nature cell biology

Proteomic and genomic approaches reveal critical functions of H3K9 methylation and heterochromatin protein-1γ in reprogramming to pluripotency

Rupa Sridharan[1,4], Michelle Gonzales-Cope[2,5], Constantinos Chronis[1,5], Giancarlo Bonora[1], Robin McKee[1], Chengyang Huang[1], Sanjeet Patel[1], David Lopez[1], Nilamadhab Mishra[3], Matteo Pellegrini[1], Michael Carey[1], Benjamin A. Garcia[2,6] and Kathrin Plath[1,6]

Reprogramming of somatic cells into induced pluripotent stem cells (iPSCs) involves a marked reorganization of chromatin. To identify post-translational histone modifications that change in global abundance during this process, we have applied a quantitative mass-spectrometry-based approach. We found that iPSCs, compared with both the starting fibroblasts and a late reprogramming intermediate (pre-iPSCs), are enriched for histone modifications associated with active chromatin, and depleted for marks of transcriptional elongation and a subset of repressive modifications including H3K9me2/me3. Dissecting the contribution of H3K9 methylation to reprogramming, we show that the H3K9 methyltransferases Ehmt1, Ehmt2 and Setdb1 regulate global H3K9me2/me3 levels and that their depletion increases iPSC formation from both fibroblasts and pre-iPSCs. Similarly, we find that inhibition of heterochromatin protein-1γ (*Cbx3*), a protein known to recognize H3K9 methylation, enhances reprogramming. Genome-wide location analysis revealed that Cbx3 predominantly binds active genes in both pre-iPSCs and pluripotent cells but with a strikingly different distribution: in pre-iPSCs, but not in embryonic stem cells, Cbx3 associates with active transcriptional start sites, suggesting a developmentally regulated role for Cbx3 in transcriptional activation. Despite largely non-overlapping functions and the predominant association of Cbx3 with active transcription, the H3K9 methyltransferases and Cbx3 both inhibit reprogramming by repressing the pluripotency factor *Nanog*. Together, our findings demonstrate that Cbx3 and H3K9 methylation restrict late reprogramming events, and suggest that a marked change in global chromatin character constitutes an epigenetic roadblock for reprogramming.

Reprogramming of somatic cells into iPSCs by overexpression of the transcription factors Oct4, Sox2, Klf4 and cMyc is a fascinating, but inefficient process, with only a small subset of starting cells converting to a pluripotent state after 1–2 weeks[1,2]. Mechanistic insights into how chromatin regulators and chromatin states control reprogramming are only now beginning to be explored[3–11]. To gain insight into global chromatin changes that occur during reprogramming to iPSCs, we were interested in quantifying the post-translational modifications (PTMs) of histones. We reasoned that histone PTMs with marked changes in global levels during reprogramming could be important for the suppression or promotion of the process.

Investigations of the role of histone PTMs during reprogramming have typically relied on the use of site-specific histone antibodies in immunostaining and chromatin immunoprecipitation (ChIP) experiments[3,12–15]. Although very insightful, these methods are reliant on the availability of cognate antibodies, and epitope recognition can be affected by modifications on neighbouring residues or interacting factors. To circumvent these issues, we used label-free mass spectrometry (qMS)-based proteomics[16,17] as an alternative approach to quantify alterations in histone PTMs during reprogramming, which is independent of antibodies (Supplementary Fig. S1A).

[1]University of California Los Angeles, David Geffen School of Medicine, Department of Biological Chemistry, Jonsson Comprehensive Cancer Center, Molecular Biology Institute, Bioinformatics Interdepartmental Degree Program, Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research, Los Angeles, California 90095, USA. [2]Epigenetics Program, Department of Biochemistry and Biophysics, Perelman School of Medicine, University of Pennsylvania, 1009C Stellar-Chance Laboratories, 422 Curie Boulevard, Philadelphia, Pennsylvania 19104, USA. [3]Section on Rheumatology, Department of Internal Medicine, Wake Forest University School of Medicine, Winston-Salem, North Carolina 27157, USA. [4]Present address: Wisconsin Institute for Discovery, Department of Cell and Regenerative Biology, University of Wisconsin, 330 N. Orchard Street, Room 2118, Madison, Wisconsin 53715, USA. [5]These authors contributed equally to this work. [6]Correspondence should be addressed to B.A.G. or K.P. (e-mail: bgarci@mail.med.upenn.edu or kplath@mednet.ucla.edu)

Received 20 September 2012; accepted 24 April 2013; published online 9 June 2013; DOI: 10.1038/ncb2768

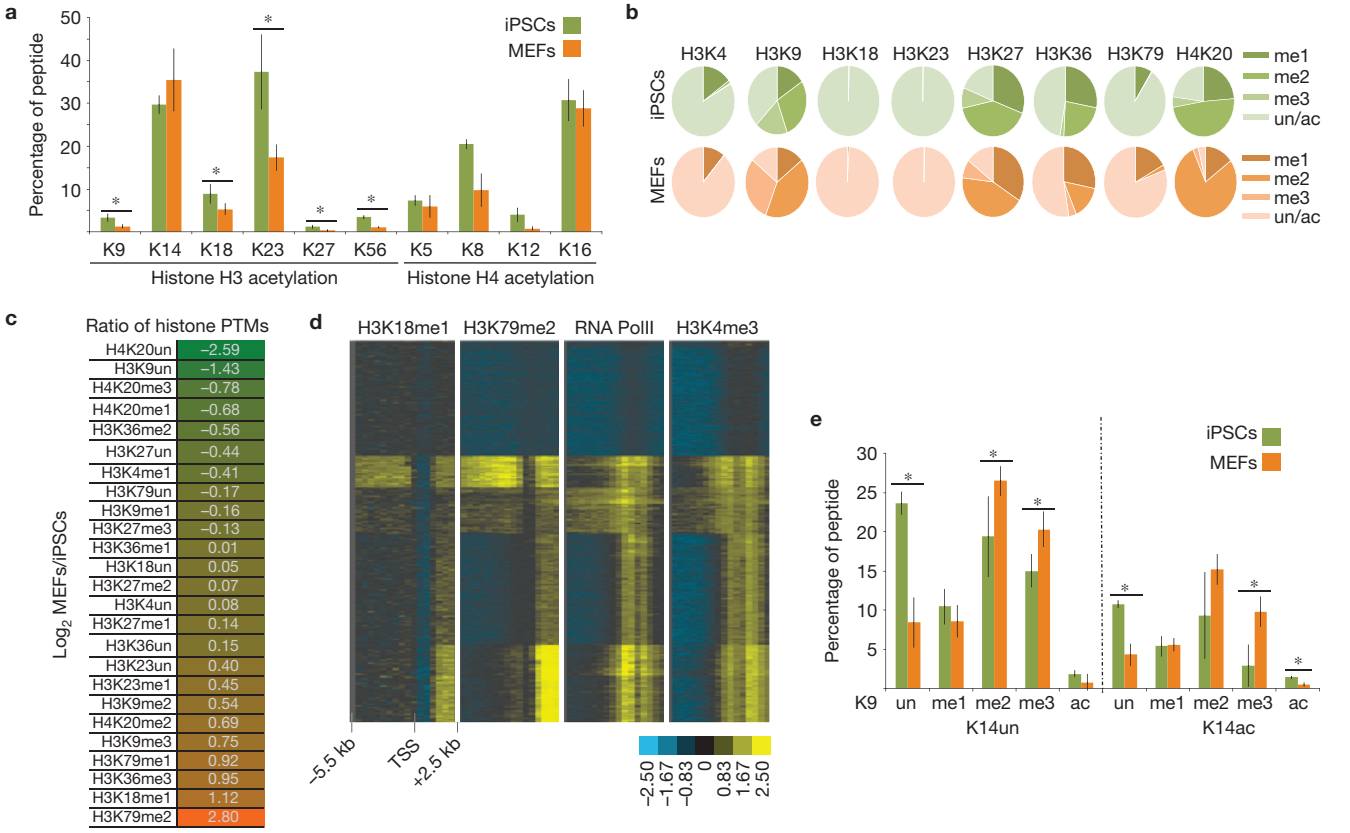

Figure 1 Global levels of most histone PTMs differ between MEFs and iPSCs. (**a**) Graph showing the percentage of acetylated (ac) lysine (K) residues in female iPSCs and MEFs. Data presented are the mean of 6 replicate mass spectrometry experiments obtained from 2 biological replicates, each analysed 3 times. The asterisks indicate significantly different levels between iPSCs and MEFs at $P < 0.01$ determined by the t-test. Error bar = s.d. of 6 replicates. Source data are provided in Supplementary Table S1. (**b**) Pie charts indicating the percentage of mono- (me1), di- (me2) and tri- (me3) methylated and unmethylated (un) forms of the indicated lysine residues in female iPSCs and MEFs. The unmethylated portion represents the sum of acetylated and unacetylated isoforms. Standard deviations for 6 replicates and significant differences in these PTMs between iPSCs and MEFs are provided in Supplementary Fig. S1B. (**c**) Log_2 ratio of the percentage of methylated and unmethylated lysines between female MEFs and iPSCs. Histone PTMs that are lower in MEFs than iPSCs are highlighted in green and those that are higher in MEFs than iPSCs in orange. The unmethylated state can be acetylated and/or unacetylated. (**d**) Hierarchical clustering of ChIP-chip data for H3K18me1, H3K79me2, RNA polymerase II and H3K4me3 for 18,300 genes. Each row represents the promoter region of a gene spanning −5.5 kb to +2.5 kb relative to the transcriptional start site (TSS), divided into sixteen 500 bp fragments that exhibit the average log ratio of probe signal intensity with blue, yellow and grey representing lower-than-average, higher-than-average and missing values (mostly due to lack of probes in those regions) for enrichment relative to input. (**e**) The same as in **a**, except that the combination of PTMs on the histone H3 peptide containing amino acids 9–17 was quantified. Data presented are the mean of 6 mass spectrometry values obtained from 2 biological replicates, each analysed 3 times.The K9 residue was detected in the ac, me1, me2, me3 or the unmethylated/unacetylated (un) forms, and K14 in the ac or un forms. Each combination of histone PTMs on K9 and K14 is referred to as a peptide isoform. Source data are provided in Supplementary Table S1.

RESULTS

Global levels of many histone PTMs change during reprogramming

To begin with, we determined the abundance of acetylation or methylation modifications at individual lysine residues of histones H3 and H4, at the start and endpoint of reprogramming, that is, in mouse embryonic fibroblasts (MEFs) and iPSCs derived from these cells. Two iPSC and two MEF lines were subjected to six independent qMS reactions per sample (see Methods), generating a highly reproducible quantification of histone PTMs in these cell types, which is summarized in Supplementary Table S1.

We found a wide variation in the abundance of histone acetylation across lysine (K) residues in both histone H3 and H4. Within histone H3, acetylation was most abundant on residues K14 and K23 irrespective of the cell type analysed (Fig. 1a). Acetylation of H3K9, a mark associated with transcriptionally active or bivalent promoters and enhancers[18,19], of H3K27, a mark characteristic of active enhancers[20], or of H3K56, which overlaps with the binding of OCT4, SOX2 and NANOG in human embryonic stem cells[21] (ESCs), was present on less than 5% of histone H3 molecules (Fig. 1a). These results potentially reflect the preferential association of these acetylation events with regulatory genomic elements compared with a broader chromatin role for H3K14 and H3K23ac. For example, H3K14 acetylation has already been implicated in DNA damage responses[22]. Similar variations in the extent of acetylation were found for the lysine residues of the amino-terminal tail of histone H4 (Fig. 1a). Importantly, almost all acetylated lysines on histone H3 and H4 were more prevalent in iPSCs than MEFs (Fig. 1a). As acetylation of lysines has been associated with active chromatin states and transcription[18], our findings extend the conclusion that pluripotent cells are more euchromatic than differentiated cells[23,24].

Histone methylation patterns are more complex owing to the presence of mono-, di- and tri-methylation states. As for acetylation, there is a marked variation in the abundance of methylation across lysine residues in MEFs and iPSCs (Fig. 1b and Supplementary Fig. S1B). For methylated lysines related to transcriptional repression, such as H3K9 and H3K27, between 60 and 80% of the respective lysines are methylated in both MEFs and iPSCs, revealing an unexpected coverage of the genome by histones carrying these methylated residues. Methylation at lysine residues known to be associated with enhancers and promoters, for example at H3K4, is much less abundant in both cell types. Surprisingly, methylation associated with transcriptional elongation, particularly H3K36 methylation, is relatively abundant in the genome.

Analysing differences in global histone methylation profiles between iPSCs and MEFs (Fig. 1c), we found that H3K79me2 and H3K36me3, two methylation marks associated with transcriptional elongation[25,26], and the relatively uncharacterized mark H3K18me1 (ref. 27), were the top three marks that are more abundant in MEFs than iPSCs (Fig. 1c). Notably, the reduction of global levels of H3K79me2 and H3K36me3 during reprogramming may be important for the generation of iPSCs because the inhibition of Dot1L, the enzyme responsible for H3K79 methylation, and overexpression of H3K36me2/me3 demethylases enhance iPSC formation[7–9]. To better understand the function of H3K18me1, we performed ChIP with an antibody specific for H3K18me1 (Supplementary Fig. S1C) in combination with promoter microarrays. We found that H3K18me1 is enriched in coding regions with a pattern similar to that of H3K79me2 (Fig. 1d). These findings identify the association of a previously uncharacterized histone modification with transcriptional elongation. They also indicate that the most downregulated histone modifications during reprogramming are all linked to transcriptional elongation, which is surprising given that pluripotent cells have been argued to be transcriptionally more permissive compared with differentiated cells[28]. Therefore, these modifications may have different functions in pluripotent and differentiated cells, which will be interesting to study in the future.

Among methylation marks associated with transcriptional silencing, H3K27 methylation states were not very different between iPSCs and MEFs; H3K9me2 and H3K9me3 levels were higher in MEFs than iPSCs; and H4K20me3 and H4K20me1 were more abundant in iPSCs than MEFs (Fig. 1b,c and Supplementary Fig. S1B). We also noted a strong increase in unmethylated H3K9 and H4K20 residues from MEFs to iPSCs that was higher than that of any methylation mark (Fig. 1c), suggesting that the unmethylated state of these lysine residues is an important feature of the pluripotent state. Together, these data indicate that not all repressive methylation histone marks are depleted in iPSCs compared with MEFs, although pluripotent cells have a more euchromatic character when compared with MEFs (refs 23,24). We also conclude that H3K9 and H4K20 methylation, in contrast to H3K27 methylation, may exert effects on reprogramming by modulation of their global levels.

Histones can be modified simultaneously on multiple amino acids. qMS offers arguably the only approach to exactly quantify combinations of PTMs that occur on the same peptide. We therefore examined the combination of acetylation and methylation that occurred within each tryptic histone peptide (that is, in the peptides of H3 containing either K9/K14, K18/K23 or K27/K36, or in the H4 peptide carrying K5/K8/K12/K16; Fig. 1e and Supplementary Fig. S1D–F). All examined histone peptides accommodate modifications at neighbouring amino acids, highlighting the complex control of histone modifications and functional output. For example, for the H3K9/K14 peptide, we found that repressive K9 methylation marks and the activating K14ac were often present on the same histone molecule (Fig. 1e). Furthermore, although total levels of H3K14ac were similar between iPSCs and MEFs (Fig. 1a), H3K14ac was significantly higher in iPSCs than MEFs when H3K9 was unmodified or acetylated on the same histone molecule (Fig. 1e), indicating that modifications on the K9 residue affect the acetylation status of K14 in a cell-type-specific manner. In addition, the unmodified form of the H3K9/K14 peptide (K9un/K14un) was the most prevalent isoform of this peptide in iPSCs, perhaps enabling the rapid acquisition of various modifications in response to differentiation cues (Fig. 1e).

A global change in chromatin character occurs late in reprogramming

The differences in global levels of histone PTMs between MEFs and iPSCs prompted us to determine when they occur during reprogramming, and whether iPSCs are similar to ESCs in their global histone PTM profile. Although reprogramming is inefficient, intermediate stages of the process have been described[1,3,11,15,29,30]. To examine the global chromatin state in a late intermediate of reprogramming, we took advantage of pre-iPSCs rather than determining histone PTM profiles in heterogeneous reprogramming cultures. These cells can be isolated from reprogramming cultures as a clonal population of cells with an ESC-like morphology that have efficiently repressed the somatic gene expression program but lack the expression of most pluripotency factors[3,15,31]. These cells are commonly obtained when reprogramming is induced with retrovirally expressed Oct4, Sox2, Klf4 and cMyc (refs 3,15,31). We reasoned that the analysis of the histone PTM profile in pre-iPSC lines would allow us to determine when global chromatin character changes occur in the reprogramming process relative to known transcriptional changes. Hence, we performed label-free qMS analysis for histone PTMs on one male ESC line, and one male and one female pre-iPSC line, with at least five replicate qMS data sets per cell line, and compared them with the iPSC and MEF data. Quantitative differences in histone PTMs between two cell types were confirmed directly by using chemical stable isotope labelling and subsequent mixing of the histone samples from the two cell types before MS analysis[17] (Supplementary Fig. S2).

Unsupervised hierarchical clustering of histone PTM levels for all replicate data sets, based on combinations of modifications per tryptic histone peptide (summarized in Supplementary Table S1), demonstrated that the global chromatin character of pre-iPSCs is more similar to that of MEFs (Fig. 2a). Furthermore, pre-iPSCs and MEFs cluster away from both ESCs and iPSCs, which in turn are more related to each other in chromatin state (Fig. 2a). In pre-iPSCs a few histone PTMs were present at an intermediate level between iPSCs and MEFs, as for instance H3K18ac/K23ac (Fig. 2b), or less abundant than in any of the other cell lines, as for instance H4K5acK8acK12acK16un (Fig. 2c). Irrespective of these differences, a shift towards the pluripotency profile is not evident for most histone PTMs in pre-iPSCs. Together these findings demonstrate that ESCs and iPSCs share a similar global histone modification profile revealing a pluripotency-specific global chromatin structure. Furthermore, the transition from the MEF-like

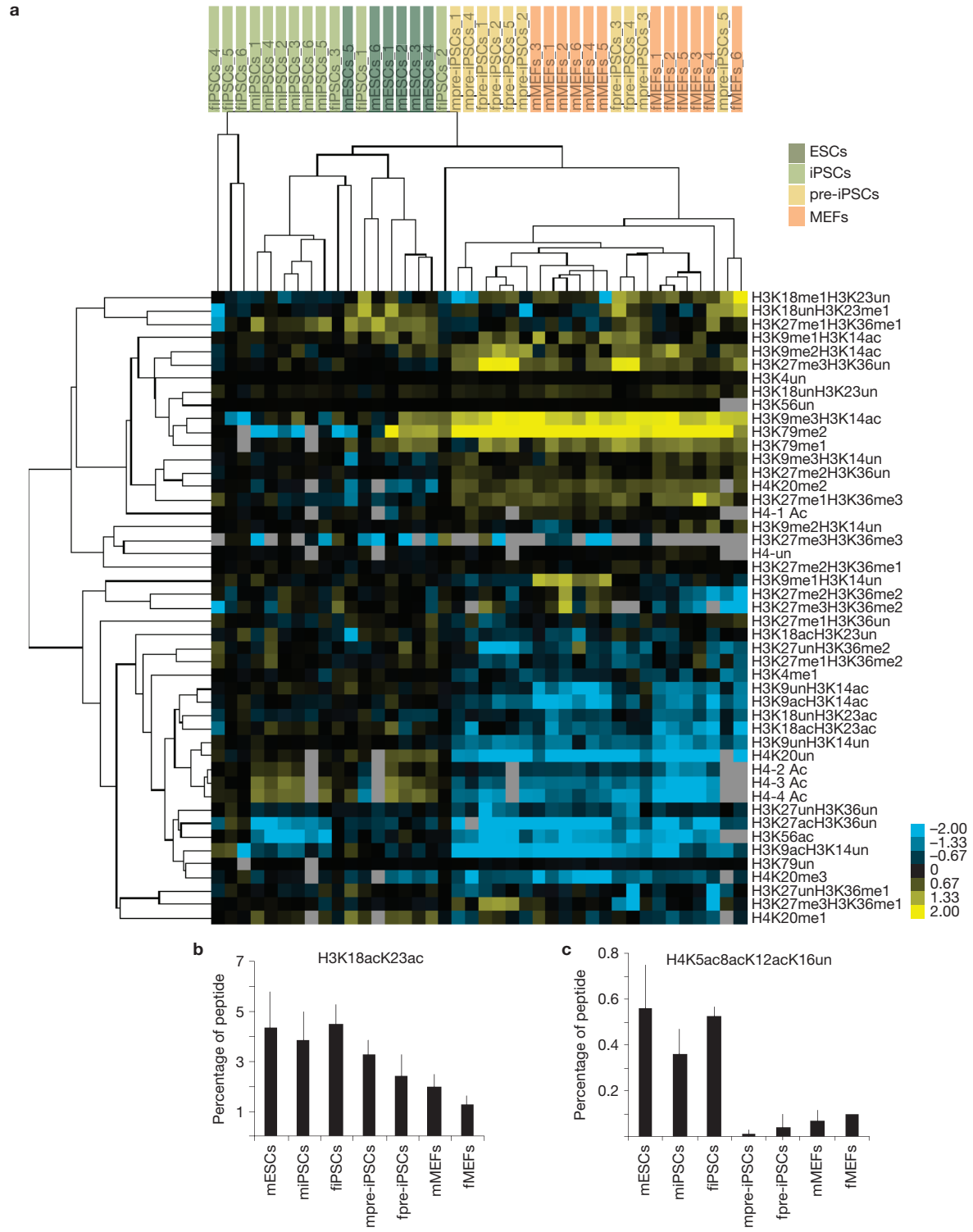

Figure 2 Comparison of histone PTM profiles between ESCs, iPSCs, pre-iPSCs and MEFs. (**a**) Unsupervised hierarchical clustering of the $\log_2$ ratio of the percentage of histone isoforms for each qMS replicate in the indicated cell types relative to the average percentage of the same isoform in female iPSCs. Cell types included are: female iPSCs (fiPSCs, 6 replicates), male iPSCs (miPSCs, 6 replicates), male ESCs (mESCs, 6 replicates), female pre-iPSCs (fpre-iPSCs, 5 replicates), male pre-iPSCs (mpre-iPSCs, 5 replicates), male fibroblasts (mMEFs, 6 replicates) and female fibroblasts (fMEFs, 6 replicates). In the heat map, yellow represents isoforms that are more abundant than in fiPSCs and blue those that are less abundant. For the H4 peptide including amino acids 4–17, 1ac refers to the sum of isoforms that have a single acetylated lysine among K5, K8, K12 or K16; 2ac and 3ac refers to the sum of all isoforms with 2 and 3 acetyl marks, respectively; and 4ac refers to the completely acetylated peptide. (**b**) The graph shows the percentage of the isoform of the H3 peptide containing amino acids 18–26 that is acetylated at both K18 and K23, for the indicated cell types. Data presented are the mean of 6 mass spectrometry values obtained from 2 biological replicates, each analysed 3 times. Error bars indicate standard deviation. Source data are provided in Supplementary Table S1. (**c**) The same as in **b**, except for the isoform of the H4 peptide containing amino acids 4–17 with acetylation at positions K5, K8 and K12, but not K16.

to the pluripotency-specific global chromatin character occurs late in reprogramming, after the state represented by pre-iPSCs, rather than gradually throughout the entire reprogramming process. On the basis of these data, we propose that the establishment of the global pluripotency-specific chromatin state may constitute an epigenetic barrier that contributes to the reprogramming block encountered by pre-iPSCs and the low overall efficiency of reprogramming to iPSCs.

H3K9 methylation regulators control the efficiency of reprogramming

Our qMS approach demonstrated that pre-iPSCs and MEFs are more enriched for repressive H3K9 methylation marks than iPSCs. We therefore directed our efforts on further deconvoluting reprogramming barriers associated with the H3K9 site. Specifically, we investigated whether the depletion of the writers, the histone methyltransferase (HMTase) genes *Ehmt1*, *Ehmt2* and *Setdb1*, or the readers, the heterochromatin protein 1 (HP1) family members *Cbx1*, *Cbx3* and *Cbx5*, or overexpression of the H3K9 demethylase gene *Jmjd2c* could modulate the efficiency of reprogramming. HP1 proteins are small proteins that have been shown to bind specifically to methylated histone H3K9 through their chromodomain in biochemical assays[32,33]. However, although initially identified as evolutionarily conserved regulators of heterochromatin formation, recent progress suggests further roles for HP1 proteins in the regulation of active gene expression in euchromatic regions[34,35].

As the knockout of some of these genes, such as *Setdb1*, is known to lead to loss of pluripotency[36,37], we only transiently depleted them during reprogramming by means of short interfering RNA (siRNA)-mediated knockdown. Reprogramming was induced by overexpression of Oct4, Sox2 and Klf4, that is, in the absence of ectopic cMyc, in MEFs containing GFP reporters linked to pluripotency promoters (*Oct4* or *Nanog*). Efficient knockdown for all target genes was confirmed in each reprogramming experiment (Supplementary Fig. S3A). The depletion of any of the three H3K9 HMTases consistently increased the number of *Oct4*–GFP-positive colonies at least twofold, and simultaneous knockdown of all three HMTases (henceforth called 3XHMT) enhanced reprogramming even more efficiently (Fig. 3a(i), Supplementary Fig. S3Ai). GFP-positive colonies also expressed the endogenous pluripotency marker Esrrb, indicating that the increase in colony number did not simply represent GFP-reporter activation (Supplementary Fig. S3B). Depletion of *Cbx3* consistently increased the number of *Oct4*–GFP-positive colonies, whereas the effects of *Cbx1* and *Cbx5* interference were positive but milder and not always reproducible (Fig. 3a(ii) and Supplementary Fig. S3Aii). 3XHMT or *Cbx3* knockdown also had a positive effect on reprogramming when cMyc was included as a reprogramming factor, and reprogrammed colonies appeared at least a day earlier than under control conditions, indicating an improvement of the kinetics of the process (Fig. 3b). In addition, overexpression of *Jmjd2c*, a H3K9me2/me3 demethylase gene[38] important for the maintenance of the pluripotent state[39], enhanced reprogramming about twofold (Fig. 3c and Supplementary Fig. S3Aiii). Together, these results extend recent findings that the H3K9 methylation machinery impairs reprogramming to iPSCs and cell-fusion-mediated gain of pluripotency[9–11,40], and indicate that not only the H3K9 methylation enzymes but also members of the HP1 family act as critical barriers of reprogramming.

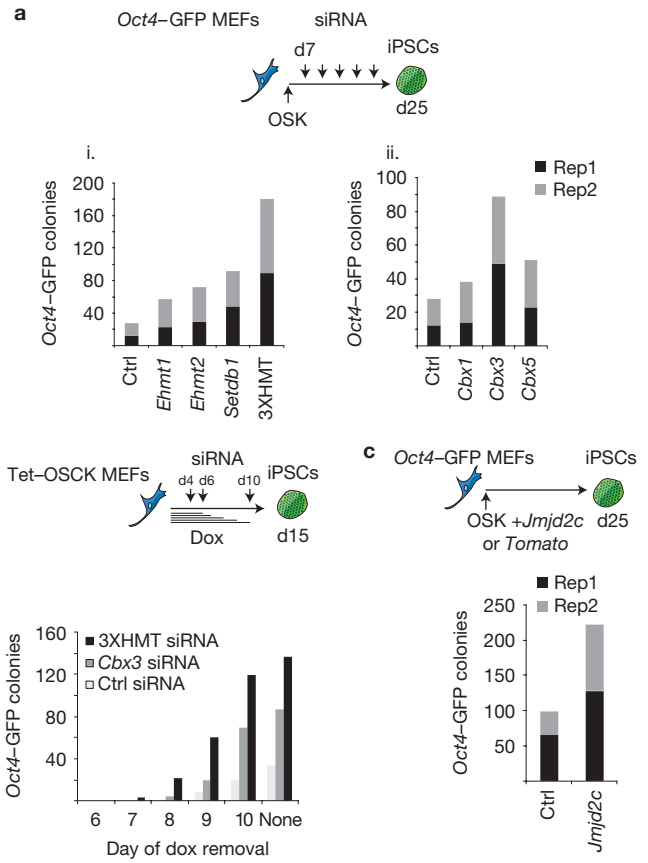

Figure 3 H3K9 methyltransferases, demethylases and methyl binders control the efficiency and kinetics of reprogramming. (**a**) Depletion of H3K9 HMTases and *Cbx3* enhances OSK reprogramming. *Oct4*–GFP reporter MEFs were infected with individual retroviruses expressing *Oct4*, *Sox2* and *Klf4* (OSK) and transfected with siRNAs targeting the indicated genes every 3 days starting on day 7 (d7) post-infection. The number of GFP-positive colonies was determined on day 25 of reprogramming: (i) for the knockdown of H3K9 methyltransferase genes *Ehmt1*, *Ehmt2* and *Setdb1*, either singly or combined (indicated as 3XHMT); (ii) for the knockdown of the heterochromatin protein 1 family members *Cbx1* = HP1β, *Cbx3* = HP1γ, or *Cbx5* = HP1α. Data from 2 technical replicates (Rep1 and Rep2) from 1 out of 3 representative experiments are presented. (**b**) Depletion of the H3K9 HMTases or *Cbx3* enhances the efficiency and kinetics of OSCK reprogramming. *Oct4*–GFP reporter MEFs carrying a single, doxycycline (dox)-inducible polycistronic transgene encoding *Oct4, Sox2, cMyc* and *Klf4* (OSCK) were induced to reprogram by the addition of dox. On days 4, 6 and 10 post-induction, control siRNAs or the siRNA cocktail targeting 3XHMT or *Cbx3* was transfected. To determine when reprogramming-factor-independent GFP-positive colonies can be obtained, dox was withdrawn on days 6, 7, 8, 9 and 10, respectively, and in all cases the number of GFP-positive colonies was determined on day 15. On dox withdrawal, siRNA treatments were not continued. (**c**) Overexpression of the histone H3K9 demethylase *Jmjd2c* enhances reprogramming. *Oct4*–GFP reporter MEFs were reprogrammed with OSK as in **a** with the addition of *Tomato* (control) or *Jmjd2c*-expressing retroviruses. The number of GFP-positive colonies obtained was quantified on day 25. Data from 2 technical replicates (Rep1 and Rep2) from 1 out of 3 representative experiments are presented.

To more directly address whether the reprogramming phenotypes observed above are linked to late reprogramming steps, we next investigated whether pluripotency could be induced in pre-iPSCs by modulating HP1 or H3K9-HMTase levels. Knockdown of any of the three H3K9-HMTases or of any of the HP1 proteins in

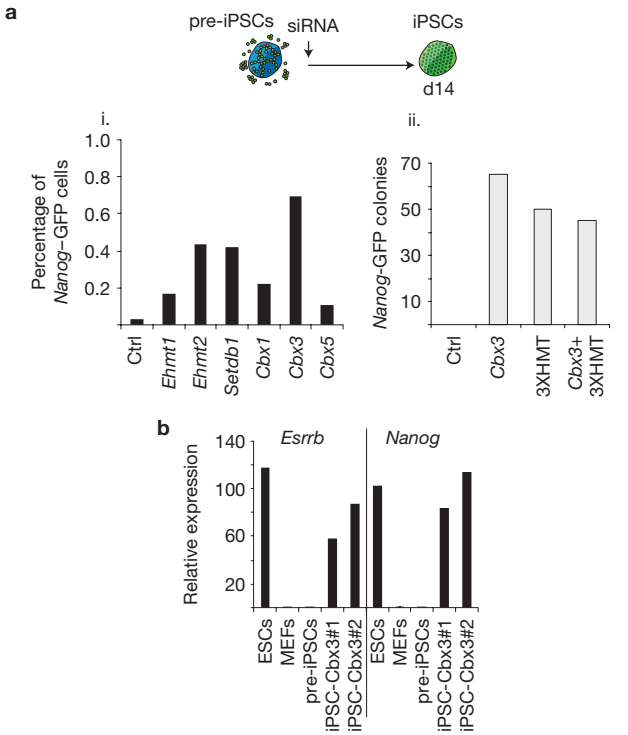

Figure 4 Interference with H3K9 methyltransferases or Cbx3 induces reprogramming in pre-iPSCs. (**a**) pre-iPSCs derived from *Nanog*–GFP reporter MEFs by retroviral expression of *Oct4, Sox2, Klf4* and *cMyc* were transfected once with the indicated siRNAs. Reprogramming efficiency was determined by FACS analysis of GFP-positive cells on day 14 post siRNA transfection (i) or by counting the number of GFP-positive colonies on day 15 (ii). Note that colony numbers presented for *Cbx3* and 3XHMT single knockdowns are the same as in Fig. 6d(iii), tetO–Nanog(-)/dox(-) condition. Repeated transfection of the siRNAs did not result in an increase in the number of GFP-positive cells. (**b**) *Nanog*–GFP-positive ESC-like colonies obtained on *Cbx3* knockdown in pre-iPSCs were expanded and two lines (iPSCs-Cbx3 #1 and #2) analysed by quantitative real-time PCR for transcript levels of the pluripotency genes *Esrrb* and *Nanog*. For comparison, ESCs, MEFs and the starting pre-iPSCs, which were set to 1, are included in the analysis. Data presented are the mean of three technical replicates from one experiment.

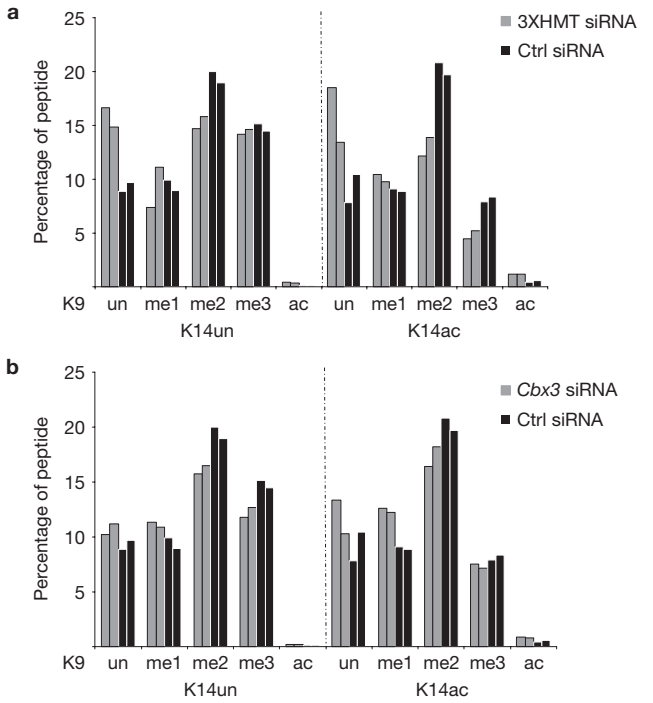

Figure 5 Depletion of the H3K9 HMTases but not of *Cbx3* induces a change in PTMs on the histone H3K9/K14 peptide in pre-iPSCs. (**a**) Percentages of the different isoforms of the H3 peptide containing amino acids 9 to 17 (in which K9 and K14 are modified) in pre-iPSCs, 3 days post-transfection of control and 3XHMT siRNAs, determined by qMS. The data for two biological replicates are given. Source data are provided in Supplementary Table S2. (**b**) The same as in **a**, except that the qMS approach was performed on *Cbx3* knockdown.

pre-iPSCs led to an increase in the number of *Nanog*–GFP-positive cells (Fig. 4a(i)). Among the HP1 family members the promoting effect was again most pronounced for *Cbx3* depletion (Fig. 4a(i)). *Cbx3* knockdown enhanced the appearance of GFP-positive colonies with an ESC-like morphology to a similar extent as the knockdown of all three H3K9 HMTases together (Fig. 4a(ii) and Supplementary Fig. S3C). However, combined knockdown of *Cbx3* and 3XHMT did not enhance colony formation further (Fig. 4a(ii)), suggesting that at least some of the events regulated by Cbx3 and the HMTases during reprogramming are overlapping. GFP-positive colonies isolated from the *Cbx3*- siRNA-treated pre-iPSC cultures were expanded and exhibited high expression levels of pluripotency markers such as *Esrrb* and *Nanog* (Fig. 4b) and silencing of the retrovirally encoded *Oct4* and *Sox2* transgenes (Supplementary Fig. S3D), satisfying hallmarks of pluripotency. The finding that decreasing expression levels of *Cbx3* or H3K9 HMTases enhance iPSC formation from pre-iPSCs indicates that these proteins constitute a barrier to late reprogramming events.

Functions of H3K9-HMTases and *Cbx3* during reprogramming

Next, we performed qMS analysis of histone PTMs in pre-iPSCs three days after initiation of 3XHMT or *Cbx3* knockdown to gain insight into the molecular mechanisms of how these regulators promote late reprogramming events (MS data are summarized in Supplementary Table S2). We reasoned that the analysis of histone PTMs shortly after initiation of knockdown but before *Nanog*–GFP expression was detectable (which first appeared seven days after the introduction of siRNAs (Supplementary Fig. S3C)), would reveal direct effects on global chromatin character due to depletion of the factors.

On 3XHMT knockdown, H3K9me2 levels decreased strongly, irrespective of the presence or absence of H3K14 acetylation on the same histone molecule, whereas H3K9me3 decreased only when K14 was acetylated as well (Fig. 5a), indicating that the 3XHMT knockdown has specific effects in the context of combinatorial histone modifications. The decrease in H3K9me2/me3 was accompanied by a significant gain in the four unmethylated isoforms of the H3K9/K14 peptide: K9un/K14un, K9un/K14ac, K9ac/K14un and K9ac/K14ac (Fig. 5a). These changes in global chromatin state on the H3K9/K14 peptide trended towards the pattern seen in iPSCs (compare Fig. 5a with Fig. 1e). With the exception of a few low-abundance histone PTMs, we detected no other marked changes in the histone PTM profile on 3XHMT knockdown (Supplementary Table S2), indicating that the PTM state of the H3K9 site does not immediately affect most other histone PTMs. On *Cbx3* knockdown most isoforms of the H3K9/K14 peptide did not change significantly in abundance, except those

containing H3K9ac with and without K14ac (Fig. 5b). We conclude that *Ehmt1*, *Ehmt2* and *Setdb1*, but not *Cbx3*, directly contribute to the regulation of global H3K9me2/me3 levels in pre-iPSCs, and that a change in global H3K9me levels itself is not sufficient for the induction of pluripotency as additional time in culture is required for the efficient activation of the pluripotency network.

We also performed genome-wide transcriptional profiling on pre-iPSCs three days after transfection of the siRNAs, to further understand the role of the H3K9-HMTases and Cbx3 in reprogramming. Relatively few genes were differentially expressed in pre-iPSCs depleted for 3XHMT or *Cbx3* (3XHMT siRNA: 222 genes 1.5-fold up and 261 genes 1.5-fold down; *Cbx3* siRNA: 352 genes 1.5-fold up and 368 genes 1.5-fold down), and about a fifth of the up- and downregulated genes, respectively, changed their expression in the same direction between *Cbx3* or 3XHMT knockdown (Fig. 6a and Supplementary Table S3). Further analysis demonstrated that the 3XHMT knockdown drives the gene expression program of pre-iPSCs more strongly towards the iPSC expression pattern than *Cbx3* depletion (Fig. 6b). Accordingly, genes upregulated on 3XHMT knockdown are more highly expressed in pluripotent cells than pre-iPSCs and downregulated genes are expressed at significantly lower levels in pluripotent cells than pre-iPSCs (Fig. 6c). For the *Cbx3* knockdown this trend was seen only for the downregulated genes (Fig. 6c). Interestingly, the 56 genes downregulated both in the 3XHMT and *Cbx3* knockdown included *Tgfβ2* and 49 of these genes were also expressed at lower levels in iPSCs than pre-iPSCs (Fig. 6a), suggesting that the suppression of these genes may be important for the reprogramming enhancement observed on these knockdowns. Consistent with this, TGFβ signalling is already known to inhibit reprogramming[41,42]. Inspecting the differentially expressed genes for other critical regulators of reprogramming, we found the pluripotency factor *Nanog*, which has previously been shown to be essential for the final reprogramming stage and to enhance reprogramming when overexpressed[43,44], to be among the most upregulated genes in both 3XHMT- and *Cbx3*-depleted pre-iPSCs (Fig. 6a,b; 7- and 10-fold up, respectively). Additional pluripotency factors including *Gdf3*, *Zfp42*, *Dppa4* and *Lin28* were among the upregulated genes in pre-iPSCs specifically on the 3xHMT knockdown (Fig. 6a and Supplementary Table S3). We conclude that depletion of *Cbx3* or 3XHMT yields partially overlapping gene expression changes in pre-iPSCs that converge on the induction of *Nanog* and the downregulation of genes that become reduced during the transition to the pluripotent state. Our finding that the *Cbx3* and 3XHMT knockdowns are not additive in their reprogramming enhancement (Fig. 4b) is consistent with the idea that these knockdowns may enhance iPSC formation through overlapping transcriptional responses.

We therefore explored the contribution of *Nanog* upregulation to the reprogramming enhancement on 3xHMT or *Cbx3* knockdown. As loss-of-function of *Nanog* prevents the establishment of iPSCs (ref. 43), we did not test the consequence of 3XHMT or *Cbx3* knockdown in pre-iPSCs lacking Nanog, but instead combined the knockdowns with *Nanog* overexpression. pre-iPSCs carrying a doxycycline-inducible *Nanog* transgene were transfected with siRNAs targeting 3XHMT or *Cbx3* (Fig. 6d(i)). Immunostaining indicated that over 90% of the infected cells expressed *Nanog* on addition of doxycycline, and no expression in the absence of doxycycline (Fig. 6d(ii)). By itself, overexpression of *Nanog* resulted in a strong induction of reprogrammed colonies similar to that seen on 3XHMT or *Cbx3* knockdown, indicating that high Nanog levels can efficiently convert our pre-iPSCs to iPSCs (Fig. 6d(iii)). Importantly, the 3XHMT or *Cbx3* knockdowns conferred only a further twofold enhancement in iPSC colony formation to *Nanog*-expressing pre-iPSCs (Fig. 6d(iii)), consistent with the interpretation that *Nanog* upregulation is a key downstream event in the enhancement of reprogramming on 3XHMT or *Cbx3* depletion.

Nanog is a direct target of Cbx3 in pre-iPSCs

Our data predicted that *Nanog* is a target of Cbx3 and H3K9 methyltransferases during reprogramming. To this end, we determined the genomic Cbx3 binding sites in pre-iPSCs and ESCs using ChIP-seq (Supplementary Table S4). For each cell type, data from two biological replicates were merged for further analysis because they correlated well. We found that Cbx3 occupies regions upstream of the transcriptional start site (TSS) at the repressed *Nanog* locus in pre-iPSCs, overlapping with known upstream regulatory sites. In ESCs, where Nanog is strongly expressed, Cbx3 binding is absent from these regulatory regions (Fig. 7a). Given that *Nanog* is the most upregulated gene on *Cbx3* knockdown in pre-iPSCs, these data indicate that *Cbx3* directly represses *Nanog* in pre-iPSCs. Notably, the genomic region upstream of the TSS of *Nanog* is also enriched for H3K9me3 in pre-iPSCs, partially overlapping with Cbx3 occupancy (Supplementary Fig. S4A), in agreement with findings that showed H3K9 methylation upstream of the *Nanog* promoter in differentiating ESCs (ref. 39). Together, these findings indicate that Cbx3 functions with H3K9me3 to repress *Nanog* in the reprogramming process, strengthening our conclusion that *Nanog* is an important downstream target on which Cbx3 and the regulation of H3K9 methylation converge during reprogramming.

Cbx3 predominantly binds highly expressed genes

Although the *Nanog* locus is bound by Cbx3 at upstream regulatory regions in pre-iPSCs, genome-wide Cbx3 predominantly occupies genic regions in both pre-iPSCs and ESCs (Fig. 7b). The number of target genes with significant Cbx3 enrichment was markedly lower in ESCs compared with pre-iPSCs (Supplementary Fig S4b), and, across genes, Cbx3 exhibited a distinct binding pattern between these two cell types (Fig. 7c–e). Specifically, in ESCs, Cbx3 binding within genes increases from the TSS throughout the gene to the 3′ end, a pattern consistent with recent reports on Cbx3 binding in cancer cell lines[35]. Most of the target genes of Cbx3 in ESCs are also bound in pre-iPSCs (Supplementary Fig. S4B), typically associate with Cbx3 across the gene body (Fig. 7e), and function in ribosome biogenesis, gene expression and nucleosome assembly based on GO analysis. However, in pre-iPSCs Cbx3 also occupies the TSS regions of a large number of genes (Fig. 7d,e). Despite the different binding pattern, Cbx3-bound genes are on average expressed at significantly higher levels than their unbound counterparts in both cell types (Fig. 7f). Grouping Cbx3-bound genes in ESCs and pre-iPSCs into expression tiers also demonstrated that strong gene body and TSS occupancy of Cbx3 favours highly expressed genes (Supplementary Fig. S4C,D). On the basis of these findings and published reports[35], we conclude that Cbx3 typically associates with gene bodies of highly transcribed genes. In addition our data uncover an unexpectedly strong association of Cbx3 with the TSS of expressed genes specifically in pre-iPSCs, which we also observed in early reprogramming intermediates.

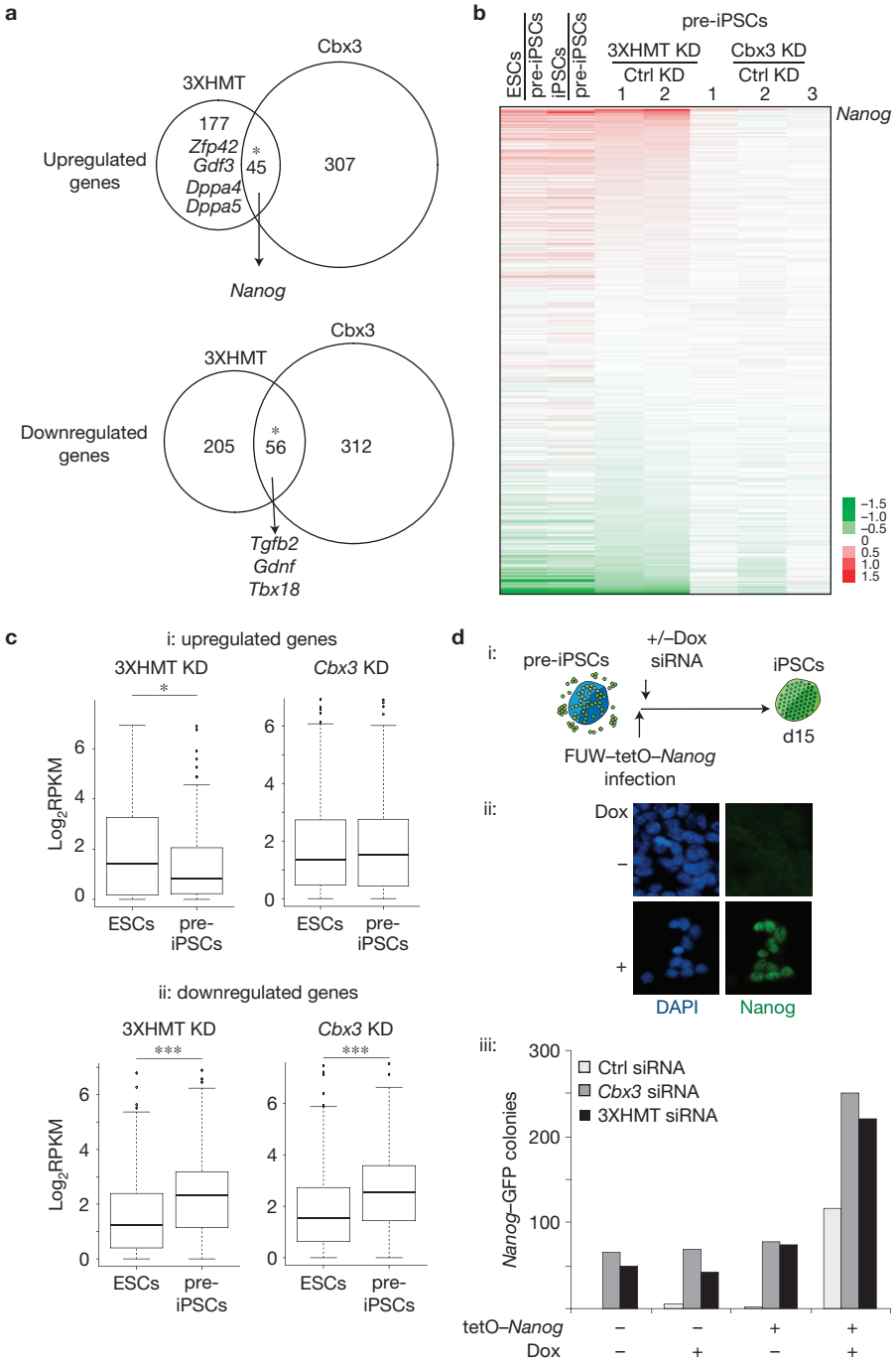

Figure 6 Depletion of H3K9 HMTases and *Cbx3* promotes reprogramming of pre-iPSCs by inducing *Nanog* expression. (**a**) pre-iPSCs were transfected with control siRNAs or siRNAs targeting the three H3K9 HMTases (3XHMT) or *Cbx3* and genome-wide expression was determined 3 days post-transfection. The Venn diagrams show the overlap of genes that are up- or downregulated in pre-iPSCs following 3XHMT and *Cbx3* knockdown, respectively, compared with control transfected cells. The asterisks denote significant overlap determined by the hypergeometric test, with a P value of 5.19×10^{-31} and 5.84×10^{-39} for up- and downregulated genes, respectively. Interesting differentially expressed genes are indicated. Source data are provided in Supplementary Table S3. (**b**) Expression changes in pre-iPSCs on 3XHMTase or *Cbx3* knockdowns in relation to expression in pluripotent cells. The heat map shows the differential expression ($\log_2$) for all genes (16,333) between ESCs, iPSCs, two replicates of the 3XHMT knockdown in pre-iPSCs, and three replicates of *Cbx3* knockdown in pre-iPSCs, each relative to control siRNA-transfected pre-iPSCs. Genes are ranked on the basis of average expression differences caused by the 3XHMTase knockdown. Among the most upregulated genes is *Nanog*. Source data are provided in Supplementary Table S3. (**c**) (i) Box plot of the absolute gene expression levels in ESCs and pre-iPSCs for genes 1.5-fold upregulated on 3XHMT knockdown (left) and *Cbx3* knockdown (right) in pre-iPSCs; (ii) for genes 1.5-fold downregulated. Significance of difference in expression was determined by the Wilcoxon test, $^{*}P < 0.05$, $^{***}P < 0.01$. Source data are provided in Supplementary Table S3. (**d**) Combination of *Nanog* overexpression and 3XHMT or *Cbx3* knockdowns in pre-iPSCs. (i) pre-iPSCs derived from *Nanog*-GFP reporter MEFs that carried M2rtTA (reverse tetracycline transactivator) were infected with a doxycycline (Dox)-inducible lentivirus encoding *Nanog* cDNA and transfected once with siRNAs as indicated. (ii) Immunostaining for Nanog protein (green) of lentivirally infected cells in the presence or absence of dox. DAPI (blue) marks nuclei. (iii) Quantification of the number of GFP-positive colonies on day 15 post siRNA treatments for pre-iPSCs with and without the *Nanog* virus (tetO–*Nanog*), in the presence or absence of dox.

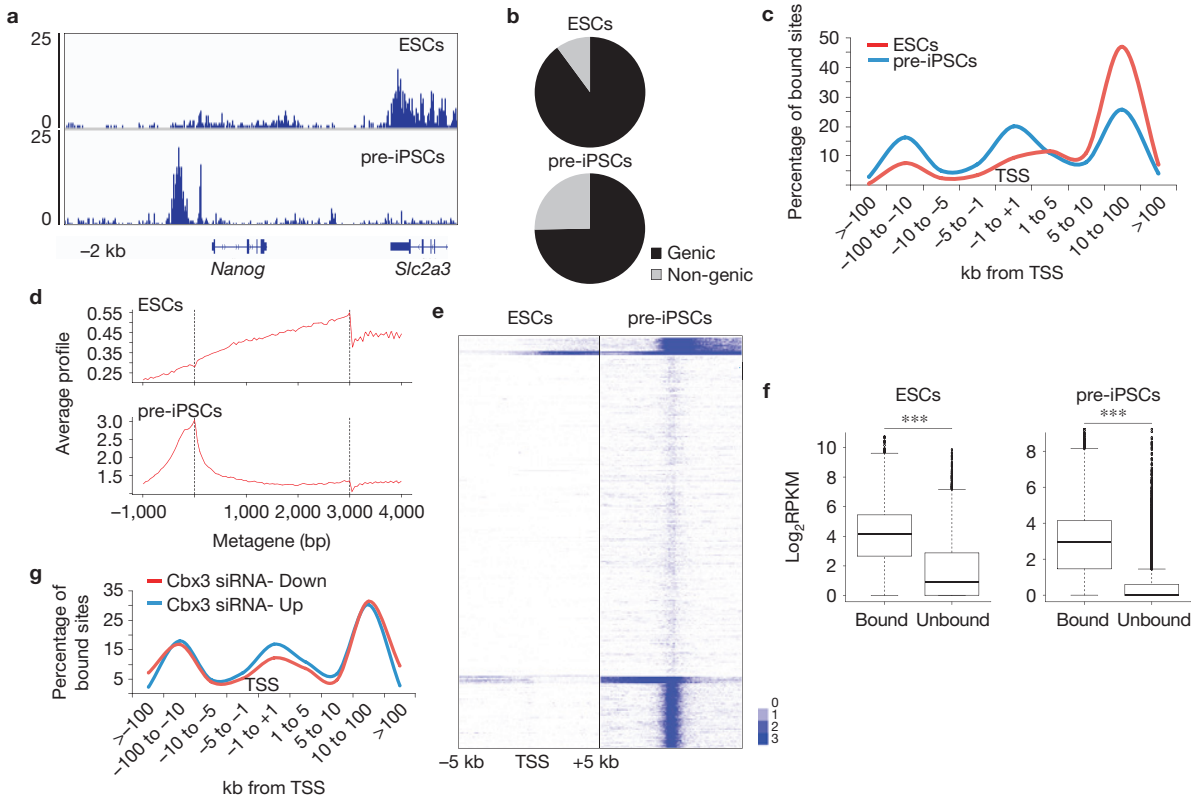

Figure 7 Cell-type-specific occupancy of Cbx3 at the TSS. (**a**) Snapshot of Cbx3 ChIP-seq data at the *Nanog* locus in pre-iPSCs and ESCs. Note that the neighbouring gene *Slc2a3* exhibits Cbx3 binding in the coding region in ESCs but not in pre-iPSCs, whereas *Nanog* exhibits the opposite behaviour. (**b**) Distribution of significant Cbx3 peaks between genic regions and non-genic regions, in both pre-iPSCs and ESCs, based on ChIP-seq. A Cbx3 peak was assigned to genic regions when it overlapped with a region from the TSS (transcriptional start site) to the 3′ end of the gene. (**c**) Frequency of sites bound significantly by Cbx3 with respect to the TSS is depicted for both pre-iPSCs and ESCs. (**d**) Average gene profiles for Cbx3 enrichment in pre-iPSCs and ESCs. The *y* axis represents average signal (tag count) from 100 bp windows. (**e**) Heat map of Cbx3 enrichment for all genes from 5 kb upstream to 5 kb downstream of the TSS in pre-iPSCs and ESCs, divided into five groups based on K-means clustering. Blue denotes enrichment in a 100 bp bin normalized to 1 million reads subtracted from normalized input values; values were $\log_2$ transformed. (**f**) Cbx3-occupied genes are more highly expressed than unbound genes. Box plot of absolute gene expression levels in ESCs and pre-iPSCs for Cbx3-bound and unbound genes. $\log_2$ expression levels are based on reads per kilobase per million mapped reads (RPKM) from RNA-seq data for these cell types. In this figure, genes were considered bound when Cbx3 was enriched anywhere in a region encompassing 10 kb upstream of the TSS and the coding region. $^{***}P < 0.01$, Wilcoxon test. (**g**) Distribution of Cbx3-binding sites with respect to distance to the TSS for genes that are up- and downregulated, respectively, in pre-iPSCs on *Cbx3* knockdown. All source data for this figure are provided in Supplementary Table S4.

Despite the widespread binding of Cbx3 to actively transcribed genes in pre-iPSCs only a relatively small number of genes were differentially expressed on *Cbx3* depletion (Fig. 6a). A greater proportion of downregulated than upregulated genes was directly bound by Cbx3 (256 and 196 genes, respectively). Therefore, Cbx3 can both positively and negatively affect its target genes, probably depending on the exact context of its binding. Consistent with this, upregulated genes have slightly more pronounced binding of Cbx3 at the TSS and upstream region than downregulated genes (Fig. 7g).

Cbx3 associates with the transcriptional initiation complex in a cell-type- and activator-dependent manner

Exploring the nature of the Cbx3 TSS enrichment in pre-iPSCs further, we found that the mediator co-activator complex, which is part of the RNA polymerase II pre-initiation complex (PIC; ref. 45), mimics the Cbx3-binding pattern at the TSS (Fig. 8a). The association of Cbx3 with the TSS in pre-iPSCs but not ESCs suggested that Cbx3 and mediator might interact in differentiated cells but not in pluripotent cells. To test this hypothesis, we immunoprecipitated the mediator complex from ESCs and ESC-derived differentiated cells, taking advantage of a tetracycline-inducible transgene encoding a Flag-tagged subunit of mediator (Flag–Med29), and determined whether Cbx3 could be detected in the immunoprecipitate (Fig. 8b). The results showed that Med29 can co-precipitate Cbx3 much more strongly in differentiated cells than ESCs, indicating that the interaction of mediator and Cbx3 is regulated in a cell-type-dependent fashion. We also studied the recruitment of Cbx3 to the PIC *in vitro*, employing nuclear extract, the model transcriptional activator GAL4–VP16 and an immobilized GAL4-responsive DNA template. The results show that Cbx3 is recruited from nuclear extracts to the chromatin template, but only on activator addition, as for mediator (Fig. 8c). These data suggest that Cbx3 is recruited to the TSS of transcribed genes in an activator-dependent manner by associating with the PIC, specifically in differentiated cells. It is conceivable that differential post-translational modifications, the presence of different splice forms of mediator subunits, or a different chromatin composition at the TSS in ESCs versus differentiated cells[46] could explain the cell type specificity of the Cbx3 association with the PIC. As the association of Cbx3 with the TSS in pre-iPSCs seems

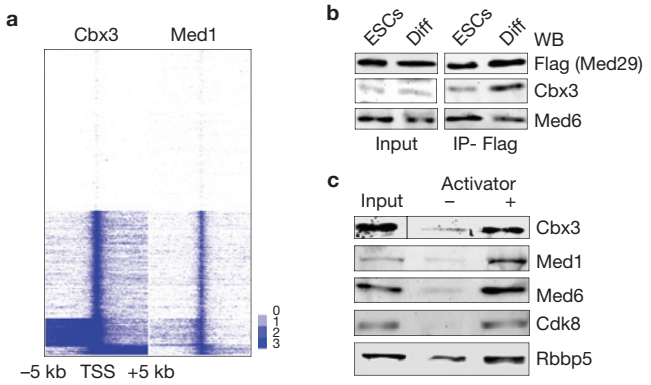

Figure 8 Cbx3 association with mediator components. (**a**) Cbx3 co-localizes with mediator at the TSS. Heat map of Cbx3 and mediator (Med1 subunit) enrichment in pre-iPSCs for all genes from 5 kb upstream to 5kb downstream of the TSS, divided into five groups based on K-means co-clustering, based on ChIP-seq. Blue denotes enrichment in a 100 bp bin normalized to 1 million reads subtracted from normalized input values; values were $\log_2$ transformed. (**b**) Cbx3 precipitates more efficiently with Med29 in differentiated cells than ESCs. Immunoprecipitates (IP) of Flag–Med29 from ESCs and differentiated cells (neural precursor cells (NPCs) differentiated from ESCs) were probed with antibodies against Cbx3, Flag or Med6 by western blotting (WB). 1% of input and 5% of IP are shown. Equal amounts of Cbx3 and Med6 are present in inputs from both cell types. Cbx3 but not the control Med26 is enriched in the Med29 IP in NPCs. (**c**) Cbx3 is recruited to the PIC in an activator-dependent manner. Transcriptional pre-initiation complexes (PICs) were formed in the presence and absence of a model activator GAL4–VP16 in nuclear extract. Note that Cbx3 was incorporated in the PIC only when the VP16 transcriptional activator was present, as for mediator (subunits Med1, Med6, Cdk8) and the MII complex component Rbbp5. Uncropped scans are available in Supplementary Fig. S5.

to occur independently of H3K9me3 (Supplementary Fig. S4E), the interaction with PIC components is probably responsible for the Cbx3 recruitment to the TSS. Taken together, our findings reveal a function of Cbx3 in the context of transcriptional regulation at the TSS that seems to be specific for differentiated cells and reprogramming intermediates, as Cbx3 occupancy at the TSS is lacking in pluripotent cells.

DISCUSSION

In summary, our qMS approach has yielded a quantitative and comprehensive analysis of global histone PTMs in differentiated and pluripotent cells, and during reprogramming. Our data indicate that the global histone PTM profile changes late in the reprogramming process, which may be associated with the efficient activation of the pluripotency network, major replication timing changes within early embryonic genes[24], and the reactivation of the inactive X chromosome[47]. We speculate that global chromatin and replication-timing reorganization are key aspects of the final reprogramming stage, required for establishment of the self-sustaining pluripotency network.

The quantitative repository of global histone modification changes generated here can be used as a starting point for further dissection of reprogramming roadblocks and epigenetic differences between pluripotent and differentiated cells. On the basis of this idea, we determined the role of H3K9 methylation during reprogramming by analysing proteins involved in the regulation of this histone modification. Despite the fact that our work reveals various distinct functions for Cbx3 and the methyltransferase genes *Ehmt1*, *Ehmt2* and *Setdb1* during reprogramming, reflected in the differential control of global H3K9 methylation levels or association with the basic transcriptional machinery, the removal of the three H3K9-HMTases or *Cbx3* elicited partially overlapping transcriptional responses including the reactivation of the silent *Nanog* locus late in reprogramming. Our data suggest that these common expression changes cause reprogramming enhancement.

By examining the role of Cbx3, we found a remarkable switch in the location of Cbx3 between pluripotent and non-pluripotent cells, and a physical association of Cbx3 with the mediator complex that could be responsible for targeting Cbx3 to the TSS in non-pluripotent cells. As only a small subset of its target genes become differentially expressed in pre-iPSCs on knockdown of *Cbx3*, we speculate that Cbx3 binding at the TSS mediates more subtle functions and acts perhaps to maintain or restore nucleosome density near the promoter during transcription to protect the transcribed DNA, or to control aspects of transcriptional elongation. Alternatively, different HP1 family members may also act at the TSS and mask the effects of Cbx3. Consistent with this, another HP1 family member, HP1α, has been shown to localize to the TSS of transcribed genes in *Drosophila*[48], suggesting an evolutionarily conserved role of HP1 family members in the PIC. Understanding the mechanistic basis for the switch in Cbx3 localization between pluripotent and non-pluripotent cells will reveal further insight into the nature of the pluripotent state. □

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Supplementary Information is available in the online version of the paper

ACKNOWLEDGEMENTS

We thank V. Pasque for critical reading of the manuscript and M. Grunstein (UCLA) for providing antibodies. K.P. is supported by the Eli and Edythe Broad Center of Regenerative Medicine and Stem Cell Research at UCLA, NIH (DP2OD001686 and P01 GM099134) and CIRM (RN1-00564); R.S. was supported by the Jonsson Comprehensive Cancer Center, C.C. by a Leukaemia and Lymphoma Research Grant (10040), G.B. by the Whitcome Pre-doctoral Training Program, B.A.G. by a National Science Foundation Early Faculty CAREER award, an NIH Innovator award (DP2OD007447) and NIH (P01 GM099134) and M.C. by the NIH (GM074701). R.S., C.C. and S.P. were supported by CIRM training grants.

AUTHOR CONTRIBUTIONS

R.S., K.P. and B.A.G. planned the project. R.S. and K.P. wrote the manuscript. The following performed experiments, analysed and interpreted data: R.S., C.C., G.B., R.M., S.P. under K.P.'s supervision, M.G-C. under B.A.G.'s supervision, C.H. under M.C.'s supervision, and D.L. under M.P.'s supervision. N.M. generated H3K18me1 antibody.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Published online at www.nature.com/doifinder/10.1038/ncb2768
Reprints and permissions information is available online at www.nature.com/reprints

1. Plath, K. & Lowry, W. E. Progress in understanding reprogramming to the induced pluripotent state. *Nat. Rev. Genet.* **12,** 253–265 (2011).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
3. Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454,** 49–55 (2008).
4. Ang, Y-S. *et al.* Wdr5 mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145,** 183–197 (2011).
5. Gaspar-Maia, A. *et al.* Chd1 regulates open chromatin and pluripotency of embryonic stem cells. *Nature* **460,** 863–868 (2009).
6. Singhal, N. *et al.* Chromatin-remodeling components of the baf complex facilitate reprogramming. *Cell* **141,** 943–955 (2010).

7. Wang, T. *et al.* The histone demethylases Jhdm1a/1b enhance somatic cell reprogramming in a vitamin-C-dependent manner. *Cell Stem Cell* **9,** 575–587 (2011).
8. Liang, G., He, J. & Zhang, Y. Kdm2b promotes induced pluripotent stem cell generation by facilitating gene activation early in reprogramming. *Nat. Cell Biol.* **14,** 457–466 (2012).
9. Onder, T. T. *et al.* Chromatin-modifying enzymes as modulators of reprogramming. *Nature* **483,** 598–602 (2012).
10. Chen, J. *et al.* H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. *Nature Genet.* **45,** 34–42 (2013).
11. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151,** 994–1004 (2012).
12. Mattout, A., Biran, A. & Meshorer, E. Global epigenetic changes during somatic cell reprogramming to iPS cells. *J. Mol. Cell Biol.* **3,** 341–350 (2011).
13. Maherali, N. *et al.* Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution. *Cell Stem Cell* **1,** 55–70 (2007).
14. Chin, M. H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5,** 111–123 (2009).
15. Sridharan, R. *et al.* Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136,** 364–377 (2009).
16. Garcia, B. A. *et al.* Chemical derivatization of histones for facilitated analysis by mass spectrometry. *Nat. Protoc.* **2,** 933–938 (2007).
17. Plazas-Mayorca, M. D. *et al.* One-pot shotgun quantitative mass spectrometry characterization of histones. *J. Proteome Res.* **8,** 5367–5374 (2009).
18. Wang, Z. *et al.* Combinatorial patterns of histone acetylations and methylations in the human genome. *Nat. Genet.* **40,** 897–903 (2008).
19. Karmodiya, K. *et al.* H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genom.* **13,** 424 (2013).
20. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–283 (2011).
21. Xie, W. *et al.* Histone h3 lysine 56 acetylation is linked to the core transcriptional network in human embryonic stem cells. *Mol. Cell* **33,** 417–427 (2009).
22. Wang, Y. *et al.* Histone H3 lysine 14 acetylation is required for activation of a DNA damage checkpoint in fission yeast. *J. Biol. Chem.* **287,** 4386–4393 (2012).
23. Meshorer, E. *et al.* Hyperdynamic plasticity of chromatin proteins in pluripotent embryonic stem cells. *Dev. Cell* **10,** 105–116 (2006).
24. Hiratani, I. *et al.* Genome-wide dynamics of replication timing revealed by *in vitro* models of mouse embryogenesis. *Gen. Res.* **20,** 155–169 (2010).
25. Steger, D. J. *et al.* DOT1L/KMT4 recruitment and H3K79 methylation are ubiquitously coupled with gene transcription in mammalian cells. *Mol. Cell. Biol.* **28,** 2825–2839 (2008).
26. Krogan, N. J. *et al.* Methylation of histone H3 by Set2 in *Saccharomyces cerevisiae* is linked to transcriptional elongation by RNA polymerase II. *Mol. Cell. Biol.* **23,** 4207–4218 (2003).
27. Zee, B. M. *et al.* *In vivo* residue-specific histone methylation dynamics. *J. Biol. Chem.* **285,** 3341–3350 (2010).
28. Efroni, S. *et al.* Global transcription in pluripotent embryonic stem cells. *Cell Stem Cell* **2,** 437–447 (2008).
29. Golipour, A. *et al.* A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network. *Stem Cell* **11,** 769–782 (2012).
30. Polo, J. M. *et al.* A molecular roadmap of reprogramming somatic cells into iPS cells. *Cell* **151,** 1617–1632 (2012).
31. Silva, J. *et al.* Promotion of reprogramming to ground state pluripotency by signal inhibition. *PLoS Biol.* **6,** e253 (2008).
32. Bannister, A. J. *et al.* Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **410,** 120–124 (2001).
33. Lachner, M., O'Carroll, D., Rea, S., Mechtler, K. & Jenuwein, T. Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **410,** 116–120 (2001).
34. Kwon, S. H. & Workman, J. L. The changing faces of HP1: from heterochromatin formation and gene silencing to euchromatic gene expression. *Bioessays* **33,** 280–289 (2011).
35. Smallwood, A. *et al.* CBX3 regulates efficient RNA processing genome-wide. *Genome Res.* **22,** 1426–1436 (2012).
36. Bilodeau, S., Kagey, M. H., Frampton, G. M., Rahl, P. B. & Young, R. A. SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* **23,** 2484–2489 (2009).
37. Yuan, P. *et al.* Eset partners with Oct4 to restrict extraembryonic trophoblast lineage potential in embryonic stem cells. *Genes Dev.* **23,** 2507–2520 (2009).
38. Cloos, P. A. C. *et al.* The putative oncogene GASC1 demethylates tri- and dimethylated lysine 9 on histone H3. *Nature* **442,** 307–311 (2006).
39. Loh, Y. H., Zhang, W., Chen, X., George, J. & Ng, H. H. Jmjd1a and Jmjd2c histone H3 Lys 9 demethylases regulate self-renewal in embryonic stem cells. *Genes Dev.* **21,** 2545–2557 (2007).
40. Ma, D. K., Chiang, C-H. J., Ponnusamy, K., Ming, G-L. & Song, H. G9a and Jhdm2a regulate embryonic stem cell fusion-induced reprogramming of adult neural stem cells. *Stem Cell.* **26,** 2131–2141 (2008).
41. Li, R. *et al.* A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. *Cell Stem Cell* **7,** 51–63 (2010).
42. Ichida, J. K. *et al.* A small-molecule inhibitor of Tgf-β; signaling replaces Sox2 in reprogramming by inducing nanog. *Cell Stem Cell* **5,** 491–503 (2009).
43. Silva, J. *et al.* Nanog is the gateway to the pluripotent ground state. *Cell* **138,** 722–737 (2009).
44. Hanna, J. *et al.* Direct cell reprogramming is a stochastic process amenable to acceleration. *Nature* **462,** 595–601 (2009).
45. Thomas, M. C. & Chiang, C-M. The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* **41,** 105–178 (2006).
46. Griffiths, D. S. *et al.* LIF-independent JAK signalling to chromatin in embryonic stem cells uncovered from an adult stem cell disease. *Nat. Cell Biol.* **13,** 13–21 (2010).
47. Stadtfeld, M., Maherali, N., Breault, D. T. & Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. *Cell Stem Cell* **2,** 230–240 (2008).
48. Yin, H., Sweeney, S., Raha, D., Snyder, M. & Lin, H. A high-resolution whole-genome map of key chromatin modifications in the adult *Drosophila melanogaster*. *PLoS Genet.* **7,** e1002380 (2011).

METHODS

Histone sample preparation for quantitative mass spectrometry (qMS). Cell pellets were lysed, nuclei isolated and histones extracted as previously described[17]. For each sample, approximately 100 µg of extracted histones was re-suspended in 30 µl of 100 mM ammonium bicarbonate, at pH 8.0. Chemical propionylation derivatization, digestion and desalting of histones was performed as described previously[27], except that histones were digested for 6 h. We performed both label-free and isotopically labelled peptide relative quantification. For isotopically labelled peptide comparative MS analysis, d_0- and d_{10}-propionic anhydride were used as previously described[16]. All proteomics data are available at the Stem Cell Omics Repository at http://scor.chem.wisc.edu/.

Mass spectrometry and data analysis. Samples were analysed by LC–MS and MS/MS as described previously[16]. In brief, digested samples were loaded by an Eksigent AS2 autosampler onto 75-µm-ID fused-silica capillary columns packed with 12 cm of C18 reversed-phase resin (Magic C18, 5 µm particles, Michrom BioResources), constructed with an electrospray ionization tip. Peptides were separated by nanoflowLC and introduced into a hybrid linear quadrupole ion trap-Orbitrap mass spectrometer (ThermoElectron), and resolved with a gradient from 5 to 35% buffer B in a 110-min gradient (buffer A: 0.1 M acetic acid; buffer B: 70% acetonitrile in 0.1 M acetic acid) with a flow rate of 150 nl min^{-1} on an Agilent 1200 binary HPLC system. The Orbitrap was operated in data-dependent mode essentially as previously described[16]. Relative abundances of peptide species were calculated by chromatographic peak integration of full MS scans using an in-house developed computer program. Peptide identity and modifications were verified by manual inspection of MS/MS spectra. Cluster 3.0 was used to create hierarchical clustering of ratio data and Java Treeview for visualization of the output.

Cell lines used for analysis of histone PTMs. The following cell lines were used for histone PTM qMS analysis in Figs 1 and 2: a female iPSC (2D4) line generated by retroviral expression of Oct4, Sox2, cMyc and Klf4 (ref. 13); a male iPSC line (C3) obtained on retroviral expression of Oct4, Sox2 and Klf4 (that is, in the absence of cMyc); a female pre-iPSC line (1A2; ref. 13) and a male pre-iPSC line (12-1) both obtained on retroviral expression of Oct4, Sox2, Klf4 and cMyc in Nanog–GFP reporter MEFs. In addition, we used the male ESC line V6.5, and male and female wild-type MEFs from day 13.5 embryos. ESCs, iPSCs and pre-iPSCs were grown in standard mouse ESC media and MEFs in the same media lacking LIF.

Reprogramming experiments. Reprogramming experiments were carried out from *Oct4*–GFP (ref. 49) or *Nanog* GFP (ref. 13) reporter MEFs using pMX retroviruses encoding Oct4, Sox2 and Klf4 as described previously[15] and conducted in media containing 15% serum (FBS). MEFs containing a single polycistronic, tet-inducible cassette carrying the four reprogramming factors Oct4, Sox2, Klf4 and cMyc in the Col1A locus, the tet-transactivator M2rtTA in the R26 locus, and the *Oct4*–GFP reporter were generated as described previously[50], and induced to reprogram with 2 µg ml^{-1} doxycycline. Reprogramming was scored by counting the number of GFP-positive ESC-like colonies at indicated days. All reprogramming experiments from fibroblasts and pre-iPSC experiments were done in biological triplicates, and for each figure, error bars represent standard deviation from two technical replicates of a representative experiment. For pre-iPSC experiments, reprogramming to iPSCs on siRNA knockdown was assessed by counting Nanog–GFP-positive colonies or quantifying GFP-positive cells by FACS at the indicated days. For FACS analysis, 12-1 pre-iPSCs were collected with trypsin, passed through a 40 µm cell strainer to obtain single-cell suspensions and analysed on an LSR cytometer (BD Biosciences). Data were analysed using the FlowJo software (TreeStar). Fuw–tetO–loxp–mNANOG was created by ligation-independent cloning (Infusion, Clontech Mountain View) by digesting the vector Fuw–tetO–loxp–hKLF4 (Addgene 20727) with EcoRI and PCR-amplifying mouse Nanog from pMX–mNANOG (Addgene 13354). The vector was co-transfected with pCMV–delta8.9 and pCAGS–VSVg (gift from D. Kohn, UCLA, USA) into HEK293T cells and viral conditioned media collected 48 h post transfection in serum-free media (Ultraculture, Lonza). Expression was confirmed by immunostaining. For overexpression, Jmjd2c was cloned from a PCR product using gene-specific primers (forward: 5′-ATGCGAATTCATGGAGGTGGTGGAGGTG-3′; reverse: 5′-ATGCGCGGCCGCCTACTGTCTCTTCTGACA-3′) into the EcoRI and NotI sites of the pMX vector. Expression was confirmed by quantitative real-time PCR performed with gene-specific primers RNA three days after transduction (forward: 5′-GGCCATGGAAGTAACCTTGA-3′; reverse: 5′-GAGGCTTACCAAGTGGATGG-3′).

siRNA transfection. Sets of four different siRNAs were purchased from Dharmacon and transfected using Lipofectamine RNAiMAX (Life Technologies) according to the manufacturers instructions. Of the set of four siRNAs, the one producing the most efficient knockdown was used in reprogramming experiments at a final concentration of 20 µM: Cbx1, MU-060281-01 #2; Cbx3, MU-044218-01 #2; Cbx5, MU-040799-01 #2; Setdb1, MU-040815-01 #4; Ehmt1, LU-059041-01 #3; Ehmt2, MU053728-00 #3. For control siRNA treatments, we used the non-targeting luciferase control D-001210-02. The timing of siRNA transfections is indicated in each figure. For pre-iPSC reprogramming experiments, reverse transfection of siRNAs was performed once, on 200,000 cells of the 12-1 pre-iPSC line, plated on gelatin. To test knockdown efficiency during reprogramming, RNA was collected three days after the first transfection and on day 22, that is, 3 days after the last transfection, and quantitative real-time PCR was performed with the gene-specific primers listed below. Cbx1 forward: 5′-GGAGAGGAAAGCAAACCAAA-3′, reverse: 5′-AGCTCCAATAATCCGCTCTG-3′; Cbx3 forward: 5′-CTGGACCGTCGTGTAGTGAA-3′, reverse: 5′-AAATTTTCT-TCTGGTTCCCAG-3′; Cbx5 forward: 5′-GGAAATCCAGTTTCTCCAACA-3′, reverse: 5′-GCTCCGATGATCTTTTCTGG-3′; Ehmt1 forward: 5′-TTGCTGCATG-AAAACTGAGC-3′, reverse: 5′-CAAGACCCATTTGTTTCTCCA-3′; Ehmt2 forward: 5′-CATGTCCAAACCTAGCAACG-3′, reverse: 5′-CCAGAGTTCAGCT-TCCTCCTT-3′; Setdb1 forward: 5′-GCAACTCAGAACCCGTCCTA-3′, reverse: 5′-ATAGGCTGTAGGGGCTCCAT-3′.

Expression analysis and data processing. To determine the transcriptional changes on 3XHMT and *Cbx3* knockdown in pre-iPSCs, total RNA was extracted from the pre-iPSC line 12-1 three days after the cells were subjected to transfection with control siRNAs (in biological triplicates), *Cbx3* siRNA (in biological triplicates), or a pool of *Ehmt1* siRNA, *Ehmt2*, siRNA and *Setdb1* siRNA (in biological duplicates), and analysed on an Affymetrix GeneChip Mouse Genome 430 2.0 array at the UCLA Clinical Microarray core facility. Quantile normalization was performed using the Affymetrix package (affy) from Bioconductor. To convert probe data into gene expression data, probes ending in _at and _a_at were averaged for each gene. Data at the probe level were normalized with previously published data for ESCs, iPSCs and MEFs (ref. 15). Data are deposited in the GEO database under GSE44084.

RNA-seq was performed using 4 µg of messenger RNA as starting material from ESCs and pre-IPSCs, using standard Illumina RNA-seq library construction protocols. Briefly, polyadenylated RNA was purified by two rounds of oligo-dT bead selection followed by divalent cation fragmentation under elevated temperature. Following cDNA synthesis with random hexamers, the double-stranded products were end repaired, a single A base was added, and Illumina adaptors were ligated onto the cDNA products. Ligation products with an average size of 300 base pairs (bp) were purified by means of agarose gel electrophoresis. The adaptor-ligated single-stranded complementary DNA was then amplified with 10 cycles of PCR. RNA-seq libraries were sequenced on Illumina HiSeq 2000. The RPKM (reads per kilobase of exon per million) was then computed for each gene.

Med29 immunoprecipitation and PIC capture assay. ESCs expressing FLAG–Med29 were obtained by targeting 3xFlag–Med29 under the control of a tet-inducible promoter into the ColA1 locus[51] in V6.5 ESCs carrying M2rtTA in the R26 locus. Targeting was confirmed by Southern blotting. Neural precursors were differentiated from these cells after suspension culture of embryoid bodies for four days, and selection in ITSF media for 6 days[52]. Nuclear extracts of ESCs and neural precursors were prepared and the purification of protein complexes containing Med29 was performed as previously described[53]. PIC assembly was performed from HeLa nuclear extract using the immobilized G5E4T and analysed by western blotting as described previously[54]. All primary antibodies were used at 1:1,000 dilution and secondary antibodies at 1:10,000 dilution: anti-FLAG from Sigma (F-1804), anti-Med1(sc-8898) and anti-Cdk8(sc-1521) from SantaCruz; anti-Cbx3 from Millipore; anti-Rbbp5 (Bethyl-A300-109) from Bethyl Labs.

ChIP-seq and ChIP-chip analysis. 12-1 pre-iPSCs or V6.5 ESCs were chemically crosslinked by the addition of formaldehyde to 1% final concentration for 10 min at room temperature, and quenched with 0.125 M final concentration glycine. Cells were washed twice in PBS, re-suspended in sonication buffer (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% Na-deoxycholate and 0.1% SDS), and sonicated with a Diagenode Bioruptor. Cell extracts were incubated with an antibody against Cbx3 (Millipore, 05-690; clone 42s2) or Med1 (sc-8998) overnight at 4 °C and immunoprecipitates were collected with magnetic beads. Beads were washed twice with RIPA buffer, low-salt buffer (20 mM Tris at pH 8.1, 150 mM NaCl, 2 mM EDTA, 1% Triton X-100, 0.1% SDS), high-salt buffer (20 mM Tris at pH 8.1, 500 mM NaCl, 2 mM EDTA, 1% Triton X-100 and 0.1% SDS), LiCl buffer (10 mM Tris at pH 8.1, 250 mM LiCl, 1 mM EDTA, 1% deoxycholate and 1% NP-40), and with 1× TE. Reverse crosslinking occurred overnight at 65 °C with 1% SDS and proteinase K. Illumina/Solexa sequence preparation, sequencing and

DOI: 10.1038/ncb2768

quality control were performed according to Illumina protocols, with the minor modification of limiting the PCR amplification step to 10 cycles.

Reads were mapped to the mm9 genome using the Bowtie software and only those reads that aligned to a unique position with no more than two sequence mismatches were retained for further analysis. Significant binding events were called as peaks using MACS2.0 using a false discovery rate of 0.05 and the –broadpeaks setting that allows calling of broader domains. Location analysis of called peaks was performed using the Sole-search tool. Visualization of the ChIP-seq signal around the TSS is provided by heat maps generated using Java Treeview. Briefly, enrichment is shown after normalization to 1 million reads and subtraction of normalized input values per 100 bp window. Data are deposited in the GEO database under GSE44242.

For ChIP-chip of H3K18me1 and H3K79me2, chromatin fragments (500 μg) from V6.5 ESCs were enriched with specific antibodies, labelled and hybridized, along with corresponding input fragments, to an Agilent promoter microarray (Agilent-G4490) that contains the promoter regions of 18,300 annotated mouse genes, encompassing regions 5.5 kb upstream to 2.5 kb downstream of the respective TSSs as described previously[15]. The H3K18me1 antibody was generated by N.M., and the H3K79me2 antibody was kindly provided by M. Grunstein at UCLA (USA). The ChIP-chip data sets for H3K4me3 and RNA PolII data have been previously published[5,15]. Hybridization onto the arrays, washing and scanning were done according to the manufacturer's protocols. Average probe signals were extracted in a 500 bp window-step-wise manner as described previously[15].

49. Szabó, P. E., Hübner, K., Schöler, H. & Mann, J. R. Allele-specific expression of imprinted genes in mouse migratory primordial germ cells. *Mech. Dev.* **115,** 157–160 (2002).
50. Stadtfeld, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nature Meth.* **7,** 53–55 (2009).
51. Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44,** 23–28 (2006).
52. Okabe, S., Forsberg-Nilsson, K., Spiro, A. C., Segal, M. & McKay, R. D. Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. *Mech. Dev.* **59,** 89–102 (1996).
53. Chen, X-F. *et al*. Mediator and SAGA have distinct roles in pol ii preinitiation complex assembly and function. *Cell Rep.* **2,** 1061–1067 (2012).
54. Lin, J. J. *et al*. Mediator coordinates PIC assembly with recruitment of CHD1. *Genes Dev.* **25,** 2198–2209 (2011).

DOI: 10.1038/ncb2768

nature cell biology

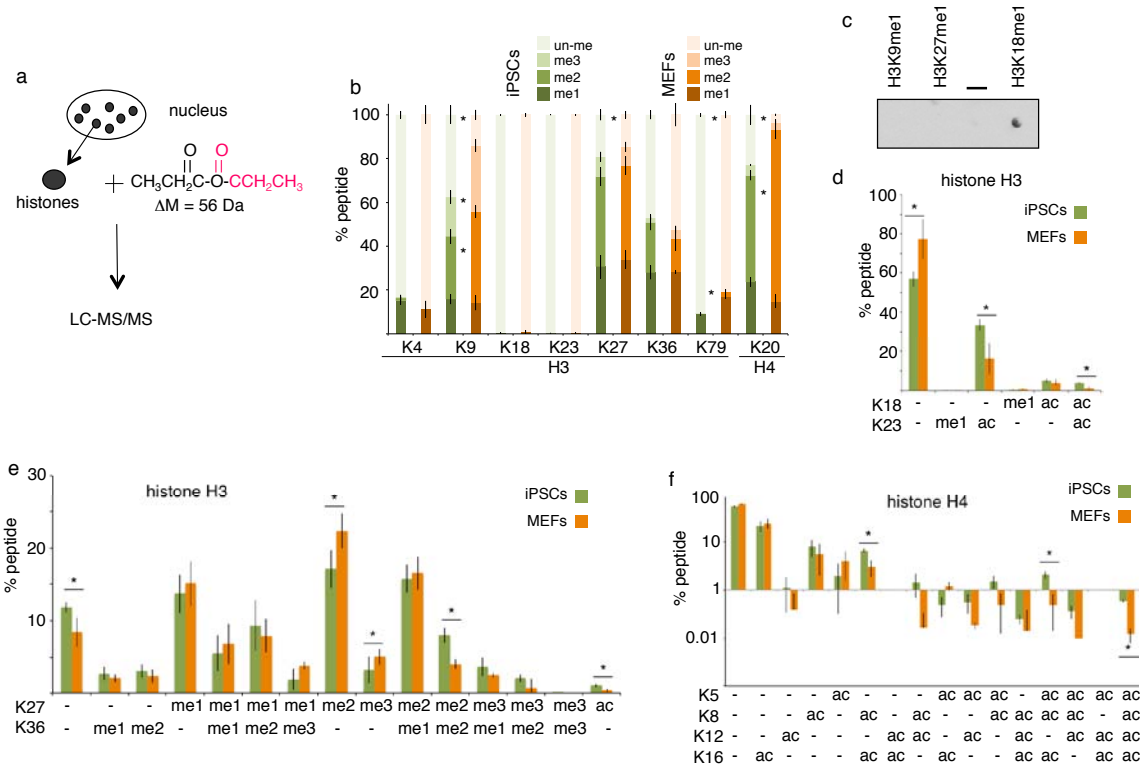

Figure S1 Characterization of changes in global histone PTM levels during reprogramming. **A)** Scheme of the qMS approach. Histones were extracted from nuclei of various cell types and derivatized individually with propionic anhydride, causing a change in mass of 56 daltons. Quantification was performed by LC MS/MS on an orbitrap mass spectrometer. Subsequent analysis of histone PTMs was limited to acetylation and methylation on histone H3 and H4. For each tryptic peptide of H3 and H4, the sum of the observed combinations of acetyl, methyl, or unmethylated/unacetylated PTMs (i.e. isoforms) was set to 100%. **B)** The graph supports Figure 1B of the main text and shows the percentage of the mono- (me1), di- (me2), tri-methylated (me3), and unmethylated forms of the lysines of histone H3 and H4 that exhibit methylation in female iPSCs and MEFs. The unmethylated portion represents the sum of acetylated and unacetylated isoforms (ac/un). Data presented is the mean of 6 values obtained from 2 biological replicates, each analyzed three times, and error bars represent the standard deviation, based on data presented in Table S1. Asterisks indicate significant differences between MEFs and iPSCs for the respective isoform, at $p<0.01$ as determined by the t-test. **C)** Characterization of the H3K18me1 antibody. A dot-blot containing H3K18me1, H3K9me1, and H3K27me1 peptide spots was incubated with a newly generated polyclonal H3K18me1 antibody to demonstrate the specificity of the antibody raised against H3K18me1 peptide. **D)** Graph displaying the percentages of all detected isoforms of the histone H3 peptide containing amino acid 18 to 26, in female iPSCs and MEFs. Data presented is the mean of 6 values obtained from 2 biological replicates, each analyzed three times based on data presented in Table S1. The asterisks indicate significantly different levels between iPSCs and MEFs at $p<0.01$, as determined by the t-test. Error bars = standard deviation of replicates. In this case, the lysine residues 18 and 23 can be acetylated (ac), methylated, or be in the un-acetylated/un-methylated (un) form. Notably, not all theoretically possible combinations of the modifications on K18 and K23 were found in our mass-spec analysis. Specifically, the K18me1/K23me1 isoform and the higher degrees of H3K18 or H3K23 methylation were not detected. We found a 3-fold increase of the di-acetylated H3K18ac/K23ac isoform of the peptide from MEFs to iPSCs, which was greater than that seen for the monoacetylated K18un/K23ac (2-fold) or K18ac/K23un (1.2-fold) isoforms, suggesting co-regulation of these two acetylation marks during reprogramming. **E)** As in (D), except for the percentages of the all detected isoforms of the histone H3 peptide containing amino acids 27 to 36. This peptide contains the lysine residues 27 and 36 that can be modified. H3K27ac, which is found at active enhancers, is more abundant in iPSCs than MEFs. H3K27me2 and me3, in the absence of acetylation or methylation on K36, are significantly higher in MEFs than iPSCs, but not when in the presence of the methylation marks on K36. **F)** As in (D), except for the percentages of the all detected isoforms of the histone H4 peptide containing amino acids 4 to 16, which contains the lysine residues 5,8,12, and 16, for which we detected only the acetylation PTM. Acetylation is indicated by "ac" and the unacetylated form by "-". 50% of the peptide was completely unacetylated. Among the mono-acetylated forms of this peptide, the acetylation levels decrease from K16 to K8, followed by K5, and minimal acetylation occurred at K12, with relatively small differences between iPSCs and MEFs. For isoforms containing dual acetylation, the acetylation of both K8 and K16 is as abundant as the isoform that contains acetylation on K8 only, and significantly more prominent in iPSCs than MEFs. Though the triple acetylated isoforms of this peptide are of relatively low abundance, the isoform carrying K8ac, K5ac, and K16ac is almost as abundant as the isoform carrying only K5ac and is significantly more prominent in iPSCs than MEFs. Similarly, the completely acetylated isoform of this peptide is found at higher levels in iPSCs than MEFs. Together, the analysis of the histone PTM isoforms of this peptide suggests co-regulation of specific acetylation PTMs during reprogramming.

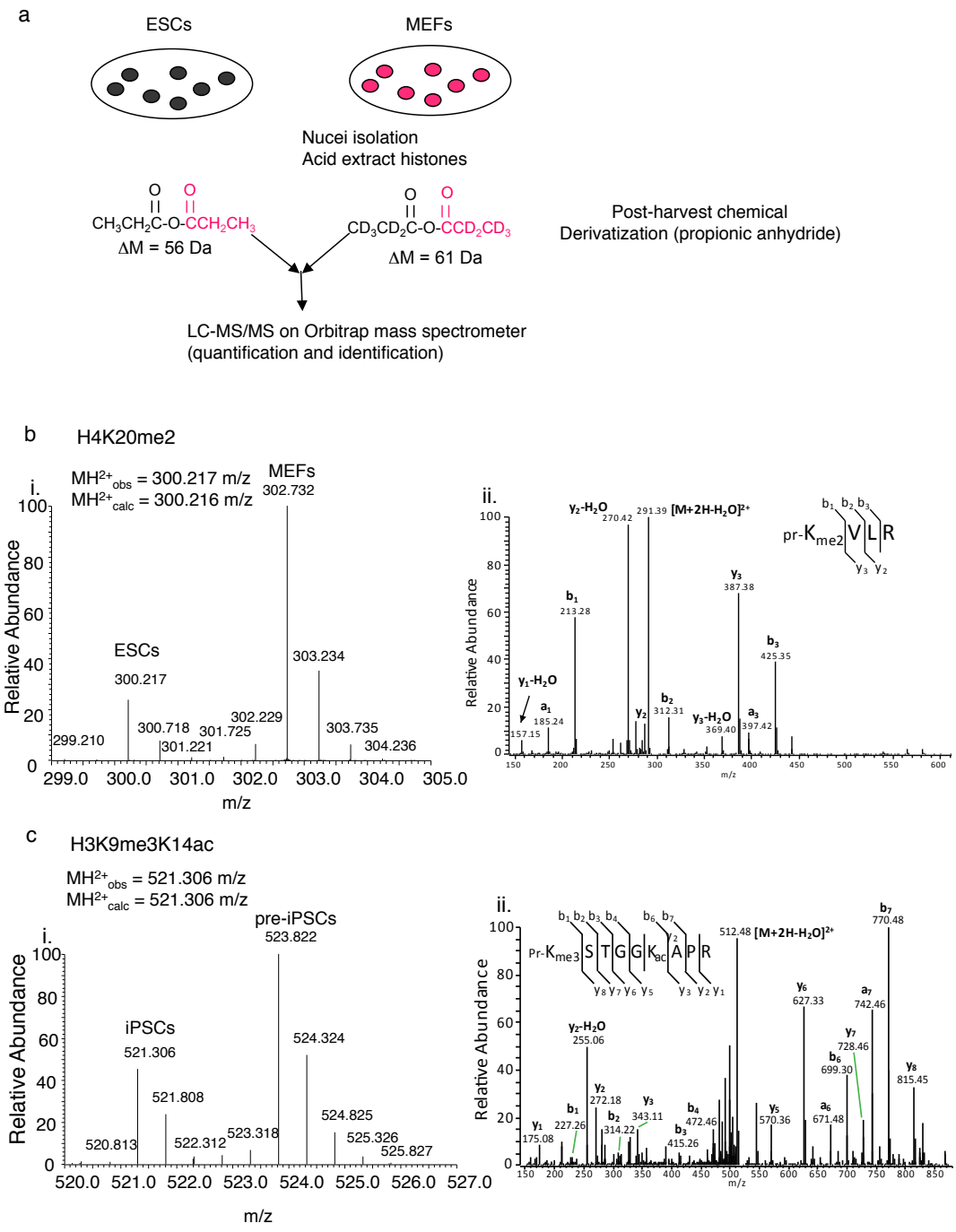

Figure S2 Direct comparison of histone PTM abundance between two cell types. **A)** To compare the levels of histone PTMs between two different cell types directly, we quantified histone PTMs with a second proteomics approach. Here, histones were individually extracted from each of the two cell types to be compared, samples were derivatized with D0 and D5 propionic anhydride (for chemical isotopic labeling), respectively, yielding an additional mass of 61 and 56 dalton (Da). The samples from both cell types were then mixed equally and analyzed by high-resolution nanoLC-MS/MS experiments on an Orbitrap mass spectrometer. As a result, peptides derived from different cell types can be distinguished from their masses post identification. **B)** Quantitative mass spectrometry comparison of the histone H4K20me2 peptide extracted from ESCs and MEFs. i) Full mass spectrum showing a pair of labeled peptide ions at 300.217 (ESCs) and 302.732 (MEFs) m/z, respectively. The peptide ion at 302.732 m/z (MEFs) is ~3-fold higher in abundance than the peptide at 300.217 m/z (ESCs). ii) MS/MS interrogation of the peptide ion at 300.217 m/z in (i) identifies this peptide as the H4 20-23 residue peptide (KVLR) with K20 dimethylation. **C)** Quantitative mass spectrometry comparison of the histone H3K9me3K14ac peptide extracted from iPSCs and pre-iPSCs. i) Full mass spectrum showing a pair of labeled peptide ions at 521.306 (iPSCs) and 523.822 (pre-iPSCs) m/z, respectively, indicating the differential abundance of the peptide between these cell types. ii) MS/MS interrogation of the peptide shown in (i) at 521.306 m/z confirms that the peptide contains H3K9me3K14ac.

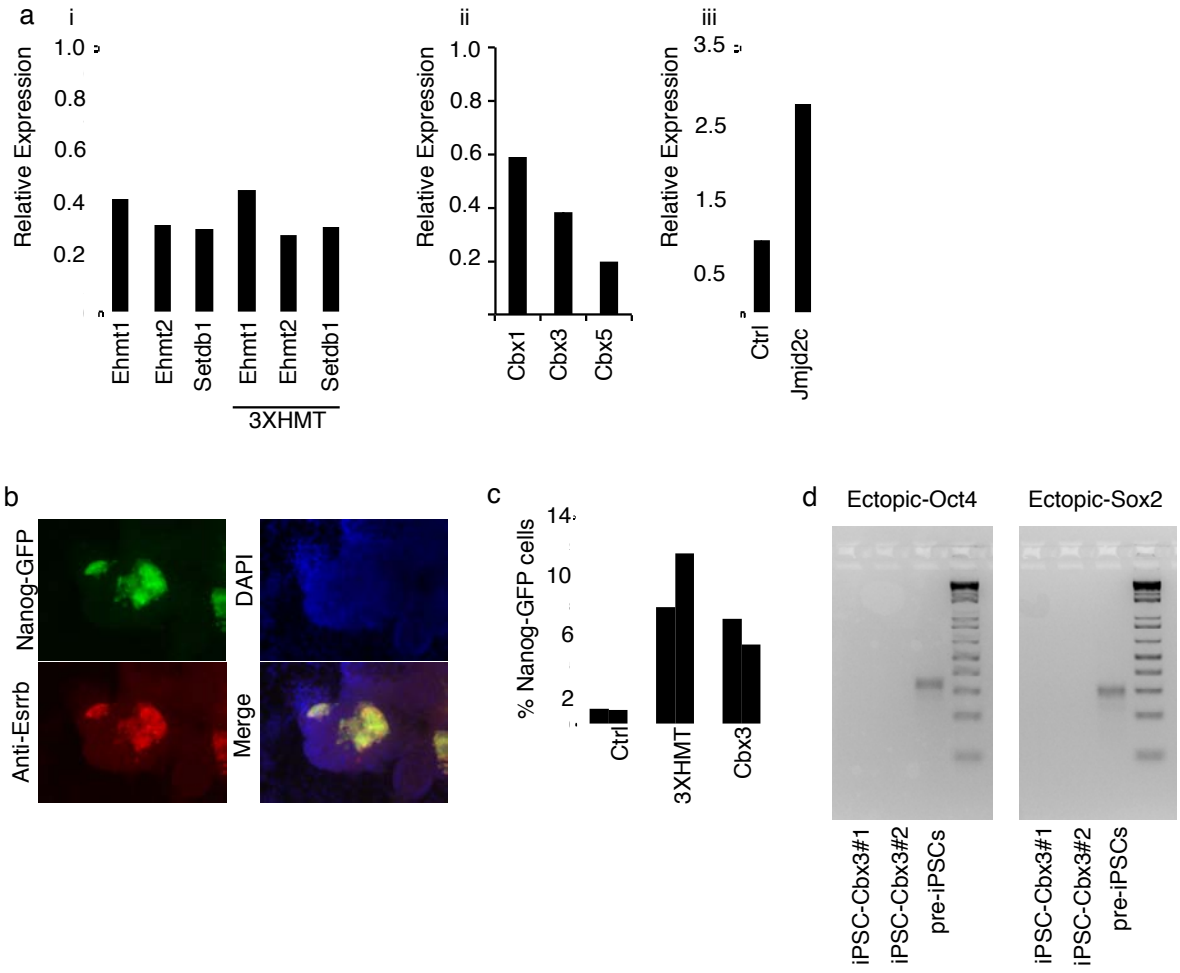

Figure S3 Additional characterization of fibroblast and pre-iPSC reprogramming experiments. **A)** Confirmation of the efficient knockdown of *Ehmt1, Ehmt2*, *Setdb1,* or *Cbx3*, *Cbx5*, and *Cbx1* transcript levels and of overexpression of *Jmjd2c* in the reprogramming experiments shown in Figure 3 of the main manuscript. qRT-PCR was performed three days after administering the first siRNA transfection or at the end of the overexpression experiment. For each graph, expression is plotted relative to transcript levels in control siRNA treatments, which was set to 1. Data presented is the mean of three technical replicates from one out of three representative experiments. i) Confirmation of knockdown for the experiment shown in Fig 3Ai. For the 3XHMT knockdown, all three H3K9 HMTases were efficiently knocked down. ii) Confirmation of knockdown for the experiment shown in Fig 3Aii. iii) Confirmation of *Jmjd2c* overexpression for the experiment shown in Fig 3C. **B)** Immunostaining analysis of a reprogrammed colony obtained on day 25 of a reprogramming experiment in which *Nanog*-GFP reporter MEFs were infected with retroviruses expressing the reprogramming factors *Oct4*, *Sox2,* and *Klf4*, and the three H3K9 HMTases were simultaneously knocked down by siRNAs. Shown is an immunostaining for the pluripotency marker Esrrb (red), *Nanog*-driven GFP-fluorescence (green), and DAPI staining to mark the nuclei (blue). We conclude that reprogrammed colonies do not only exhibit the activation of the *Nanog* GFP reporter but also of other pluripotency markers. **C)** Quantification of *Nanog*-GFP-positive cells from 3XHMT or *Cbx3* knockdown in pre-iPSCs carrying the *Nanog*-GFP reporter by FACS analysis. Data from duplicate experiments is presented. **D)** *Nanog*-GFP-positive, ESC-like colonies obtained upon *Cbx3* knockdown in pre-iPSCs have silenced the Oct4 and Sox2 retroviruses. Semi-quantitative RT-PCR to determine expression of the retrovirally encoded *Oct4* and *Sox2* in pre-iPSCs and the iPSC lines Cbx3 #1 and #2 derived from these pre-iPSCs upon *Cbx3* knockdown. We conclude that pre-iPSCs but not the iPSCs generated upon *Cbx3* knockdown express the reprogramming factor-encoding retroviruses.

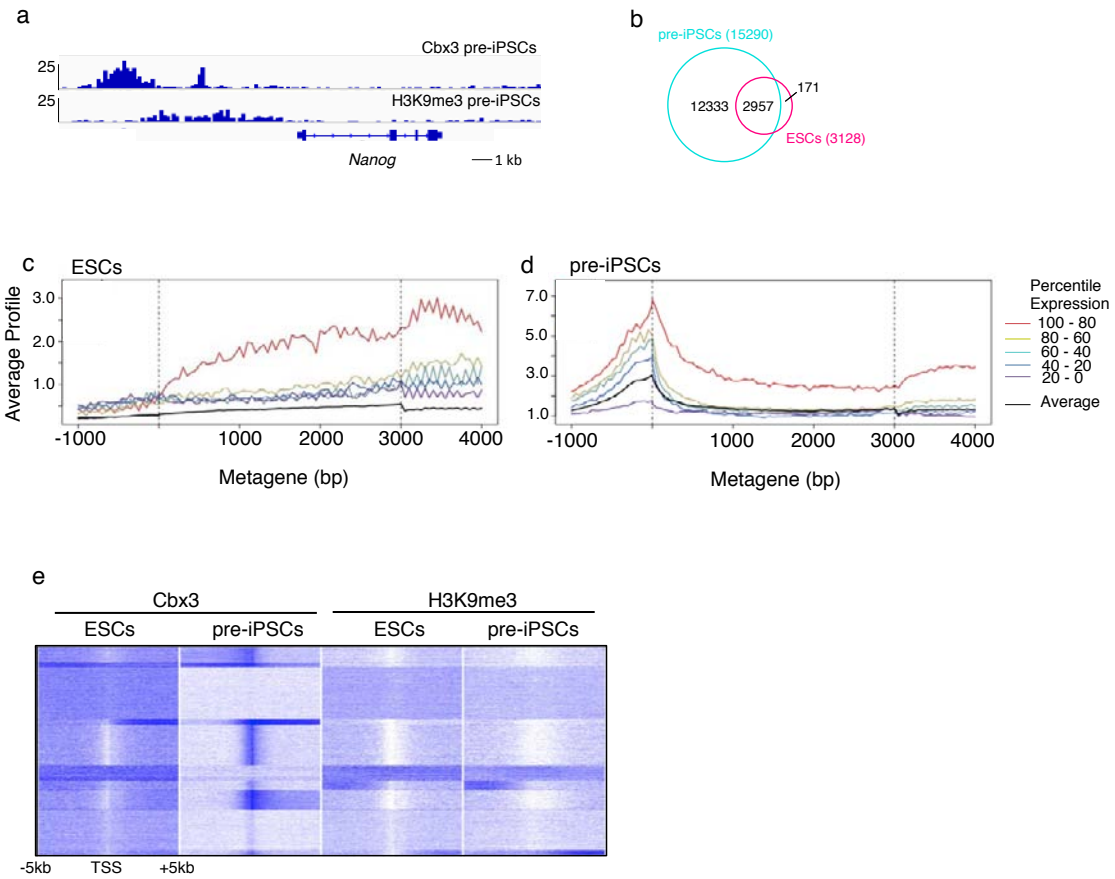

Figure S4 Characterization of Cbx3 occupancy in pre-iPSCs and ESCs. **A)** H3K9me3 and Cbx3 are enriched at the *Nanog* locus in pre-iPSCs. Snapshot of Cbx3 and H3K9me3 ChIP-seq data at the *Nanog* locus in pre-iPSCs. **B)** Venn diagram showing the overlap of Cbx3-bound genes between pre-iPSCs and ESCs. Notably, Cbx3 occupies many more genes in pre-iPSCs than ESCs. A Cbx3 binding peak was assigned to a gene when it overlapped with the region encompassing 10kb upstream of the TSS and the 3' end of a gene. **C)** Cbx3 occupancy correlates with gene expression. Average gene profiles for Cbx3 enrichment within genes bound by Cbx3 in ESCs as defined in (B), grouped into five expression tiers. The average gene profiles of Cbx3 enrichment were compressed into a metagene format for each of these groups. For comparison, the average profile of Cbx3 occupancy across all genes (bound and unbound) is presented (average). Note that genes in the 100-80th percentile (most highly expressed bound genes) have the highest Cbx3 occupancy, particularly in the gene body increasing towards the 3' end. **D)** As in (C) except for pre-iPSCs. Similar to ESCs, the most highly expressed Cbx3-bound genes are most strongly enriched for Cbx3 within gene bodies. However, in pre-iPSCs there is an additional enrichment at around the TSS that also correlates with expression. **E)** Lack of overlap between Cbx3 and H3K9me3 around the TSS. Heatmap of Cbx3 and H3K9me3 enrichment in ESCs and pre-iPSCs, respectively, within a region encompassing 5kb upstream and downstream of the transcription start site (TSS), based on ChIP-seq analysis. Genes were grouped by K-means clustering (10 clusters) of all four data sets. Blue denotes enrichment in a 100 bp bin. Chip-Seq reads were mapped to mm9 genome using the Bowtie software and only those reads that aligned to a unique position with no more than two sequence mismatches were retained for further analysis. Multiple reads mapping to the same location in the genome were collapsed to a single read to account for clonal amplification effects and the start site of each read was shifted by 75 bases in the 3' direction to estimate the actual midpoint of the DNA fragment that produced the read (or tag). The genome was tiled into 100-bp windows (or bins) and the total number of tags falling into each of these bins was tallied for both the ChIP and input samples. The tag counts based on the scale-normalized input sample were used to estimate the expected number of counts within each window. Those windows with zero counts in the input were assigned the average value for all windows across the chromosome. The Poisson distribution was used to estimate the probability of observing the ChIP counts within a window given the expected counts estimated by the input sample (λ). All windows with p-value < 1.0E-5 were considered to represent peaks of significant ChIP-enrichment relative to input. The heatmap was generated using Java Treeview. Strikingly, there is very low TSS enrichment of H3K9me3 in either cell type, indicating rather limited overlap of Cbx3 and H3K9me3, particularly in pre-iPSCs. ChIP-seq for H3K9me3 was performed with a specific antibody (Active motif-39161).

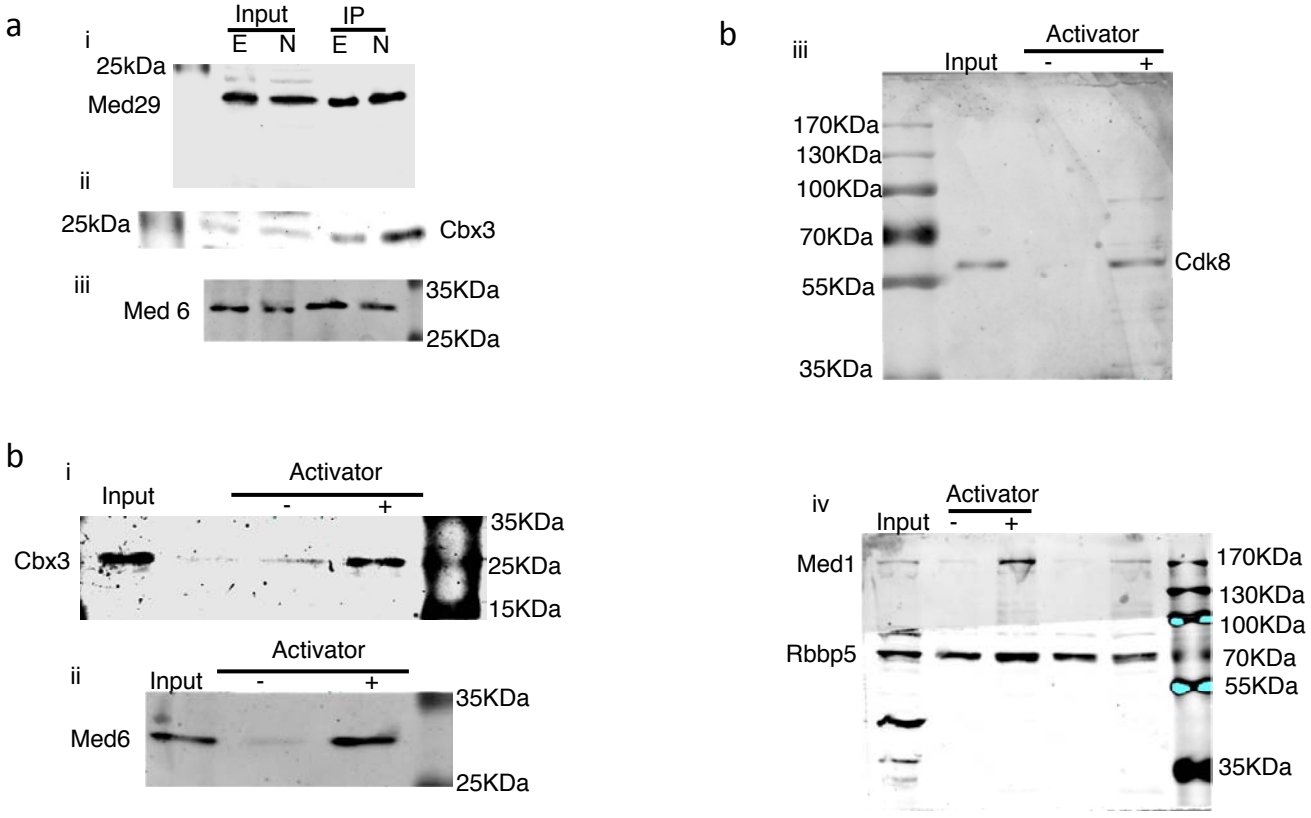

Figure S5 Additional characterization of the Mediator/Cbx3 interaction (unprocessed Western Blots) **A)** Uncropped image of the western blots presented in Fig 8B; i) Flag (Med29); ii) Cbx3; and iii) Med6. E= ESCs and N= ESC-derived neural precursors. Note that the membrane was cut before exposing to antibody to save on reagents and samples are derived from the same experiment and that blots were processed in parallel. **B)** Uncropped images of western blots presented in Fig 8C; i) Cbx3; ii) Med6; iii) Cdk8; and iv) Med1, and Rbbp5. Samples were derived from the same experiment and processed in parallel.

Supplementary Table Legends

Supplementary Table 1 (related to Figures 1 and 2)
Summary of the levels of histone PTMs in all cell types analyzed. The information for all analyzed peptides and the relative level of all detected isoforms for each tryptic histone peptide are given. Fig 1a, 1b, 1c, 1e, 2a, 2b, 2c, S1b, S1d-f are based on these values. The average value denotes the mean of the histone PTM isoform data from two biological replicates and three technical replicates and "SD" the standard deviation.

Supplementary Table 2 (related to Figure 5)
Summary of histone PTM levels upon 3XHMT, *Cbx3*, or control knockdown in pre-iPSCs, three days post siRNA transfection. For the 3XHMT and Cbx3 siRNA treatments, histone PTMs were determined in duplicate reactions, for control siRNA treatments in triplicate. Sheet 1 gives the relative levels of all detected isoforms for each histone peptide for each technical replicate. Sheet 2 contains the averaged data and standard deviation and, in addition, identifies significant differences in histone PTM levels between the control siRNA treatments and the si*Cbx3* or si3XHMT conditions. The data in this table are the source for the graphs in Figure 5

Supplementary Table 3 (related to Figure 6)
Summary of microarray gene expression data for pre-iPSCs depleted for *Cbx3* or 3XHMT. Sheet 1 contains the averaged expression data for ESCs, iPSCs and MEFs (taken from Sridharan et al 2009), and for all the replicates for pre-iPSCs treated with control siRNAs (3 biological replicates), si*Cbx3* (3 biological replicates), or si3XHMT (2 biological replicates) generated in this study. Sheet 2 and 3 contain 1.5 fold up and down regulated genes, respectively, for si3XHMT and si*Cbx3,* respectively and are the source data for Fig 6a, 6b and 7g. Sheet 4 contains RNA-seq based expression data in ESCs and pre-iPSCs, for the set of genes differentially expressed in the 3XHMT and *Cbx3* knockdown, representing the source data for Fig 6c.

Supplementary Table 4 (related to Figure 7)
Summary of Cbx3 ChIP-seq data. Sheet 1 contains the locations of all peaks that are bound by Cbx3 in pre-iPSCs and ESCs, respectively and is the source data for Fig 7b, 7c and 7e Sheet 2 contains the list of genes bound by Cbx3 in pre-iPSCs and ESCs (from -10kb upstream to the transcription termination site) and their expression levels and is the source data for 7d and 7f.