

Chromatin-modifying enzymes as modulators of reprogramming

Tamer T. Onder^{1,2,3,4}, Nergis Kara⁵, Anne Cherry^{1,2,3,4}, Amit U. Sinha^{6,7}, Nan Zhu^{3,6,7}, Kathrin M. Bernt^{3,6,7}, Patrick Cahan^{1,2,3,4}, B. Ogan Mancarci⁸, Juli Unternaehrer^{1,2,3,4}, Piyush B. Gupta^{9,10}, Eric S. Lander^{9,11,12}, Scott A. Armstrong^{3,6,7} & George Q. Daley^{1,2,3,4,6,13,14}

Generation of induced pluripotent stem cells (iPSCs) by somatic cell reprogramming involves global epigenetic remodelling¹. Whereas several proteins are known to regulate chromatin marks associated with the distinct epigenetic states of cells before and after reprogramming^{2,3}, the role of specific chromatin-modifying enzymes in reprogramming remains to be determined. To address how chromatin-modifying proteins influence reprogramming, we used short hairpin RNAs (shRNAs) to target genes in DNA and histone methylation pathways, and identified positive and negative modulators of iPSC generation. Whereas inhibition of the core components of the polycomb repressive complex 1 and 2, including the histone 3 lysine 27 methyltransferase EZH2, reduced reprogramming efficiency, suppression of SUV39H1, YY1 and DOT1L enhanced reprogramming. Specifically, inhibition of the H3K79 histone methyltransferase DOT1L by shRNA or a small molecule accelerated reprogramming, significantly increased the yield of iPSC colonies, and substituted for KLF4 and c-Myc (also known as MYC). Inhibition of DOT1L early in the reprogramming process is associated with a marked increase in two alternative factors, NANOG and LIN28, which play essential functional roles in the enhancement of reprogramming. Genome-wide analysis of H3K79me2 distribution revealed that fibroblast-specific genes associated with the epithelial to mesenchymal transition lose H3K79me2 in the initial phases of reprogramming. DOT1L inhibition facilitates the loss of this mark from genes that are fated to be repressed in the pluripotent state. These findings implicate specific chromatin-modifying enzymes as barriers to or facilitators of reprogramming, and demonstrate how modulation of chromatin-modifying enzymes can be exploited to more efficiently generate iPSCs with fewer exogenous transcription factors.

To examine the influence of chromatin modifiers on somatic cell reprogramming, we used a loss-of-function approach to interrogate the role of 22 select genes in DNA and histone methylation pathways. We tested a pool of three hairpins for each of 22 target genes and observed knockdown efficiencies of >60% for 21 out of 22 targets (Supplementary Fig. 1). We infected fibroblasts differentiated from the H1 human embryonic stem cell (ESC) line (dH1fs) with shRNA pools, transduced them with reprogramming vectors expressing OCT4 (also known as POU5F1), SOX2, KLF4 and c-Myc (OSKM), and identified the resulting iPSCs by Tra-1-60 staining (Fig. 1a)⁴. Eight shRNA pools reduced reprogramming efficiency (Fig. 1b). Among the target genes were *OCT4* (included as a control), and *EHMT1* and *SETDB1*, two H3K9 methyltransferases whose histone mark is associated with transcriptional repression. The remaining five shRNA

pools targeted components of polycomb repressive complexes (PRC), major mediators of gene silencing and heterochromatin formation⁵. Inhibition of PRC1 (BMI1, RING1) and PRC2 components (EZH2, EED, SUZ12) significantly decreased reprogramming efficiency while having negligible effects on cell proliferation (Fig. 1c and Supplementary Fig. 2). This finding is of particular significance given that EZH2 is necessary for fusion-based reprogramming⁶ and highlights the importance of transcriptional silencing of the somatic cell gene expression program during generation of iPSCs.

In contrast to genes whose functions seem to be required for reprogramming, inhibition of three genes enhanced reprogramming: *YY1*, *SUV39H1* and *DOT1L* (Fig. 1b, d). YY1 is a context-dependent transcriptional activator or repressor⁷, whereas SUV39H1 is a histone H3K9 methyltransferase implicated in heterochromatin formation⁸. Interestingly, enzymes that modify H3K9 were associated with both inhibition and enhancement of reprogramming, which suggested that unravelling the mechanisms for their effects might be challenging. Thus, we focused on DOT1L, a histone H3 lysine 79 methyltransferase that has not previously been studied in the context of reprogramming⁹. We used two hairpin vectors that resulted in the most significant downregulation of DOT1L and concomitant decrease in global H3K79 methylation levels (Supplementary Fig. 3a, b). Fibroblasts expressing *DOT1L* shRNAs formed significantly more iPSC colonies when tested separately or in a context where they were fluorescently labelled and co-mixed with control cells (Fig. 2a and Supplementary Fig. 4). This enhanced reprogramming phenotype could be reversed by overexpressing an shRNA-resistant wild-type DOT1L, but not a catalytically inactive DOT1L, indicating that inhibition of catalytic activity of DOT1L is key to enhance reprogramming¹⁰ (Fig. 2a). Our findings with dH1fs were applicable to other human fibroblasts, as IMR-90 and MRC-5 cells also showed threefold and sixfold increases in reprogramming efficiency, respectively, upon DOT1L suppression (Supplementary Fig. 5). To validate our findings independently of shRNA-mediated knockdown, we used a recently discovered small molecule inhibitor of DOT1L catalytic activity. EPZ004777 (ref. 11, referred to as iDot1L) abrogated H3K79 methylation at concentrations ranging from 1 μ M to 10 μ M and increased reprogramming efficiency three- to fourfold (Fig. 2b and Supplementary Fig. 6a, b). Combination of inhibitor treatment with DOT1L knockdown did not further increase reprogramming efficiency, reinforcing our previous observation that inhibition of the catalytic activity of DOT1L is key to reprogramming (Supplementary Fig. 6c). iPSCs generated through DOT1L inhibition showed characteristic ESC morphology, immunoreactivity for SSEA4, SSEA3, Tra-1-81, OCT4 and NANOG,

¹Stem Cell Transplantation Program, Division of Pediatric Hematology and Oncology, Manton Center for Orphan Disease Research, Children's Hospital Boston and Dana Farber Cancer Institute, Boston, Massachusetts, 02115, USA. ²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, 02115, USA. ³Harvard Stem Cell Institute, Cambridge, Massachusetts, 02138, USA. ⁴Stem Cell Program, Children's Hospital Boston, Boston, Massachusetts, 02115, USA. ⁵German Cancer Research Center, Heidelberg, 69120, Germany. ⁶Division of Hematology/Oncology, Children's Hospital, Harvard Medical School, Boston, Massachusetts, 02115, USA. ⁷Department of Pediatric Oncology, Harvard Medical School, Boston, Massachusetts, 02115, USA. ⁸Department of Molecular Biology and Genetics, Bilkent University, Ankara, 06800, Turkey. ⁹Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts, 02142, USA. ¹⁰Whitehead Institute for Biomedical Research, Cambridge, Massachusetts, 02142, USA. ¹¹The Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts, 02142, USA. ¹²Department of Systems Biology, Harvard Medical School, Boston, Massachusetts, 02115, USA. ¹³Division of Hematology, Brigham and Women's Hospital, Boston, Massachusetts, 02115, USA. ¹⁴Howard Hughes Medical Institute, Chevy Chase, Maryland, 20815, USA.

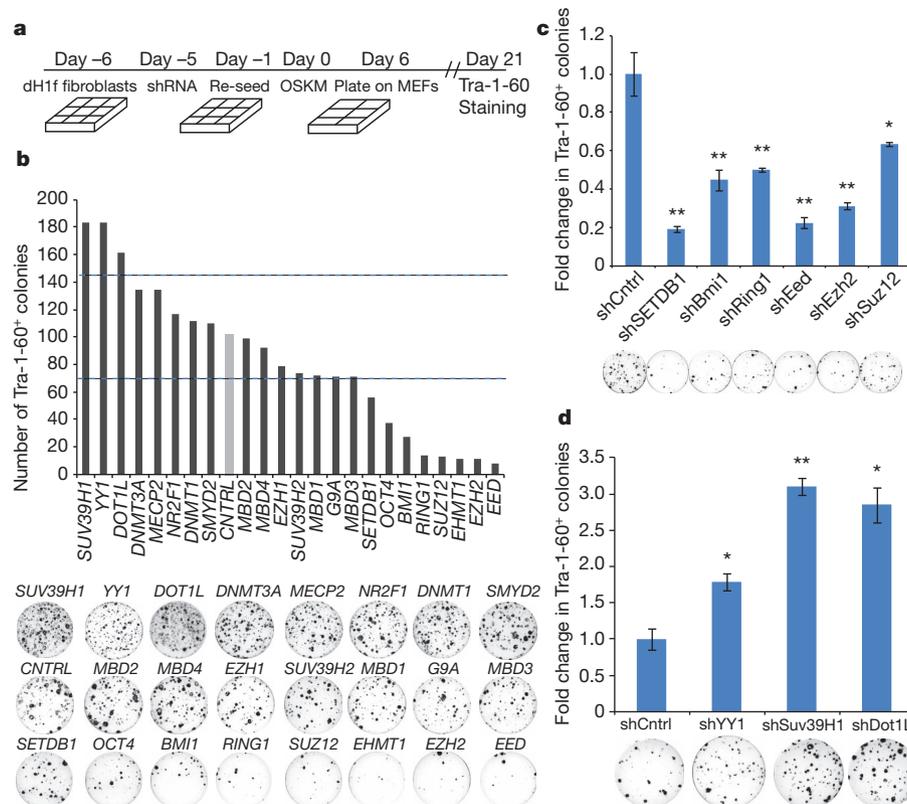


Figure 1 | Screening for inhibitors and enhancers of reprogramming.

a, Timeline of shRNA infection and iPSC generation. **b**, Number of Tra-1-60⁺ colonies 21 days after OSKM transduction of 25,000 dH1f cells previously infected with pools of shRNAs against the indicated genes. Representative Tra-1-60-stained reprogramming wells are shown. The dotted lines indicates 3 standard deviations from the mean number of colonies in control wells. **c**, Validation of primary screen hits that decrease reprogramming efficiency.

Fold change in Tra-1-60⁺ iPSC colonies relative to control cells. * $P < 0.05$, ** $P < 0.01$ compared to control shRNA-expressing fibroblasts ($n = 4$; error bars, \pm s.e.m.). Representative Tra-1-60-stained wells are shown. **d**, Validation of primary screen hits that increase reprogramming efficiency. Fold change in Tra-1-60⁺ iPSC colonies relative to control cells. * $P < 0.05$, ** $P < 0.01$ compared to control shRNA-expressing fibroblasts ($n = 4$; error bars, \pm s.e.m.). Representative Tra-1-60-stained wells are shown.

and differentiated into all three embryonic germ layers *in vitro* and in teratomas (Supplementary Fig. 7a–c). Therefore, iPSCs generated following DOT1L inhibition display all of the hallmarks of pluripotency.

We next assessed DOT1L inhibition in murine reprogramming. iDot1L treatment led to threefold enhancement of reprogramming of mouse embryonic fibroblasts carrying an *OCT4-GFP* (green fluorescent protein) reporter gene (*OCT4-GFP* MEFs; Fig. 2c). Reprogramming of tail-tip fibroblasts (TTFs) derived from a conditional knockout *DOT1L* mouse strain yielded significantly more iPSC colonies upon deletion of *DOT1L*¹² (Supplementary Fig. 8a). Cre-mediated excision of both floxed *DOT1L* alleles in iPSC clones derived from homozygous TTFs was confirmed by genomic PCR (Supplementary Fig. 8b). DOT1L inhibition also increased reprogramming efficiency of MEFs and peripheral blood cells derived from an inducible secondary iPSC mouse strain¹³ (Supplementary Fig. 8c, d). Taken together, these results demonstrate that DOT1L inhibition enhances reprogramming of both mouse and human cells.

We next examined the cellular mechanisms by which DOT1L inhibition promotes reprogramming. DOT1L inhibition affected neither retroviral transgene expression nor cellular proliferation (Supplementary Fig. 9a–c). Although previous studies indicated that *DOT1L*-null cells have increased apoptosis and accumulation of cells in G2 phase⁹, we failed to observe a significant increase in apoptosis or change in the cell cycle profile of DOT1L-inhibited fibroblasts (Supplementary Fig. 9d, e). In human iPSC clones derived from shDot1L fibroblasts, DOT1L inhibition was no longer evident, reflecting the known silencing of retroviruses that occurs during reprogramming (Supplementary Fig. 10a). Quantitative PCR (qPCR) analysis

revealed that the silencing occurred by day 15 after OSKM transduction (Supplementary Fig. 10b, c). To define the crucial time window for DOT1L inhibition, we treated fibroblasts with iDot1L at 1-week intervals during reprogramming. iDot1L treatment in either the first or second week was sufficient to enhance reprogramming, whereas treatment in the third week or a 5-day pretreatment had no effect (Supplementary Fig. 10d, e). Immunofluorescence analysis revealed significantly greater numbers of Tra-1-60-positive cell clusters on day 10 and day 14 in shDot1L cultures (Supplementary Fig. 11a, b), indicating that the emergence of iPSCs is accelerated upon DOT1L inhibition. When we extended the reprogramming experiments by 10 more days, shDot1L cells still yielded more iPSC colonies than controls (Supplementary Fig. 11c). Taken together, these findings indicate that DOT1L inhibition acts in early to middle stages to accelerate and increase the efficiency of the reprogramming process.

To assess whether DOT1L inhibition could replace any of the reprogramming factors, we infected control and DOT1L-inhibited fibroblasts with three factors, omitting one factor at a time. In the absence of OCT4 or SOX2 no iPSC colonies emerged (Fig. 2d). When we omitted either KLF4 or *c-Myc*, DOT1L-inhibited fibroblasts gave rise to robust numbers of Tra-1-60-positive colonies, whereas control cells generated very few colonies, as reported previously⁴ (Fig. 2d–f and Supplementary Fig. 12a). Importantly, DOT1L-inhibited fibroblasts transduced with only OCT4 and SOX2 gave rise to Tra-1-60-positive colonies, whereas control fibroblasts did not (Fig. 2d–f). These two-factor iPSCs showed typical ESC morphology, silenced the reprogramming vectors and had all of the hallmarks of pluripotency as gauged by endogenous pluripotency factor expression and the ability to form all three embryonic germ layers *in vitro* and in teratomas (Supplementary

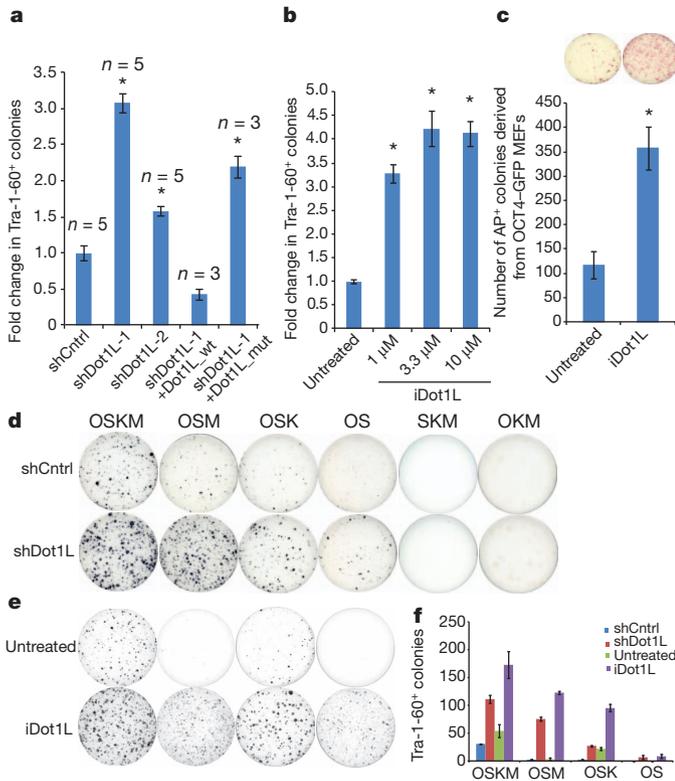


Figure 2 | DOT1L inhibition enhances reprogramming efficiency and substitutes for KLF4 and Myc. **a**, Fold change in the reprogramming efficiency of dH1f cells infected with two independent *DOT1L* shRNAs or co-infected with shRNA-1 and a vector expressing an shRNA-resistant wild-type or catalytically dead mutant *DOT1L*. Data correspond to the average and s.e.m.; $n = 5$ independent experiments. * $P < 0.01$ compared to control shRNA-expressing fibroblasts. **b**, Fold change in the reprogramming efficiency of dH1f cells treated with iDot1L at the indicated concentrations for 21 days. Data correspond to the mean \pm s.d.; $n = 3$. * $P < 0.001$ compared to untreated fibroblasts. **c**, Number of alkaline-phosphatase-positive (AP⁺) colonies derived from OSKM-transduced untreated or iDot1L-treated (10 μ M) OCT4-GFP MEFs. * $P < 0.001$ compared untreated MEFs ($n = 4$; error bars, \pm s.d.). Representative AP-stained wells are shown. **d**, Tra-1-60 stained of plates of shCntrl and shDot1L fibroblasts in the absence of each factor or both KLF4 and c-Myc. **e**, Tra-1-60-stained plates of untreated and iDot1L treated (3.3 μ M) fibroblasts in the absence of each factor or both KLF4 and c-Myc. **f**, Quantification of the Tra-1-60⁺ colonies in Fig. 2d, e representing mean and s.d. of two independent experiments done in triplicate.

Figs 7a–c and 12b). PCR on genomic DNA isolated from expanded colonies confirmed the absence of integrated KLF4 and c-Myc transgenes (Supplementary Fig. 12c). Thus, we were able to generate two-factor iPSCs either by suppression of *DOT1L* expression or chemical inhibition of its methyltransferase activity.

To gain insights into the molecular mechanisms of how *DOT1L* inhibition promotes reprogramming and replaces KLF4 we performed global gene-expression analyses on control and shDot1L fibroblasts before and 6 days after OSKM and OSM transduction, along with cells that were treated with iDot1L. Relatively few genes were differentially expressed in shDot1L cells on day 6 of reprogramming (22 up, 23 down; Supplementary Table 3). Inhibitor-treated cells showed broader gene expression changes (405 up and 175 down; Supplementary Table 3), presumably due to more complete inhibition of K79me2 levels (Fig. 3a). In the absence of KLF4, 94 genes were differentially upregulated in shDot1L cells; intersection of this set of genes with the set differentially upregulated in four-factor reprogramming of *DOT1L*-inhibited cells yielded only five common genes (Fig. 3a, b). We were particularly intrigued to find *NANOG* and *LIN28* upregulated in all three instances of *DOT1L* inhibition, because these two genes are part of the core pluripotency network of human ESCs^{14,15} and can reprogram human

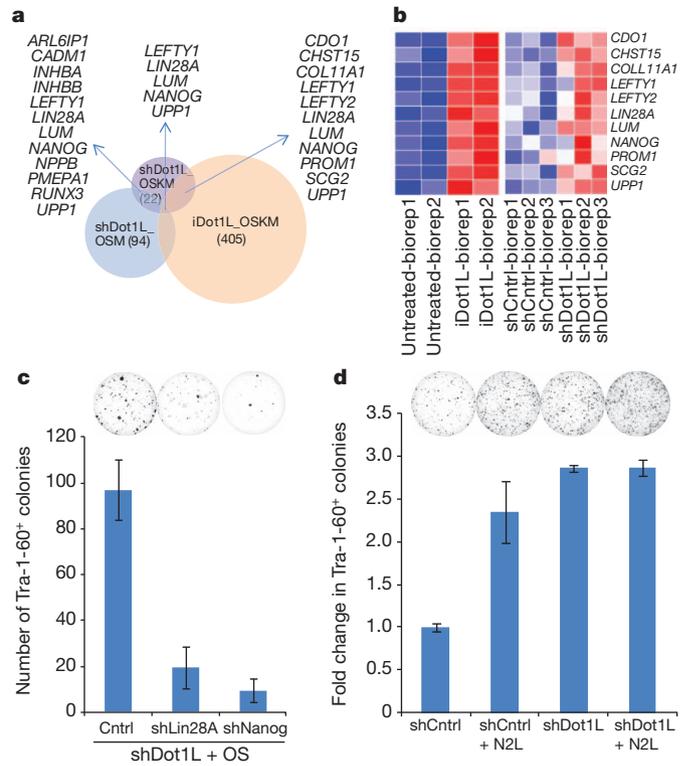


Figure 3 | NANOG and LIN28 are required for enhancement of reprogramming by DOT1L inhibition. **a**, Overlap of differentially upregulated genes in shDot1L cells 6 days post-OSKM and OSM transduction with the genes upregulated in OSKM-transduced iDot1L-treated cells. **b**, Heat maps showing differential expression levels of commonly upregulated genes in OSKM-transduced *DOT1L*-inhibited cells. **c**, Number of Tra-1-60⁺ iPSC colonies upon knockdown of Nanog or Lin28 in 2-factor reprogramming of shDot1L cells. Data represent mean and s.e.m. of 2 independent experiments done in triplicate. **d**, Fold-change in Tra-1-60⁺ iPSC colonies in 4-factor (OSKM) and 6-factor (OSKMNL) reprogramming of shCntrl and shDot1L fibroblasts. Data represent mean and s.e.m. of two independent experiments done in duplicate. Representative Tra-1-60-stained wells are shown above.

fibroblasts into iPSCs when used in combination with *OCT4* and *SOX2* (ref. 16).

We explored the possibility that *NANOG* and *LIN28* upregulation might account for the enhanced reprogramming observed following *DOT1L* inhibition, and validated their upregulation in shDot1L fibroblasts upon OSM or OS transduction (Supplementary Fig. 13a, b). Interestingly, at this early time point *REX1* (also known as *ZFP42*) and *DNMT3B*, two other well-characterized pluripotency genes, were not upregulated, indicating that *DOT1L* inhibition does not broadly upregulate the pluripotency network. Suppression of either Nanog or Lin28 abrogated the two-factor (OS) reprogramming of shDot1L fibroblasts, indicating the essential roles of *NANOG* and *LIN28* in this process (Fig. 3c and Supplementary Fig. 13c). *DOT1L* inhibition also led to increased *NANOG* expression in the context of *OCT4*, *SOX2* and *LIN28* (OSL) and *LIN28* expression in the context of *OCT4*, *SOX2* and *NANOG* (OSN) (Supplementary Fig. 14a). Furthermore, *DOT1L* inhibition significantly increased the efficiency of three-factor reprogramming in the context of OSN and OSL (Supplementary Fig. 14b). Finally, inclusion of *NANOG* and *LIN28* in the OSKM reprogramming cocktail did not confer any additional enhancement to shDot1L cells (Fig. 4d and Supplementary Fig. 14c). Taken together, these data implicate *NANOG* and *LIN28* in the enhancement of reprogramming and replacement of KLF4 and c-Myc with *DOT1L* inhibition.

To gain insight into the genome-wide chromatin changes that are facilitated by *DOT1L* inhibition, we performed chromatin immunoprecipitation followed by DNA sequencing (ChIP-seq) for H3K79me2 and H3K27me3 in human ESCs as well as fibroblasts undergoing

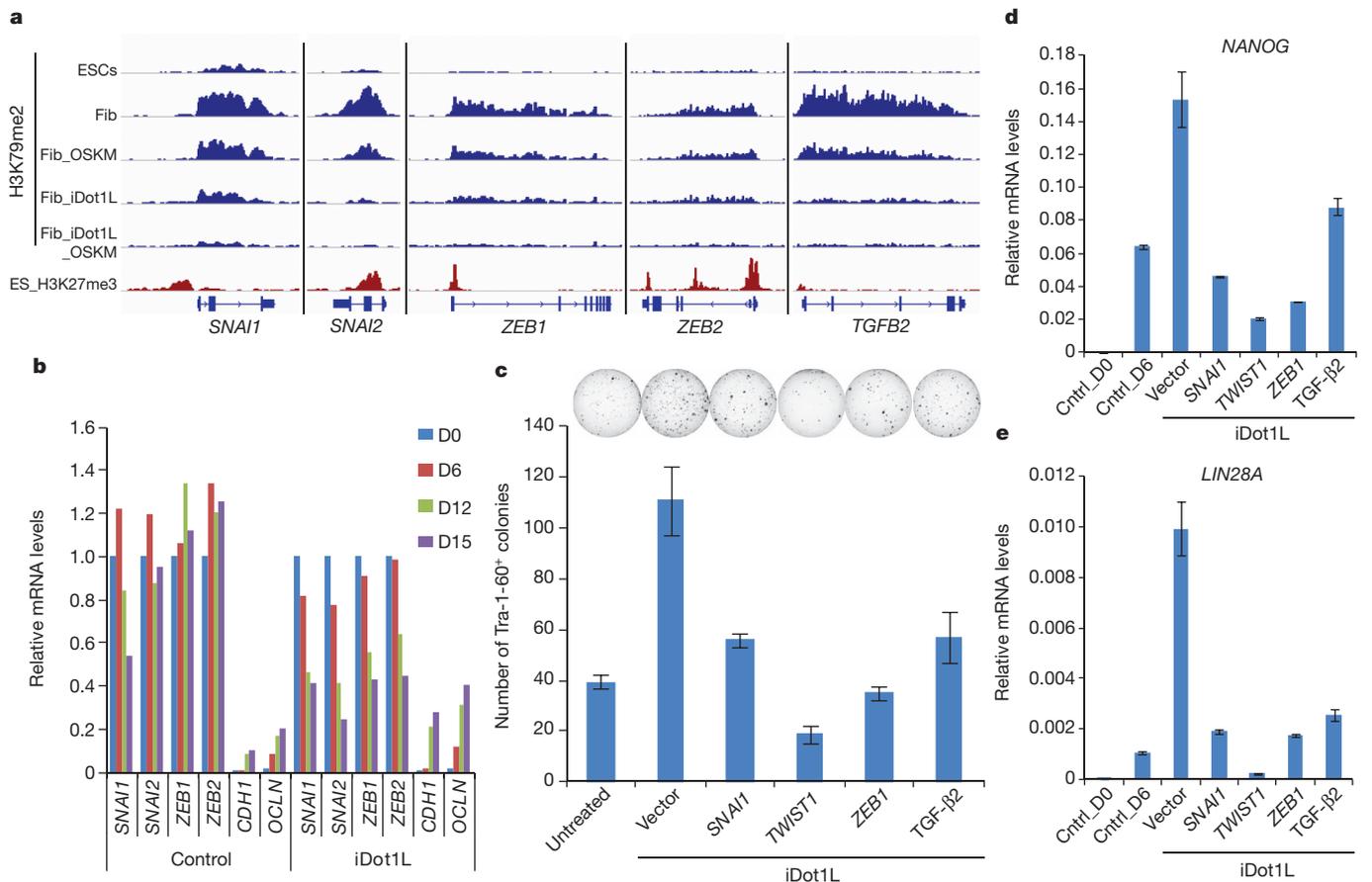


Figure 4 | Genome-wide analysis of H3K79me2 marks during reprogramming. **a**, H3K79me2 ChIP-seencing tracks (blue) for select EMT-associated genes in fibroblasts (Fib) and ESCs along with the corresponding H3K27me3 tracks in ESCs (red). **b**, Expression of EMT-associated transcription factors (EMT-TF) and epithelial genes in control and iDot1L-treated fibroblasts at the indicated time points during reprogramming. qPCR was normalized to uninfected fibroblasts for EMT-TFs and H1 ESCs for *CDH1* and *OCLN*. **c**, Number of Tra-1-60⁺ colonies derived from untreated and iDot1L-treated (3.3 μ M) dH1f cells that are either infected with *SNAI1*,

TWIST1 or *ZEB1* expression vectors or treated with soluble TGF- β 2 (2 ng ml⁻¹) ($n = 3$; error bars, \pm s.d.). Representative Tra-1-60-stained wells are shown. **d**, qRT-PCR quantification of *NANOG* mRNA level on day 6 of OSKM-expressing untreated or iDot1L-treated (3.3 μ M) fibroblasts expressing the indicated EMT-factors. Expression levels were normalized to those observed in H1 ESCs. **e**, qRT-PCR quantification of *LIN28A* mRNA level on day 6 of OSKM-expressing untreated or iDot1L-treated (3.3 μ M) fibroblasts expressing the indicated EMT-factors. Expression levels were normalized to those observed in H1 ESCs.

reprogramming, with or without iDot1L treatment (Supplementary Fig. 15). In both ESCs and fibroblasts, H3K79me2 is positively associated with transcriptionally active genes and negatively associated with genes marked by H3K27me3 (Supplementary Fig. 16a–c). ESC-specific genes marked by H3K79me2 included pluripotency factors, a subset of their downstream targets, and genes involved in epithelial cell adhesion such as *CDH1* (E-cadherin) (280 genes; Supplementary Fig. 17a, b and Supplementary Tables 4, 5). In contrast, in fibroblasts, genes marked by H3K79me2 were significantly enriched in genes induced during the epithelial to mesenchymal transition (EMT) (377 genes; Supplementary Fig. 17a).

Among the 348 genes that showed reduced H3K79me2 6 days after OSKM expression, we likewise found a significant enrichment of gene sets associated with the induction of a mesenchymal state, including *SNAI2*, *TGFB2* and *TGFBRI* (Supplementary Fig. 18a)^{17,18}. Only a few of these genes showed decreased expression at day 6 (12 out of 348), but the vast majority of them lacked this mark in the pluripotent state (272 out of the 348 devoid of H3K79me2 in ESCs), suggesting they were destined for transcriptional silencing during reprogramming. This finding prompted us to ask whether DOT1L inhibition results in the removal of H3K79me2 from such fibroblast-specific, EMT-associated genes. Upon DOT1L inhibitor treatment, H3K79me2 levels were reduced on almost all loci, with the exception of a subset comprised mostly of housekeeping genes that also had high levels of H3K79me2 in ESCs (Supplementary Fig. 19a). Strikingly, the genes

that lost proportionally the most H3K79me2 in inhibitor-treated fibroblasts during reprogramming (eightfold or more) were again highly enriched in genes induced in EMT (Supplementary Fig. 19b). Mesenchymal master regulators such as *SNAI1*, *SNAI2*, *ZEB1*, *ZEB2* and *TGFB2* were among these genes (Fig. 4a)¹⁹. In the presence of the DOT1L inhibitor, these regulators were more strongly repressed during reprogramming, whereas epithelial genes such as *CDH1* and *OCLN* were more robustly upregulated (Fig. 4b). The extinction of fibroblast gene expression was accompanied by increased deposition of the repressive H3K27me3 mark on the majority of fibroblast-specific regulators examined (Supplementary Fig. 20). In contrast, H3K27me3 was depleted to a greater extent on *SOX2* and E-cadherin promoters, reflecting their activation during reprogramming. Finally, the H3K27me3 status of master regulators of other lineages, such as *OLIG2*, *MYOD1*, *NKX2-1* and *GATA4*, remained unchanged upon DOT1L inhibitor treatment, indicating that the deposition of H3K27me3 was specific to fibroblast-specific regulators.

To test the functional importance of downregulation of mesenchymal regulators in the iDot1L-mediated enhancement of reprogramming, we overexpressed *TWIST1*, *SNAI1* and *ZEB1* or added soluble TGF- β 2 to cells undergoing reprogramming in the presence of the DOT1L inhibitor. All of these perturbations significantly counteracted the enhancement observed with DOT1L inhibition (Fig. 4c). Interestingly, expression of these factors also abrogated the iDot1L-mediated upregulation of *NANOG* and *LIN28*, suggesting that the effect of DOT1L

inhibition on these two pluripotency genes is likely to be indirect (Fig. 4d, e). Conversely, we tested whether destabilization of the mesenchymal state by inhibition of TGF- β signalling would be redundant with DOT1L inhibition. A small molecule inhibitor of TGF- β signalling (SB431542) increased reprogramming efficiency, but in combination with the DOT1L inhibitor, showed no significant further increase in iPSC colonies (Supplementary Fig. 21). Taken together these data indicate that in fibroblasts, downregulation of the mesenchymal gene expression program is critical to enhancement of reprogramming by DOT1L inhibition.

Our loss-of-function survey indicates that chromatin-modifying enzymes play critical roles for both reactivating silenced loci as well as reinstating closed domains of heterochromatin during the global epigenetic remodelling of differentiated cells to pluripotency, thus implicating specific enzymes as facilitators or barriers to cell fate transitions. DOT1L inhibition seems to enhance reprogramming at least in part by facilitating loss of H3K79me2 from fibroblast genes whose silencing is required for reprogramming (Supplementary Fig. 22). Interestingly, KLF4, which can be replaced by DOT1L inhibition, has been shown to facilitate a mesenchymal to epithelial transition (MET) by inducing E-cadherin expression²⁰. Persistent H3K79me2 at the fibroblast master regulators during the initial phases of reprogramming seems to prevent shutdown of these genes, thus hindering the acquisition of an epithelial phenotype concomitant with delayed activation of NANOG and LIN28. In this regard H3K79me2 acts as a barrier to efficient repression of the somatic program by the reprogramming factors. This notion is consistent with the role of Dot1 in yeast, where it antagonizes gene repression²¹. As reprogramming of blood cells is also enhanced by DOT1L inhibition, we speculate that DOT1L inhibition may enhance reprogramming in a broad range of cell types by facilitating the silencing of lineage-specific programs of gene expression. Finally, our results also demonstrate that specific chromatin modifiers can be modulated to generate iPSCs more efficiently and with fewer exogenously introduced transcription factors.

METHODS SUMMARY

shRNAs were designed using the RNAi Codex²². 97-mer oligonucleotides (Supplementary Table 1) were PCR-amplified and cloned into the MSCV-PM²³ vector. Reprogramming assays were carried out with either retroviral⁴ or lentiviral¹⁶ reprogramming vectors. dH1f cells were previously described⁴. For gene expression analyses, total RNA was extracted from two or three independent culture plates for each condition and transcriptional profiling was performed using Affymetrix U133A microarrays. ChIP-seq was performed as described with slight modifications¹².

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 16 May 2011; accepted 16 February 2012.

Published online 4 March; corrected 28 March 2012 (see full-text HTML version for details).

1. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
2. Hawkins, R. D. *et al.* Distinct epigenomic landscapes of pluripotent and lineage-committed human cells. *Cell Stem Cell* **6**, 479–491 (2010).
3. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).

4. Park, I.-H. *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* **451**, 141–146 (2008).
5. Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**, 343–349 (2011).
6. Pereira, C. F. *et al.* ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. *Cell Stem Cell* **6**, 547–556 (2010).
7. Shi, Y. *et al.* Transcriptional repression by YY1, a human GLI-Krüppel-related protein, and relief of repression by adenovirus E1A protein. *Cell* **67**, 377–388 (1991).
8. Schotta, G., Ebert, A. & Reuter, G. S. U. (VAR)3–9 is a conserved key function in heterochromatic gene silencing. *Genetica* **117**, 149–158 (2003).
9. Jones, B. *et al.* The histone H3K79 methyltransferase Dot1L is essential for mammalian development and heterochromatin structure. *PLoS Genet.* **4**, e1000190 (2008).
10. Okada, Y. *et al.* hDOT1L links histone methylation to leukemogenesis. *Cell* **121**, 167–178 (2005).
11. Daigle, S. R. *et al.* Selective killing of mixed lineage leukemia cells by a potent small-molecule DOT1L inhibitor. *Cancer Cell* **20**, 53–65 (2011).
12. Bernt, K. M. *et al.* MLL-rearranged leukemia is dependent on aberrant H3K79 methylation by DOT1L. *Cancer Cell* **20**, 66–78 (2011).
13. Carey, B. W. *et al.* Single-gene transgenic mouse strains for reprogramming adult somatic cells. *Nature Methods* **7**, 56–59 (2010).
14. Boyer, L. A. *et al.* Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**, 947–956 (2005).
15. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
16. Yu, J. *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**, 1917–1920 (2007).
17. Charafe-Jauffret, E. *et al.* Gene expression profiling of breast cell lines identifies potential new basal markers. *Oncogene* **25**, 2273–2284 (2006).
18. Onder, T. T. *et al.* Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. *Cancer Res.* **68**, 3645–3654 (2008).
19. Taube, J. H. *et al.* Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proc. Natl Acad. Sci. USA* **107**, 15449–15454 (2010).
20. Samavarchi-Tehrani, P. *et al.* Functional genomics reveals a BMP-driven mesenchymal-to-epithelial transition in the initiation of somatic cell reprogramming. *Cell Stem Cell* **7**, 64–77 (2010).
21. Stulemeijer, I. J. *et al.* Dot1 binding induces chromatin rearrangements by histone methylation-dependent and -independent mechanisms. *Epigenetics Chromatin* **4**, 2 (2011).
22. Olson, A. *et al.* RNAi Codex: a portal/database for short-hairpin RNA (shRNA) gene-silencing constructs. *Nucleic Acids Res.* **34**, D153–D157 (2006).
23. Schlabach, M. R. *et al.* Cancer proliferation gene discovery through functional genomics. *Science* **319**, 620–624 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank G. Hu and S. J. Elledge for providing the MSCV-PM vector, K. Ng and M. W. Lensch for teratoma injections and assessment and S. Loewer for discussions. We also thank E. Olhava and Epizyme Inc. for synthesizing and providing the DOT1L inhibitor, EPZ004777. G.Q.D. is an investigator of the Howard Hughes Medical Institute. Research was funded by grants from the US National Institutes of Health (NIH) to S.A.A. (CA140575) and G.Q.D., and the CHB Stem Cell Program.

Author Contributions T.T.O. performed project planning, experimental work, data interpretation and preparation of the manuscript. N.K., A.C., N.Z., J.U. and B.O.M. performed experimental work. P.C. and A.U.S. participated in data analysis. K.M.B. and S.A.A. provided critical materials and participated in the preparation of the manuscript. P.B.G. and E.S.L. participated in data acquisition, data interpretation and preparation of the manuscript. G.Q.D. supervised research and participated in project planning, data interpretation and preparation of the manuscript.

Author Information The microarray and ChIP-seq data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) and are accessible through GEO Series accession numbers GSE29253 and GSE35791. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to G.Q.D. (George.Daley@childrens.harvard.edu).

METHODS

shRNA cloning. shRNAs were designed using the RNAi Codex²². 97-mer oligos (Supplementary Table 1) were amplified with the following primer pair: Forward: GATGGCTGCTCGAGAAGGTATATTGCTGTTGACAGTGAGCG, reverse: GTCTAGAGGAATTCCGAGGAGTAGGC. PCR products were gel-purified, digested with EcoRI and XhoI and ligated into the MSCV-PM vector. Clones were verified by sequencing. shRNA targeting the firefly luciferase was used as a control²³. NANOG shRNA was previously described²⁴.

Production of viral supernatants. 293T cells were plated at a density of 2.5×10^6 cells per 10-cm dish. The next day, cells were transfected with 2.5 μ g viral vector, 2.25 μ g Gag-Pol vector and 0.25 μ g VSV-G plasmid using 20 μ l Fugene 6 (Roche Applied Science) in 400 μ l DMEM per plate. Supernatant was collected 48 h and 72 h post-transfection and filtered through 45- μ m pore size filters. For concentration, viral supernatants were mixed with PEG3350 solution (Sigma P3640, dissolved in PBS, 10% final concentration) and left overnight at 4 °C. The next day, supernatants were centrifuged at 2,500 r.p.m. for 20 min, and the pellets were re-suspended in PBS. Titering was performed on 293Ts. For shRNA infections, 500 μ l of unconcentrated viral supernatant was used to infect 25,000 cells in the presence of 10 μ g ml⁻¹ protamine sulphate. For fluorescent labelling of dH1fs, we used lentiviruses PRRL-GFP (Addgene catalogue no. 12252) and FUDGW-Tomato (Addgene catalogue no. 22771).

Reprogramming assays. dH1f cells were first infected with shRNA viruses at high multiplicity of infection (m.o.i.) to ensure all cells received at least one vector (gauged by puromycin resistance of parallel infected wells). 25,000 shRNA-infected dH1f cells were then plated per well in 12-well plates and infected overnight with either retroviral (m.o.i. 2.5)²⁵ or lentiviral (Addgene catalogue no. 21162, 21164; 100–200 μ l supernatant)²⁶ reprogramming factors. For two-factor reprogramming, OCT4 and SOX2 viruses were used at an m.o.i. of 5. Six days later, cells were trypsinized and re-plated 1:4 or 1:6 onto six-well plates. Medium was changed to hES medium daily until day 21 when plates were fixed. Small molecule inhibitor of DOT1L, EPZ004777 (a gift from Epizyme, Inc.) was dissolved in DMSO as a 10 mM stock and was added at the indicated concentrations. For DOT1L rescue experiments, an MSCV-based retroviral vector encoding human DOT1L with or without mutations in the SAM-binding site (gifts of Y. Zhang) were mutagenized at the shRNA target site using a QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies). In certain experiments, NANOG and LIN28 expression was achieved using lentiviruses (Addgene catalogue no. 21163). IMR-90 and MRC5 human diploid fibroblasts were purchased from ATCC and 50,000 cells were used in reprogramming experiments.

SB431542 (Stemgent) was used at a final concentration of 2 μ M. TGF- β 2 (R&D Systems) was added daily at a concentration of 2 ng ml⁻¹. Twist1 (Addgene catalogue no. 1783), Snai1 (Addgene catalogue no. 23347) and Zeb1 (a gift of R. A. Weinberg) were overexpressed using retro- or lentiviruses. Statistical analysis was performed using a Student's *t*-test.

Microarray analysis. Total RNA was extracted from two or three independent culture plates for each condition with an RNeasy Mini kit (Qiagen). Synthesis of complementary RNA from total RNA and hybridization/scanning of microarrays were performed with Affymetrix GeneChip products (HGU133A) as described in the GeneChip manual. Normalization of the raw gene expression data, quality control checks and subsequent analyses were done with the open-source R-project statistical software (<http://www.r-project.org/>) together with Bioconductor packages. Raw data files (.CEL) were converted into probe set values by RMA (robust multi-array averaging) normalization. Genes were selected at a threshold of log ratio 0.4. The microarray data have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE29253.

SYBR-Green real-time RT-PCR. Total RNA was extracted using an RNeasy Mini kit coupled with an RNase-free DNase set (Qiagen) and reverse transcribed with Hexanucleotide Mix (Roche). The resulting complementary DNAs were used for PCR using SYBR-Green Master PCR mix (Applied Biosystems) in triplicates. All quantifications were normalized to an endogenous β -actin control. The relative quantification value for each target gene compared to the calibrator for that target is expressed as $2^{-(C_t - C_c)}$ (C_t and C_c are the mean threshold cycle differences after normalizing to β -actin). List of primers can be found in Supplementary Table 2.

Immunostaining. Immunostaining of reprogramming plates was performed as described²⁷. Briefly, cells were fixed with 4% paraformaldehyde and stained with biotin-anti-Tra-1-60 (eBioscience, catalogue no. 13-8863-82, 1:250) and streptavidin horseradish peroxidase (Biolegend, catalogue no. 405210, 1:500) diluted in PBS (3%), FCS (0.3%) Triton X-100. Staining was developed with the Vector labs DAB kit (catalogue no. SK-4100), and iPSC colonies quantified with ImageJ software.

For the characterization of shDot1l-iPSC cells, we picked single colonies onto MEF-coated 96-well plates. The plates were fixed for 20 min with 4% paraformaldehyde/phosphate-buffered saline with calcium and magnesium (PBS (+/+)), washed

several times with PBS (+/+) and incubated overnight at 4 °C with primary antibody and Hoechst stain diluted in 3% donkey serum/3% BSA Fraction VII/0.01% Triton X-100/PBS (+/+); Hoechst, Invitrogen catalogue no. H3570 (1:20,000), Tra-1-81/A488 (BD catalogue no. 560174), SSEA-4/A647 (BD catalogue no. 560219), Tra-1-60/A647 (BD catalogue no. 560122), Nanog, rabbit polyclonal (Abcam catalogue no. ab21624), OCT4, rabbit polyclonal (Abcam catalogue no. ab19857). For Nanog and OCT4, donkey anti-rabbit IgG/A555 (Molecular Probes catalogue no. A31572) secondary antibody was used. After several washes with PBS (+/+), images were acquired using a BD Pathway 435 imager equipped with a $\times 10$ objective.

Teratoma formation assay. iPSCs grown on MEFs were harvested with Collagenase IV (1 mg ml⁻¹ in DMEM/F12). Cell clumps from one six-well plate were resuspended in 50 μ l DMEM/F12, 100 μ l collagen I (Invitrogen catalogue no. A1064401) and 150 μ l hESC-qualified Matrigel (BD Biosciences#354277). Cell clumps were then injected into the hind limb femoral muscles (100 μ l suspension per leg) of Rag2 γ /c mice. After 6–8 weeks, teratomas were harvested and fixed in Bouin's solution overnight. Samples were then embedded in paraffin, and sections were stained with haematoxylin/eosin (Rodent Histopathology Core, Harvard Medical School).

Characterization of iPSC cells. Embryoid body differentiation was performed as described²⁸. To check for the presence of the reprogramming transgenes, genomic DNA was isolated using DNeasy Blood & Tissue Kit (Qiagen) and PCR was performed with specific primers to the endogenous or the viral transgenes⁴.

ChIP-sequencing. ChIP-seq was performed as described with slight modifications¹². 300,000 cells were fixed at room temperature in PBS 1% formalin (v/v) for 10 min with gentle agitation. Fixation was stopped by the addition of glycine (125 mM final concentration) and agitation for 5 min at room temperature. Fixed cells were washed twice in ice-cold PBS, resuspended in 100 μ l of SDS lysis buffer (1% SDS, 10 mM EDTA, 50 mM Tris-HCl, pH 8.1). Chromatin was sheared by sonication to about 100–500 base pair fragments using a Bioruptor (diagenode) and diluted tenfold with dilution buffer (0.01% SDS, 1.1% Triton-X100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, 167 mM NaCl). Antibodies against specific histone modifications were added to sonicated chromatin solution and incubated at 4 °C overnight with gentle agitation. The antibodies used were anti-H3K27me3 (Millipore 07-449) and anti-H3K79me2 (abcam 3594). Immune complexes were collected by incubation with 20 μ l of protein A/G agarose beads (Millipore) for an hour at 4 °C with gentle agitation. Precipitates were washed sequentially with ice-cold low-salt wash (0.1% SDS, 1% Triton-X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 150 mM NaCl), high-salt wash (0.1% SDS, 1% Triton-X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, 500 mM NaCl), LiCl wash (0.25 M LiCl, 1% IGEPAL CA-630, 1% deoxycholic acid, 1 mM EDTA, 10 mM Tris-HCl, pH 8.1) and TE wash (1 mM EDTA, 10 mM Tris-HCl, pH 8.1) for 5 min each at 4 °C with gentle agitation. Samples were centrifuged briefly in between washes to collect the beads. Immunoprecipitated DNA was eluted by incubating beads with 150 μ l elution buffer (1% SDS, 0.1 M NaHCO₃) with gentle agitation for 15 min at room temperature. Elution was repeated once and eluates were combined, sodium chloride (final concentration of 0.2 M) was added to the eluate and eluates were incubated at 65 °C overnight to reverse crosslinking. DNA was purified using PCR purification spin columns (Qiagen). For ChIP sequencing, ChIP DNA libraries were made following Illumina ChIP-seq library preparation kit and subjected to Solexa sequencing (Illumina) at the Center for Cancer Computational Biology, Dana Faber Cancer Institute. Sequencing was performed on Illumina HiSeq2000. The reads were aligned to the human genome hg18 using Bowtie²⁹ and the reads that mapped to multiple locations in the genome were discarded. We quantified the histone modification level as the number of reads per million per kilobase in a window of interest. The window was 1 kb upstream to 1 kb downstream from the transcription start site (TSS) for H3K27me3 and 1 kb upstream to 2 kb downstream of the TSS for H3K79me2. To determine the significance of signal at a gene, an empirical background model was estimated. Genes that showed interesting pattern of histone methylation change were identified using iCanPlot (<http://www.icanplot.org>). Geneset Overlap Analysis was performed by finding the overlap of a set of genes of interest with the gene sets in the collections c2.all and c5.all in MSigDB (total number of genesets in these collections is 5,562)³⁰. Hypergeometric test, with Bonferroni correction for multiple hypothesis testing, was performed to generate the *P* values associated with gene set overlap analysis. ChIP-Seq data have been deposited at the NCBI Gene Expression Omnibus with accession number GSE35791.

24. Zaehres, H. *et al.* High-efficiency RNA interference in human embryonic stem cells. *Stem Cells* **23**, 299–305 (2005).

25. Park, I.-H. *et al.* Generation of human-induced pluripotent stem cells. *Nature Protocols* **3**, 1180–1186 (2008).

26. Yu, J. *et al.* Human induced pluripotent stem cells free of vector and transgene sequences. *Science* **324**, 797–801 (2009).

27. Chan, E. M. *et al.* Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nature Biotechnol.* **27**, 1033–1037 (2009).
28. Loewer, S. *et al.* Large intergenic non-coding RNA-RoR modulates reprogramming of human induced pluripotent stem cells. *Nature Genet.* **42**, 1113–1117 (2010).
29. Langmead, B. *et al.* Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
30. Subramanian, A. *et al.* GSEA-P: a desktop application for Gene Set Enrichment Analysis. *Bioinformatics* **23**, 3251–3253 (2007).