Cell Reprogramming Requires Silencing of a Core Subset of Polycomb Targets

Giulia Fragola1,2, Pierre-Luc Germain1, Pasquale Laise1,9, Alessandro Cuomo1, Alessandro Blasimme1,9, Fridolin Gross1, Elena Signaroldi1, Gabriele Bucci1,2, Cesar Sommer3, Giancarlo Pruneri4, Giovanni Mazzaroli4, Tiziana Bonaldi1, Gustavo Mostoslavsky3, Stefano Casola1,2*, Giuseppe Testa1,2*

European Institute of Oncology, IFOM-IEO Campus, Milan, Italy, 1 European Institute of Oncology, IFOM-IEO Campus, Milan, Italy, 2 IFOM Foundation, FIRC Institute of Molecular Oncology Foundation, IFOM-IEO Campus, Milan, Italy, 3 Boston University School of Medicine, Boston, Massachusetts, United States of America, 4 European Institute of Oncology, Milan, Italy

Abstract

Transcription factor (TF)-induced reprogramming of somatic cells into induced pluripotent stem cells (iPSC) is associated with genome-wide changes in chromatin modifications. Polycomb-mediated histone H3 lysine-27 trimethylation (H3K27me3) has been proposed as a defining mark that distinguishes the somatic from the iPSC epigenome. Here, we dissected the functional role of H3K27me3 in TF–induced reprogramming through the inactivation of the H3K27 methylase EZH2 at the onset of reprogramming. Our results demonstrate that surprisingly the establishment of functional iPSC proceeds despite global loss of H3K27me3. iPSC lacking EZH2 efficiently silenced the somatic transcriptome and differentiated into tissues derived from the three germ layers. Remarkably, the genome-wide analysis of H3K27me3 in Ezh2 mutant iPSC cells revealed the retention of this mark on a highly selected group of Polycomb targets enriched for developmental regulators controlling the expression of lineage specific genes. Erasure of H3K27me3 from these targets led to a striking impairment in TF–induced reprogramming. These results indicate that PRC2-mediated H3K27 trimethylation is required on a highly selective core of Polycomb targets whose repression enables TF–dependent cell reprogramming.

Introduction

The realization that the expression of few transcription factors can reassign cell fate has been a paradigm-shifting insight for biology and medicine, from the pioneering conversion of fibroblasts into myoblasts [1], to the achievement of inducing pluripotency from adult somatic cells [2]. In particular, for the derivation of induced pluripotent stem cells (iPSC), the enticing medical opportunities during the past few years have focused most efforts on improving the efficiency and safety of transcription factor (TF)-induced reprogramming, and on defining the molecular and functional equivalence between iPSC and embryonic stem cells (ESC). Substantial progress has been made in the characterization of distinct stages of the reprogramming process (reviewed in [3]) as well as of its key features, such as the mesenchymal-to-epithelial transition [4,5], and the importance of cell-cycle progression [6–11]. Yet, in terms of molecular understanding, we still lack a mechanistic insight of how the somatic gene expression program is silenced in order to enable the reacquisition of the pluripotent state. In particular, the functional relevance of defined chromatin modifications has only recently started to be explored [12–15], especially as far as those histone marks are concerned that orchestrate genome programming during development.

Genome-wide analyses have started to provide a detailed description of the chromatin changes that underlie TF-induced reprogramming to pluripotency. The very initial stages of the process entail the widespread acquisition of dimethylation of lysine-4 of histone H3 (H3K4me2) at promoters and enhancers of pluripotency genes that will be reactivated however only at later stages, indicating that this modification primes but does not directly trigger gene activation [16]. Conversely, analyses of established iPSC at the end of the reprogramming process revealed that changes in Polycomb-mediated trimethylation of lysine-27 of histone H3 (H3K27me3) represent a key feature that distinguishes the fibroblast from the iPSC epigenomes [17].

The Polycomb axis is organized around two families of protein complexes, Polycomb Repressive Complex 1 (PRC1) and Poly-
Author Summary

Multicellular organisms are composed of a variety of cell types. Over the last years we have learned that cell differentiation is fully reversible and that it takes few specific transcription factors (proteins that bind to DNA and regulate gene expression) to convert one cell type into another. The most dramatic example is the reprogramming of somatic cells into stem cells that reacquire the potential to give rise to all cell types of the body. This process entails the resetting of the gene expression program of the somatic cells necessary to acquire a pluripotent state but remains poorly understood. Here, we define the role in cell reprogramming of the Polycomb axis, one of the key effectors of gene silencing that operates through a specific chemical modification of histones (the trimethylation of lysine-27 on histone H3) that is stably inherited from one cell generation to the next. Surprisingly, we uncovered a selective requirement for Polycomb silencing during somatic cell reprogramming. Upon inactivation of the essential Polycomb protein Ezh2, histones trimethylated on lysine-27 were retained on a selected subset of physiological Polycomb target genes whose products control normal development, and erasure of the histone mark from these critical genes abolished reprogramming.

Polycomb Repressive Complex 2 (PRC2) (reviewed in [18]). Both contain enzymatic activities that catalyze, respectively, mono-ubiquitylation of lysine-119 on histone H2A (H2AK119Ub) (carried out by RING1B and, to a lesser extent, RING1A) and di- and tri-methylation of lysine-27 of histone H3 (catalyzed by EZH2 and, to a lesser extent, EZH1) [18]. H3K27me3 constitutes a docking site for the chromodomain of CBX proteins, members of PRC1, leading to a cascade model for Polycomb action, in which PRC2 deposits H3K27me3 that then recruits PRC1 to a docking site for the chromodomain of CBX proteins, members of PRC1, leading to a cascade model for Polycomb action, in which PRC2 deposits H3K27me3 that then recruits PRC1 to

Results

Derivation of iPS from mouse embryonic fibroblasts upon Ezh2 inactivation

In order to investigate the function of H3K27 trimethylation (H3K27me3) in TF-induced cell reprogramming, we derived embryonic fibroblasts (MEF) from mutant mice harboring a conditional Ezh2 knock-out allele carrying loxP sites (Ezh2fl/fl) flanking exons coding for the catalytic SET domain [30] (Figure 1 and Table 1). Experimental mice carried also an Oct4-GFP knock-in reporter gene [31] to monitor endogenous Oct4 gene re-activation during reprogramming. We expressed Oct4, Klf4, c-Myc and Sox2 reprogramming factors from a single doxycycline-regulated bicistronic lentiviral vector (STEMCCA), in which each cistron codes for a fusion of two factors which are then released through, respectively, F2A and E2A self-cleaving peptide signals [32]. As depicted in Figure 1A, our reprogramming protocol entailed: i) treatment of experimental (Ezh2fl/fl; Oct4-GFP) and control (Ezh2+/fl; Oct4-GFP) MEF with cell-permeable TAT-Cre recombinase to inactivate Ezh2; ii) infection with STEMCCA and reverse tetracycline transactivator (tTA) lentiviruses, followed by replating at clonal density and switch to a chemically-defined embryonic stem cell (ESC) medium (KSR) and administration of doxycycline for 23 days; iii) withdrawal of doxycycline to select cells that re-expressed the endogenous pluripotency factors; and iv) isolation, clonal propagation and molecular and functional characterization of individual iPS clones. As controls, we used interchangeably iPS from Ezh2fl/fl; Oct4-GFP and control (Ezh2+/fl; Oct4-GFP) MEF to confirm that cell-permeable TAT-Cre recombinase was able to inactivate Ezh2 in vivo. In addition, we replicated this experiment in two different batches of MEF, we found no significant difference in the number of iPS colonies generated from mutant (Ezh2fl/fl; Oct4-GFP) and control (Ezh2+/fl; Oct4-GFP) fibroblasts (Figure 1C). As genome-wide chromatin analyses have postulated a critical role for H3K27me3 in the resetting of transcriptional programs during iPSC derivation [17], we tested whether this unexpected finding resulted from selective reprogramming of Ezh2fl/fl MEF that had escaped Cre-mediated recombination, thereby preserving H3K27me3-dependent reprogramming proficiency. We ruled out this possibility as the majority of iPS clones analyzed carried the SET-deleted Ezh2 allele regardless of the genotype (Figure S1A).

Representative iPS clones of either genotype were analyzed by Western blot and found to be devoid of functional EZH2 and with undetectable H3K27me3 (Figure 1D). Levels of H3K27 dimethylation (H3K27me2) were also decreased in mutant iPS clones, whereas H3K27 monomethylation (H3K27me1) remained unaltered (Figure 1D). Furthermore, global levels of the other major repressive histone modification, histone H3K9 trimethylation (H3K9me3), remained unaltered upon Ezh2 inactivation (Figure S1C).
Global loss of H3K27me3 was confirmed by high performance liquid chromatography (HPLC) coupled to tandem mass spectrometry (MS/MS) analysis of the histone H3 fraction purified from control and mutant iPSC clones. Specifically, we determined the relative abundance of the peaks corresponding to the various combinations of modifications harbored by the H3 peptide spanning lysine-27 through arginine-40 (H327–40). A significant change in the abundance and distribution of the various forms of methylated H327–40 was observed (Figure 1E). We then carried out a detailed MS/MS fragmentation analysis to pinpoint methylation at specific residues, which indicated that H3K27me3 was undetectable in \( \text{Ezh2}^{DSEL/DSET} \) iPSC clones, confirming that changes in the relative abundance of methylated H3 27–40 are mainly attributed to loss of this modification (Table 1 and Figure S1E).

Finally, we measured by qRT-PCR the expression levels, in both mutant and control iPSC, of the exogenously provided reprogramming factors, to exclude that reprogramming in the absence of H3K27me3 had selected iPSC clones with a pronounced leakiness of the doxycycline-inducible transgenes, whose sustained expression enabled the maintenance of H3K27me3-depleted iPSC. As shown in Figure S1D, we excluded this possibility since EZH2-proficient and EZH2-deficient iPSC showed equivalent, minimal levels of transgene expression following doxycycline withdrawal.

**Self-renewal and pluripotency of \( \text{Ezh2}^{DSEL/DSET} \) iPSC**

Having determined that \( \text{Ezh2} \) inactivation in MEF is compatible with TF-induced cell reprogramming, we performed a comprehensive functional characterization of representative control and mutant iPSC clones. \( \text{Ezh2}^{DSEL/DSET} \) and \( \text{Ezh2}^{+/+} \) control MEF yielded iPSC colonies: i) with distinctive iPSC/ESC morphology (Figure 2A, left panel), ii) that stained positive for AP (Figure 2A, middle panel); and iii) that had reactivated the endogenous Oct4 gene (as assessed by GFP fluorescence, Figure 2A, right panel). iPSC clones of either genotype had the same percentage of cells

**Figure 1. Derivation and biochemical analysis of iPSC upon conditional \( \text{Ezh2} \) inactivation.** A. Diagram of the reprogramming protocol of MEF. B. Alkaline phosphatase staining of control (lower row) and mutant (upper row) primary iPSC colonies one week following doxycycline withdrawal. C. Number of AP-positive primary iPSC colonies obtained upon infection of, respectively, \( 2 \times 10^3, 5 \times 10^3, 1 \times 10^4 \) or \( 6 \times 10^4 \) MEF in two experiments performed with two biological replicates per genotype. D. \( \text{Ezh2}, \) H3K27me1, H3K27me2 and H3K27me3 protein levels assessed by Western blot in two representative \( \text{Ezh2}^{+/+} \) and mutant (\( \text{Ezh2}^{DSEL/DSET} \)) iPSC clones. Vinculin and Histone H3 were used as loading controls for, respectively, EZH2 and methylated forms of H3K27. E. Relative abundance in control (upper row) and mutant (lower row) iPSC clones of the six possible methylolation isomers of the Histone H3 peptide spanning amino acids 27–40, as determined by mass spectrometry.

doi:10.1371/journal.pgen.1003292.g001
co-expressing the pluripotency markers OCT4 and SSEA1, as measured by flow cytometry (Figure 2B). Next we assessed control and mutant iPSC clones for the two cardinal features that define the pluripotent state: self-renewal and the ability to differentiate into cell types of the three germ layers.

Self-renewal is currently best assessed by growing ESC and iPSC under chemically defined conditions that entail the dual inhibition (2i) of mitogen activated protein kinase (MAPK) and glycogen synthase kinase 3 (GSK3) signaling, in the presence of leukemia inhibitory factor (LIF) (commonly referred to as 2i/LIF medium). These highly selective culture conditions were shown to capture the ground state of pluripotency exhibited by cells of the epiblast inner cell mass (ICM) [33]; importantly, they were also shown to selectively promote the full reprogramming of partially reprogrammed cells to the authentic pluripotent state [34]. As shown in Figure 2C, Ezh2 mutant and control (Ezh2+/−) iPSC clones grew readily in 2i/LIF and have been cultivated for over 20 passages. Finally, we tested the ability of Ezh2-proficient (Ezh2+/+) and mutant (Ezh2+/−) iPSC to differentiate along the three embryonic lineages through the formation of teratomas. iPSC clones of either genotype (2 clones for each genotype) injected into embryonic lineages through the formation of teratomas. iPSC passages. Finally, we tested the ability of clones grew readily in 2i/LIF and have been cultivated for over 20 passages. These highly selective culture conditions were shown to capture the ground state of pluripotency exhibited by cells of the epiblast inner cell mass (ICM) [33]; importantly, they were also shown to selectively promote the full reprogramming of partially reprogrammed cells to the authentic pluripotent state [34].

Next, we asked to which extent the transcriptome of Ezh2 depleted iPSC retain H3K27me3 on selected targets. To address this point we performed chromatin immunoprecipitation coupled with STEMCCA and doxycycline administration (Figure 3B). The finding that bulk H3K27me3 was apparently dispensable for reprogramming even when erased at the onset of the reprogramming process (Figure 3A). We therefore asked whether the unexpected possibility to reprogram despite inactivation of Ezh2 was due to residual H3K27me3 that could have still ensured, during the very first days of reprogramming, a sufficient degree of repression of lineage specific genes. To this end, we aimed at erasing the H3K27me3 mark completely by dilution, through serial passage of TAT-Cre treated MEF before the start of reprogramming. In MEF, however, PRC2 is a direct repressor of Cdkn2a, a locus encoding three key cell cycle regulators (p16, p19 and p15) whose activation promotes senescence [35]. Importantly, expression of this locus, and in particular of p19/Arf, was shown to hinder iPSC reprogramming [9], Thus, to prevent senescence driven by Cdkn2a de-repression following Ezh2 inactivation (Figure S3A), we resorted to compound primary tail tip fibroblasts (TTF) harboring both the conditional Ezh2 allele and the Ink4A/Arf knockout allele [36]. We subjected TTF to two sequential rounds of TAT-Cre transduction and passaged them 5 times before infection with STEMCCA and doxycycline administration (Figure 3B). Following confirmation that H3K27me3 was undetectable by Western blot on the day of infection with the reprogramming lentivirus (Figure 3C), and that efficiency of infection was equivalent for control and mutant TTF (Figure S3B), we went on to measure the efficiency of TF-dependent reprogramming under these most stringent conditions. As shown in Figure 3D, AP staining revealed that TTF starting off with undetectable H3K27me3 and controls yielded iPSC colonies with similar efficiency. This indicates that bulk levels of H3K27me3 are not required to prime silencing during the first days of reprogramming.

**Ezh2 depleted iPSC retain H3K27me3 on selected Polycomb targets**

The finding that bulk H3K27me3 was apparently dispensable for reprogramming even when erased at the onset of the process was at odds with its purported role as the critical mark that distinguishes MEF from iPSC epigenomes as well as, more broadly, with its pivotal role in the maintenance of gene repression through embryogenesis and adulthood [17]. We therefore asked whether, upon Ezh2 inactivation, residual levels of H3K27me3 below the threshold of Western blot and mass spectrometry sensitivity, could still be deposited on selected targets. To address this point we performed chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq). We generated ChIP-seq profiles from two independent control and mutant iPSC clones for both H3K27me3 and H3K27me2 with highly specific

---

**Table 1. Mass spectrometry analysis: H3K27me3 levels below the limit of detection in Ezh2+/−/− iPSCs.**

| Peptide Forms | Modification Sites | Mascot Score* | PTM Score** |
|---------------|--------------------|---------------|-------------|
|               |                    | −/− | Ezh2+/−/− | −/− | Ezh2+/−/− |
| me3           | K27me3             | 48.72 | ND | 104.43 | ND |
| me3           | K27me1-K36me2      | ND | 38.55 | ND | 92.16 |
| me3           | K27me2-K36me1      | 52.3 | 20.02 | 117.35 | 32.14 |
| me4           | K27me2-K36me2      | 73.57 | 56.77 | 145.18 | 117.35 |
| me4           | K27me3-K36me1      | 47.61 | ND | 65.14 | ND |
| me5           | K27me3-K36me2      | 39.75 | ND | 92.16 | ND |

*Perkins et al 1999. **Cox and Mann 2010.

Mascot and PTM scores attributed in control (Ezh2+/+) and mutant (Ezh2+/−/−) iPSC clones to each combination of amino acid modifications on the H327–40 tri-, tetra- and penta-methylated peptide.
monoclonal antibodies. Consistently with our prediction, the higher sensitivity of ChIP-qseq did reveal the presence of residual H3K27me3 in mutant iPSC clones. Specifically, the mark was retained on 2477 genes (with an enriched region overlapping a +/-3kb region interval around the transcriptional start site, TSS), comprising close to half of all H3K27me3 targets retrieved from wild type iPSC. Mutant clones showed a preferential retention of H3K27me3 proximal to the TSS of target genes. In comparison to the full complement of PRC2 targets in control iPSC, mutant clones displayed a clear tripartition in the genome-wide distribution of H3K27me3 and H3K27me2 marks (Figure 4A). 47% of genes retained both H3K27me3 and H3K27me2, 39% of genes were marked only by H3K27me2 and 13.7% of genes lost both marks. Importantly, the complement of genes enriched for H3K27me3 in mutant iPSC clones was almost entirely comprised within the group of H3K27me3 targets found in control iPSC cells, thus excluding a significant redistribution of the mark to new targets in cells reprogrammed in the absence of functional Ezh2 (Figure 4B). Furthermore, we found only a small overlap in the distribution of H3K27me2-only targets between control and mutant iPSC clones. Instead, the complement of genes marked only by H3K27me2 in mutant iPSC clones was to a good extent comprised within the subset of genes that are H3K27 trimethylated in control iPSC (Figure 4C). Thus, we conclude that during reprogramming in the absence of Ezh2, i) H3K27me3 is selectively retained on a subset of the targets that are normally H3K27 trimethylated in iPSC, where it coexists with H3K27me2; ii) H3K27me2 is lost at targets that are normally carrying only this mark in iPSC; and iii) H3K27me2 is retained in 86% of the targets that are normally H3K27 trimethylated in iPSC, coexisting, in half of these, with residual H3K27me3. We validated these findings through individual ChIP-qPCR on genes selected among those that were downregulated in the MEF to iPSC transition (Figure 4D and Table S2). We confirmed the sharp distinction between a group of genes that retained both H3K27me3 and H3K27me2 and those that only retained H3K27me2, irrespective of the level of transcriptional repression that was equivalent for the two groups between control and mutant iPSC (Figure 4D and Table S2). Interestingly, we found a stronger enrichment for PRC2 on the genes that selectively retained the H3K27me3 mark, likely reflecting its ability to act as docking site for the EED subunit of PRC2.

In the absence of Ezh2, EZH1 is the only other known enzymatic component of PRC2, where its enzymatic activity is however twenty fold weaker than that of EZH2 [37]. Our findings support therefore a model in which, in the presence of EZH1-only PRC2 complexes, genes that are normally more enriched for PRC2 will be more likely to accumulate H3K27me3 because they will have a higher chance of having the weak activity of EZH1 result in productive trimethylation. In turn this will further enhance PRC2 recruitment, establishing a feed-forward loop that guarantees on selected targets residual levels of H3K27me3. We tested this hypothesis by analyzing, in control iPSC, the average enrichment for H3K27me3 in genes that, in mutant iPSC, respectively retain or lose this mark. As shown in Figure 4E and Figure S4B, genes that retain H3K27me3 in mutant iPSC showed indeed a highly significant stronger enrichment for this mark in control iPSC cells, both in terms of maximum height and average read density (p-value<2.2e-16, two tailed t-Test). In almost all cases however, targets that retain H3K27me3 in mutant iPSC also feature H3K27me2 (contrary to what happens in control iPSC), indicating that EZH1 does not reach H3K27 trimethylation: 1) in all cells of the population, 2) on both allels of the same cell, or 3) on both histone H3 tails of the same nucleosome. The non-randomness of H3K27me3 marking in mutant iPSC was confirmed by the virtually identical genome-wide distribution of H3K27me3 and H3K27me2 in two independent iPSC mutant clones (Figure S4D), as well as by the striking similarity between the distributions of, respectively, H3K27me3 in control and H3K27me2 in mutant iPSC (Figure S4C).

**Targets of H3K27me3 in Ezh2 mutant iPSC are enriched for transcriptional regulators and developmental determinants**

Finally, we asked whether the H3K27me3 tripartition of the mutant iPSC epigenome identified functionally relevant classes of genes. To this end we analyzed the three groups of PRC2 targets differentially methylated in mutant iPSC (H3K27me3+/H3K27me2-; H3K27me3-/H3K27me2+; and H3K27me3+/H3K27me3-) in terms of the Gene Ontology (GO) of their members (Figure 4F). Remarkably, the subset of genes that retained H3K27me3 in mutant iPSC showed a distinct GO profile when compared to the other two classes, with a clear enrichment for categories linked to development, cellular differentiation and transcriptional regulation. This unique signature was confirmed when the GO enrichment in H3K27me3-retaining genes was probed against the sole complement of all Polycomb targets, further underscoring the functional partition of the Polycomb epigenome. The H3K27me3+/H3K27me2- and H3K27me3+/H3K27me2+ subsets were instead strongly enriched for genes involved in metabolic homeostasis and cellular transport.

Remarkably, while genes that retained H3K27me3 in mutant iPSC comprised 47% of the physiological PRC2 epigenome, they were disproportionately enriched for genes preferentially expressed in MEF versus iPSC (comprising 69.2% of the total, Figure 4G). This suggested that retention of H3K27me3 in Ezh2 mutant iPSC occurred on selected targets to enable silencing of the MEF-specific gene expression program. Specifically, we found 550 genes, among those that retained H3K27me3 in mutant iPSC,
that were downregulated in the MEF to iPSC transition (Figure S3A). Among these, 175 acquire H3K27me3 de novo in the MEF to iPSC transition and are enriched in GO categories related to transcriptional regulation (Figure S3B). We hypothesized the presence within this group of one or more MEF-specific master regulators that were silenced during reprogramming through H3K27me3. We thus performed a master regulator analysis (MRA) to identify the TFs whose targets (predicted on the basis of their consensus binding sites) are over-represented in the differentially expressed genes (DEGs) between MEF and iPSC. To this end we applied the transcription factor binding site over-representation analysis algorithm [30]. This analysis uncovered four TFs, Egr1, Ets1, Prxl1, Prxl2, whose binding sites were significantly over-represented (FDR<5%) among the 175 DEGs that acquire de novo H3K27me3 in the transition from MEF to iPSC, and are predicted to control the bulk (84%) of the up-regulated genes in MEF with respect to iPSC (Figure S3C). Notably, expression of Egr1 and Ets transcript levels showed a strong reduction within the first week of reprogramming, corroborating their role in the silencing of the MEF-specific program (Figure S3D).

**Discussion**

Trimethylation on lysine-27 of histone H3 by Polycomb proteins has been characterized as a critical mechanism that ensures the cell-type-specific establishment and maintenance of gene silencing in multicellular organisms [19,20,40,41]. From ESC to tissue-specific stem cells, this chromatin mark has been associated to the timed repression of developmental regulator genes, which underlies in turn the physiological unfolding of cell fate transitions [19–22,42–44]. Consistent with its role in the physiology of genome programming during development, the Polycomb axis has been hypothesized as a key player in the reprogramming of somatic cells epigenomes. In particular TF-induced reprogramming, through its relatively high efficiency coupled with the stability of the reprogrammed state and the ability to study defined stages of the process, enabled the first

PLOS Genetics | www.plosgenetics.org 7 February 2013 | Volume 9 | Issue 2 | e1003292
insight into the chromatin changes that underlie cell fate reassignment on a genome-wide scale. Among 16,500 promoters that were analyzed by ChIP-on-chip, roughly 1000 differed sharply in the H3K4me3 and H3K27me3 status when MEF were compared to ESC [17]. Nearly all of these signature genes displayed H3K4me3 and H3K27me3 patterns that were identical between ESC and iPSC, thus providing the first evidence that the transition from MEF to iPSC recapitulated, in addition to the transcriptome, to a large extent also the epigenome of ESC. Importantly, however, a comparison of methylation maps between MEF, ESC and iPSC revealed that H3K4me3 changed considerably less between the three cell types than H3K27me3, whose distribution was instead remarkably different between MEF, on the one hand, and ESC and iPSC on the other. This observation suggested that changes in H3K27me3 were more relevant to reprogramming than those in H3K4me3, and led to posit for the Polycomb pathway a central role in the conversion of MEF into iPSC [17]. A similar approach was used to investigate the earliest stages of the reprogramming process, uncovering the acquisition of H3K4 dimethylation (H3K4me2) at several pluripotency loci and developmental regulators that precedes their transcriptional activation [16]. More recently the inactivation of the H3K27 demethylase UTX revealed the essential role of H3K27 dimethylation in TF-induced reprogramming, through the reactivation of a handful of pluripotency genes during the initial stages of the process [13]. Similarly, Polycomb affected the ability of ESC to reassign fate in cell fusion-based short term reprogramming [45]. Thus, while these data reveal the importance, in the early phase of reprogramming, of the reactivation of the pluripotency network through widespread H3K4me2 acquisition and loss of H3K27 methylation, the relevance of H3K27 trimethylation activity in silencing somatic gene expression remains still ill defined.

Here we provide functional validation for the role of H3K27me3 in TF-induced reprogramming. Our findings reveal that functional inactivation of EZH2, the main catalyst of H3K27 trimethylation within PRC2, is surprisingly compatible with TF-cell reprogramming. In its absence, reprogramming proceeds through the deposition of H3K27me3 by the alternative EZH1-PRC2, on a highly selected core of Polycomb targets controlling lineage determination and transcriptional regulation. Moreover, our data show that correct reprogramming requires an exquisitely low amount of H3K27me3 below the limit of detection of mass spectrometry. Bioinformatics analysis of ChIP-seq data revealed that H3K27me3 is retained in Ezh2 mutant iPSC on about 2500 genes, representing close to half of all PRC2 targets in iPSC and ESC. These include 2190 genes that are already marked by H3K27me3 in MEF, with the remaining ones acquiring the modification during the MEF to iPSC transition, thus indicating that residual PRC2 is able through cell division to both propagate the mark at pre-existing loci and redistribute it to new targets. The correct reprogramming guaranteed by selective retention of H3K27me3 was also consistent with the normal differentiation of Ezh2-mutant iPSC in teratoma assays, defining the core subset of functionally relevant Polycomb targets. Similar results were obtained also when Ezh2 was inactivated in established iPSC (data not shown). These data exclude a functional difference between iPSC reprogrammed 
ab initio in the absence of Ezh2 or acutely depleted of the enzyme following reacquisition of pluripotency. In turn, this is consistent with the observation that H3K7me3 is partially retained upon Ezh2 inactivation.

Importantly, fully independent mutant iPSC clones were indistinguishable in their patterns of residual H3K27me3, indicating that retention of the mark either follows an instructive mechanism or, if stochastic, is subjected to a strong selection pressure that limits the range of residual targets. The former possibility appears more likely on the basis of two related observations: first, that residual H3K27me3 targets in mutant iPSC are preferentially associated to CpG islands, and second, that these residual H3K27me3 targets also correspond to genes with higher levels of H3K27me3 in normal iPSC. Combined, these evidences suggest that the basic mode of PRC2 recruitment is conserved in Ezh2-mutant cells, allowing prevalent targets to recruit enough mutant PRC2 for EZH11 to catalyze productive H3K27 trimethylation.

In mutant iPSC, further inhibition of PRC2 through knockdown of its critical subunit EED resulted in loss of H3K27me3 and re-expression of developmental regulator genes, indicating that maintenance and/or de novo acquisition of H3K27me3 is critical to ensure silencing of the MEF-specific gene expression program. Consistently, Eed knock-down in MEF prevented reprogramming. This result highlights the essential contribution of H3K27me3 to the repression of developmental regulator genes that enables successful cell fate reassignment.

Materials and Methods

Mice

Ezh2sp, OCT4-GFP and Cdkn2a−/− mice have been previously described [30,31,36]. Compound mutants were generated intercrossing individual strains. Primers used for genotyping are listed in Table S5.

Derivation of primary MEF and TTF

MEF were produced from E13.5 compound mutants following standard procedures. Briefly, embryos were harvested from pregnant females and MEF were isolated by enzymatic dissociation of tissues. Cells were cultured in MEF medium (DMEM high-glucose, 10% fetal calf serum, 2mM L-glutamine, 50 units/ml penicillin, 50 µg/ml streptomycin) at 37°C, 5% CO2 for up to

Figure 4. Genome-wide distribution of H3K27me3 in Ezh2SET/SET iPSC revealed through ChIP-seq. A. Pie chart showing partition of Polycomb targets based on H3K27 methylation status in Ezh2SET/SET iPSC. B. Venn diagram displaying overlap between H3K27me3 genes in Ezh2SET/SET (grey) and Ezh2SET/SET (purple) iPSC. C. Venn diagram showing overlap between H3K4me3+ genes in Ezh2SET/SET (grey), H3K27me3* genes in Ezh2SET/SET (orange) and H3K27me2+/H3K27me3− in Ezh2SET/SET (light blue). D. Analysis of transcript levels measured by qRT-PCR, and status of SUZ12 binding, H3K27me2 and H3K27me1 enrichment revealed by ChIP q-PCR at promoters of 7 genes overexpressed in that residual PRC2 is able through cell division to both propagate modification during the MEF to iPSC transition, thus indicating H3K27me3 in MEF, with the remaining ones acquiring the modification during the MEF to iPSC transition, thus indicating
passage-3 before TAT-Cre transduction and/or infection with reprogramming viruses. For the preparation of adult tail tip fibroblasts (TTF), 0.5 cm tail tips from 8 to 16-weeks old compound mutants were biopsied aseptically, subjected to enzymatic dissociation and cultured in MEF medium at 37°C, 5% CO2 for one week.

Figure 5. Effect of PRC2 inactivation on established Ezh2ΔSET/ΔSET iPSC clones and TF–induced reprogramming. A. Western blot analysis of EED protein levels in two Ezh2ΔSET/ΔSET iPSC clones infected with control virus (empty) or lentiviruses expressing independent short hairpin (sh) RNAs targeting Eed (sh19 and sh21) (left). Quantification of EED protein levels in infected cells after normalization based on Vinculin levels (right). B. H3K27me3 status (left panel) and expression levels (right panel) measured respectively by ChIP-qPCR and qRT-PCR, of 4 representative genes up regulated in MEF relative to iPSC, in two Ezh2ΔSET/ΔSET iPSC clones. Ezh2-mutant iPSC were infected with viruses expressing two independent hairpins for Eed or with a control virus. Status of H3K27me3 (±SEM) is represented as enrichment relative to input, after normalization for H3 density within the same amplicon. Expression levels are shown as fold change relative to iPSC infected with the empty vector. Error bars refer to qPCR triplicates. C. Western blot analysis of EED and H3K27me3 protein levels at day-6 of puromycin selection on Cdkn2a−/− Ezh2-proficient TTF expressing three independent Eed hairpins (lines 1, 2 and 3) or infected with a control virus (line 4). Quantification of protein levels relative to Vinculin or Histone H3 are shown in the right panel. D. AP staining of primary iPSC colonies obtained upon reprogramming of 1×10³ (upper panel) Cdkn2a−/− Ezh2-proficient TTF expressing three independent Eed hairpins (lines 1, 2 and 3) or infected with an empty lentiviral vector (line 4) used as control. E. Quantification of AP⁺ iPSC colonies. Column height represents number of AP⁺ iPSC colonies obtained from 1×10³ TTF expressing either one of the three independent Eed hairpins (lines 1, 2 and 3) or infected with an empty lentivirus vector (line 4) as control. Data are representative of two independent experiments performed using three different shRNAs.

doi:10.1371/journal.pgen.1003292.g005
Lentivirus production

Lentiviral constructs coding for doxycycline-inducible STEMCCA, reverse tetracycline transactivator (rtTA) and the Zs-green reporter gene have been previously described [32]. Production of lentiviral particles was performed as previously described [32]. Briefly, plasmids expressing viral proteins GAG, POL, REV, TAT, and the vesicular stomatitis virus envelope glycoprotein (VSV-G) were co-transfected with either STEMCCA or rtTA vectors into semi-confluent 293T cells by calcium phosphate precipitation in the presence of 25 μM chloroquine. Supernatant of transfected cells were collected every 12 hours during 2 consecutive days and concentrated by centrifugation. Viral particles were resuspended in MEF medium and either used freshly for infection or frozen at −80°C.

Infection of MEF and derivation of Oct4-GFP; Ezh2fl SET iPSC

MEF at passage 1 to 3 carrying either one (control) or both (mutant) Ezh2 floxed alleles were seeded at a density of 2×10^5 cells/well in a 6-well tissue culture dish. 24 hr later, MEF were transfused in a 1:1 PBS/DMEM/Glutamax (Invitrogen) solution with 50 μg/ml of recombinant Tat-Cre recombinase [46]. 16 hour after transduction, medium was replaced, and MEF were infected with the STEMCCA and rtTA lentiviruses. 24 hours following infection with STEMCCA and rtTA, MEF were harvested and seeded at clonal density on a mytomycin-treated MEF feeder layer. One day after, MEF medium was replaced with ESC medium (high glucose DMEM, 15% knockout serum replacement-Invitrogen-, 2 mM L-glutamine, 50 units/ml penicillin, 50 μg/ml streptomycin, 0.1 mM non-essential amino acids, 1/500 home-made leukemia inhibitory factor, 0.1 mM 2-mercaptoethanol) supplemented with 1 μg/ml doxycycline to induce expression of the reprogramming factors. Doxycycline treatment was replaced every other day for 23 days. Four days after doxycycline withdrawal, iPSC colonies were stained for alkaline phosphatase (AP), counted to assess the efficiency of reprogramming or picked and expanded on a mitomycin-treated MEF feeder layer. One day after, iPSC clones were grown in 2i/LIF medium (DMEM/F12+GlutaMAX, Neurobasal Medium, N2 Supplement-Invitrogen-, B27 supplement-Invitrogen-, 1.5 mM Heps, 2 mM Glutamine, 0.1 mM β-mercaptoethanol, 1/500 home-made LIF, 3 μM CHIR99021, 1 μM PD0325901.

Infection of TTF and derivation of Cdkn2afl−/−; Ezh2fl SET iPSC

Control Cdkn2afl−/−; Ezh2fl SET and experimental Cdkn2afl−/−; Ezh2fl SET TTFs at passage-1 were treated with 50 μg/ml of Tat-Cre recombinase and passage-3 days later. At 80% confluency, cells underwent a second round of Tat-Cre transduction and were further expanded to allow the dilution of H3K27me3, before infection with STEMCCA and rtTA lentiviruses, as described above.

RNA interference

Puromycin-resistant lentiviral constructs expressing Eed shRNAs were purchased from Open Biosystem (TRCN0000095719, TRCN0000095721, TRCN0000095722). Lentiviruses were produced as described above. Cdkn2afl−/− TTFs were infected with viruses expressing either shRNA # TRCN0000095719, TRCN0000095721, TRCN0000095722 or an empty pLKO.1 vector and selected with 2 μg/ml of puromycin. After 6 days of selection, resistant TTFs were reprogrammed by infection with Tet-STEMCCA as described above. Ezh2fl SET iPS cells were infected with viruses expressing either shRNA # TRCN0000095719, TRCN0000095721 or as control, an empty pLKO.1 vector and expanded for two weeks in 0.7 μg/ml of puromycin. RNA, proteins and chromatin were collected on the same day.

Alkaline phosphatase staining

Alkaline phosphatase staining was performed using the Leukocyte Alkaline Phosphatase kit (Sigma Aldrich) following manufacturer’s instructions.

Flow cytometry

One million cells were stained with phycoerythrin-conjugated anti-mouse/human SSEA1 (eBioscience, 12-8813) in FACS buffer (1% bovine serum albumin (BSA), 0.05% NaN3 in PBS), washed and acquired on a FACS Calibur instrument (BD Biosciences). Data were analyzed using FlowJo software (Tree Star inc.).

Teratoma assay

iPSC were cultured in 2i/LIF medium for 3 passages. 2×10^6 cells were injected subcutaneously into NOD-SCID Iβ2c−/− mice. 2-3 weeks after injection, mice were sacrificed and tumors were isolated and fixed in 4% formaldehyde for immunohistochemistry.

RNA extraction and cDNA synthesis

RNA was extracted using TRIzol Reagent (Invitrogen) and purified with RNeasy mini kit (QIAGEN) following manufacturer’s instructions. cDNA was prepared using SuperScript VILO cDNA Synthesis Kit (Invitrogen) following manufacturer’s instructions.

Quantitative gene expression analysis

Quantitative real-time PCR analysis for the expression of fibroblast specific genes on MEF, 2 Ezh2fl SET iPSC clones, 2 Ezh2fl SET iPSC clones, and 1 E14tg2a E S cell line was performed using pre-designed Taqman gene expression assays following manufacturer’s instructions. RNA was extracted using TRIzol Reagent (Invitrogen) and purified with RNeasy mini kit (QIAGEN) following manufacturer’s instructions. (Applied Biosystems: CtlAmMm 00438094_g1, Cehpa Mm00514283_s1, Dok1 Mm00438352_m1, Enpp2 Mm00901709_m1, Ezh2 Mm00501097_m1, Esa2 Mm00514283_s1, Etf1 Mm00441447_m1, Fads3 Mm00517643_m1, Bmp1 Mm00802220_m1, Ndg1 Mm00440447_m1, Snai2 Mm00441531_m1, Sox9 Mm00440840_m1, Tsirn1, Mm00525200_s1, Vangp5 Mm00441444_m1, Bmp7 Mm00802220_m1, Cap2 Mm00482643_m1, Vcan Mm01283063_m1, Ptpom Mm00436095_m1, Lox Mm00495386_m1, Pdk1 Mm00481111_m1). Quantitative real-time PCR analysis for the expression of the STEMCCA vector was performed using a custom-designed Taqman gene expression assay, previously described [32].

cDNA arrays

iPSC were cultured in 2i/LIF medium for 3 passages before RNA extraction. Expression was assessed using the TaqMan Stem Cell Pluripotency Array (Applied Biosystem 4385363) following the manufacturer’s instructions. dCT values were calculated subtracting each single CT value to the geometrical mean of the two reference genes (Actb, Raf1, Ctnnb1, Gapdh and Etf1a). Hierarchical clustering and heat map were obtained using Gene Cluster 3.0 software (Michael Eisen, Copyright 1998–99 Stanford University; Michiel de Hoon, 2002 University of Tokyo, Human Genome Center).
Microarray analysis

Microarray analysis was performed on 4 Ezh2+/− iPSC clones, 4 Ezh2ASET/+ASET iPSC clones, 3 independent preparation of MEFs and one sample of E1Htg2a ESC using Affymetrix Mouse Gene 1.0 ST arrays. Labeling, hybridization, and washing were performed according to Affimetrix guidelines. Data analysis was performed with R software version 2.15.0 (http://www.r-project.org) starting from raw data (CEL files). Data were normalized using robust multichip analysis (RMA) algorithm [47] as implemented in [48] and made available through Bioconductor (http://www.bioconductor.org/). To perform differential expression analysis, we started from the log2 normalized value relative to each gene. Then, we used a t-test to identify significant changes in gene expression between wild type and mutant samples, and a P value was calculated for each gene. A false discovery rate (FDR) procedure [49] was applied to take into account multiple testing correction. To identify differentially expressed genes (DEGs) we considered a threshold of 0.05 on the corrected P value, along with a cut-off of 1.5-fold changes.

Gene Ontology analysis

We used Cytoscape software [50] and the BiNGO plugin [51] to analyze Gene Ontology terms significantly enriched in the various gene sets. The enrichment for each term was tested using hypergeometric test and P values were corrected using FDR procedure. All terms with a FDR <0.01 were considered enriched.

Protein extraction and Western blot analysis

iPSC were cultured in 2i/LIF medium for 3 passages before protein extraction. Cell pellets were lysed with urea buffer (8 M Urea, 0.1 M Na2HPO4, 0.01 M Tris base diluted in water, pH 8.0) at room temperature for 30 minutes on a rotating wheel. Lysates were sonicated with a Bioruptor Sonication System (UCD200) (3 cycles of 30 seconds with one minute breaks, high power). Lysates were centrifuged at 13000 rpm for 15 minutes and supernatants were transferred to a new tube. Protein quantification was performed using Bio-Rad protein assay and following manufacturer’s instructions. For the detection of histone modifications 40 μg of total protein extracts were loaded into a 12% acrylamide gel. For the measurement of EZH2 levels, 80 μg of total protein lysates were loaded onto an 8% acrylamide gel. Western blot was performed using standard procedures. Intensities of Western blot bands were determined using ImageJ software (rsweb.nih.gov/ij/). Antibodies used for Western blot are listed in Table S3.

Mass spectrometry

A detailed description of core histone extraction and mass spectrometry analysis can be found in Text S1.

 Histology and immunohistochemistry

Teratoma tissues were washed in PBS buffer for 30 minutes, fixed in 4% buffered formalin for 4 hours and paraffin embedded. Samples were processed in consecutive 3 μm thick sections and stained with haematoxylin and eosin. Immunostaining for desmin, protein S-100 and cytokeratin were performed using an automated immunostainer (Autostainer, DakoCytomation, Glostrup, Denmark) and a commercially available detection kit (DakoEnVision Plus-HRP), according to the manufacturer’s instructions. Detailed information about antibodies and immunostaining procedure are listed in Table S4. After immunostaining, sections were counterstained with 1% modified Harris hematoxylin, dehydrated and mounted. Standard reference positive and negative controls were run simultaneously. Haematoyxlin and eosin and immunohistochemistry images were taken with an Olympus Upright BX 51 optical microscope equipped with a Nikon digital color camera. Digital images were processed with Adobe Photoshop CS3.

Chromatin immunoprecipitation

iPSC cultured in 2i/LIF medium were trypsinized and resuspended in 1% formaldehyde/PBS solution. Cross-linking was allowed to proceed for 10 min at room temperature and stopped by addition of glycine at a final concentration of 0.125 M, followed by an additional incubation for 5 min. Fixed cells were washed twice with PBS and resuspended in SDS buffer (SDS 0.5%, Tris-Cl pH 8.1, 50 mM, NaCl 100 mM, EDTA pH 8.5 mM, NaNO3 0.02%, diluted in ddH2O) and stored at −80°C. SDS cell suspensions were thawed at room temperature using a water bath for 20 minutes, centrifuged at 2000 rpm and resuspended in ice cold IP buffer (SDS buffer/Triton dilution buffer = 2:1; Triton dilution buffer: Triton X-100 5%, Tris-Cl pH 8.6 100 mM, NaCl 100 mM, EDTA pH 8.5 mM, NaNO3 0.02%, diluted in ddH2O). Fixed cells were sonicated yielding genomic DNA fragments with a bulk size of 400 to 1000 bp. Sonicated material was centrifuged at 13000 rpm for 30 minutes at 4°C and supernatants were transferred into a new tube. Chromatin was quantified using Bio-Rad protein assay following manufacturer’s instructions. For immunoprecipitation using antibodies against H3 and specific histone modifications, 100 μg of chromatin were used. For immunoprecipitation of SUZ12, 500 μg of chromatin were used. For each immunoprecipitation assay, chromatin was diluted in 1 ml IP buffer and 10 μl were taken and stored at −80°C as 1% of input. Primary antibodies were incubated overnight at 4°C on a rotating platform. To each sample, 50 μl of 50% slurry of protein A-Sepharose (Amersham) beads were added for 2–3 h. Beads were washed three times in 150 mM wash buffer and one time in 500 mM wash buffer (Triton-X 1%, NaCl 150 mM or 500 mM, Tris-Cl pH 8.0 20 mM, SDS 0.1%, EDTA pH 8.2 mM diluted in water). Beads (and input samples) were resuspended in 120 μl of 0.1% SDS, 0.1 M NaHCO3 buffer and de-cross-linked at 65°C overnight. DNA was purified using PCR purification kit (QIAGEN) following the manufacturer’s instructions and eluted in 100 μl of water. 1 μl of eluted material was used for each real-time quantitative PCR (qPCR) reaction. Quantitative real-time PCR analysis was performed in triplicate using Fast SYBR Green master mix (Applied Biosystems) in a 7500 Fast Real-Time PCR instrument (Applied Biosystems). Antibodies used for ChIP are listed in Table S3. Primers used for ChIP-qPCR are listed in Table S5.

ChIP-seq

ChIP-seq analysis was performed on two Ezh2+/− iPSC clones and two Ezh2ASET/+ASET iPSC clones. 10 ng of sheared DNA obtained before (input) and after chromatin IP using anti-H3K27me3 and -H3K27me2 specific antibodies, were prepared with the Illumina ChIPSeq sample prep kit and multiplexing oligonucleotide kit. DNA libraries were quantified using a high sensitivity Chip on Bioanalyzer (Agilent) and diluted to a concentration of 16 pM. Diluted libraries were used for cluster generation and sequencing on a HiSeq 2000 instrument (Illumina) following manufacturer’s protocol.

Bioinformatics analysis of ChIP-seq data

Mapping of reads. After filtering for artifacts with FASTX- Toolkit v.0.0.13, reads were aligned to the mm9 genome using Bowtie v.0.12.7 [52], allowing up to two mismatches per read and discarding multiply-aligning reads. Each ChIP-seq experiment had

Selectivity of H3K27me3 in Cell Reprogramming
a total of 18 to 28 millions of uniquely mapped reads. Aligned reads of a MEF H3K27me3 ChIP-seq dataset were downloaded from the Gene Expression Omnibus (accession id GSM656316) and processed similarly to the other datasets (see below).

**Identification of enriched domains.** We used two different strategies to identify enriched domains using, respectively, MACS v.1.4.0 [53] and RSEG v.0.4.8 [54]. For MACS analysis, we disabled the shifting model and the dynamic lambda (which are inappropriate for histone modifications) and used a stringent p-value threshold of 10e-10. A more generous identification of enriched regions was done using RSEG, which was specifically designed to identify extended regions marked by histone marks (RSEG was independently benchmarked in [55]). We used Hideaki’s empirical method to determine bin size, 20 iterations for the training, and provided deadzones (regions where no read can map uniquely) as described in the documentation. Domains called as unconfident were discarded. In both approaches, we compared the signal over the input.

On the basis of a calibration performed in ESC and of biological and technical considerations (see Text S1 for further details), we applied the MACS method to identify H3K27me3 enriched genomic regions, while RSEG was used to identify domains of enriched H3K27me2.

**Annotation.** For both H3K27me3 and H3K27me2, we assigned enriched regions to genes by determining for each RefSeq transcript, whether an enriched domain overlapped with a +/-5 kb region around the transcription start site. We also tested a +/-2.5 kb interval, which resulted in a marginal decrease in the number of gene annotations. Transcripts were then collapsed to gene symbols using BioMart [56], and only the genes marked in both biological replicates (Figure S4D) were retained for further analysis.

Heat map and clustering of H3K27me3 and H3K27me2 distributions around the TSS were produced with Seqminer [57] using K-means linear normalization.

Venn diagrams showing overlap between different datasets were prepared using BioVenn (by Tim Hulsen http://www.cmbi.ru.nl/cdd/biovenn/).

**Ethics statement**

This project involved the minimum number of mice required to fulfill the research objectives. Experiments involving animals were performed in accordance with the Italian Laws (D.Lvo 116/92 and following additions), which enforces EU 86/609 Directive (Council Directive 86/609/EEC of 24 November 1986 on the approximation of laws, regulations and administrative provisions of the Member States regarding the protection of animals used for experimental and other scientific purposes). The authority responsible for ensuring compliance with the provisions of EU 86/609 Directive is the Italian Ministry of Health. Our mouse facility is authorized by the Ministry of Health (DM N°86/2005 - 17/06/2005) and a veterinarian is responsible for the well-being of the experimental animals.

An Institutional Animal Care and Use Committee (IACUC) of the IFOM Foundation-FIRC Institute of Molecular Oncology Foundation, supervises the ethical conduct of research involving non-human vertebrates.

**Supporting Information**

**Figure S1** Molecular and biochemical characterization of Ezh2 mutant iPSC clones. A. Flow cytometric analysis of Ezh2 control (+/ASET, upper row) and mutant (ASET/ASET, lower row) MEF infected with STEMCCA, rtTA and GFP-expressing lentiviruses in two replicate experiments (left and right panel). MEF are gated according to cell size (ESC) and GFP expression. Numbers within dot plots indicate percentage of gated cells. B. Table indicating the status of the Ezh2 floxed allele in individual iPSC clones isolated upon reprogramming of TAT-Cre transduced Ezh2fl/+ and Ezh2fl/fl MEF. Numbers of iPSC clones of the indicated Ezh2 genotypes and total number of isolated clones are shown. C. Western blot analysis of H3K9me3 levels in embryonic stem cells (ESC), control (+/+ and ASET/+ and mutant (ASET/ ASET) iPSC (two clones/genotype). H3 levels were used as control for protein loading. D. Transcription from the STEMCCA vector as revealed by qRT-PCR in representative control (n = 4) and mutant (n = 3) iPSC clones. Measurements are relative to STEMCCA transcript levels detected in infected MEF treated for 10 days with doxycycline. Uninfected MEF were used as negative control. Standard deviations refer to replicates of the qPCR reaction. E. Annotated MS/MS spectra of H3 peptide 27–40 species, with one or more co-existing post-translational modifications detected in Ezh2 control (+/+; left spectra) and mutant (ASET/ASET) iPSC. Spectra are displayed according to whether the peptide contained a total of 3 (upper row), 4 (middle row) or 5 (lower row) methyl groups. The m/z ratio of b- and y- product ions identified are annotated in the spectrum and also reported along the amino acid sequence in blue and red, respectively. (TIF)

**Figure S2** Identification of MEF specific genes by cDNA microarray analysis. Heat map representation of the average expression profile of MEF coming from 3 different embryos, 4 Ezh2 control (+/+), 4 Ezh2 mutant (ASET/ASET) and 4 Ezh2 mutant (ASET/ASET) iPSC clones. Shown are the expression levels of 3644 genes differentially expressed between iPSCs and MEFs (p-value = 0.05; f.c. = 1.5). Expression ranges from lower (green) to higher levels (red). (TIF)

**Figure S3** Establishment of iPSC clones upon genome-wide erasure of H3K27me3 at the onset of reprogramming. A. H3K27me3 and p16ink4a protein levels measured by Western blot analysis in representative populations of TAT-Cre transduced Ezh2 control (+/ASET) and mutant (ASET/ASET) MEF after two passages and 11 days of culture. As comparison, representative control (+/+; left column) and mutant (ASET/ASET) iPSC clones were analyzed. Vinculin protein levels were used as loading control. B. Flow cytometric assessment of the infection efficiency of Ezh2 control (Ezh2ASET/+; right columns) and mutant (Ezh2ASET/ASET, left columns) Cdki2a−/−/TTFs (3 independent batches/genotype). Cells were infected with STEMCCA, rtTA and GFP expressing lentiviruses. Numbers within plots indicate percentage of gated cells. (TIF)

**Figure S4** Epigenetic characterization of Ezh2ASET/ASET iPSCs. A. Analysis of transcript levels (qRT-PCR) and status of PRC2 (SUZ12), H3K27me2 and H3K27me1 (ChIP-qPCR) enrichment at promoters of 17 genes overexpressed in MEF vs. iPSC. For all analyses, two Ezh2 control (+/+; grey) iPSC clones were compared to two mutant (ASET/ASET; purple) counterparts. Levels of expression are shown as ddCt (log2 scale) relative MEF. Status of a particular histone modification (±SEM) is represented as enrichment relative to input, after normalization for H3 density within the same amplicon. SUZ12 enrichment at promoters of the indicated genes is assessed comparing it to that of unrelated IgG. Error bars referred to qPCR triplicates. B. Distribution of the maximum height of H3K27me3 signal in Ezh2ASET/ASET iPSC according to H3K27me status in 2 Ezh2ASET/ASET iPSC clones: genes that
lose H3K27me3 (red lines), genes that retain H3K27me3 (green lines). P-value<2.2e-16 (two-sided t-test). C. Heat map representation of H3K27me3 and H3K27me2 distributions in a ±5 Kb window around the TSS of genes marked by at least one of the two methylation states. Data from 2 representative clones per genotype are shown. Regions were inverted depending on the direction of transcription. Genes were clustered according to the combination of the two marks. D. Venn diagrams showing overlap of H3K27me3 targets (upper row) or H3K27me2 targets (lower row) between 2 iPSC clones of the same genotype (Ezh2+/−, left; Ezh2−/−, right).

**Figure S5** Targets of H3K27me3 in Ezh2 mutant iPSC are enriched for transcriptional regulators and developmental determinants. A. Tree diagram representing the main steps that lead to the identification of 175 genes that acquired H3K27me3 de novo in the MEF to iPSC transition. See main text for further explanation. B Gene ontology analysis of de novo H3K27 trimethylated genes in the MEF to iPSC transition. Bars represent P values in −Log2 scale of the corresponding biological process. Dashed line indicates significance threshold. C. Network showing protein-DNA interactions (blue lines) between genes that change expression upon reprogramming (blue circles) and the four identified master regulators of fibroblast expression program (yellow circles). D. qRT-PCR analysis of *Ets1* and *Egr1* transcript levels at day 0, day 3, day 6, day 10 and day 14 of MEF reprogramming. mRNA levels are normalized to *Tbp*. Error bars represent qPCR triplicates.

**Figure S6** TF-induced reprogramming in the absence of PRC2. A. Experimental time line showing the main experimental steps: Infection of Ezh2 proficient Cola2tg−/− TTMs with control virus (empty) or lentiviruses expressing independent short hairpin (sh) RNAs targeting *Eod1*; puromycin selection; Infection with Tet-STEMCCA and rtTA lentiviruses; replating of infected cells at cloning dilution and addition of ESC medium supplemented with doxycycline; Doxycycline withdrawal and scoring of AP-positive colonies.

**Table S1** qRT-PCR analysis of transcript levels in two representative Ezh2+/− (# 40A7 and # 40A8) and Ezh2−/− (# 4B3 and # 4B7) iPSC clones, of a selected list of genes controlling stem cell pluripotency, self renewal and differentiation. Transcript levels were also measured in ESC and MEF. Data are represented as dCts relative to the geometric mean of 5 housekeeping genes.
15. Wang T, Chen K, Zeng X, Yang J, Wu Y, et al. (2011) The histone demethylases Jhd1m1a/b enhance somatic cell reprogramming in a vitamin-C-dependent manner. Cell Stem Cell 9: 575–587.

16. Koche RP, Smith ZD, Adli M, Gu H, Ku M, et al. (2011) Reprogramming factor protein initiates widespread targeted chromatin remodeling. Cell Stem Cell 8: 96–105.

17. Maherani N, Sritharan R, Xie W, Utikal J, Emili S, et al. (2007) Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue repair. EMBO J 26: 4061–4071.

18. Morey L, Helin K (2010) Polycomb group protein-mediated repression of transcription. Trends Biochem Sci 35: 323–332.

19. Mohr F, Weber M, Rehban M, Roloff TC, Richter J, et al. (2008) Lineage-specific polycomb targets analyze novel Ezh2 methyltransferase define restriction and potential of neuronal progenitors. Mol Cell 30: 755–766.

20. Mohr F, Schubeler D (2009) Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet 25: 129–136.

21. Boyer LA, Plath K, Zeitlinger J, Brambink T, Medeiros LA, et al. (2006) Polycomb complexes repress developmental regulators in marine embryonic stem cells. Nature 441: 349–353.

22. Lee TI, Jenner RG, Boyer LA, Gurunther MG, Levine SS, et al. (2006) Control of developmental regulators by Polycomb in human embryonic stem cells. Cell 125: 301–315.

23. Leeb M, Pasini D, Novatchkova M, Jaritz M, Helin K, et al. (2010) Polycomb complexes act redundantly to repress genomic repeats and genes. Genes Dev 24: 2935–2976.

24. O’Carroll D, Ezardth S, Pagani M, Barton SC, Surani MA, et al. (2001) The polycomb group gene Ezh2 is required for early mouse development. Mol Cell Biol 21: 4330–4336.

25. Pasini D, Bracken AP, Hansen JB, Capillo M, Helin K (2007) The polycomb group protein Suz12 is required for embryonic stem cell remodelling. Mol Cell Biol 27: 3769–3779.

26. Pasini D, Bracken AP, Jensen MR, Lasserri Denchi E, Helin K (2004) Suz12 is essential for mouse development and for Ezh2 histone methyltransferase activity. EMBO J 23: 4061–4071.

27. Shen X, Liu Y, Hsu YJ, Fujiwara Y, Kim J, et al. (2008) EZH1 mediates methylation on histone H3 lysine 27 and complements EZH2 in maintaining pluripotency. Cell Stem Cell 2: 347–356.

28. Smoother LA, Zemelman BV, Iavarone A, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. Nature 469: 343–349.

29. Hirabayashi Y, Suzuki N, Tuob M, Endo TA, Toyoda T, et al. (2009) Polycomb limits the neurogenic competence of neural precursor cells to promote astroglial fate transition. Neuron 63: 600–613.

30. Testa G (2011) The time of timing: How Polycomb proteins regulate neurogenesis. Bioessays 33: 519–529.

31. Mohn F, Schubeler D (2009) Genetics and epigenetics: stability and plasticity during cellular differentiation. Trends Genet 25: 129–136.

32. Simon JA, Kingston RE (2009) Mechanisms of polycomb gene silencing: knowns and unknowns. Nat Rev Mol Cell Biol 10: 697–708.

33. Margueron R, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. Nature 469: 343–349.

34. Micsinai M, Parisi F, Strino F, Asp P, Dynlacht BD, et al. (2012) Picking ChIP-seq MINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res 40: e70.

35. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T, et al. (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27: 431–432.

36. Reiner A, Yekutieli D, Benjamini Y (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. Bioinformatics 19: 368–375.

37. Song Q, Smith AD (2011) Identifying dispersed epigenomic domains from ChIP-seq data. Bioinformatics 27: 670–671.

38. Smoother LA, Zemelman BV, Iavarone A, Reinberg D (2011) The Polycomb complex PRC2 and its mark in life. Nature 469: 343–349.

39. Pirruccello BF, Tsukazaki M, Schuchardt U, Sauer S, Ryan NK, et al. (2010) ESCs require PRC2 to direct the successful reprogramming of differentiated cells toward pluripotency. Cell Stem Cell 3: 531–541.

40. Ye T, Krebs AR, Wu X, Choukral MA, Krimer C, Plevnik F, et al. (2011) seqMINER: an integrated ChIP-seq data interpretation platform. Nucleic Acids Res 39: e35.