was not sufficient to induce iPSC formation in the absence of exogenous OKSM overexpression (Fig. S5F). Gatad2a ablation did not replace exogenous

OSK expression neither to give rise to iPSCs and the presence of exogenous cMyc was essential to obtain iPSCs at up to 100% efficiency (Fig. S5F). Last, in order to validate the specificity of Gatad2a function, we reconstituted Gatad2a expression in Gatad2a$^{-/-}$ MEFs using viral infection. Subsequent reprogramming resulted in a significant decrease of reprogramming efficiency comparing to Gatad2a$^{-/-}$ cells (Fig. 3D). Gatad2b over expression failed to achieve the same phenotype and had no influence on reprogramming efficiency of Gatad2a$^{-/-}$ MEFs. These results suggest a non-redundant role for Gatad2a which cannot be compensated for by Gatad2b (Fig. 3D). The latter is consistent with recent observation that Gatad2a and Gatad2b form mutually exclusive complexes (Spruijt et al., 2016) and that Gatad2a$^{-/-}$ mice are embryonically lethal by E8.5-10.5 despite residual Gatad2b expression (Marino and Nusse, 2007).

While Gatad2a$^{-/-}$ MEFs had epigenetic and functional features of somatic WT MEFs (including RNA-seq, WGBS and H3K27ac) (Fig. 3A-C), in order to exclude a possibility that a boost in iPSCs efficiency is partially reliant on incomplete differentiation of somatic cells generated while Gatad2a was completely absent during the differentiation process, we generated Gatad2a$^{flox/flox}$ Secondary iPSCs (Fig. 3E). Subsequently, the *Rosa26* locus was correctly targeted with a knock-in CRE-Ert allele. Secondary Gatad2a$^{flox/flox}$ MEFs were used for reprogramming efficiency measurement with the presence or absence of Cre recombinase or Tamoxifen (4OHT) which were both efficient in ablating Gatad2a protein expression (Fig. 3F-G). Reprogramming efficiency and kinetics from Tat-CRE (HTNC) treated cells (Gatad2a$^{\Delta/\Delta}$) were equivalent to Gatad2a$^{-/-}$ cells, while reprogramming efficiency of isogenic paternal isogenic Gatad2a$^{flox/flox}$ somatic cells were below 18% (Fig. 3H). Finally, siRNA mediated depletion of GATAD2A in primary human adult fibroblasts yielded up to 9 fold increase in iPSC recovery only 9 days following sendai virus delivery of OSKM factors in naïve growth conditions (Fig. S5G-I) (Gafni et al., 2013).

Gatad2a restrains naive pluripotency maintenance and lineage priming

In order to evaluate Gatad2a influence on naïve pluripotency maintenance, Gatad2a was knocked out in a murine ESC line derived from mice harboring the OG2-vPE-Oct4-GFP single copy transgenic reporter that exclusively marks the naïve pluripotent state (Bao et al., 2009; Yoshimizu et al., 1999) (Fig. 4A). The Gatad2a$^{-/-}$ ESC stained positive for pluripotency markers tested (Fig. 4B) and exhibited a normal protein expression profile (Fig. 4C), suggesting that Gatad2a is dispensable for naïve pluripotency maintenance. Mbd3 protein levels were slightly reduced in Gatad2a KO ESCs, and the opposite pattern was observed for Gatad2a levels in Mbd3 hypomorphic mouse cells (Fig. 4C, Fig. S1E). The latter is consistent with previous report indicating reduced protein stability of NuRD complex partners when one of the components is deleted (Kaji et al., 2006). OG2- PE-Oct4-GFP reporter levels were similar in WT and Gatad2a-KO, as shown by FACS analysis (Fig. 4A-B, D). Some naïve pluripotency markers were slightly, yet significantly, higher in Gatad2$^{-/-}$ cells in FBS/LIF or 2i/LIF conditions (Fig. S6A). More strikingly, Gatad2a$^{-/-}$, but not WT cells, could be maintained in serum only (FBS only) conditions without supplementation of LIF for >12 passages while maintaining >PE-Oct4-GFP and naïve marker expression (Fig. 4D, Fig. S6C)), consistent with an increase in their naïve

pluripotency stability and similar to previously obtained results with Mbd3$^{-/-}$ murine ESCs (Kaji et al., 2006; O'Shaughnessy-Kirwan et al., 2015; Rais et al., 2013).

We next set out to find whether Gatad2a deletion has a functional effect on the cell ability to undergo lineage priming, by examining conversion of naïve ESC into primed Epiblast like cells (EpiLC) *in vitro* (Weinberger et al., 2016) after 60 hours expansion in primed Fgf2/Activin conditions (Fig. 4E). Gatad2a$^{+/+}$ cells showed a rapid decrease in PE-Oct4-GFP signal and lost their domed-like shape morphology, while Gatad2a$^{-/-}$ cells retained their naïve morphology and had a negligible decrease in vPE-Oct4-GFP (Fig. 4E). RT-PCR analysis revealed that conversion of Gatad2a$^{-/-}$ caused a smaller downregulation of naïve pluripotency related genes compared to isogenic WT cells (Fig. 4F, Fig. S6B). In addition, KO cells were deficient in up-regulating early development genes such as Otx2 and Fgf5 (Fig. 4F, Fig. S6B). Notably, this delay in undergoing priming is resolved later on, as Gatad2a$^{-/-}$ are able to differentiate despite of the latter delay and form mature teratomas upon microinjection into immune deficient mice *in vivo* (Fig. 4G) as similarly described for Chd4$^{-/-}$ and Mbd3$^{-/-}$ ESCs. Collectively, these observed phenotypes resemble those previously validated with Mbd3 and Chd4 depleted naïve ESCs (Kaji et al., 2006; O'Shaughnessy-Kirwan et al., 2015), and underlines the common functionality of these three NuRD components.

To evaluate whether Gatad2a depletion promotes reversion of primed cells to naïve pluripotency, Epiblast stem cells (EpiSCs) were established and validated from Gatad2a$^{flox/flox}$ ESCs that harbor naïve specific OG2-vPE-Oct4-GFP reporter (Fig. 4H). In comparison to isogenic Gatad2a$^{flox/flox}$ EpiSCs, single cell clonal analysis for epigenetic reversion of EpiSCs demonstrated >95% >PE-Oct4-GFP+ single cell reversion efficiency in Gatad2a depleted cells (Fig. 4I). These efficiencies are similar to those obtained following reversion of Mbd3$^{flox/-}$ EpiSCs (Rais et al., 2013). Deleting Gatad2a in EpiSCs while maintaining them in primed Fgf2/Activin conditions for extended passages (Fig. S6D) partially compromised their primed identity towards naivety. While OG2-iPE-Oct4-GFP remained negative in both early and late passage Gatad2a$^{\Delta/\Delta}$ cells expanded in primed conditions, late passage Gatad2a$^{\Delta/\Delta}$ had reduced expression of primed markers such as Brachyury and upregulated expression of naïve markers in comparison to early passage Gatad2a$^{\Delta/\Delta}$ cells (Fig. S6E-F). These results demonstrate that reduction of Gatad2a levels compromises the identity of the primed pluripotent state and renders nearly complete reversion of all donor primed EpiSCs to ground state pluripotency upon exposure to robust naïve conditions.

Mbd3, Gatad2a and Chd4 form a key molecular axis within the NuRD complex that restrains naïve pluripotency

Gatad2a inhibition was a key and dominant contributor to the radically efficient regression towards naïve pluripotency reported herein (Fig. 2). Thus, we aimed to define the mechanisms by which Gatad2a influences Mbd3/NuRD to facilitate inhibition of naïve iPSC reprogramming. We established an ESC line that harbors a DOX inducible allele of 2XFlag-Mbd3 (Fig. S2Aiii). Subsequently, we generated an isogenic Gatad2a$^{-/-}$ clone by CRISPR/Cas9 mediated targeting (Fig. S2). In order to estimate the influence of Gatad2a-KO on

Mbd3 interactome we applied LC-MS/MS analysis of Mbd3 protein interactions in Gatad2a-KO line and its isogenic WT control. When focusing the analysis on NuRD components, we noted that all canonical members of the NuRD complex were identified in both samples in high abundance and high correlation, except for Gatad2a and Chd4, which were missing from the Gatad2a-KO sample (Fig. 5A).

We then set out to establish whether and how Mbd3, Chd4 and Gatad2a constitute a biochemical axis within the NuRD complex. Co-immunoprecipitation (Co-IP) analysis for Mbd3 in ESCs and MEFs showed that in the absence of Gatad2a, Chd4 could not be immunoprecipitated with Mbd3 (Fig. 5B). The latter principle was validated following IP of endogenous Chd4 in MEFs expressing OKSM undergoing reprogramming, showing that upon depletion of Gatad2a via siRNA, Chd4 can no longer directly interact with Mbd3 (Fig. 5C). Transgenic ectopic reconstitution of Gatad2a in Gatad2a$^{-/-}$ MEFs, reestablished specific interaction between Mbd3 and Chd4 (Fig. 5D).

We next wanted to determine which domains of Gatad2a and Mbd3 regulate this specific interaction. Mbd3, which is a member of the Methyl-binding-domain (MBD) proteins retains two domains – the Methyl binding domain, which has been shown to mediate its interaction with transcription factors like OKSM during iPSC reprogramming (Rais et al., 2013), and the highly conserved Coiled coil region whose precise function in Mbd3 is not fully characterized (Fig. 5E). 2xFlag-tagged Mbd3 (2XFlag-WT-Mbd3) expression vector was generated together with two mutant versions of Mbd3, one of which lacks the coiled-coil region (2xFlag-ΔCCR-Mbd3) and the other one lacks the Methyl binding domain (2XFlag-ΔMBD-Mbd3) (Fig. 5F). Co-IP experiments showed that while the elimination of the coiled-coil region of Mbd3 (2XFlag-ΔCCRMbd3) does not interrupt binding to reprogramming factors (e.g. Klf4), it does prevent exclusively Gatad2a and Chd4 binding (Fig.5G). On the contrary, ΔMBD-Mbd3 was unable to bind Klf4, but maintained its ability to bind to different NuRD components, like Gatad2a and Chd4, as similarly shown in previous publications (Rais et al., 2013) (Fig. 5G). Reconstitution of Mbd3$^{flox/-}$ cells with ΔCCR-Mbd3 transgene, which is missing the coiled-coil domain, did not inhibit reprogramming efficiency, consistent with its inability to recruit Chd4 (Fig. 5H-I).

In order to further characterize Mbd3-Gatad2a-Chd4 axis, we tried to interrupt the axis assembly by creating a competition over the binding site of Gatad2a to Mbd3. We over-expressed Gatad2a-coiled coil region (Gatad2a-CCR1) in a truncated construct or as a short peptide (Fig. S7A-C) together with 2XFlag-Mbd3. Indeed, the over-expression of Gatad2a Coiled-coil region reduced the endogenous Gatad2a and Chd4 binding to Mbd3 (Fig. S7B-C). Overexpression of the MBD domain of Mbd3 as an independent peptide directly interacted with Oct4 and Hdac2, showing again that the MBD is required and sufficient for the binding of the pluripotency factors (Rais et al., 2013), but not to Gatad2a and Chd4 (Fig. S7D). Collectively these results establish a function of Mbd3-CCR as the mediator of Gatad2a and Chd4 binding, and corresponds with previous publications showing that the highly conserved Mbd2-CCR forms a heterodimer with Gatad2a and thus to modulate the complex's function (Gnanapragasam et al., 2011).

Interactions and assembly of Mbd3/NuRD are context dependent and can be modified post-translationally

Reprogramming factors directly interact and co-IP with Mbd3/NuRD complex following reprogramming initiation in somatic cells (Rais et al., 2013). By assessing these interactions in other cell states (representing different differentiation conditions), we noted that some of these interactions do not exist in naïve ESCs expanded in 2i/LIF conditions despite abundant expression of Oct4 and Klf4 (Fig. 6A) as well as expression of all known NuRD components under these conditions (Fig. 6A). However, upon brief 48h priming of naïve conditions and turning them into EpiLCs (Fig. 6A) or long term EpiSCs (Fig. 6B), interactions between Oct4/Klf4 and Mbd3/NuRD became rapidly and strongly evident. These results are consistent with the notion that Mbd3/NuRD promotes ESC differentiation (Reynolds et al., 2012), indicating that this propensity collates with increased NuRD-TF interactions upon pluripotency priming, or in somatic cell state conditions experiencing non-physiologic ectopic expression of pluripotency factors like OSK (Fig. 6A-B).

While the nature and molecular basis of these context specific interaction remain to be explored, we asked whether enhancing naïve pluripotency conditions by blocking other signaling pathways may deplete certain Mbd3-Chd4-Gatad2a/Complex components. Remarkably, of the several small molecule inhibitors previously published to promote naïve ESC maintenance (SRCi, TGFRi, PKCi) (Dutta et al., 2011; Shimizu et al., 2012), the broad spectrum PKC inhibitor Go6983 dramatically depleted Mbd3 protein expression within 24 hours of treatment in different naïve conditions (Fig. 6C-F). This effect was seen in the context of pluripotent cells but not significantly when Go6983 was applied on somatic cells (Fig. 6G) suggesting that the Mbd3 depletion effect was not by specific and direct action of this small molecule on the NuRD complex. Finally, this reduction in Mbd3 protein was not accompanied by a change of Mbd3 mRNA transcript levels (Fig. 6H).

We next tested the ability of Go6983 to boost iPSC efficacy formation from WT secondary reprogrammable MEFs. We observed an increase of reprogramming efficiency up to 45% compared to DMSO treated cells (Fig 6I-J). It is possible that the efficiencies were not nearing 90-100%, because Go6983 does not deplete Mbd3 in somatic cells and only later after initiation of the process when cells start becoming ES-like (Fig. 6I), which helps further boost reprogramming efficiency. However, in the context of EpiSC reversion, we were able to obtain up to 85% EpiSC reversion efficiency in combination with LIF/ERKi/ROCKi/DBZ/Vitamin C were supplemented with Go6983, while without PKCi or when using SRCi instead, reversion efficiencies remained below 25% (Fig. 6K). Consistent with the observed Go6983 effect on depletion of Mbd3, but not other NuRD components (Fig. 6F), only ectopic transgenic expression of Mbd3 inhibited the boost in EpiSC reversion efficiency we observed in PKCi containing conditions (Fig. S7E). Finally, while WT ESCs cannot maintain their naïve pluripotency in N2B27 LIF unless an additional component like 2i or PKCi are provided (Dutta et al., 2011; Ying et al., 2008), Gatad2a$^{-/-}$ ESCs were fully stable for many passages in N2B27 LIF only conditions (in the absence of Go6983) as determined by the stringent naïve specific vPE-Oct4-GFP reporter and RT-PCR for naïve pluripotency factor markers (Fig. 6L, Fig. S6C). The latter results underscore a previously

unidentified link between naïve PSC booster Go6983 and depletion of Mbd3/NuRD in the context of murine pluripotency and reprogramming.

Given the above result showing context- and signaling dependent influence on NuRD complex stability and function, we wondered whether some of other recently identified genetic perturbations shown to boost iPSC reprogramming efficiency had direct or indirect effects on NuRD assembly and its ability to be loaded on the chromatin. Serendipitously, we analyzed the effect of depleting global SUMOylation on iPSC reprogramming as it was shown to be an efficient booster (Borkent et al., 2016; Cheloufi et al., 2015). Moreover, SUMOylation on Gatad2a protein has been shown to be essential for its direct binding within NuRD (Gong et al., 2006). We validated that knockdown of Ubc9 boosts WT MEF reprogramming up to ~40% following OKSM induction as previously reported (Fig 6J). The ability of Chd4 to Co-IP with Gatad2a was specifically decreased following Ubc9 knockdown in ES, but not in control shRNA (Fig. 7A). Notably, Ubc9 depletion did not affect abundance of NuRD components in these cells (Fig. 7B). Similar results were obtained when using a specific small molecule inhibitor (2-D08) to inhibit SUMOylation (Fig. 7C). Cellular fractionation, followed by chromatin isolation in control vs. Ubc9 depleted samples undergoing pluripotency reprogramming showed specific depletion of Gatad2a/Mbd3 from the purified chromatin fraction, but not DNMT1 or Mta2 (used as controls) (Fig. 7D-E), supporting that Gatad2a-Mbd3/NuRD assembly and loading on the chromatin is compromised following depleting SUMOylation and may contribute to the increase of iPSC efficiency obtained under these conditions.

The critical amino acid residues in GATAD2A within canonical SUMOylation motifs have been previously characterized (Gong et al., 2006). In order to evaluate the contribution of SUMOylation to Gatad2a function in mouse reprogramming, we cloned the WT and validated mutant form of GATAD2A at the two critical SUMOylated residues (K30R, K485R) (Gong et al., 2006) (Fig. 7F), which were subsequently introduced into Gatad2a-KO secondary MEF as lentiviruses (Fig. 7G). MEFs were then subjected to iPSC reprogramming. While cells infected with Gatad2a-WT showed a dramatic decrease in reprogramming efficiency, comparing to the Mock infected cells, the Gatad2a mutant at the SUMOylation residues had a negligible effect on blocking the cells' reprogramming ability (Fig. 7H-I). In sum, we identified a link between inhibition of SUMOylation that boosts iPSC efficiency, to NuRD stability and assembly in the context of reprogramming, specifically through Gatad2a-Chd4 interaction.

Discussion

Here we provide an additional example showing that radically efficient and deterministic direct induction of pluripotency is feasible with modified direct *in vitro* reprogramming approaches. The findings related to Gatad2a depletion and iPSC reprogramming characterized herein, expand previous work demonstrating that optimized depletion of the NuRD complex component Mbd3, can alter this process and shift it toward deterministic dynamics and high efficiency (Rais et al., 2013). Nonetheless, complete depletion of Mbd3 during somatic state rapidly yields cell cycle malfunctions, thus technically restricting the usage of this platform to a carefully engineered set of mutant lines. In this work, we have

further elucidated the mechanisms of Mbd3 repressive effect during reprogramming and underlined the importance of NuRD complex recruitment. Mbd3-Gatad2a-Chd4 is a triple component axis within Mbd3/NuRD complex which mediates the inhibitory function of NuRD in the context of reprogramming and maintenance of naïve pluripotency. Disassembly of this sub-complex by different approaches (Fig. S7F), leads to deterministic reprogramming.

Results obtained herein underline the importance of the NuRD structure and sub-complexes; Chd4, Mbd3 and Gatad2a are all mutually exclusive in the NuRD complex and each can be replaced by a highly homologous protein (Chd3, Mbd2, Gatad2b, respectively). However, the specific combination of all three seems to be crucial in order to execute the inhibitory function related to pluripotency and differentiation. Emerging evidence suggests that this effect can be also seen during development (Bode et al., 2016)(Nitarska et al., 2016). Further, whether certain conformations of NuRD complex can also enhances direct trans-differentiation between somatic cell types remains to be further explored (Becker et al., 2017).

Although the optimized depletion of Mbd3, Gatad2a or Chd4 result in a similar effect on iPSC reprogramming efficiency, these proteins hold differential roles, possibly in NuRD sub-complexes, or even in a NuRD-independent manner. This notion is supported by their different effect on cell proliferation, and also expressed in their slightly different KO phenotypes during embryonic development, as both Chd4 and Mbd3 are embryonic lethal at an early stage (E4.5-E6.5) (Kaji et al., 2006; O'Shaughnessy-Kirwan et al., 2015), while Gatad2a-KO embryos are viable until E8.5-10.5 (Marino and Nusse, 2007). Thus, further work is needed to fully understand the mechanisms underlying Mbd3 functions (including NuRD dependent and independent ones) and specific role for Gatad2a, but not Gatad2b which can also assemble with Chd4/Mbd3. The findings that recruitment of OKSM by Mbd3/NuRD is prominent in primed or somatic state, but not in naïve pluripotency, highlight the context dependent ability of NuRD interaction in different states whose molecular basis is future research interest.

Although this study shows that neutralizing Mbd3-Gatad2a-Chd4/NuRD complex facilitates up to 100% reprogramming efficiency, it remains to be seen whether all the cells follow the same molecular trajectory during the 8 days course or, alternatively, can achieve the final naïve state by distinct parallel paths. Our results do not necessarily suggest that WT cells that succeed in generating iPSCs must neutralize this pathway in order to succeed but may suggest that some cells succeed despite the negative effect of the presence of this complex. Gatad2a-Mbd3/NuRD may act like a negative rheostat rather than an absolute blocker for iPSC formation, where it severely reduces the chance of reactivation of a conducive reprogramming trajectory in Gatad2a/Mbd3 WT cells, since it represses the same genes OKSM are trying to reactivate (Rais et al., 2013). Upon dismantling this negative effector, nearly all cells can embark correctly on a conducive reprogramming trajectory towards naïve pluripotency.

A recent study has provided an NuRD independent example of the "gas and brakes" mechanisms for iPSC formation (Rais et al., 2013), where the OKSM reprogramming

factors not only recruit complexes that positively promote reactivation of naïve pluripotency, but also recruit negative factors, in this case the NCoR/SMRT-HDAC3 co-repressor complex (Zhuang et al., 2018). Up to 70% Oct4-GFP reactivation over a reprogramming period of 12 days by OKSM upon depletion of NCoR/SMRT. The latter is recruited directly by OKSM and negatively represses key loci required for successful reprogramming. The above results suggest that just like in the case of reprogramming factors that can act synergistically and occasionally substitute for each other to induce iPSCs, multiple co-repressor complex like Mbd3/NURD, NCoR/SMRT, and potentially others may have overlapping effect in restraining the awakening of pluripotency is somatic cells.

Star * Methods

Key Resources Table

List of RT-PCR primers used in this study.

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Mbd3	Bethyl	Cat# A302-529; RRID:AB_1998976
Mbd2	Bethyl	Cat# A301-633A; RRID:AB_1211480
Gatad2a (H-162)	Santa Cruz	Cat# Sc-134712; RRID:AB_11150493
Gatad2b	Bethyl	Cat# A301-281A; RRID:AB_937934
Hdac2 (c-8)	Santa Cruz	Cat# Sc-9959; RRID:AB_627704
Mta2 (c-20)	Santa Cruz	Cat# Sc-9447; RRID:AB_214965
Chd4	Abcam	Cat# Ab70469; RRID:AB_2229454
Oct4 (H-134)	Santa Cruz	Cat# Sc-9081; RRID:AB_2167703
Klf4 (H-180)	Santa Cruz	Cat# Sc-20691; RRID:AB_669567
Nanog	Behtyl	Cat# A300-398A; RRID:AB_386109
Ha.11 16B12	Covance	Cat# MMS-101R; RRID:AB_10063630
Gapdh	Epitomics	Cat# 2251-1; RRID:AB_1267174
Hsp90β	Epitomics	Cat# 1492-1; RRID:AB_562129
Prmt5	Millipore	Cat# 07-405; RRID:AB_310589
Cdk2ap1 (H-4)	Santa Cruz	Cat# sc-390283
RbAp48 (N-19)	Santa Cruz	Cat# sc-8270; RRID:AB_2300829
Sall4	Abcam	Cat# ab29112; RRID:AB_777810
Ubc9	Cell signaling	Cat# #4918; RRID:AB_2210600
Dnmt1	Abcam	Cat# ab87654; RRID:AB_2041077
H3K27ac (ChIP)	Abcam	Cat# ab4729; RRID:AB_2118291
Oct4 (ChIP)	Santa Cruz	Cat# SC8628; RRID:AB_655982
Klf4 (ChIP)	R&D	Cat# AF3158; RRID:AB_2130245

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Sox2 (ChIP)	Millipore	Cat# AB5603; RRID:AB_2286686
c-Myc (ChIP)	Santa Cruz	Cat# sc764; RRID:AB_631276
Oct4 (H-134) - Immunostainings	Santa-Cruz	Cat# SC9081; RRID:AB_2167703
SSEA-1 (MC-480)	Developmental Studies Hybridoma Bank	N/A
Esrrb	R&D	Cat# PP-H6705-00; RRID:AB_
Flag-M2	Sigma	Cat# F3165; RRID:AB_259529
ICAM-1 (CD54)	Biolegend	Cat# 116119; RRID:AB_10613645
CD44 Monoclonal antibody (IM7), PE	eBioscience	Cat# 12-0441-81; RRID:AB_465663
Chemicals, Peptides, and Recombinant Proteins		
FGF2	Peprotech	100-18C
ACTIVIN A	Peprotech	120-14
PD0325901	Axon Medchem	#1408
CHIR99021	Axon Medchem	#1386
Recombinant human LIF	Peprotech	300-05
N-Ethylmaleimide	Sigma-Aldrich	E3876
3Xflag peptide	Sigma-Aldrich	F4799
HA- peptide	Sigma-Aldrich	I2149
Ubc9-inhibitor 2-D08	Sigma-Aldrich	SML1052
Protease inhibitor cocktail (Sigma)	Sigma-Aldrich	P8340
PKC inhibitor GO6983	Axon Medchem	#2285
SRC inhibitor- CPG77675	Axon Medchem	#2097
ROCK inhibitor- Y27632	Axon Medchem	#1683
DBZ, γ-Secretase Inhibitor	Axon Medchem	#1488
TGF-βRI A83-01	Tocris	#2939
PKC inhibitor- GFX GO6850	Tocris	#0741
Critical Commercial Assays		
Alkaline Phosphatase Kit	Millipore	SCR004
Lipofectamine RNAiMAX	ThermoScientific	#13778075
Annexin-PI Kit	ThermoScientific	V13245
TruSeq RNA Sample Preparation Kit v2	Illumina	RS-122-2001
EZ DNA Methylation-Gold kit	Zymo	D5005
EpiGenome Methyl-Seq	Illumina	EGMK81312
Deposited Data		
ATAC-Seq, ChIP-Seq, RNA-Seq, WGBS	This Paper	GSE102518
Experimental Models: Cell Lines		
Mbd3$^{flox/-}$ cell lines – ES and Secondary MEF	Rais et al. (2013)	N/A
KH2 Flp-in murine ES cell line	Beard et al. (2006)	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FUW-M2RtTA; FUW-TetO-STEMCCA, OG2-ΔPE-Oct4-GFP cells (both Gatad2a WT and KO)	This paper	N/A
M.Col1a:STEMCCA-OKSM$^{+/-}$; Rosa26-M2rtTA; ΔPE-GOF18-Oct4-GFP cells (both Gatad2a WT and KO)	This paper	N/A
FUW-TetO-STEMCCA-OKSM R26-rtTA Cola:tetO-3XFlag-Mbd3 (both Gatad2a WT and KO)	This paper	N/A
V6.5 murine ES cell line	Beard et al. (2006)	N/A
Experimental Models: Organisms/Strains		
R26-M2rtTA knock-in Mouse strain	Jackson	#006965; RRID:IMSR_JAX:006965
Col1a1-tetO-OKSM Mouse strain	Jackson	#011001; RRID:IMSR_JAX:011001
OG2 mouse strain	Jackson	#004654; RRID:IMSR_JAX:004654
Oligonucleotides		
siRNA targeting Mouse-Mbd3	Invitrogen	MSS-237238
siRNA targeting Mouse-Chd3	Invitrogen	MSS238382, MSS238383, MSS238384
siRNA targeting Mouse-Gatad2a	Invitrogen	MSS- 200894, MSS-200895, MSS-200896
siRNA targeting Mouse-Chd4	Invitrogen	MSS239240, MSS239241, MSS239242
siRNA targeting Mouse-Zmynd8	Invitrogen	MSS213659, MSS213660, MSS213661
siRNA targeting Mouse- Gatad2b	Invitrogen	MSS213785, MSS213786, MSS213787
siRNA targeting Mosuse-Mta2	Invitrogen	MSS215840, MSS215841, MSS215842
siRNA targeting Mouse-Mbd2	Invitrogen	MSS205071, MSS205072, MSS205073
siRNA targeting Mouse-Hdac2	Invitrogen	MSS206540, MSS206541, MSS206542
siRNA targeting Mouse-Prkcz	Invitrogen	S71714
siRNA targeting Stealth RNAi™ siRNA Scrambled Negative Control, Med GC	Invitrogen	12935300
siRNA targeting Human GATAD2A	Invitrogen	HSS147779, HSS147780, HSS182660
rtPCR Primers, See Supplementary Excel Sheet 2	This paper	*N/A*
Recombinant DNA		
FUW-Flag-Mbd3	Rais et al. (2013)	Addgene #52371
FUW-Flag-Mbd3 delta1-70	Rais et al. (2013)	Addgene #52372
ΔPE-Oct4-GOF18-GFP Zeomycin-Resistant	Rais et al. (2013)	Addgene # 52382
FUW-mCherry-NLS cassette	Rais et al. (2013)	Addgene #52409
FUW-TetO-STEMCCA-OKSM	Sommer et al., (2009)	
pLM-mCerulean-cMyc	Papapetrou et al. (2009)	Addgene #23244

REAGENT or RESOURCE	SOURCE	IDENTIFIER
FUW-STEMCCA-OKS-mCherry	Sommer et al., (2009)	N/A
px330 Cas9-mGatad2a	This paper	N/A
px330 Cas9-sgRNA for conditional mGatad2a targeting	This Paper	N/A
FUW-2XFlag-ΔCCR-Mbd3 (Deletion of 220-279)	This Paper	N/A
FUW-Gatad2a	This Paper	N/A
FUW-Gatad2b	This Paper	N/A
pCAG-HA-Gatad2a-CCR (140-174 of Gatad2a	This Paper	N/A
pCAG-HA-(MBD)Mbd3 (1-170 of Mbd3)	This Paper	N/A
pCAG-HA-GFP	This Paper	N/A
FUW-Flag-Gatad2a_Mutated SUMO sites	This Paper	N/A
Software and Algorithms		
Expressionist software	Genedata	https://www.genedata.com/
Mascot	Matrix Science	http://www.matrixscience.com/server.html
Tophat 2.0.10	Trapnell et al, 2009	https://ccb.jhu.edu/software/tophat/index.shtml
Cufflinks 2.2.1	Trapnell et al, 2012	http://cole-trapnelllab.github.io/cufflinks/
R pheatmap package	https://cran.rproject.org/web/packages/pheatmap/index.html	https://cran.rproject.org/web/packages/pheatmap/index.html
R prcomp package	https://stat.ethz.ch/R-manual/R-devel/library/stats/html/prcomp.html	https://stat.ethz.ch/R-manual/R-devel/library/stats/html/prcomp.html
Bowtie 2	Langmead B & Salzberg S. 2012	http://bowtiebio.sourceforge.net/bowtie2/index.shtml
Picard tools	http://broadinstitute.github.io/picard/	http://broadinstitute.github.io/picard/
MACS 1.4.2	Zhang Y, et al, 2008	http://liulab.dfci.harvard.edu/MACS/
bedtools	Quinlan et al, 2010	http://bedtools.readthedocs.io/en/latest/
samtools	Li H. et al, 2009	http://www.htslib.org/doc/samtools.html
IGV	James et al, 2011	https://software.broadinstitute.org/software/igv/
Python misha package	Amos Tanay lab (Weizmann Institute, Israel),	https://bitbucket.org/tanaylab/mishapackage
	Nagano et al, 2017	
MATLAB	MathWorks	https://www.mathworks.com/products/matlab.html
Prism	GraphPad software	https://www.graphpad.com/scientific-software/prism/
FlowJo	FlowJo	https://www.flowjo.com/

REAGENT or RESOURCE	SOURCE	IDENTIFIER
ZEN Software	Zeiss	https://www.zeiss.com/microscopy/int/products/microscope-software/zen-lite.html
Other Link to related raw datasets: https://data.mendeley.com/datasets/vyjpgt3cc8/draft?a=093a2a8c-4a24-4e55-a621-6f82c34f2e2d		

Contact for Reagent and Resources Sharing

Further information, and requests for reagents will be fulfilled by the Lead Contact, Dr. Jacob H. Hanna (Jacob.hanna@weizmann.ac.il).

Experimental Model And Subject Details

Mice—6-20 week old OG2-ΔPE-Oct4-GFP transgenic allele-carrying mice (CBA/CaJ X C57BL/6J) were from The Jackson Laboratory (also known as OG2 strain - Jackson 004654, kind gift by H. Scholer). 6-20 week old Female R26-M2RtTA homozygous mice (Jackson #006965) were mated with 6-20 week old m.Col1a-TetO-STEMCCA-OKSM (Jackson #011001, kindly provided by K. Hochedlinger) male homozygous mice, in order to produce M2rtTA-OKSM mice double positive mice. The F1 mice (6-12 weeks old) were mated in order to create double homozygous offspring (RTTA $^{+\backslash+}$ OKSM $^{+\backslash+}$). F2 offspring (Double homozygous rtTA$^{+\backslash+}$ OKSM$^{+\backslash+}$) were mated with previously generated male or female OG2-ΔPE-Oct4-GFP homozygous reporter transgenic mice, in order to generate a double heterozygote with Oct4 reporter offspring (M2rtTA $^{+\backslash-}$ OKSM $^{+\backslash-}$OG2-ΔPE-Oct4-GFP$^{+\backslash-}$. All animal experiments were performed according to the Animal Protection Guidelines of Weizmann Institute of Science, Rehovot, Israel. All animal experiments described herein were approved by relevant Weizmann Institute IACUC (#00330111-Hanna). All mice were housed in a 12-hour light/12-hour dark cycle conditions in a specialized and certified animal facility.

Cell culture—WT or Mutant mouse ESC/iPSC lines and sub-clones were routinely expanded in mouse ES medium (mESM) consisting of: 500ml DMEM-high glucose (ThermoScientific), 15% USDA certified Fetal Bovine Serum (Biological Industries), 1mM L-Glutamine (Biological Industries), 1% nonessential amino acids (Biological Industries), 0.1mM β-mercaptoethanol (Sigma), penicillin-streptomycin (Biological Industries), 10μg recombinant human LIF (Peprotech). For ground state naïve conditions (N2B27 2i/LIF), murine naïve pluripotent cells (iPSCs and ESCs) were conducted in serum-free chemically defined N2B27-based media: N2B27-based media: 250ml Neurobasal (ThermoScientific), 250ml DMEM:F12 (ThermoScientific) 5ml N2 supplement (Invitrogen; 17502048), 5ml B27 supplement (Invitrogen; 17504044), 1mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1mM β-mercaptoethanol (Sigma), penicillin-streptomycin (Invitrogen), 5mg/ml BSA (Sigma), small-molecule inhibitors CHIR99021 (CH, 3 μM - Axon Medchem) and PD0325901 (PD, 1 μM - Axon Medchem). When used Go6983 was applied at 2 μM final concentration. Primed N2B27 media for murine cells (EpiSCs or EpiLCs) contained 12ng/ml recombinant human FGF2 (Peprotech) and 20ng/ml recombinant human ACTIVIN A (Peprotech) (instead of 2i/LIF). Mycoplasma detection

tests were conducted routinely every month with MycoALERT ELISA based kit (Lonza) to exclude mycoplasma free conditions and cells throughout the study. Cell line identity was validated by routine genotyping validation via PCR analysis, and correct targeting was validated via Southern Blot analysis.

Method Details

Knockout by CRISPR/Cas9 targeting—Different PSC lines were genetically manipulated by CRISPR/Cas9, in order to achieve Gatad2a-KO on different genetic backgrounds. Cells were transfected (Xfect, Clontech) with Cas9-sgRNA Gatad2a (CGCCTGATGTGATTGTGCTCT) on px330 backbone, alongside mCherry-NLS plasmids. The cells were sorted after 72 hours (mCherry positive cells) under sterile conditions via ND FACS ARIA III, and seeded on irradiated MEFs. After 8 days, colonies were picked and examined by HRM analysis (MeltDoctor HRM master mix, Life technologies, #4415440. HRM Primers- TTCTGAAGCATCCATGCTACACCTG, GCTCTCGTTCTTCTGGGCTGCTT). Candidate colonies were analyzed by western blot and confirmed by sequencing. For generation of floxed conditional knockout Gatad2a cells, CRISPR/Cas9 and donor vector were used as described (Fig 3E). Briefly, cells were transfected by electroporation with both donor vector containing Gatad2a Exon2 flanked by lox sequences, and px330 Cas9-sgRNA (gatacccatggtgcccacgg). Cells were subjected to selection by G418 and subsequent analysis of selected clones by Southern Blot analysis. Southern Blot analysis was used to confirm correct targeting. 10-15μg of genomic DNA was digested with HindIII restriction enzyme for 5 hours and separated by gel electrophoresis. The DNA was transferred to a nitrocellulose membrane that was next hybridized with a radioactive labeled probe and developed using ECL (ThermoScientific). Positive clones were transfected with Flippase to excise the neomycin selection cassette. Genotyping for the Gatad2a floxed allele and region was carried using the following primers: mGatad2a_dLox_F – AGCACTGAGGCGCTCTTGGTGA; mGatad2a_dLox_R – GCTCTCGTTCTTCTGGGCTGCTT; mGatad2a_ko_Frw – TGACAGAGCTCTGGGGTCGAGTA, resulting in one of three possible genotypes- WT allele (133bp), floxed (180bp) or deleted (240bp).

Secondary reprogrammable cells and systems—Please see Fig. S2A for schematics summarizing the three isogenic systems detailed below:

System i: FUW-M2RtTA; FUW-TetO-STEMCCA, ΔPE-Oct4-GFP cells: 129*Jae* MEFs were transduced with constitutively expressed FUW-M2RtTA and with FUW-TetO-STEMCCAOKSM (Sommer et al., 2009) via standard lentiviral preparation and infection. Cells were then reprogrammed in mESM + DOX for 14 days, and number of clonal iPSC line were established. A randomly selected clone was then transduced with ΔPE-Oct4-GFP carrying Zeomycin resistance (Addgene # 52382) (Gafni et al., 2013; Rais et al., 2013). Randomly selected male clone validated for specific activity of the ΔPE-Oct4-GFP reporter (downregulated upon priming and completely shut off upon differentiation 10 day EBs), was next labeled with a constitutively expressed mCherry-NLS cassettes using electroporation (Addgene #52409) (Rais et al., 2013). Only then, a sub cloned clone was subjected to

CRISPR/Cas9 targeting in order to generate and isogenic Gatad2a-KO cell line which was used as a source for secondary somatic cells or EpiSCs and used for conversion.

System ii: m.Col1a:STEMCCA-OKSM$^{+\backslash -}$; Rosa26-M2rtTA; ΔPE-GOF18-Oct4-GFP cells: R26-M2RtTA homozygous mice (Jackson #006965) were mated with m.Col1a-TetO-STEMCCAOKSM (Jackson #011001, kindly provided by K. Hochedlinger) homozygous mice, in order to produce M2rtTA-OKSM mice double positive mice. The F1 mice were mated in order to create double homozygous offspring (RTTA $^{+\backslash +}$ OKSM $^{+\backslash +}$). F2 offspring (Double homozygous M2rtTA$^{+\backslash +}$ OKSM$^{+\backslash +}$) were mated with previously generated OG2-ΔPE-Oct4-GFP homozygous reporter transgenic mice (also known as OG2 strain and line) (Jackson 004654, kind gift by H. Scholer), in order to generate a double heterozygote with Oct4 reporter offspring (M2rtTA $^{+\backslash -}$OKSM $^{+\backslash -}$ OG2-ΔPE-Oct4-GFP $^{+\backslash -}$). Subsequently, mouse ESCs were derived in mESM-MEF conditions and a male cell line carrying the above genotype was transduced with a constitutively expressed mCherry-NLS cassettes using electroporation. Only then, a sub cloned clone was subjected to CRISPR/Cas9 targeting in order to generate and isogenic Gatad2a-KO cell line which was used as a source for secondary somatic cells or EpiSCs and used for conversion.

System iii: FUW-TetO-STEMCCA-OKSM R26-rtTA m.Col1a:tetO-3XFlag-Mbd3: MEFs generated via microinjection of R26-rtTA col1a:TetO-3XFlag-Mbd3 ESCs, were introduced with OKSM (FUW-TetO-STEMCCA-OKSM) using a lentiviral transduction. A male iPSC cell line was established in mESM + DOX conditions, and subsequent Gatad2a-KO was achieved using CRISPR/Cas9 genome editing. In all of the above, both the Gatad2a-KO and the isogenic Gatad2a-WT were then injected to a blastocyst, and MEF were harvested at E12.5-E13.5. Chimeric MEF were then isolated using Puromycin resistance.

Secondary mouse embryonic fibroblast (MEF) from Mbd3$^{flox/-}$ cell line (A12 clone: *Mbd3*$^{flox/-}$ cell lines that carries the Oct4-GFP transgenic reporter (complete *Oct4* enhancer region with distal and proximal enhancer elements) (Addgene plasmid #60527)) and WT* cell line (WT 1 clone that carries the ΔPE-Oct4-GFP reporter (Addgene plasmid#52382) were previously described (Rais et al., 2013). Note that we do not use Oct4-GFP or any other selection for cells before harvesting samples for conducting genomic experiments. Throughout the 8-day course of the reprogramming experiment the cells were not passaged to avoid any biases. All animal studies were conducted according to the guideline and following approval by the Weizmann Institute IACUC (approvals #33550117-2 and 33520117-3). Cell sorting and FACS analysis were conducted on 4 lasers equipped FACS Aria III cells sorter (BD). Analysis was conducted with either DIVA software or Flowjo.

Deterministic somatic cell to iPSC secondary reprogramming experiments— Deterministic or near-deterministic iPSC reprogramming experiments were executed according to the following protocol. Reprogramming was initiated in mESM, which contained 500 ml DMEM-high glucose (ThermoScientific), 15% USDA certified fetal calf serum (Biological Industries), 1 mM L-glutamine (BI), 1% non-essential amino acids (BI), 1% penicillin– streptomycin (BI), 0.1 mM β-mercaptoethanol (Sigma), 10 μg recombinant human LIF (Peprotech). MES medium for reprogramming was supplemented with

Doxycycline (1 μg/ml), and the MEF[2] were seeded on irradiated DR4 feeders. After 3.5 days, the medium was changed to serum-free media: referred to as KSR-2i/LIF media, which was also supplemented with Doxycycline, as well as 2i/LIF. The medium contained-500 ml DMEM – high glucose (ThermoScientific), 15% Knockout Serum Replacement (KSR - ThermoScientific; 10828), 1 mM L-Glutamine (BI), 1% non-essential amino acids (BI), 0.1 mM β-mercaptoethanol (Sigma), 1% penicillin–streptomycin (BI), 10 μg recombinant human LIF (Peprotech), CHIR99021 (3 μM; Axon Medchem), PD0325901 (PD, 1μM; Axon Medchem). Cells were seeded upon irradiated MEF on 0.2% gelatin coated plates, at a low density of approximately 100-150 cells/cm. Reprogramming was conducted at 5% O2 5% Co2 and 37C in the first 2.5 days, and then moved to 20% O2 5%Co2 and 37C. Media replaced every other day (0, 2, 3.5, 6, 8). No blinding was conducted when testing outcome of reprogramming experiments throughout the study. For all mouse iPSC reprogramming experiments intended for genomics analysis, irradiated human foreskin fibroblasts were used as feeder cells (rather than mouse DR4 MEFs), as any sequencing input originating from the use of human feeder cells cannot be aligned to the mouse genome and is therefore omitted from the analysis. Isolation of secondary B-cells, macrophages and NPCs was conducted as previously described (Rais et al., 2013; Di Stefano et al., 2013). Throughout the 8-day course of the reprogramming experiment the cells were not passaged to avoid any biases.

For single cell plating experiments, MEFs were seeded on feeders coated 96-wells plates, using either manual seeding or sort-based seeding with FACS ARIA III. The cells were (both Gatad2a-KO and isogenic WT) were subjected to reprogramming protocol indicated above. After 8 days, cell viability and the efficiency of seeding were then measured by mCherry positive signal (for cell lines harboring mCherry-NLS), or by positive Flag staining (e.g. for cell lines harboring Flag-Mbd3). Reprogramming efficiency was evaluated by ΔPE-Oct4-GFP or Nanog-GFP expression, or by positive staining for different pluripotency markers as indicated. Throughout the 8-day course of the reprogramming experiment the cells were not passaged to avoid any biases.

Knockdown by siRNA transfection—siRNA transfections were carried using Lipofectamine RNAiMAX (ThermoScientific #13778075), according to manufacturer protocol. Briefly, cells were seeded 24 hours before the transfection took place, in order to achieve approximately 70% confluent adherent cells. Next, both siRNA and Lipofectamine were diluted in Opti-MEM, in accordance to plate size (for a 6-wells plate well, 20pmole of siRNA were applied). Both the reagent and siRNA were mixed and vortexed, then left on RT for 5 minutes; the mixture was subsequently added to cells. Cell growth Media was changed after 24 hours, and cells for protein samples were harvested 48 hours after transfections. When siRNA transfection was a part of a reprogramming experiment, cells were retransfected every 48 hours, in order to maintain the knockdown, unless stated otherwise. Knockdown was confirmed by Western blot analysis. In Fig. 1, we tested known NuRD components for which both validated antibodies and siRNAs were available. The siRNA used are indicated in KEY RESOURCES TABLE.

Generation of lentiviruses for primary infections—For primary lentiviral infections, ~2.5x10^6 293T cells in a 10cm culture dish were transfected with JetPEI® (Polyplus) 20ul reagent for 10ug DNA as follow: pPAX (3.5 μg), pMDG (1.5 μg) and 5μg of the lentiviral target plasmid (e.g. FUW-TetO-STEMCCA). Viral supernatant was harvest 48 and 72 hours post transfection, filtered through 0.45micron sterile filters (Nalgene) and added freshly to MEF or EpiSCs.

Conversion of naïve ESC to EpiLCs or EpiSCs—Naïve ESCs were seeded on Matrigel coated plates, at 5X10^5 cells per well (6-wells plate), in naïve N2B27 2i/LIF conditions after being contained in those conditions for at least 10 days. After 24 hours, medium was changed to N2B27 supplemented with FGF2 (12ng/ml, Peprotech) and ACTIVIN A (20ng/ml, Peprotech). EpiLCs were harvested and validated at 48-96 post induction of priming. EpiSCs, which are lines stably expanded in FGF2/ACTIVIN A conditions were maintained on Matrigel coated plates and passaged with collagenase every 4-5 days as previously described (Brons et al., 2007).

Epigenetic reversion of primed EpiSCs into naïve iPSCs/ESCs—Validated Gatad2a WT, flox/flox or KO male EpiSCs were treated with ROCKi (10 Y27632) 24 hours before initiation of experiments and were harvested with TrypLE (Thermo Fisher Scientific) as single cells. Cells were either single cells sorted in Gelatin/DR4 MEF coated 96 well plates or bulk plated on 6 well Gelatin/DR4 coated plates. Cells were immediately plated in N2B27 based media with the indicated supplements used at these concentrations: human LIF (20ng/ml), CHIR99021 (CH, 3 μM or 1 μM for titrated version (t2i)- Axon Medchem), PD0325901 (PD, 1 μM - Axon Medchem), PKCi (Go6983, 2 μM – Axon Medchem), SRCi (CGP77675, 1.2 μM – Axon Medchem), ROCK inhibitor (Y27632, 2.5 μM – Axon Medchem), γ-Secretase Inhibitor (DBZ 0.35 μM – Axon Medchem). Following approximately 6 days after plating, reversion efficiency was quantified by microscope or FACS imaging of ΔPE-Oct4-GFP. Constitutive mCherry was used for normalization when found in the cell lines used.

Human fibroblasts reprogramming—4 *10^5 Human adult female dermal fibroblasts obtained from Coriell repository (GM07982) were plated on 10c”m Matrigel coated plates and used for iPSC generation with sendai virus infection (CytoTune2-iPS – ThermoFisher Scientific) according to manufacturer’s instructions. Scheme for siRNA transfection and sendai virus infection is illustrated in Fig. S5H. For the first 3 days after the sort, cells were grown in serum/LIF mES based growth media and treated with either siScramble or siGATAD2A. Later, cells were expanded in commercialized NHSM media (RseT – Stem Cell Technologies) supplemented with 10ng/ml ACTIVIN A and 2.5microM Y27632. Reprogramming was evaluated using TRA1-60 immunostaining or alkaline phosphatase staining at day 9.

DNA plasmids—DNA plasmids used in this study for ectopic expression included: Fuw-Flag-Mbd3 (Addgene 52371), FUW-Flag-Mbd3 delta1-70 (Addgene 52372), FUW-2XFlag-ΔCCR-Mbd3 (Deletion of 220-279), FUW-Gatad2a, FUW-Gatad2b, pCAG-HA-Gatad2a-

CCR (140-174 of Gatad2a), pCAG-HA-(MBD)Mbd3 (1-170 of Mbd3), pCAG-HA-GFP, FUW-Gatad2a-mutatnt ((K30R, K485R) SUMOylation-sites).

Western blot and dot blot analysis—Cells were harvested, and whole cell protein was extracted by lysis buffer, containing 150 mM NaCl, 150 mM Tris-Hcl (PH=7.4), 0.5% NP40, 1.5 mM MgCl2, 10% Glycerol. Protein's concentration was determined by BCA Kit (ThermoScientific). Blots were incubated with the different primary antibodies (diluted in 5% BSA in PBST) as indicated in KEY RESOURCES TABLE. 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 (ThermoScientific, #34080). Dot-blot method was used to identify small peptides expression. Briefly, protein lysate was loaded to a nitrocellulose membrane and was left to air-dry. Membrane was then blocked by 5% milk, and later incubated with the antibody.

Real Time (RT)-PCR analysis—Total RNA was extracted from the cells using Trizol. 1 μg of RNA was then reverse transcribed using high-capacity cDNA reverse Transcription kit (Applied Biosystems). Quantitative PCR was performed with 10ng of cDNA, in triplicates, on Viia7 platform (Applied Biosystems). Error bars indicate standard deviation of triplicate measurements for each measurement. The primers used are indicated in KEY RESOURCES TABLE.

Alkaline phosphatase (AP) staining—Alkaline phosphatase (AP) staining was performed with AP kit (Millipore SCR004) according to manufacturer protocol. Briefly, cells were fixated using 4% PFA for 2 minutes, and later washed with TBST. The reagents were then added to the wells, followed by an incubation of 10 minutes in RT. Stained plates were scanned, and positive AP+ colonies were counted to evaluate reprogramming efficiency at differential conditions.

Cell viability measurements by BrdU—Cells were treated with growth media supplemented with 10 μM BrdU for 6 hours at 37ºC. After the incubation, the growth media was removed; cells were washed with PBS and were later subjected to standard immunostaining.

Apoptosis detection by Annexin-PI stain—Annexin-PI staining was performed with 'The dead cell Apoptosis kit with Annexin V Alexa Fluor 488 and Propidium Iodide' (ThermoScientific V13245), according to the manufacturer protocol. Examined cells were stained in both Annexin and PI in order to evaluate the rate of apoptotic and dead cells in the population. The cells were then analyzed by FACS (BD FACS ARIA III) in order to detect cell mortality and apoptosis rates.

Co-immunoprecipitation (co-IP) and proteomics analysis—When used for overexpression experiments over-expression, HEK293T were transfected with our plasmids using JetPEI Transfection reagent (Polyplus), according to the manufacturer's protocol. Briefly, for a 10 cm plate containing ~2.5x106 HEK293T cells, 20ul JetPEI® reagent and 10ug DNA were diluted in 150mM NaCl, vortexed and incubated in room temperature for 20 minutes. The mixture was then applied to the cells. The cells were harvested after 48

hours, and protein interactions were examined by co-immunoprecipitation. In addition, Flag-Tagged Mbd3 was used to establish a platform for studying Mbd3-binding proteins, by correct targeting of TetO-Mbd3-Flag into the M. Col1a locus. Isogenic Gatad2a-KO were generated from this line with CRISPR/Cas9 and used for the IP and MS analysis indicated below (See Fig. S2Aiii). Flag Co-immunoprecipitation experiments were performed using Flag magnetic beads (M2, Sigma M8823). Fresh protein extracts were incubated with the magnetic beads for 4 hours in 4 C degrees, in rotation. The beads were washed from the unbound proteins, and then incubated with elution buffer. The buffer contained 150mM Tris-Hcl (pH=7.4), 1.5 mM MgCl2, 150 mM NaCl, 0.5mg\ml 3Xflag peptide (sigma, F4799). HA co-immunoprecipitation was done with the same lysis buffer, but with Pierce anti-HA magnetic beads (ThermoScientific #88836). Elution was done with HA- peptide (Sigma I2149). Untagged and endogenous proteins coimmunoprecipitation was carried using protein-G Dynabeads (ThermoScientific); beads were incubated with the lysate for 30 minutes in RT, and elution was done by heat. The elution was analyzed by Western blot analysis as detailed above or by mass-spectrometry analysis.

Mass spectrometry sample preparation and analysis—The samples were then subjected to tryptic digestion, a process constituted from reduction, alkylation and finally digestion with trypsin (16 hours, 37 °C, at 1:50 trypsin-protein ration). Digestion was stopped, and the samples were desalted using solid-phase extraction columns. The samples were then subjected to liquid chromatography coupled to high resolution, tandem mass spectrometry (LC-MS/MS). The mobile phase used was H_2O/Acetonitrile +0.1% formic acid, with the C18 column (reverse phase column). The peptide separation was performed using T3 HSS nano-column. Peptides were analyzed by a quadrupole orbitrap mass spectrometer, using flexion nanospray apparatus. Data was acquired in data dependent acquisition (DDA) mode, using a Top12 method. MS1 resolution was set to 60,000 (at 400m/z) and maximum injection time was set to 120msec. MS2 resolution was set to 17,500 and maximum injection time of 60 msec. Raw data was then imported to the Expressionist software (Genedata), and after peak detection, the peak list was searched using the Mascot algorithm as described in (Shalit et al., 2015). Data was normalized based on total ion current. Protein abundance was obtained by summing the three most intense, unique peptides per protein. Minimal cut-off for protein inclusion was ‘Ratio Sample/Control=3’, at least 2 unique peptides, and identification in at least 2 out of 3 repeats.

Cell fractioning and Chromatin Extraction—The protocol for fractioning of proteins to Nucleoplasm, Cytoplasm and chromatin fraction was previously described (Toiber et al., 2013). Briefly, cells were resuspended in lysis buffer (10 mM HEPES pH 7.4, 10 mM KCl, 0.05% NP-40) and incubated 20 minutes on ice. The lysate was then centrifuged (14,000 RPM, 10 minutes), resulting in the separation of the nucleoplasm and the cytoplasmic fractions. Next, the pellet (containing the nuclei) is re-suspended in Low-salt Buffer (10 mM Tris-HCl pH 7.4, 0.2 mM MgCl2, 1% Triton-X 100) and incubated 15 minutes on ice. Subsequent centrifuge results in the separation of Chromatin and Nucleoplasm. The pellet (containing chromatin fraction) was then resuspended in HCl 0.2N and incubated on ice, then centrifuged. Supernatant was then kept and neutralized with Tris-HCl pH=8 1M (1:1

ratio). All buffers were supplemented with supplemented with 20mM of N-Ethylmaleimide (NEM - Sigma Aldrich E3876) and protease inhibitor cocktail (Sigma-Aldrich, P8340).

SUMOylation inhibition and analysis—ES cells were treated with different concentration of Ubc9-inhibitor 2-D08 (Sigma-Aldrich SML1052) for 48 hours. Cells were then harvested and treated with NP-40 based lysis buffer, supplemented with 20mM of N-Ethylmaleimide (NEM- Sigma E3876) and protease inhibitor cocktail (Sigma P8340).

Immunostaining of cells in culture—Cells subjected to immunostaining were 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 (ThermoScientific) and imaged. All secondary antibodies were diluted 1:200. Antibodies and their dilutions used are indicated in KEY RESOURCES TABLE.

Teratoma assay—iPSCs or ESCs grown in N2B27 2i/LIF cell lines were expended for over 8 passages 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 over-night. Slides were prepared from the paraffin embedded fixed tissue, which were next Hematoxylin & Eosin stained and inspected for representation of all three germ layers and confirmed by a pathology expert.

Mouse embryo micromanipulation—Pluripotent ESCs/iPSCs were injected into BDF2 diploid blastocysts, harvested from hormone primed BDF1 6-week-old females. Microinjection into BDF2 E3.5 blastocysts placed in M2 medium under mineral oil was done by a flat-tip microinjection pipette. A controlled number of 10-12 cells were injected into the blastocyst cavity. After injection, blastocysts were returned to KSOM media (Zenith) and placed at 37°C until transferred to recipient females. Ten to fifteen injected blastocysts were transferred to each uterine horn of 2.5 days post coitum pseudo-pregnant females. All animal experiments described herein were approved by relevant Weizmann Institute IACUC (#00330111).

Microscopy live image acquisition and analysis—Isogenic secondary OKSM inducible Gatad2a-KO and Gatad2a-WT secondary MEFs harboring ΔPE-Oct4-GFP pluripotency reporter and constitutively expressed nuclear mCherry marker (Fig. S2A), were plated in 24-well plates at low densities (~120 cells per well). Reprogramming was than imaged (beginning 72 hours after DOX administration) using AxioObserver Z1 (Zeiss) in %5 O2, %5 CO2, 37C controlled conditions. Plates were taken out at day 3.5 for media replacement (without passaging/splitting) and put back for the automated live imaging stage. Full well mosaic images were taken every 12 hours for 4 days (days 3-7 of reprogramming) at 5x magnification, including phase contrast and two fluorescent wavelength images. Notably, FACS analysis validated that the intensities are identical between KO and control isogenic PSC lines for both GFP and mCherry signals in the parental PSC lines. The latter is

expected since genetic deletion of Gatad2a was always done after all reporters were introduced, making the compared cells entirely isogenic and suitable for comparison.

PolyA-RNA-seq library preparation—Total RNA was isolated from indicated cell lines and extracted from Trizol pellets by Direct-zol RNA MiniPrep kit (Zymo), then utilized for RNA-Seq by TruSeq RNA Sample Preparation Kit v2 (Illumina) according to manufacturer's instruction.

ATAC-seq library preparation—Cells were trypsinized and counted, 50,000 cells were centrifuged at 500*g* for 3 min, followed by a wash using 50 μl of cold PBS and centrifugation at 500*g* for 3 min. 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 min using a refrigerated centrifuge. Next, the pellet was resuspended 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 min at 37 °C and immediately put on ice. Directly afterwards, the sample was purified using a Qiagen MinElute kit. Following purification, the library fragments were amplified using custom Nextera PCR primers 1 and 2 for a total of 12 cycles. Following PCR amplification, the libraries were purified using a QiagenMinElute Kit and sequenced.

Whole-Genome Bisulfite Sequencing (WGBS) Library preparation—DNA was isolated from snap-frozen cells using the Quick-gDNA mini prep kit (Zymo). DNA was then converted by bisulfite using the EZ DNA Methylation-Gold kit (Zymo). Sequencing libraries were created using the EpiGenome Methyl-Seq (Epicenter) and sequenced.

ChIP-seq library preparation—Cells were crosslinked in formaldehyde (1% final concentration, 10 min at room temperature), and then quenched with glycine (5 min at room temperature). Fixated cells were then lysed in 50 mM HEPES KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 alternative, 0.25% Triton supplemented with protease inhibitor at 4 °C (Roche, 04693159001), centrifuged at 950*g* for 10 min and re-suspended in 0.2% SDS, 10 mM EDTA, 140 mM NaCl and 10 mM Tris-HCl. Cells were then fragmented with a Branson Sonifier (model S-450D) at -4 °C to size ranges between 200 and 800 bp and precipitated by centrifugation. Antibody was prebound by incubating with Protein-G Dynabeads (Invitrogen 100-07D) in blocking buffer (PBS supplemented with 0.5% TWEEN and 0.5% BSA) for 1 h at room temperature. Washed beads were added to the chromatin lysate for incubation as detailed in the table below. Samples were washed five times with RIPA buffer, twice with RIPA buffer supplemented with 500 mM NaCl, twice with LiCl buffer (10 mM TE, 250mM LiCl, 0.5% NP-40, 0.5% DOC), once with TE (10Mm Tris-HCl pH 8.0, 1mM EDTA), and then eluted in 0.5% SDS, 300 mM NaCl, 5 mM EDTA, 10 mM Tris HCl pH 8.0. Eluate was incubated treated sequentially with RNaseA (Roche, 11119915001) for 30 min and proteinase K (NEB, P8102S) for 2 h in 65 °C for 8 h, and then. 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 (Rais et al., 2013).

Quantification and Statistical Analysis

PolyA-RNA analysis—RNA-seq data are deposited under GEO no. GSE102518. Tophat software version 2.0.10 was used to align reads to mouse mm10 reference genome (UCSC, December 2011). FPKM values were calculated over all genes in mm10 assembly GTF (UCSC, December 2011), using cufflinks (version 2.2.1). Genes with minimal expressed (FPKM>1) in at least one time point in each of the systems (Gatad2a-KO/WT/WT*) were selected for clustering and PCA, resulting in 9,756 active genes. Unit normalized FPKM was calculated using the following formula $R^i_j{}^* = R^i_j / [\max_j(R^i_j)+1]$ where j is the sample index, i is the gene index and FPKM=1 is the transcription noise threshold, and $\max_j(R^i_j)$ is the maximal level in each dataset. Hierarchical clustering was carried out using R pheatmap command. Only isogenic sample sets were included in the PCA analysis shown in Fig. 3B. PCA analysis was carried out over the same set of 9,756 genes in unit normalization by R prcomp command.

ATAC-seq 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 (≤ 100bp) that correspond to nucleosome free region. 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. To develop this blacklist, we generated 10,000,000 synthetic 34mer reads derived from the mitochondrial genome. After mapping and peak calling of these synthetic reads we found 28 high-signal peaks for the mm10 genome. For all subsequent analysis, we discarded peaks falling within these regions.

Enhancer Identification—H3K27ac peaks were detected using MACS version 1.4.2-1 and merged for all time points using bedtools merge command. All ATAC peaks were filtered to include only peaks which co-localized with the merged H3K27ac peaks, meaning only ATAC peaks that have H3K27ac mark on at least one of the time points were passed to further processing. Finally, the peaks from all samples were unified and merged (using bedtools unionbedg and merge commands), further filtered to reject peaks that co-localized with promoter or exon regions based on mm10 assembly (UCSC, December 2011). Finally, we were left with 52,473 genomic intervals which we annotated as active enhancers.

Enhancers were considered as differential if both their ATAC-seq and H3K27ac signals show significant change during reprogramming (min z-score<0.5, max z-score>1.5, for both chromatin marks). 12,153 enhancers were found to be differential and were used for correlation analysis (Fig. 3C). Z-scores were calculated as following: Shortly, the genomic interval is divided to 50bp size bins, and the coverage in each bin is estimated. Each bin is then converted to z-score by normalizing each position by the mean and standard deviation of the sample noise ($\hat{X}j=(Xj\mu_{noise})/\sigma_{noise}$). Noise parameters were estimated for each sample from $6*10^7$ random bp across the genome. Finally, the 3rd highest bin z-score of

Europe PMC Funders Author Manuscripts

each interval is set to represent the coverage of that interval. ATAC-seq data are deposited under GEO no. GSE102518.

Whole-Genome Bisulfite Sequencing (WGBS) analysis—The sequencing reads were aligned to the mouse mm10 reference genome (UCSC, December 2011), using a proprietary script based on Bowtie2. In cases where the two reads were not aligned in a concordant manner, the reads were discarded. Methylation levels of CpGs calculated by RRBS and WGBS were unified. Mean methylation was calculated for each CpG that was covered by at least 5 distinct reads (X5). Average methylation level in various genomic intervals was calculating by taking the average over all covered X5 covered CpG sites in that interval. WGBS data are available to download from NCBI GEO, under super-series GSE102518.

ChIP-seq analysis—For alignment and peak detection, we used bowtie2 software to align reads to mouse mm10 reference genome (UCSC, December 2011), with default parameters. We identified enriched intervals of all measured proteins using MACS version 1.4.2-1. We used sequencing of whole-cell extract as control to define a background model. Duplicate reads aligned to the exact same location are excluded by MACS default configuration.

Chromatin modification profile estimation in TSS, TES and in enhancers: H3K27ac modification coverage in the genomic intervals was calculated using in-house script. Shortly, the genomic interval is divided to 50bp size bins, and the coverage in each bin is estimated. Each bin is then converted to z-score by normalizing by the mean and standard deviation of the sample noise ($X\hat{}j=(Xj-\mu_{noise})/\sigma_{noise}$). Noise parameters were estimated for each sample from $6*10^7$ random bp across the genome. Finally, the 3rd highest bin z-score of each interval is set to represent the coverage of that interval.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

J.H.H is supported by a generous gift from Ilana and Pascal Mantoux, and research grants from the: ERC-consolidator, Flight Attendant Medical Research Council (FAMRI), Israel Science Foundation (ISF-ICORE, ISF-NFSC, ISF-INCPM & ISF-Morasha (also to N.N.)), Kamin-Yeda, Minerva, Israel Cancer Research Fund (ICRF), Human Frontiers Science Program (HFSP), Benoziyo Endowment fund, New York Stem Cell Foundation (NYSCF), Helen and Martin Kimmel Institute for Stem Cell Research, Keckst Center. J.H.H. is a New York Stem Cell Foundation (NYSCF)–Robertson Investigator.

Data and Software Availability

All RNA-seq, ATAC-seq, ChIP-seq and WGBS methylation data are available to download from NCBI GEO, under super-series GSE102518. Link to related raw datasets: https://data.mendeley.com/datasets/vyjpgt3cc8/draft?a=093a2a8c-4a24-4e55-a621-6f82c34f2e2d

References

Alqarni SSM, Murthy A, Zhang W, Przewloka MR, Silva APG, Watson AA, Lejon S, Pei XY, Smits AH, Kloet SL, et al. Insight into the architecture of the NuRD complex: Structure of the RbAp48-MTA1 subcomplex. J Biol Chem. 2014; 289:21844–21855. [PubMed: 24920672]

Bao S, Tang F, Li X, Hayashi K, Gillich A, Lao K, Surani MA. Epigenetic reversion of post-implantation epiblast to pluripotent embryonic stem cells. Nature. 2009; 461:1292–1295. [PubMed: 19816418]

Becker JS, McCarthy RL, Sidoli S, Donahue G, Kaeding KE, He Z, Lin S, Garcia BA, Zaret KS. Genomic and Proteomic Resolution of Heterochromatin and Its Restriction of Alternate Fate Genes. Mol Cell. 2017

Bode D, Yu L, Tate P, Pardo M, Choudhary J. Characterization of Two Distinct Nucleosome Remodeling and Deacetylase (NuRD) Complex Assemblies in Embryonic Stem Cells. Mol Cell Proteomics. 2016; 15:878–891. [PubMed: 26714524]

Boiani, M, Kehler, J, Schöler, HR. Activity of the Germline-Specific Oct4-GFP Transgene in Normal and Clone Mouse EmbryosMethods in Molecular Biology-…. New Jersey: Humana Press; 2004. 1–34.

Borkent M, Bennett BD, Lackford B, Bar-Nur O, Brumbaugh J, Wang L, Du Y, Fargo DC, Apostolou E, Cheloufi S, et al. A Serial shRNA Screen for Roadblocks to Reprogramming Identifies the Protein Modifier SUMO2. Stem Cell Reports. 2016; 6:704–716. [PubMed: 26947976]

Brons I, Clarkson A, Ahrlund-Richter L, Pedersen RA. Derivation of pluripotent epiblast stem cells from mammalian embryos. Nature. 2007

Cheloufi S, Elling U, Hopfgartner B, Jung YL, Murn J, Ninova M, Hubmann M, Badeaux AI, Ang CE, Tenen D, et al. The histone chaperone CAF-1 safeguards somatic cell identity. Nature. 2015; 528:218–224. [PubMed: 26659182]

Dutta D, Ray S, Home P, Larson M, Wolfe MW, Paul S. Self-Renewal Versus Lineage Commitment of Embryonic Stem Cells: Protein Kinase C Signaling Shifts the Balance. Stem Cells. 2011; 29:618–628. [PubMed: 21308862]

Gafni O, Weinberger L, Mansour AA, Manor YS, Chomsky E, Ben-Yosef D, Kalma Y, Viukov S, Maza I, Zviran A, et al. Derivation of novel human ground state naive pluripotent stem cells. Nature. 2013; 504:282–286. [PubMed: 24172903]

Gnanapragasam MN, Scarsdale JN, Amaya ML, Webb HD, Desai MA, Walavalkar NM, Wang SZ, Zu Zhu S, Ginder GD, Williams DC. p66-MBD2 coiled-coil interaction and recruitment of Mi-2 are critical for globin gene silencing by the MBD2-NuRD complex. Proc Natl Acad Sci. 2011; 108:7487–7492. [PubMed: 21490301]

Gong Z, Brackertz M, Renkawitz R. SUMO Modification Enhances p66-Mediated Transcriptional Repression of the Mi-2/NuRD Complex. Mol Cell Biol. 2006; 26:4519–4528. [PubMed: 16738318]

Le Guezennec X, Vermeulen M, Brinkman AB, Hoeijmakers WAM, Cohen A, Lasonder E, Stunnenberg HG. MBD2/NuRD and MBD3/NuRD, Two Distinct Complexes with Different Biochemical and Functional Properties. Mol Cell Biol. 2006; 26:843–851. [PubMed: 16428440]

Hanna J, Markoulaki S, Schorderet P, Carey BW, Beard C, Wernig M, Creyghton MPP, Steine EJ, Cassady JP, Foreman R, et al. Direct Reprogramming of Terminally Differentiated Mature B Lymphocytes to Pluripotency. Cell. 2008; 133:250–264. [PubMed: 18423197]

Hanna J, Saha K, Pando B, van Zon J, Lengner CJ, Creyghton MP, van Oudenaarden A, Jaenisch R. Direct cell reprogramming is a stochastic process amendable to acceleration. Nature. 2009; 462:595–601. [PubMed: 19898493]

Kaji K, Caballero IM, MacLeod R, Nichols J, Wilson VA, Hendrich B. The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat Cell Biol. 2006; 8:285–292. [PubMed: 16462733]

Lujan E, Zunder ER, Ng YH, Goronzy IN, Nolan GP, Wernig M. Early reprogramming regulators identified by prospective isolation and mass cytometry. Nature. 2015; 521:352–356. [PubMed: 25830878]

Luo M, Ling T, Xie W, Sun H, Zhou Y, Zhu Q, Shen M, Zong L, Lyu G, Zhao Y, et al. NuRD blocks reprogramming of mouse somatic cells into Pluripotent stem cells. Stem Cells. 2013; 31:1278–1286. [PubMed: 23533168]

Marino S, Nusse R. Mutants in the Mouse NuRD/Mi2 Component P66α Are Embryonic Lethal. PLoS One. 2007; 2

Masika H, Farago M, Hecht M, Condiotti R, Makedonski K, Buganim Y, Burstyn-Cohen T, Bergman Y, Cedar H. Programming asynchronous replication in stem cells. Nat Struct Mol Biol. 2017; 24:1132–1138. [PubMed: 29131141]

Nitarska J, Smith JG, Sherlock WT, Hillege MMG, Nott A, Barshop WD, Vashisht AA, Wohlschlegel JA, Mitter R, Riccio A. A Functional Switch of NuRD Chromatin Remodeling Complex Subunits Regulates Mouse Cortical Development. Cell Rep. 2016; 17:1683–1698. [PubMed: 27806305]

O'Malley J, Skylaki S, Iwabuchi KA, Chantzoura E, Ruetz T, Johnsson A, Tomlinson SR, Linnarsson S, Kaji K. High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature. 2013; 499:88–91. [PubMed: 23728301]

O'Shaughnessy-Kirwan A, Signolet J, Costello I, Gharbi S, Hendrich B. Constraint of gene expression by the chromatin remodelling protein CHD4 facilitates lineage specification. Development. 2015; 142:2586–2597. [PubMed: 26116663]

Rais Y, Zviran A, Geula S, Gafni O, Chomsky E, Viukov S, Mansour AA, Caspi I, Krupalnik V, Zerbib M, et al. Deterministic direct reprogramming of somatic cells to pluripotency. Nature. 2013; 502:65–70. [PubMed: 24048479]

Reynolds N, Latos P, Hynes-Allen A, Loos R, Leaford D, O'Shaughnessy A, Mosaku O, Signolet J, Brennecke P, Kalkan T, et al. NuRD suppresses pluripotency gene expression to promote transcriptional heterogeneity and lineage commitment. Cell Stem Cell. 2012; 10:583–594. [PubMed: 22560079]

Sakai H, Urano T, Ookata K, Kim MH, Hirai Y, Saito M, Nojima Y, Ishikawa F. MBD3 and HDAC1, two components of the NuRD complex, are localized at aurora-A-positive centrosomes in M phase. J Biol Chem. 2002; 277:48714–48723. [PubMed: 12354758]

Dos Santos RL, Tosti L, Radzisheuskaya A, Caballero IM, Kaji K, Hendrich B, Silva JCR. MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. Cell Stem Cell. 2014; 15:102–110. [PubMed: 24835571]

Shalit T, Elinger D, Savidor A, Gabashvili A, Levin Y. MS1-based label-free proteomics using a quadrupole orbitrap mass spectrometer. J Proteome Res. 2015; 14:1979–1986. [PubMed: 25780947]

Shimizu T, Ueda J, Ho JC, Iwasaki K, Poellinger L, Harada I, Sawada Y. Dual Inhibition of Src and GSK3 Maintains Mouse Embryonic Stem Cells, Whose Differentiation Is Mechanically Regulated by Src Signaling. Stem Cells. 2012; 30:1394–1404. [PubMed: 22553165]

Smith ZD, Sindhu C, Meissner A. Molecular features of cellular reprogramming and development. Nat Rev Mol Cell Biol. 2016; 17:139–154. [PubMed: 26883001]

Smits AH, Jansen PWTC, Poser I, Hyman AA, Vermeulen M. Stoichiometry of chromatin-associated protein complexes revealed by label-free quantitative mass spectrometry-based proteomics. Nucleic Acids Res. 2013; 41

Sommer CA, Stadtfeld M, Murphy GJ, Hochedlinger K, Kotton DN, Mostoslavsky G. Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells. 2009; 27:543–549. [PubMed: 19096035]

Spruijt CG, Luijsterburg MS, Menafra R, Lindeboom RGH, Jansen PWTC, Edupuganti RR, Baltissen MP, Wiegant WW, Voelker-Albert MC, Matarese F, et al. ZMYND8 Co-localizes with NuRD on Target Genes and Regulates Poly(ADP-Ribose)-Dependent Recruitment of GATAD2A/NuRD to Sites of DNA Damage. Cell Rep. 2016; 17:783–798. [PubMed: 27732854]

Stadtfeld M, Maherali N, Borkent M, Hochedlinger K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. Nat Methods. 2010; 7:53–55. [PubMed: 20010832]

Di Stefano B, Sardina JL, Van Oevelen C, Collombet S, Kallin EM, Vicent GP, Lu J, Thieffry D, Beato M, Graf T. C/EBPa poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature. 2013; 506:1–17.

Takahashi K, Yamanaka S. A decade of transcription factor-mediated reprogramming to pluripotency. Nat Rev Mol Cell Biol. 2016; 17:183–193. [PubMed: 26883003]

Toiber D, Erdel F, Bouazoune K, Silberman DM, Zhong L, Mulligan P, Sebastian C, Cosentino C, Martinez-Pastor B, Giacosa S, et al. SIRT6 recruits SNF2H to DNA break sites, preventing genomic instability through chromatin remodeling. Mol Cell. 2013; 51:454–468. [PubMed: 23911928]

Weinberger L, Ayyash M, Novershtern N, Hanna JH. Dynamic stem cell states: Naive to primed pluripotency in rodents and humans. Nat Rev Mol Cell Biol. 2016; 17:155–169. [PubMed: 26860365]

Ying Q-L, Wray J, Nichols J, Batlle-Morera L, Doble B, Woodgett J, Cohen P, Smith A. The ground state of embryonic stem cell self-renewal. 2008; 453:519–523.

Yoshimizu T, Sugiyama N, De Felice M, Yeom Y, Ohbo K, Masuko K, Obinata M, Kuniya A, Schöler HR, Matsui Y. Germline-specific expression of the Oct-4/green fluorescent protein (GFP) transgene in mice. Dev Growth Differ. 1999; 41:675–684. [PubMed: 10646797]

Zhuang Q, Li W, Benda C, Huang Z, Liu P, Guo X, Luo Z, Yang Z, Yang J, Huang Y, et al. NCoR / SMRT co-repressors cooperate with c-MYC to create an epigenetic barrier to somatic cell reprogramming. Nat Cell Biol Biol (AOP). 2018

Europe PMC Funders Author Manuscripts

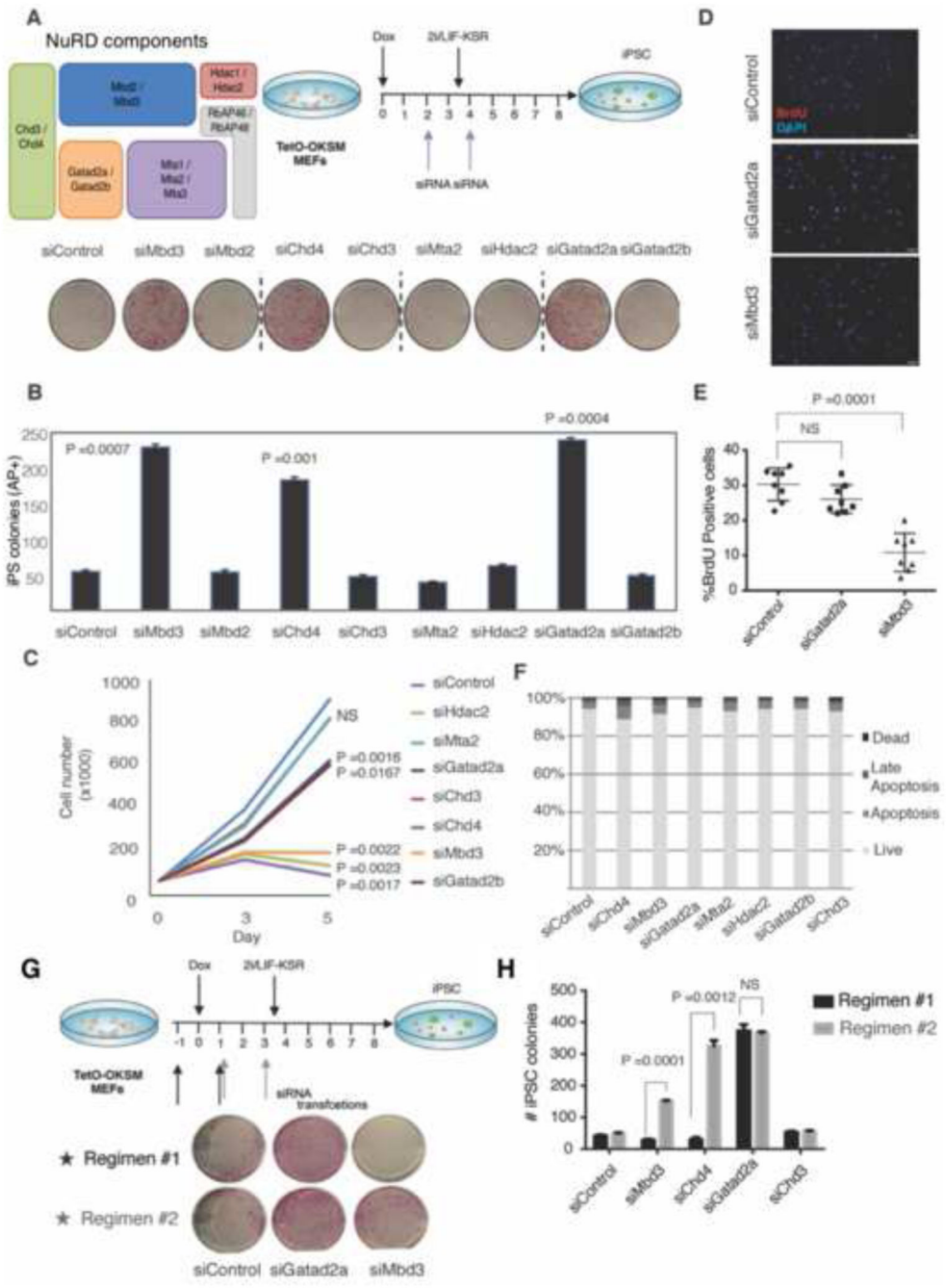

Figure 1. Gatad2a depletion increases iPSC efficiency and does not ablate somatic cell proliferation.

A. MEFs harboring TetO-OKSM and M2rtTA cassettes were transfected with siRNA targeting different canonical NuRD components (indicated in the illustration), 2 and 4 days after reprogramming initiation following DOX administration. Reprogramming was then evaluated by AP staining at day 8. **B.** Reprograming efficiency following siRNA treatments was evaluated using AP staining, after 8 days of reprogramming (n=3, two-sided Student's t-test *p* values are indicated). **C.** Cell growth curves of MEFs treated with siRNA for the

indicated NuRD components (two-sided Student's t-test *p* values are indicated). **D.** Representative images of cells treated with siRNA targeting Mbd3 or Gatad2a and exposed to BrdU in order to evaluate proliferation. **E.** Quantitative evaluation of BrdU incorporation test (n=8, two-sided Student's ttest *p* values are indicated Student's t-test). **F**. Viability and apoptosis induction were measured using FACS following Annexin-PI staining. **G.** Reprogramming efficiency following siRNA treatments targeting different NuRD components, at different time points. KD was performed at two distinct cycles: the early one (Regimen 1, marked in black) started one day prior to DOX induction, and the second one (Regimen 2, marked in grey) started one day post-DOX induction. **H.** iPSC reprogramming efficiency following different siRNA treatments. was evaluated at day 8. (n=3 per each condition, two-sided Student's t-test *p* values are indicated).

Europe PMC Funders Author Manuscripts

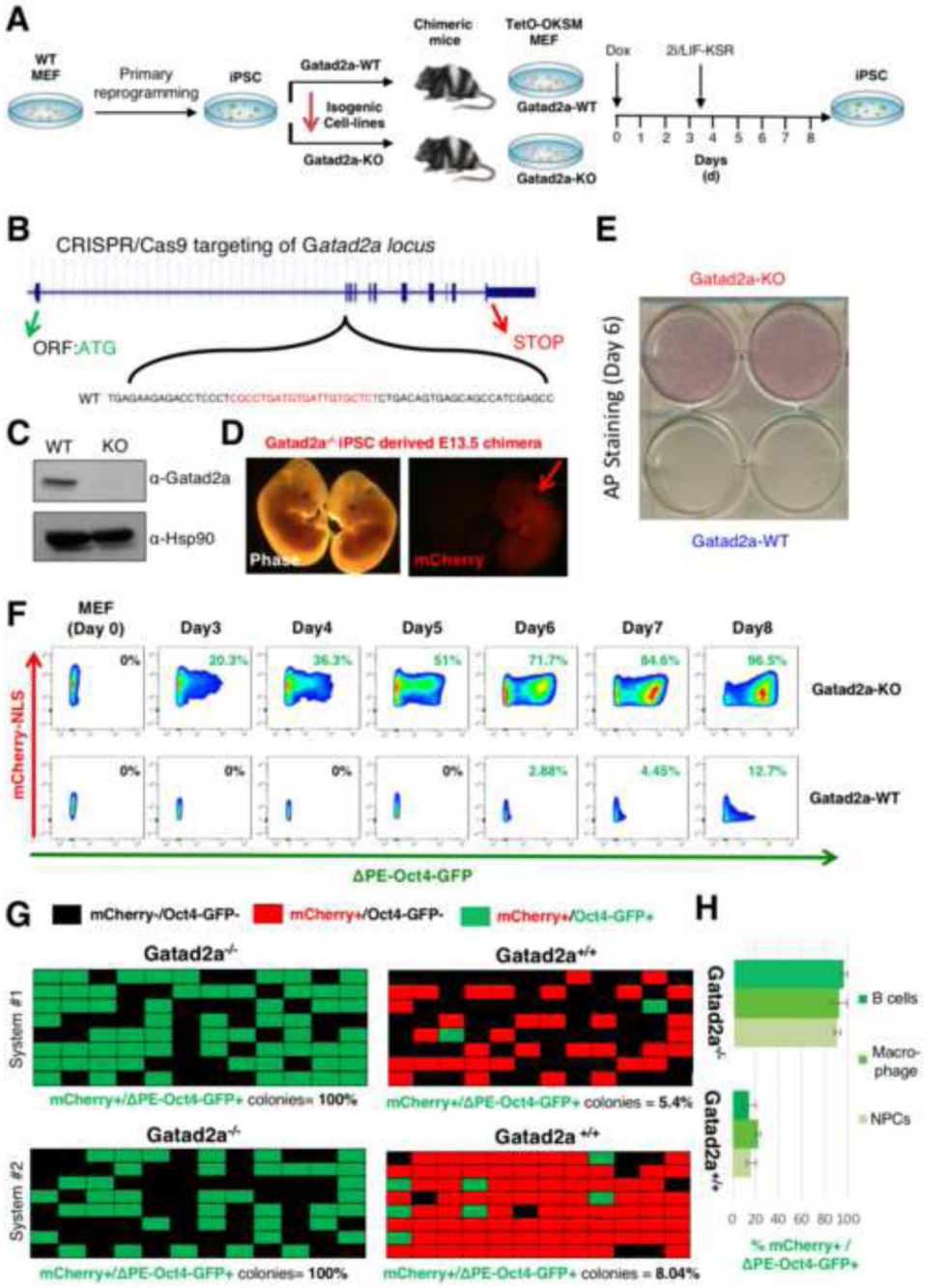

Figure 2. Deterministic induction of naïve iPSCs in Gatad2a-KO somatic cells.

A. Scheme demonstrating strategy for generating secondary isogenic Gatad2a WT and KO lines, and comparing their reprogramming efficiency side by side. 2i/LIF-KSR conditions were introduced at day 3.5 during the 8-day course. More detailed information is provided in Fig. S2A. **B**. Targeting scheme of Gatad2a locus. **C**. Western blot validation of Gatad2a levels. **D**. Representative images of Gatad2a$^{-/-}$ iPSC derived E13.5 chimera. Red arrow highlights mCherry+ chimera which originates from mCherry labeled iPSCs that were microinjected. **E.** Bulk iPSC reprogramming as in **F**, but experiment was terminated after 6

days and iPSC colony formation was evaluated by Alkaline Phosphatase staining (AP+). **F**. Representative flow cytometry measurements of ΔPE-Oct4-GFP reactivation dynamics in polyclonal/bulk Gatad2a-WT and Gatad2a-KO isogenic cell lines. **G.** Representative summaries of single-cell iPSC reprogramming efficiency experiment. Secondary isogenic Gatad2a WT and KO reprogrammable MEFs carrying constitutively expressed mCherry-NLS and naïve pluripotency specific ΔPE-Oct4-GFP reporter were sorted and seeded as single-cell per well. Reprogramming was initiated by DOX administration according to panel **A**. Reprogramming efficiency was assessed after 8 days based on the number of wells in which mCherry+ cells formed an ΔPE-Oct4-GFP positive colony. **H.** The indicated secondary Gatad2a$^{+/+}$ and Gatad2a$^{-/-}$ somatic cell types were isolated and subjected reprogramming and evaluation of iPSC efficiency following 8 days of DOX.

Europe PMC Funders Author Manuscripts

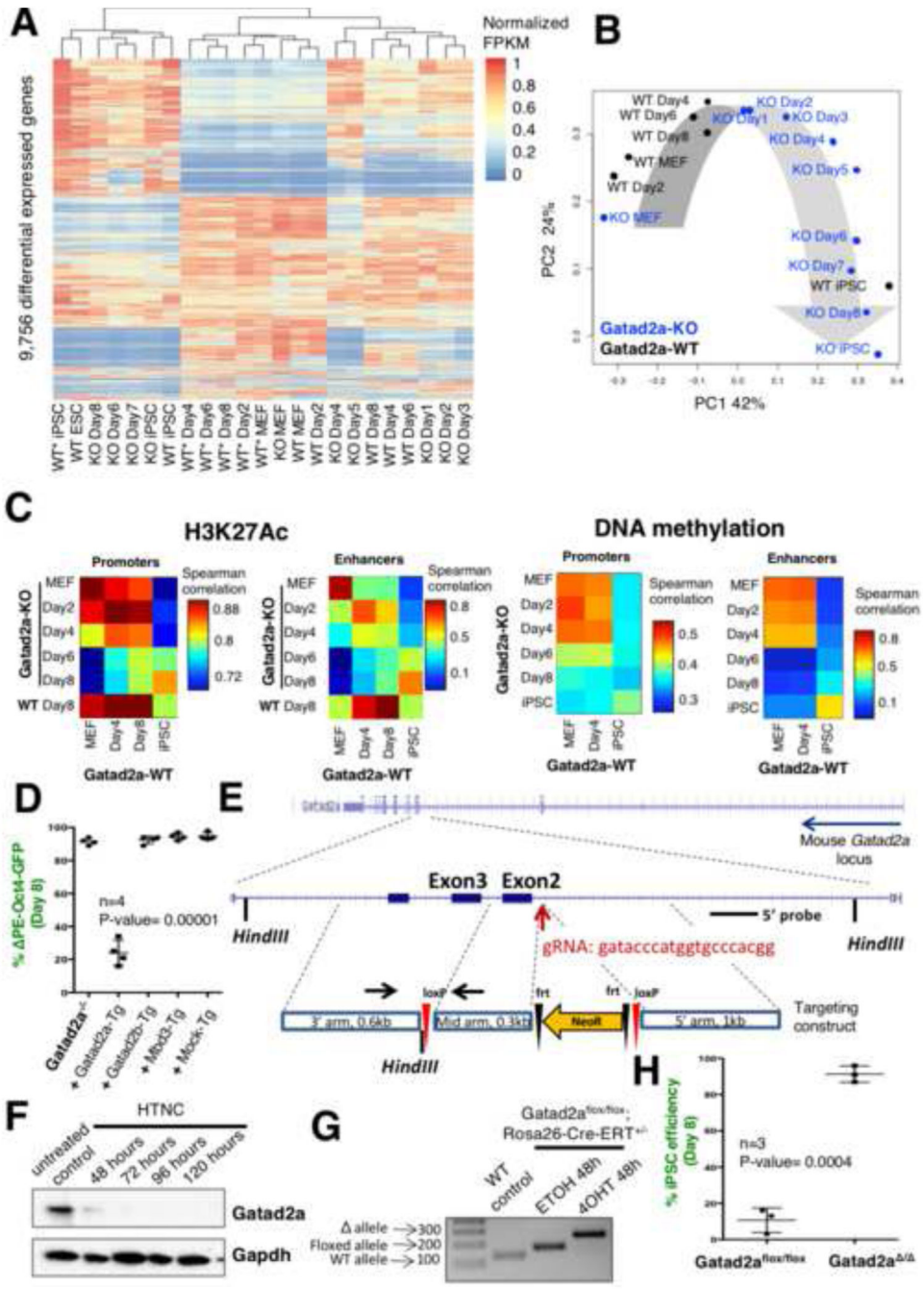

Figure 3. Naïve pluripotency establishment by OKSM following Gatad2a depletion.

A. Hierarchical clustering of transcription profiles of samples from Gatad2a$^{-/-}$ system and two WT systems, where WT are isogenic to Gatad2a-/- and WT* are non-isogenic. Values are unitnormalized FPKM. **B.** PCA of Gatad2a$^{-/-}$ and isogenic Gatad2a$^{+/+}$ WT system, showing the transformation from MEF to iPSC, and that MEF and iPSC are similar in both systems. PCA was calculated over the same set of genes shown in (**A**). **C**. Correlation matrices between Gatad2a$^{-/-}$and WT cells, based on H3K27ac signal (left), and DNA-methylation signal (right). Correlation of H3K27ac signal was calculated over promoters of

Europe PMC Funders Author Manuscripts

all differential genes (n=14,033), or over differential enhancers (n=12,153). Correlation pf DNA methylation signal was calculated over promoters of differential genes that are covered by the sequencing method (n=9,134), or over covered differential enhancers (n=11,413). **D**. Reprogramming efficiency of Gatad2a$^{-/-}$ secondary cells was evaluated before and after introducing different exogenous transgenes (Tg), including Gatad2a-Tg. **E.** Targeting strategy for generating Gatad2a conditional knockout reprogrammable system and PSCs. **F.** Western-blot time-course based validation for Gatad2a protein depletion in Gatad2a$^{flox/flox}$ MEFs following Tat-Cre treatment (HTNC). **G.** Complementing PCR based validation of Gatad2a floxed allele deletion following 4OHT treatment (the cells also harbor a correctly targeted Rosa26-CreERT allele). **H.** Gatad2a$^{fl/fl}$ reprogrammable MEF were derived and underwent Cre induction for 4 days, and only afterwards DOX reprogramming was initiated for 8 days as described in Fig. 2A, E (n=3 per conditions, student's t-test *p* value is indicated).

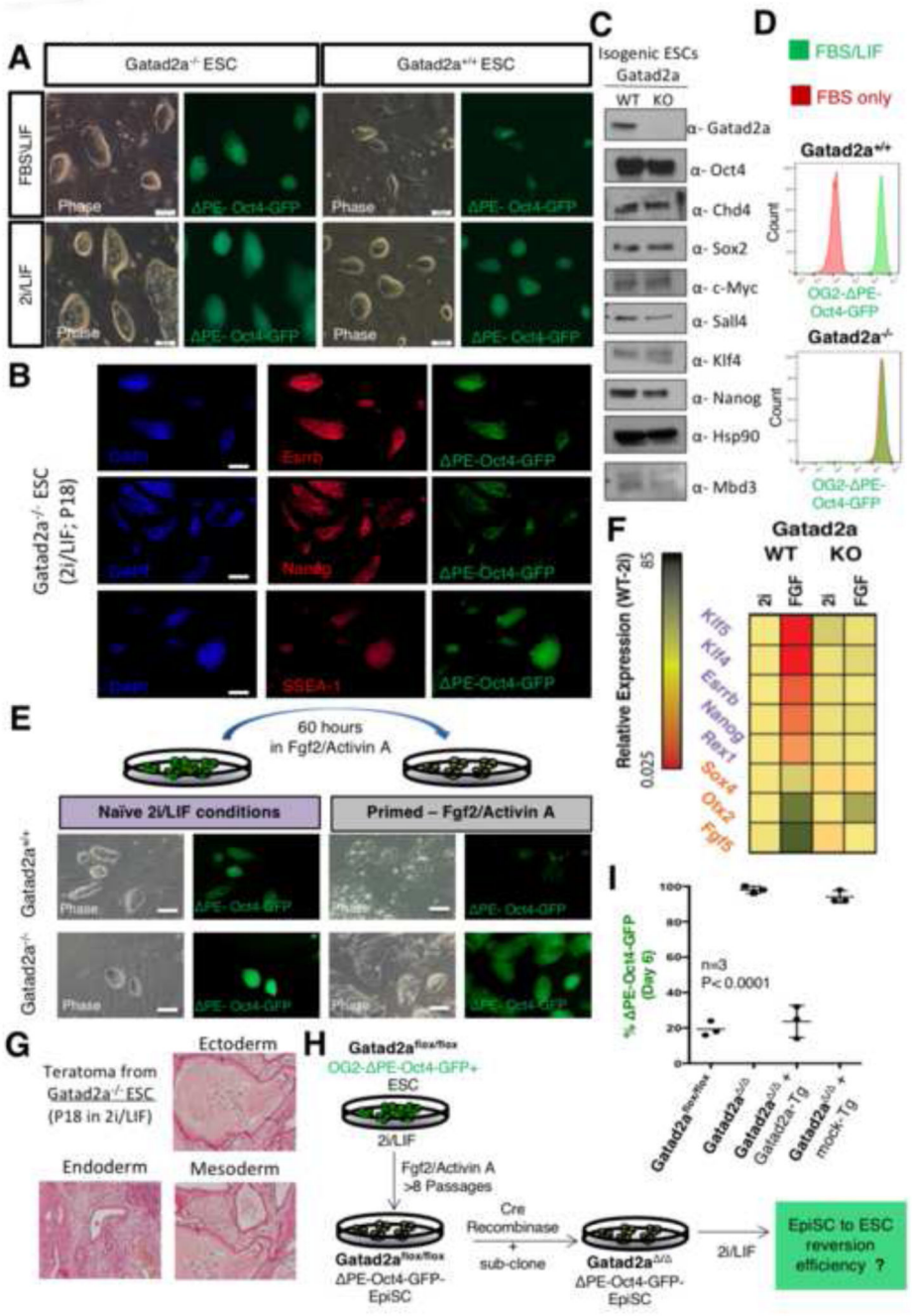

Figure 4. Gatad2a depletion enhances naïve pluripotency maintenance and induction from EpiSCs.

A. Representative images of isogenic Gatad2a$^{+/+}$; OG2-ΔPE-Oct4-GFP and Gatad2a$^{-/-}$OG2-ΔPE-Oct4-GFP ESCs FBS/LIF and 2i/LIF conditions. (scale bar=100μM). **B**. Representative immunostaining of Gatad2a-KO ES. (Scale bar=100μM). **C**. Western blot comparing pluripotency proteins' level between Gatad2a-WT and KO isogenic ES lines. **D.** Isogenic Gatad2a$^{+/+}$ and Gatad2a$^{-/-}$ ESCs were maintained on Gelatin coated plates in FBS/LIS or FBS only conditions for 5 passages and then subjected to FACS analysis for OG2-ΔPE-

Oct4-GFP levels. **E**. Priming of Gatad2a$^{+/+}$ and Gatad2a$^{-/-}$ naïve ESCs harboring OG2-ΔPE-Oct4-GFP reporter by changing conditions from 2i/LIF to Fgf2/Activin was performed. **F.** Transcriptional expression of different pluripotency and differentiation markers before and after Fgf2/Activin exposure, in isogenic Gatad2a$^{+/+}$ and Gatad2a$^{-/-}$ ESCs, presented as a relative expression column scheme. **G.** Gatad2a-KO ESCs generate mature teratomas. **H.** Strategy for generating isogenic Gatad2a$^{flox/flox}$ ΔPE-Oct4-GFP was established from parental cells expanded for 8 passages in FGF2/Activin and was validated for priming by FACS and RT-PCR analysis. Established Gatad2a$^{flox/flox}$; OG2-ΔPE-Oct4-GFP was treated with Tat-CRE, and sub cloned isogenic Gatad2a$^{\Delta/\Delta}$ EpiSC lines were derived and used for analysis within additional 5 passages of their sub cloning. **I**. Reprogramming efficiency and quantification for EpiSC reprogramming from different mutant EpiSC lines. Rescue Gatad2a-Tg or control Mock-Tg were ectopically expressed in Gatad2a$^{\Delta/\Delta}$ EpiSC and included in the analysis. (n=3 per time point, p Value<0.0001, two-sided Student's t-test).

Europe PMC Funders Author Manuscripts

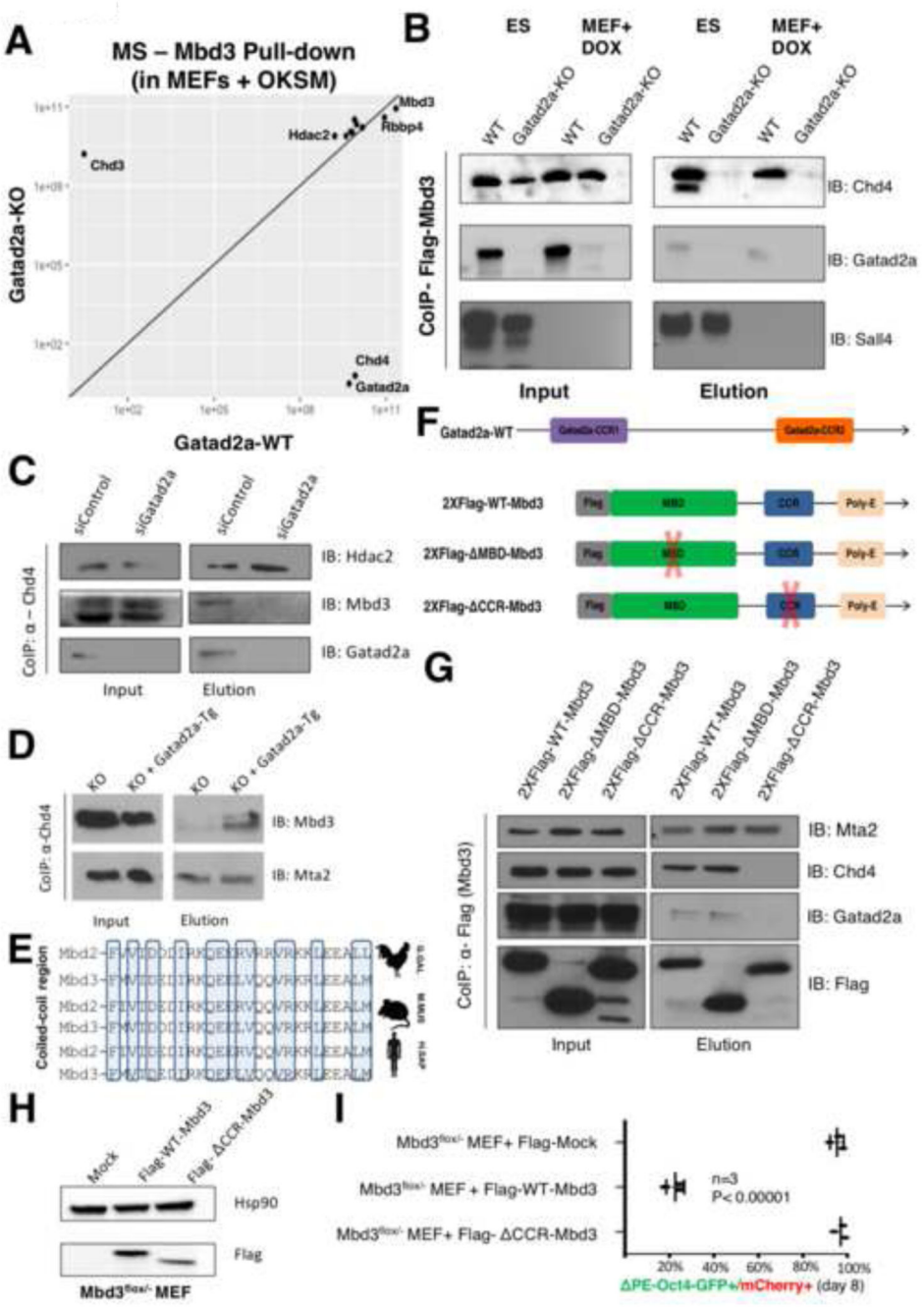

Figure 5. Gatad2a, Mbd3 and Chd4 constitute a critical axis within Mbd3/NuRD complex mediating reprogramming inhibition.

A. Mass spectrometry analysis of NuRD complex components binding efficiency to Mbd3 in Gatad2a-KO and its isogenic Gatad2a-WT line, in MEF cells during reprogramming by OKSM. Only putative NuRD components are presented. **B.** Flag-Mbd3 CoIP in Gatad2a-WT and Gatad2a-KO cells. Experiments were conducted both in MEF cells and ESs. **C.** Cells during reprogramming were treated with siRNA targeting Gatad2a, and pellets collected after four days. CoIP of Chd4 shows that siGatad2a prevents Mbd3 binding to

Europe PMC Funders Author Manuscripts

Gatad2a but also to Chd4. **D.** Gatad2a-KO MEF with overexpression of Gatad2a (transgenic recovery; abbreviated as Gatad2a-Tg) or Mock were subjected to Chd4-CoIP. **E.** The coiled coil region of Mbd3 is highly conserved between different organisms, and different proteins in the MBD family. The highlighted amino acids are crucial for Gatad2a binding to Mbd2 or Mbd3. **F.** A scheme of Flag tagged Mbd3, and mutant forms of Mbd3: lacking the Coiled coil region (ΔCCR-Mbd3) or the methyl-binding domain (ΔMBD-Mbd3). **G.** WT-Mbd3 and both mutants were over-expressed in 293T cells and were subjected to Flag-Mbd3 CoIP to examine their protein interactions. **H-I.** $Mbd3^{fl/-}$ secondary MEF, harboring ΔPE-Oct4-GFP reporter, were transfected with two different forms of Flag tagged WT-Mbd3 and ΔCCR-Mbd3. The cells were then subjected to reprogramming. (two-sided Student's t-test, n=3. *p* Value<0.0001).

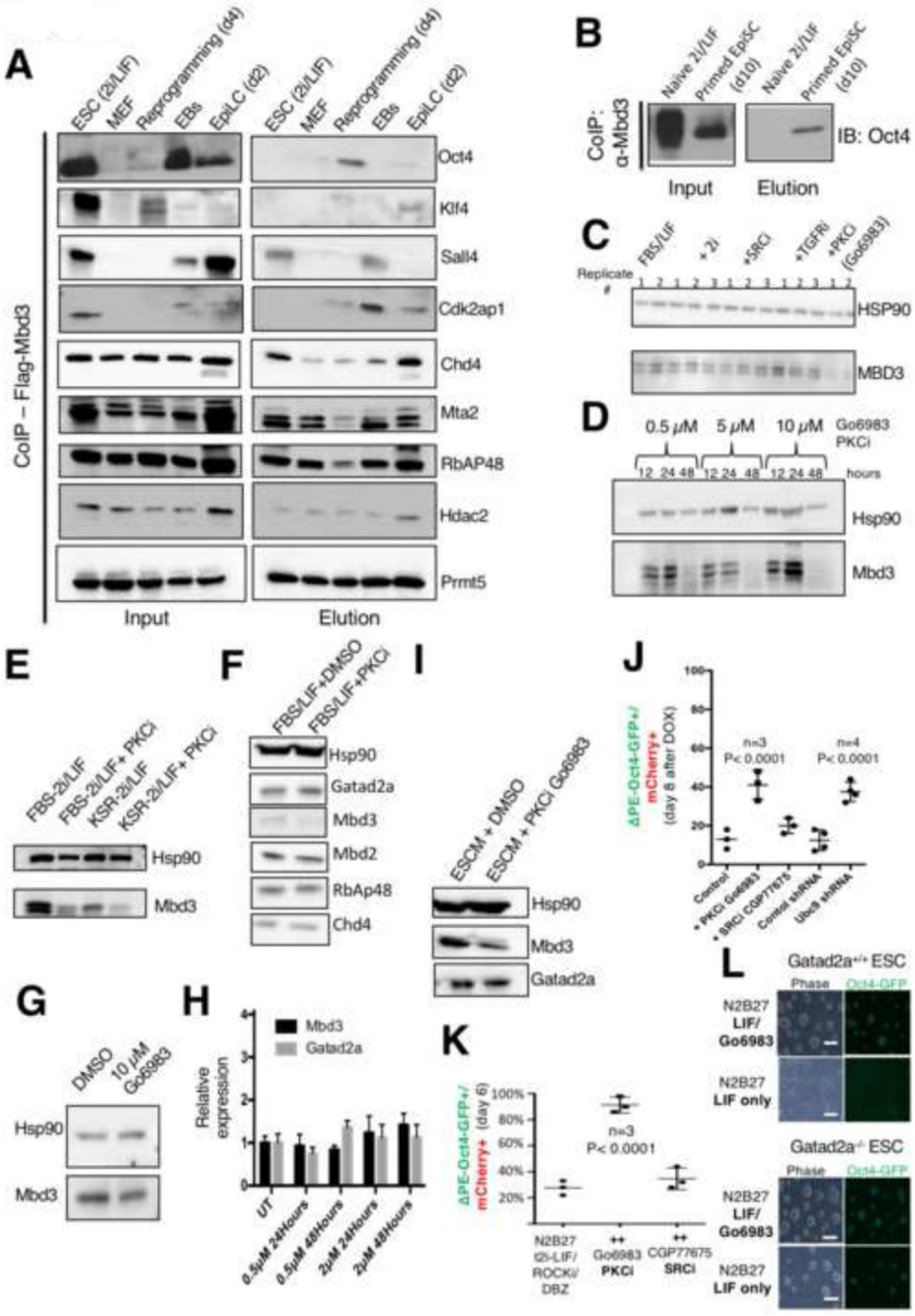

Figure 6. Interactions and assembly of Gatad2a-Mbd3/NuRD are context dependent and can be perturbed by PKC inhibition.

A. Rosa26-M2rtTA Col1a:TetO-2XFlag-Mbd3 ES cells were subjected to different differentiation protocols, and cells from 5 distinct states (naïve ESC, EpiLC, EBs, MEF, 4-day OKSM Reprogramming) were subsequently subjected to CoIP with anti-Flag-Mbd3. Lysates were then analyzed by western blot, and reacted with different antibodies against different NuRD components, pluripotency factors, and other epigenetic proteins. **B.** Rosa26-M2rtTA Col1a:TetO-2XFlag-Mbd3 ES cells were either maintained in ground state naïve

conditions or in priming condition. Lysates were subjected to Co-IP with anti-Flag-Mbd3 and examined by Western blot. Oct4 protein binding to Mbd3 can be detected only in the primed pluripotent state. **C.** Mbd3 expression in ES cells treated with growth media containing different small molecules, after 72 hours of treatment. **D.** PKCi Go6983 effect on Mbd3 level is seen after approximately 48 hours, in different concentration. **E.** Western blot showing Mbd3 protein levels in V6.5 ESCs expanded in different conditions. **F.** Western blot analysis for different NuRD components in ESC cells with and without Go6983. **G.** Mbd3 levels in MEFs following treatment with Go6983. **H.** RT-PCR analysis for Mbd3 transcript abundance following Go6983 treatment. **I.** Go6983 causes mild depletion in Mbd3 in MEFs following 3 days of OKSM expression. **J**. Gatad2a$^{+/+}$ WT MEFs carrying ΔPE -Oct4-GFP reporter and constitutive mCherry markers (secondary system i) were subjected to iPSC reprogramming protocols in in Fig. 2A and iPSC efficiency was quantified at day 8. In the last two conditions included in the panel, the MEFs were pre-treated with control and Ubc9 shRNA following with Neomycin selection (Cheloufi et al., 2015), and then subjected to DOX mediated iPSC reprogramming. **K.** WT EpiSCs reversion efficiency to naïve ESCs in different conditions. Anova test P values are indicated. **L.** Isogenic WT and Gatad2a KO ESCs were expanded on feeder free plates in N2B27 LIF only or LIF/PKCi conditions. Phase images and Oct4-GFP signal maintenance are shown after 8 passages (P8).

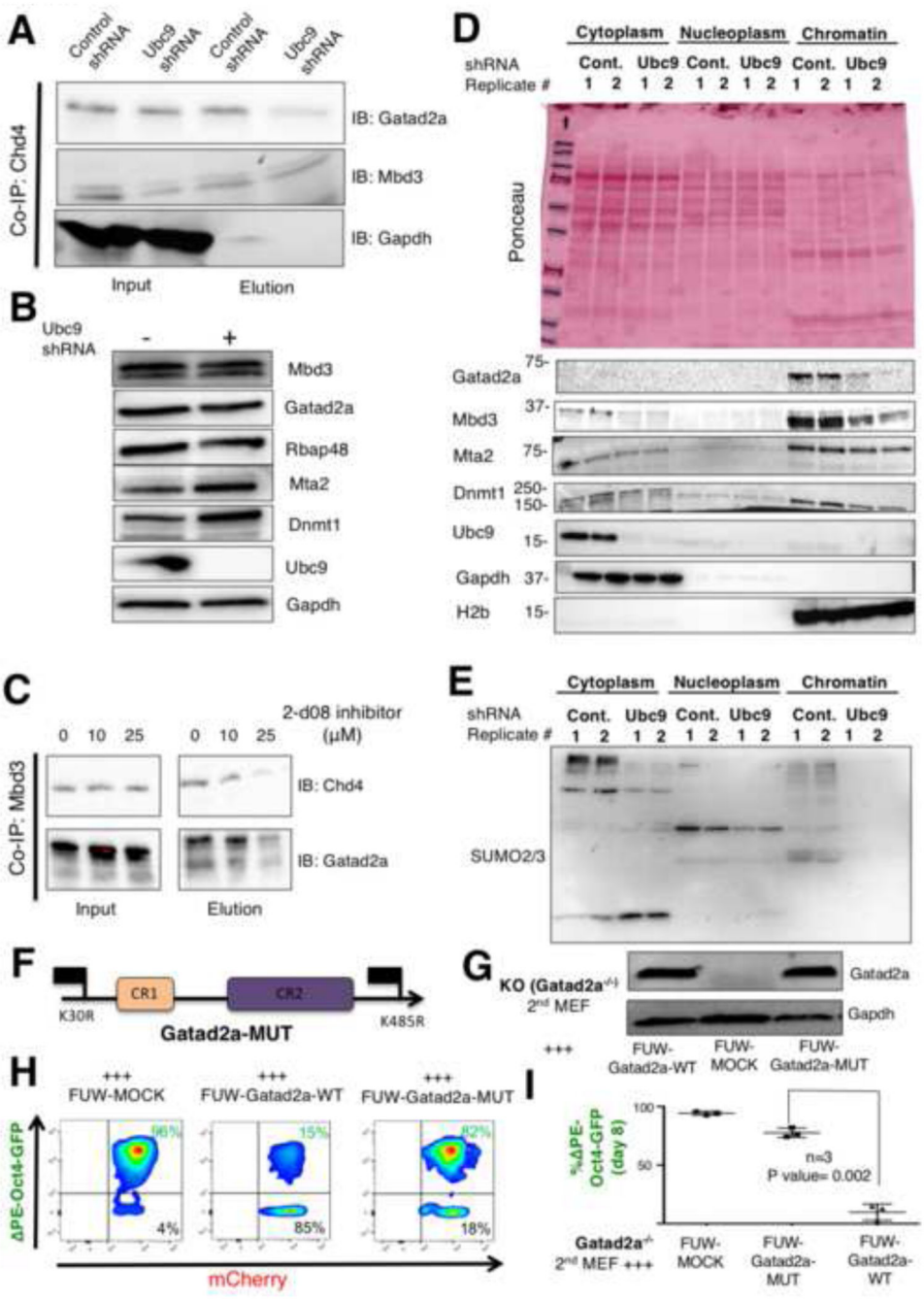

Figure 7. Assembly of Gatad2a-Mbd3/NuRD can be modified post-translationally by SUMOylation.

A. ES cells treated with naïve ground state condition were treated with shRNA targeting Ubc9 or Scramble. Cells were lysed and subsequent CoIP of Chd4 shows a decrease in Gatad2a binding. **B**. Western blot analysis for protein abundance before and after knockdown of Ubc9 in ESCs. **C.** 2-d08 specific small molecule inhibitor for SUMOylation was applied, and co-IP experiments of Mbd3 shows a decrease in Gatad2a and Chd4 binding. **D**. Cells induced in naïve ground state 2i/LIF conditions and subsequent shRNA

targeting either for Ubc9 or scramble negative control. The cells were lysed and fractioned – Cytoplasm, Nucleoplasm and Chromatin fractions, proteins were analyzed by western blot. **E.** As in **D,** with immunoblot for SUMO2/3 on the same exact gel series. **F**. Schematic representation of Gatad2a known domains and functionally validated SUMOylation sites. **G**. Western blot analysis for Gatad2a protein expression following lentiviral infection in Gatad2a$^{-/-}$ MEFs. **H**. Representative FACS images for reprogramming efficiency following Mock, Gatad2a-WT and Gatad2a-MUT (K30R, K485R) lentiviral infection of Gatad2a-/- secondary MEFs. **I**. Quantitative summary of results for experiment elucidated in **H**.

Europe PMC Funders Author Manuscripts