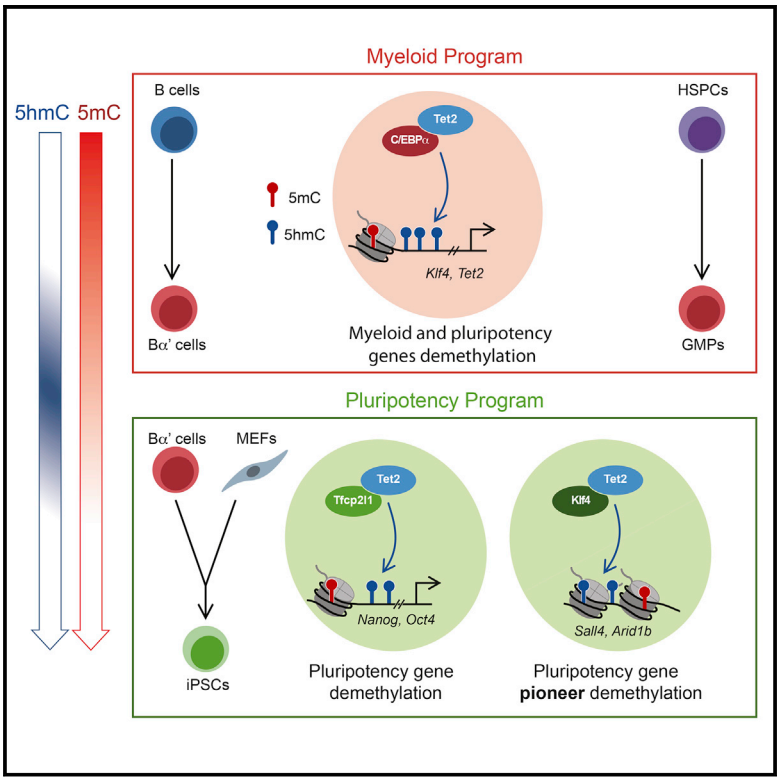
Article

Cell Stem Cell

Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate

Graphical Abstract



Highlights

- Base-resolution profiling of DNA (hydroxy)methylation during iPSC reprogramming
- Major contribution of Tet2-mediated demethylation throughout reprogramming
- C/EBPα, Klf4, and Tfcp2l1 drive Tet2-mediated enhancer demethylation and activation
- Klf4 induces enhancer demethylation in the absence of nucleosome repositioning

Authors

Jose Luis Sardina, Samuel Collombet, Tian V. Tian, ..., Konrad Hochedlinger, Denis Thieffry, Thomas Graf

Correspondence

joseluis.sardina@crg.eu (J.L.S.), thomas.graf@crg.eu (T.G.)

In Brief

Using a highly efficient reprogramming system, Sardina et al. examined the dynamics of DNA methylation and hydroxymethylation. They found that throughout the process several transcription factors can recruit Tet2 to specific sites, leading to demethylation. Some of these sites became demethylated before chromatin opening.

Sardina et al., 2018, Cell Stem Cell *23*, 1–15
November 1, 2018
https://doi.org/10.1016/j.stem.2018.08.016

CellPress

Please cite this article in press as: Sardina et al., Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate, Cell Stem Cell (2018), https://doi.org/10.1016/j.stem.2018.08.016

Cell Stem Cell
Article

Transcription Factors Drive Tet2-Mediated Enhancer Demethylation to Reprogram Cell Fate

Jose Luis Sardina,[1,10,*] Samuel Collombet,[3,10] Tian V. Tian,[1] Antonio Gómez,[1] Bruno Di Stefano,[4,5,6,7,8] Clara Berenguer,[1] Justin Brumbaugh,[4,5,6,7,8] Ralph Stadhouders,[1] Carolina Segura-Morales,[1] Marta Gut,[2,9] Ivo G. Gut,[2,9] Simon Heath,[2,9] Sergi Aranda,[1] Luciano Di Croce,[1,2] Konrad Hochedlinger,[4,5,6,7,8] Denis Thieffry,[3] and Thomas Graf[1,2,11,*]

[1]Gene Regulation, Stem Cells and Cancer Program, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Barcelona 08003, Spain
[2]Universitat Pompeu Fabra (UPF), Barcelona 08003, Spain
[3]Institut de Biologie de l'Ecole Normale Supérieure (IBENS), Paris Sciences & Lettres Université, CNRS UMR8197, INSERM U1024, Paris 75005, France
[4]Department of Molecular Biology, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Boston, MA 02114, USA
[5]Center for Regenerative Medicine, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Boston, MA 02114, USA
[6]Cancer Center, Massachusetts General Hospital, Harvard Medical School, 185 Cambridge Street, Boston, MA 02114, USA
[7]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA
[8]Harvard Stem Cell Institute, 1350 Massachusetts Avenue, Cambridge, MA 02138, USA
[9]CNAG-CRG, Centro Nacional de Análisis Genómico, Centre for Genomic Regulation (CRG), The Barcelona Institute of Science and Technology (BIST), Baldiri i Reixac 4, Barcelona 08028, Spain
[10]These authors contributed equally
[11]Lead Contact
*Correspondence: joseluis.sardina@crg.eu (J.L.S.), thomas.graf@crg.eu (T.G.)
https://doi.org/10.1016/j.stem.2018.08.016

SUMMARY

Here, we report DNA methylation and hydroxymethylation dynamics at nucleotide resolution using C/EBPα-enhanced reprogramming of B cells into induced pluripotent cells (iPSCs). We observed successive waves of hydroxymethylation at enhancers, concomitant with a decrease in DNA methylation, suggesting active demethylation. Consistent with this finding, ablation of the DNA demethylase Tet2 almost completely abolishes reprogramming. C/EBPα, Klf4, and Tfcp2l1 each interact with Tet2 and recruit the enzyme to specific DNA sites. During reprogramming, some of these sites maintain high levels of 5hmC, and enhancers and promoters of key pluripotency factors become demethylated as early as 1 day after Yamanaka factor induction. Surprisingly, methylation changes precede chromatin opening in distinct chromatin regions, including Klf4 bound sites, revealing a pioneer factor activity associated with alternation in DNA methylation. Rapid changes in hydroxymethylation similar to those in B cells were also observed during compound-accelerated reprogramming of fibroblasts into iPSCs, highlighting the generality of our observations.

INTRODUCTION

Cytosine methylation of CpGs is the major epigenetic modification of mammalian DNA and plays important roles in development and cancer (Bird, 2002; Hackett and Surani, 2014). Although DNA methylation is generally assumed to be associated with transcriptional repression, its role in gene regulation during differentiation remains poorly understood (Smith and Meissner, 2013). DNA demethylation can be either passive, by dilution of DNA methylation after each cell division, or active when initiated by Tet dioxygenases (Wu and Zhang, 2017). During active demethylation, Tet enzymes first catalyze the iterative oxidation of 5-methylcytosine (5mC) to 5-hydroxymethylcytosine (5hmC) and subsequently to higher oxidized derivatives. These oxidized forms of cytosine in turn can either be lost during replication or enzymatically removed, restoring unmodified cytosine (Wu and Zhang, 2017).

Cell-fate decisions are typically initiated by transcription factors (TFs) that regulate gene expression in concert with epigenetic modifications (Apostolou and Hochedlinger, 2013). Recent studies have shown that DNA methylation can modulate TFs binding to DNA (Domcke et al., 2015; Yin et al., 2017) and 5hmC can recruit chromatin remodeling complexes (Yildirim et al., 2011; Zhang et al., 2015). However, whether the dynamic interplay between DNA modifications and chromatin-associated proteins is a driving force of cell-fate decisions, including somatic cell reprogramming into induced pluripotent stem cells (iPSCs) (Takahashi and Yamanaka, 2006), is unknown. This is largely due to the low proportion of somatic cells that can be reprogrammed, an obstacle that was recently overcome by the development of highly efficient reprogramming protocols (Bar-Nur et al., 2014; Di Stefano et al., 2014; Rais et al., 2013; Vidal et al., 2014). In addition, it is now possible to measure not only 5mC but also 5hmC genome-wide at single-nucleotide resolution (Booth et al., 2012; Yu et al., 2012). The combination of these methodologies, together with the transposase-based ATAC assay to monitor chromatin accessibility (Buenrostro

et al., 2013), now allows studies addressing the causal relationship between DNA methylation and enhancer activity during induced cell-fate changes.

Of the three Tet enzymes described, Tet2 seems to be the family member most important for somatic cell reprogramming. Ablation of Tet2 partially inhibits the upregulation of myeloid genes during C/EBPα-induced B cell transdifferentiation into macrophages (Kallin et al., 2012). Likewise, the knockdown or knockout of Tet2 partially impairs the OSKM-induced reprogramming of B cells or MEFs, respectively (Di Stefano et al., 2016; Doege et al., 2012; Hu et al., 2014). Conversely, Tet2 overexpression increases the reprogramming efficiency of B cells (Di Stefano et al., 2014). As Tet2 lacks a DNA binding domain, it needs to be recruited by TFs to gene regulatory elements (GREs). A number of TFs interacting with Tet2 have been described, including PU.1 (de la Rica et al., 2013), Wt1 (Wang et al., 2015), Nanog (Costa et al., 2013), and Sall4 (Xiong et al., 2016). However, the role of Tet2-mediated active demethylation in cell reprogramming, and which TFs recruit the enzyme to DNA during this process, is poorly understood.

Here, we have studied the dynamics of DNA methylation and hydroxymethylation at single-nucleotide resolution during the highly efficient C/EBPα-enhanced conversion of B cells toward iPSCs (Di Stefano et al., 2016; Di Stefano et al., 2014). In parallel, we have also assessed 5hmC distribution during the rapid MEF to iPSC conversion facilitated by ascorbic acid and GSK3bi (Bar-Nur et al., 2014). Our data show that Tet2-dependent demethylation occurs throughout reprogramming and suggest that at least three TFs can recruit Tet2 to GREs of relevant target genes.

RESULTS

Dramatic Redistribution of 5mC and 5hmC during Somatic Cell Reprogramming

To study the dynamics of DNA methylation during the conversion of B cells into iPSCs, we used our previously described two-step system, consisting of pre-B cells from a reprogrammable mouse containing a tetracycline-inducible OSKM cassette, and transduced with a β-estradiol-inducible form of C/EBPα (Di Stefano et al., 2016; Di Stefano et al., 2014). Cultured pre-B cells (B cells) were first treated with β-estradiol for 18 hr, resulting in Bα′ cells, followed by doxycycline treatment for 1–4 days to initiate reprogramming to pluripotency, resulting in day 1, day 2, and day 4 cells. Embryonic stem cells (ESCs) cultured in 2i conditions (Ying et al., 2008) were used as an end-stage control (Figure 1A). We then generated genome-wide, nucleotide resolution maps for 5mC and 5hmC using high coverage bisulfite and ox-bisulfite sequencing (BS-seq/oxBS-seq) (Booth et al., 2012). In addition, to study the crosstalk between DNA methylation and chromatin accessibility we performed ATAC sequencing (ATAC-seq) (Buenrostro et al., 2013) with the same samples. 5mC and 5hmC levels were determined by computing the values of approximately 10 million CpG residues per sample at >10× coverage.

We observed a genome-wide progressive loss of 5mC as early as day 1, associated with a concomitant gain of 5hmC (Figure S1A). These changes were detectable across intergenic regions, promoters, exons, and introns (Figure S1B). By day 4 of reprogramming, ~27% of the 5mC signal and ~11% of the 5hmC signal was redistributed compared to the starting B cells (Figures 1B and 1C). In addition, principal-component analysis (PCA) for 5mC levels in our samples revealed that day 4 cells were largely similar to ESCs cultured in FBS+LIF conditions (Figure S1C). As expected, ESCs grown in 2i were almost completely devoid of 5mC as previously described (Habibi et al., 2013).

These results demonstrate the high quality of our data and reveal a dramatic redistribution of DNA methylation and hydroxymethylation signals during reprogramming of B cells to iPSCs.

Transient and Sustained Gains in Hydroxymethylation at GREs

Since the majority of ATAC peaks correspond to enhancers and promoters, we used them as a proxy for GREs. To study the dynamics of hydroxymethylation, we therefore focused on 343,249 ATAC peaks that showed >2-fold changes between any of the reprogramming time points. Next, we filtered those peaks that exhibited >20% change of 5hmC and contained >3 CpGs, resulting in 10,982 regions (Figures S1D and S1E). To explore the role of active demethylation, we compared the dynamics of 5hmC with 5mC and observed an initial negative correlation that progressively diminished during reprogramming (Figure 1D). This suggested a predominantly active demethylation at the B, Bα′, and day 1 cell stages, and an increase of passive demethylation at the day 2 and day 4 stages. Unsupervised clustering of the 10,982 5hmC regions resulted in 6 major groups, showing a sequential and mostly transient increase in hydroxymethylation (Figure 1E). The corresponding 5mC signal concomitantly decreased, supporting the notion that these regions become actively demethylated. Approximately half of the clusters gained 5hmC transiently, as exemplified for clusters 10, 2, and 15 (Figures 1F and S1F). Unexpectedly, in the remaining clusters 5hmC was either partially maintained (e.g., cluster 13) or more fully maintained during at least two consecutive time points (e.g., at the day 2 to day 4 transition in cluster 19) (Figure 1F). The latter dynamics is illustrated by −50-Kb enhancer of *Jarid2* (Figure S1G), gene encoding a histone demethylase. Of note, this family of proteins has been involved in the regulation of cell fate and reprogramming (Chen et al., 2013). Regions within the 5 clusters analyzed (2, 10, 13, 15, 19) showed a differential enrichment of motifs related to pluripotency factors (Klf and Oct/Pou) and to factors involved in genome topology (Ctcf and YY) (Figure S1H).

Our data indicate that active demethylation occurs in waves throughout reprogramming to pluripotency. In addition, unexpectedly, some regions exhibited sustained 5hmC marking during specific cell-state transitions.

Tet2 Activity Is Required for iPSC Reprogramming during Both the C/EBPα and OSKM Induction Steps

The sequential increase of the 5hmC mark observed during the B cell to iPSC conversion is likely due to the enzymatic activity of Tet proteins, known to convert 5mC into 5hmC (Pastor et al., 2013). Of the three Tet family members, only *Tet2* was found to be upregulated at the RNA level (Figure 2A). Earlier work in B cells with partial knockdown of *Tet2* using shRNAs showed a modest decrease of reprogramming efficiency (Di Stefano et al., 2016). Likewise, ablation of *Tet2* in mouse embryo fibroblasts (MEFs) partially inhibited reprogramming into iPSCs, but ablation of all

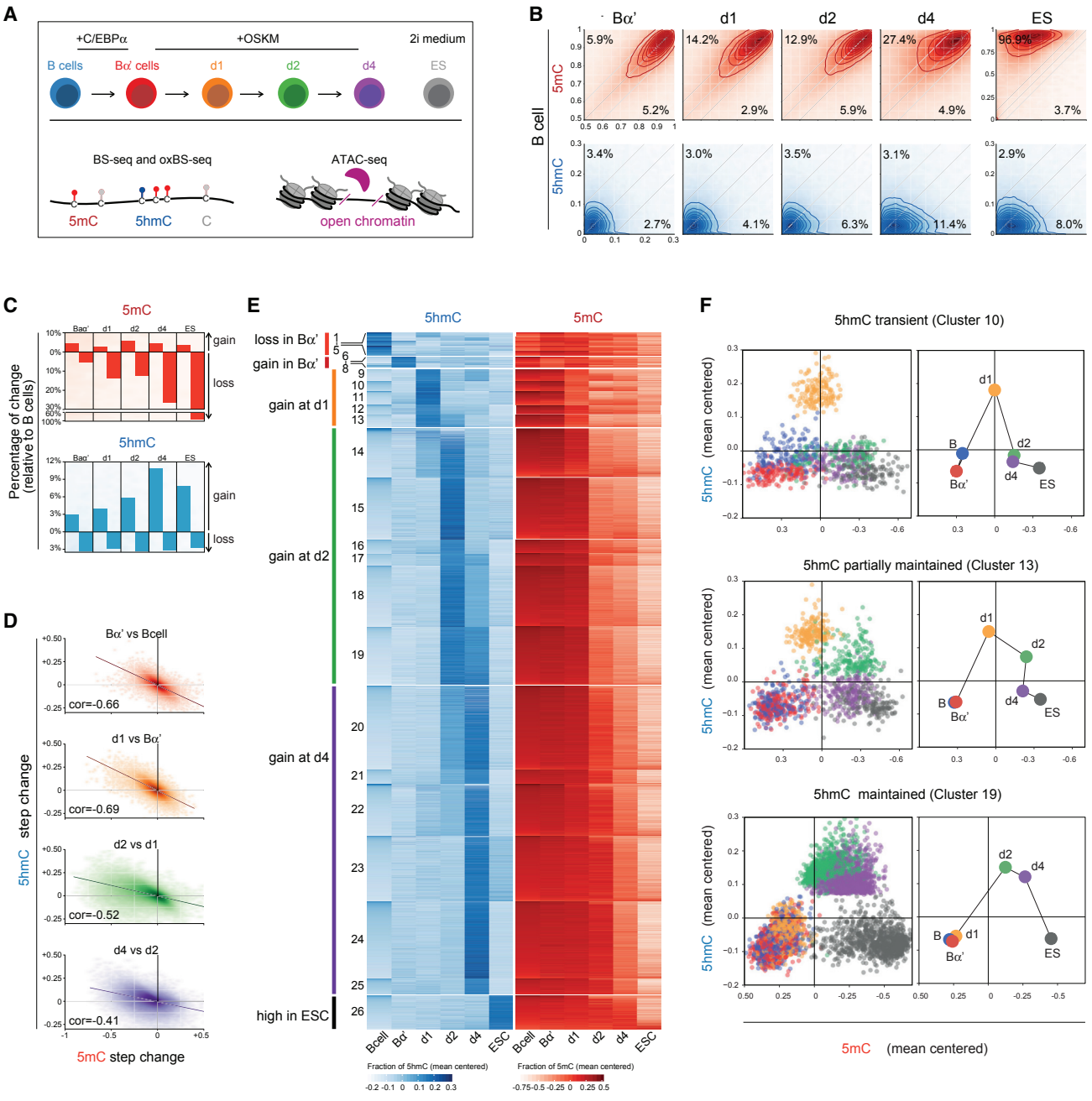


Figure 1. Active Demethylation throughout Reprogramming

(A) Schematic overview of samples and methodology used.

(B) Scatterplots representing the fraction of 5mC (red, top panels) and 5hmC (blue, bottom panels) in 1-kb bins genome-wide, in B cell (y axis) and for each subsequent time point (x axis). The diagonal lines represent no change (center), a decrease of 5mC/5hmC of more than 10% (top line), or an increase of more than 10% (bottom line). Top-left numbers: percentage of regions losing >10% 5mC/5hmC compared to B cell. Bottom-right numbers: percentage of regions gaining >10% 5mC/5hmC compared to B cell. Note that the scale for the scatterplot of 5mC in ESCs versus B cell is different (0–1 instead of 0.5–1) as the level of 5mC in ESCs is much lower than in the other conditions.

(C) Barplots displaying the number of regions changing in (B).

(D) Correlations between the step changes of 5mC and 5hmC at chromatin accessible regions (ATAC peaks), at each time point during reprogramming.

(E) Clustering of 5hmC dynamics at ATAC peaks showing at least 20% changes of 5hmC during reprogramming. Left: 5hmC signal. Right, 5mC signal.

(F) Scatterplots showing the relative dynamics of 5hmC and 5mC at selected clusters from (E). Top: transient 5hmC (cluster 10). Middle: partially maintained 5hmC (cluster 13); Bottom: maintained 5hmC (cluster 19).

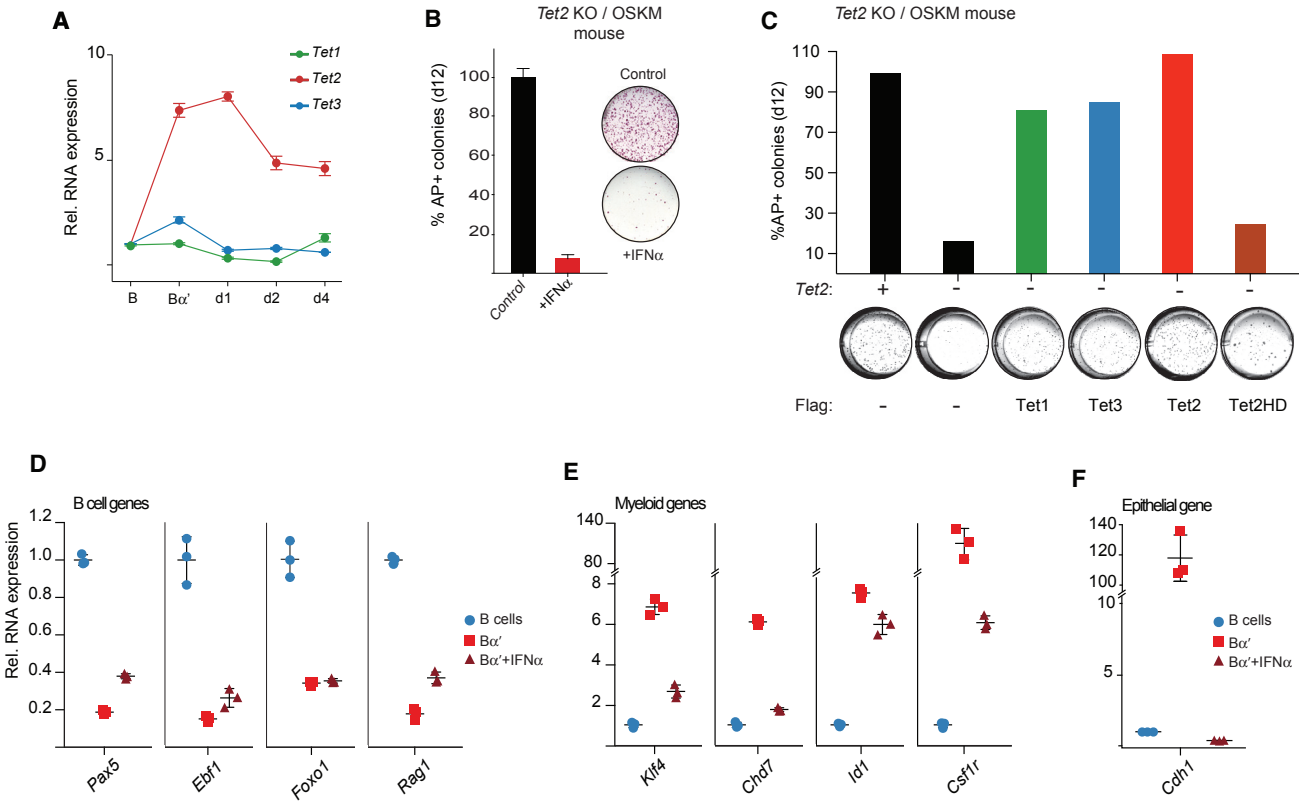


Figure 2. Tet2 Activity Is Required for Efficient B to iPSC Reprogramming

(A) Gene expression (qRT-PCR) of Tet enzymes during reprogramming. Values were normalized against *Pgk* expression. Error bars indicate SD (n = 3 biologically independent samples).

(B) Alkaline phosphatase positive (AP^+) iPSC colony counts after reprogramming of *Tet2*-inducible knockout (KO) B cells. *Tet2* excision was induced by IFN-α treatment (250 U/mL). Counts were normalized against untreated cells. Error bars indicate SD (n = 3 biological independent experiments). Right: representative alkaline phosphatase staining (AP^+) of iPS colonies after reprogramming of *Tet2*-inducible KO B cells.

(C) AP^+ iPSC colony counts after reprogramming of *Tet2*-inducible KO B cells ectopically expressing different Tet proteins. *Tet2* excision was induced by IFN-α treatment (250 U/mL). Counts were normalized against untreated cells for every experimental condition essayed. Bellow: representative AP staining of iPS colonies after reprogramming of *Tet2*-inducible KO B cells at the different experimental conditions essayed.

(D) Expression (qRT-PCR) of B cell genes upon C/EBPα induction of *Tet2*-inducible KO B cells. Values were normalized against *Pgk* expression. Error bars indicate SD (n = 3 biologically independent samples).

(E) Expression (qRT-PCR) of myeloid genes upon C/EBPα induction of *Tet2*-inducible KO B cells. Values were normalized against *Pgk* expression. Error bars indicate SD (n = 3 biologically independent samples).

(F) Expression (qRT-PCR) of epithelial gene *Cdh1* upon C/EBPα induction of *Tet2*-inducible KO B cells. Values were normalized against *Pgk* expression. Error bars indicate SD (n = 3 biologically independent samples).

three *Tet* genes completely abrogated reprogramming (Hu et al., 2014), likely because redundancy between Tet enzymes. To test the effect of a full *Tet2* knockout in B cells, we generated a mouse strain with two floxed *Tet2* alleles, an interferon (IFN)-inducible Mx1-Cre transgene, along with a doxycycline-inducible OSKM cassette, a tetracycline transactivator (rtTA), and the *Oct4*-GFP reporter transgene (Figure S2A). B cells from these mice treated with IFN-α showed negligible levels of *Tet2* mRNA (Figure S2B) without decrease of cell viability (Figure S2C). Deletion of *Tet2* by IFN-α-treatment during reprogramming of B cells (β-estradiol followed by doxycycline treatment) led to the almost complete impairment of iPSC colony formation (Figure 2B). This phenotype could be rescued by overexpression of Tet2 as well as by Tet1 and Tet3, but not by a catalytically dead version of Tet2 (HD; Ko et al., 2010) (Figures 2C and S2D–S2F). Gene expression analyses of the inducible *Tet2* knockout B cells showed silencing of the B cell program in response to C/EBPα treatment (Figure 2D), but only a partial upregulation of myeloid genes (Figure 2E). In addition, the *Tet2* knockout B cells failed to activate the epithelial gene *Cdh1* (Figure 2F), which we showed to initiate the C/EBPα-induced mesenchymal-to-epithelial transition (Di Stefano et al., 2016). This is consistent with a major role of Tet enzymes in the earliest stages of reprogramming, as described for the conversion of MEFs to iPSCs (Hu et al., 2014).

To explore the effect on reprogramming of Tet inhibition at different phases of induction, we used the broad dioxygenase inhibitor dimethyloxalylglycine (DMOG) (Amouroux et al., 2016). Fluorescence-activated cell sorting (FACS) analysis of Oct4-GFP expression in cells treated during either the C/EBPα pulse or the OSKM induction phase showed a 30%–40% reduction 4 days after OSKM induction and complete inhibition when applied during both phases (Figure S2G). When iPSC colonies

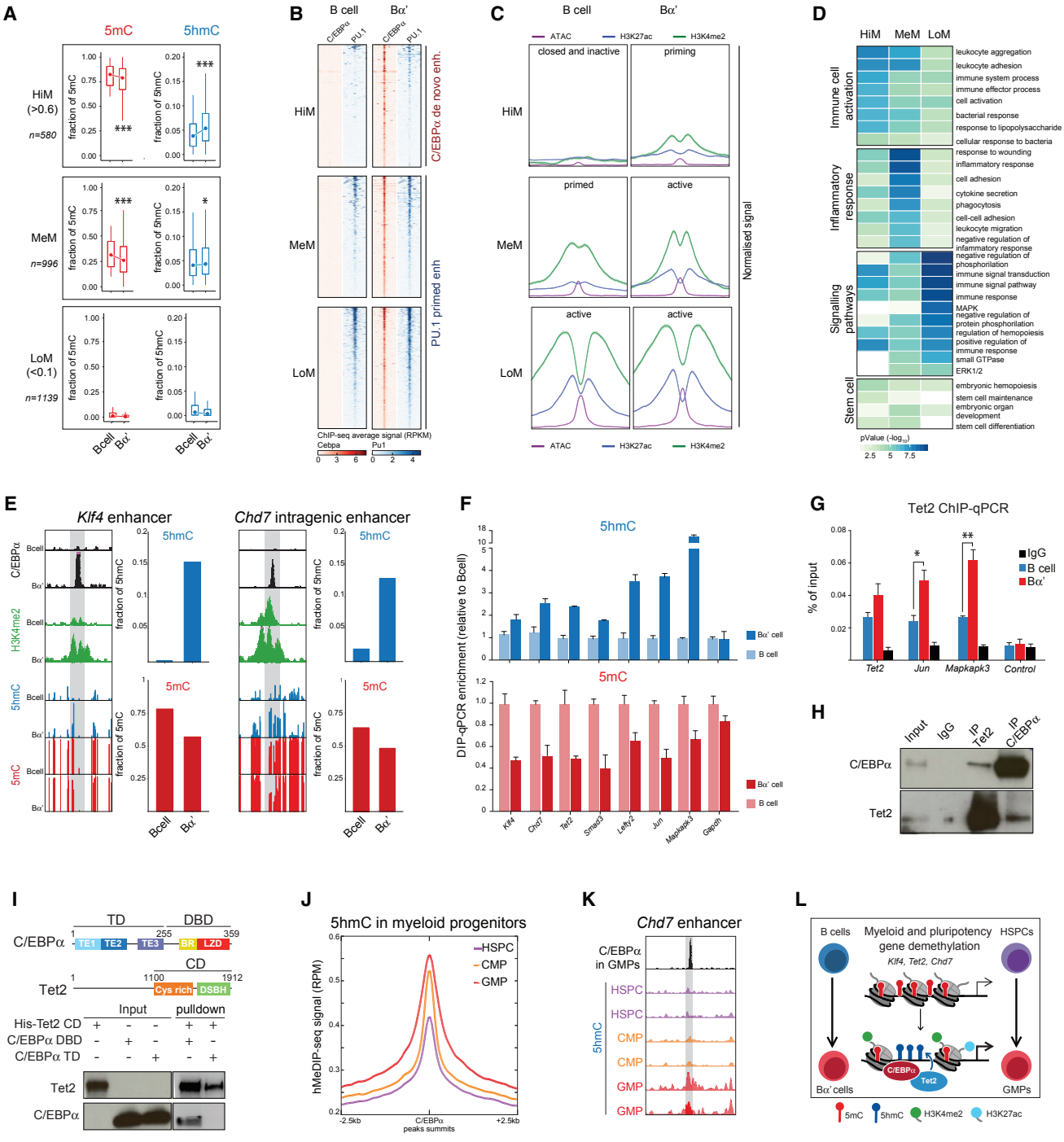


Figure 3. C/EBPα Recruits Tet2 to Activate Highly Methylated Myeloid and Pluripotency Enhancers

(A) DNA methylation dynamics at C/EBPα binding sites. Upper panels: high methylated sites (HiM, 5mC >0.6). Middle panels: medium methylated sites (MeM). Lower panels: low methylated sites (LoM, 5mC <0.1). Plots display the median (dot), inter-quartile range (box), and 95% confidence interval (vertical bar). Statistical significance was determined using a two-tailed paired Wilcoxon's test (***$p < 0.001$; *$p < 0.05$).

(B) Heatmaps of chromatin immunoprecipitation sequencing (ChIP-seq) signal for C/EBPα (red) and Pu.1 (blue) in B and Bα′ cells at the summits of C/EBPα peaks, classified as in (A).

(C) Average plots of ATAC-seq and H3K27ac and H3K4me2 ChIP-seq in B and Bα′ cells around the summits of C/EBPα peaks.

(D) Gene ontology (GO) enrichment analysis for the genes associated to C/EBPα peaks. The intensity of the color represents the p value, determined by a hypergeometric test.

(E) Genome browser snapshots showing signal for C/EBPα, H3K4me2 ChIP-seq, 5mC, and 5hmC in B and Bα′ cells, at *Klf4* enhancer (−90 kb) and *Chd7* intragenic enhancer (+150 kb). The bar plots on the right represent the average signal for 5mC and 5hmC at each region.

(legend continued on next page)

were scored after 12 days, cells treated with DMOG during the C/EBPα induction phase showed a 15-fold inhibition of iPSC colony formation, nearly as dramatic as cells treated throughout reprogramming (Figure S2H). DMOG did not have any detrimental effects on cell viability (Figure S2I).

Our data showed that ablation of Tet2 almost completely inhibits B cell to iPSC reprogramming, and that this effect can be rescued with both Tet1 and Tet3 but not with a catalytically dead Tet2. Moreover, inhibitor experiments suggest that the enzyme is required both during the C/EBPα and the OSKM induction phase. Together, our data indicate that Tet2 drives the observed hydroxymethylation waves during B cell to iPSC reprogramming.

C/EBPα Binds Highly Methylated Enhancers to License Their Demethylation

The observed requirement of Tet2 during the B to Bα′ transition suggests a role for C/EBPα in the active demethylation of GREs involved in iPSC reprogramming. To study how C/EBPα induces a re-shaping of the B cell methylome, we first grouped the C/EBPα target sites according to their methylation status in B cells: high methylation (HiM); medium methylation (MeM); and low methylation (LoM) (Figure S3A). In Bα′ cells, HiM and MeM sites lost 5mC and gained 5hmC, suggesting active demethylation (Figure 3A). We have previously shown that C/EBPα can act as a pioneer factor by binding *de novo* to GREs, as well as a secondary factor, by binding to PU.1-*primed* GREs (van Oevelen et al., 2015). Accordingly, the *de novo* enhancers were found to correlate with the HiM sites and the *primed* enhancers with MeM and LoM sites (Figure 3B). Based on the presence or absence of ATAC peaks and H3K4me2 and H3K27ac marks, HiM sites correspond to closed and inactive chromatin, MeM sites to partially accessible and poised chromatin, and LoM sites to fully accessible and active chromatin (Figure 3C).

C/EBPα Recruits Tet2 to Myeloid and Pluripotency GREs

Our previous work showed that C/EBPα links the myeloid with the pluripotency program by binding to the enhancers of *Tet2* and *Klf4* and activating the expression of these genes in Bα′ cells (Di Stefano et al., 2016). To study the role of DNA methylation in the activation of C/EBPα-bound HiM, MeM, and LoM enhancers, we first performed a gene ontology (GO) analysis (Figure 3D). Surprisingly, we found a small but significant enrichment for stem cell-related terms, mostly for genes associated with HiM and MeM sites (Figure 3D, bottom). The −90-Kb enhancer of *Klf4* and the intragenic enhancer of *Chd7*, two stem cell genes upregulated during myeloid differentiation, are examples of GREs within the HiM category (Figure 3E). These highly methylated enhancers gain 5hmC, lose 5mC, and become activated during the transition from B to Bα′ cells (Figure 3E). Demethylation of the *Klf4* and *Chd7* enhancers, as well as that of the additional C/EBPα-bound regions *Tet2*, *Smad3*, *Lefty2*, *Jun*, and *Mapkapk3*, was validated by MeDIP- and hMeDIP-qPCR (Figure 3F). C/EBPα-induced active demethylation requires DNA binding as the relevant regions remained methylated in cells expressing a C/EBPα mutant (Brm2) defective for DNA binding (Figure S3B).

The observed gain of 5hmC upon C/EBPα binding raised the possibility that this factor directly recruits Tet2. Indeed, we detected Tet2 binding at selected C/EBPα target sites in Bα′ cells (Figure 3G) and observed co-immunoprecipitation of the two proteins (Figure 3H). Pull-down assays of His-tagged peptides revealed a direct interaction between the DNA binding domain of C/EBPα and the catalytic domain of Tet2 (Figures 3I and S3C). To determine whether C/EBPα-induced demethylation is physiologically relevant, we examined the C/EBPα driven transition from hematopoietic stem-progenitor cells (HSPCs) and common myeloid progenitors (CMPs) to granulocyte macrophage progenitors (GMPs) (Zhang et al., 1997). Reprogramming partially mimics this transition, as Bα′ cells resemble GMPs both at the transcriptomic and chromatin level (Di Stefano et al., 2016). Accordingly, we found a substantial increase of 5hmC at regions bound by C/EBPα during the transition of HSPCs and CMPs to GMPs (Han et al., 2016) (Figure 3J), as exemplified for the enhancers of *Chd7*, *Jun*, and *Smad3* (Figures 3K and S3D).

Together, our data show that C/EBPα recruits Tet2 to myeloid and pluripotency associated enhancers, inducing their demethylation and activation (summarized in Figure 3L).

Identification of Regions Actively Demethylated before Chromatin Opening

To explore in an unbiased and genome-wide manner the interplay between TF binding and methylation, we first selected all dynamic ATAC regions (described in Figure 1E) displaying >20% 5mC changes and found 130,236 such regions (Figure S4A), with most of the demethylation occurring after day 2 (Figure S4B). Assuming that chromatin opening always precedes methylation changes, we expected to observe a negative correlation between the two parameters but found no such correlation

(F) 5hmC and 5mC DIP-qPCR at selected C/EBPα binding sites in B and Bα′ cells. *Gapdh* promoter is shown as a negative control region. Error bars indicate SD (n = 3 technical replicates).

(G) Tet2 ChIP-qPCR at selected C/EBPα binding sites in B and Bα′ cells. *Gapdh* promoter is shown as a negative control region. Error bars indicate SD (n = 3 biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's t test (**p < 0.01; *p < 0.05).

(H) C/EBPα-Tet2 co-immunoprecipitation. Interaction was assessed in Bα′ cells using specific antibodies against C/EBPα and Tet2.

(I) C/EBPα-Tet2 His tag pull-down assay. Top: sketch showing the domains of C/EBPα and Tet2 assayed in the pull-down. TD, transactivation domain; DBD, DNA binding and dimerization domain; CD, catalytic domain. Bottom: protein-protein interaction was assessed by western blot using specific antibodies against Tet2 and C/EBPα

(J) Average plots of 5hmC signal from hMeDIP-seq during myeloid differentiation, around C/EBPα peak summits in GMPs. HSPC, hematopoietic stem-progenitor cell; CMP, common myeloid progenitor; GMP, granulo-monocyte progenitor. Data were taken from C/EBPα ChIP-seq (GEO: GSE43007) (Hasemann et al., 2014) and hMeDIP-seq experiments (GEO: GSE77967) (Han et al., 2016).

(K) Representative genome browser snapshot at *Chd7* intragenic enhancer showing signal for C/EBPα ChIP-seq in GMPs and for 5hmC by hMeDIP-seq in HSPCs, CMPs, and GMPs.

(L) Model of C/EBPα-Tet2 interaction and induced hydroxymethylation at enhancers of myeloid and pluripotency genes.

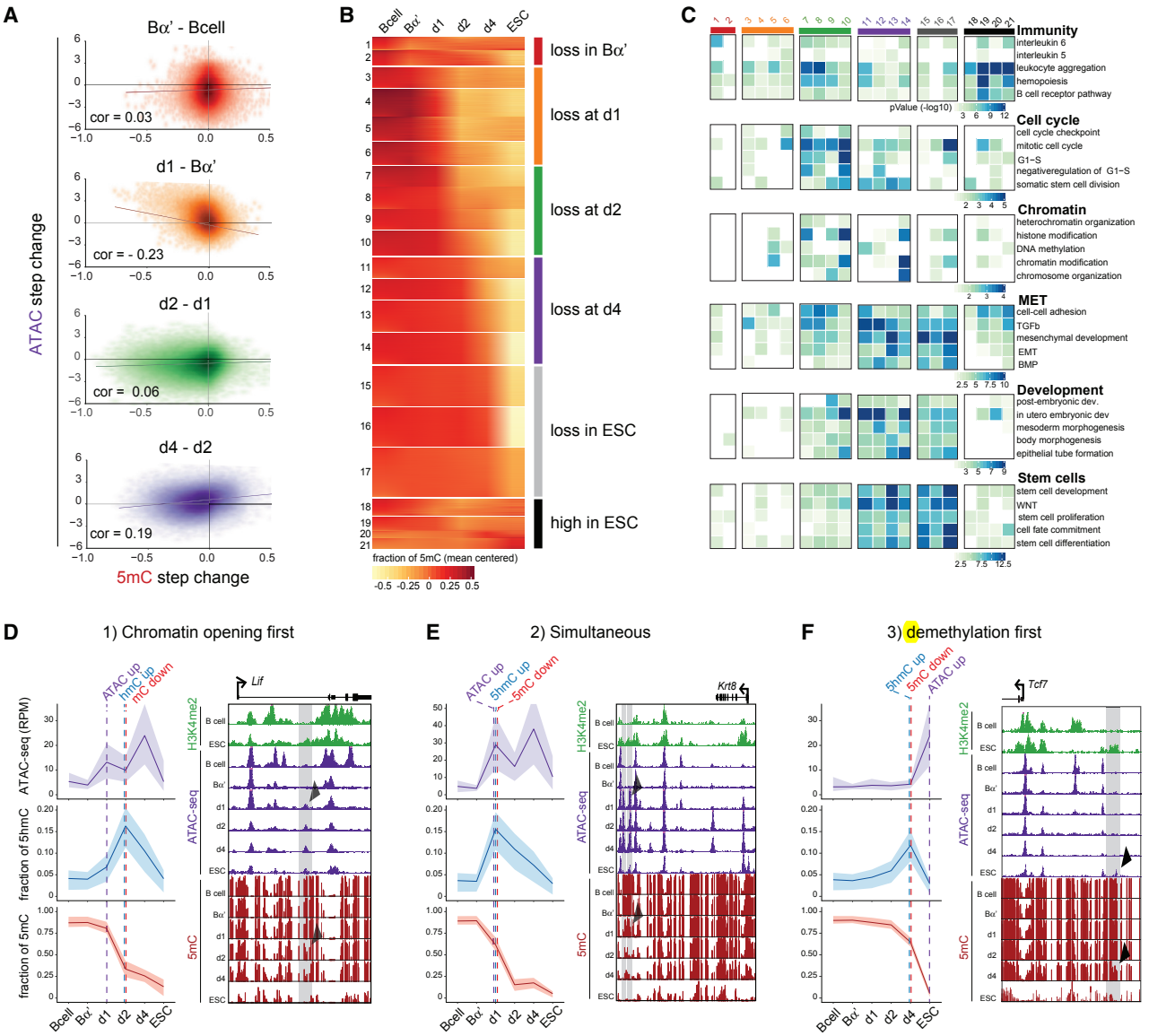


Figure 4. 5mC Dynamics at Regulatory Regions of the Genome (ATAC-seq)

(A) Correlations between the step changes of 5mC and ATAC at chromatin accessible regions (ATAC peaks), at each time point during reprogramming.

(B) Clustering of 5mC dynamics at ATAC peaks showing at least 20% changes of 5mC during reprogramming.

(C) Gene ontology (GO) enrichment analysis for the genes associated to clusters in (B). The intensity of the color represents the p value, determined by a hypergeometric test.

(D) Dynamics of ATAC, 5mC, and 5hmC at a selected cluster exhibiting chromatin opening before demethylation (cluster 9). Left: quantification of ATAC (purple), 5hmC (blue), and 5mC (red). Plots represent the mean (color line) and the inter-quartile range (shaded region). Dashed lines indicate the step of change for ATAC, 5mC, and 5hmC. Right: genome browser snapshot at a locus representative of these dynamics, showing signal for H3K4me2 ChIP-seq (B cells and ESCs), ATAC, 5mC, and 5hmC (whole reprogramming). Region following the specific dynamics is contained into the gray-shaded rectangle. Arrowheads point the step change for ATAC, 5mC, and 5hmC.

(E) Dynamics at a selected cluster exhibiting synchronous chromatin opening and methylation changes (cluster 6). Panels are similar as (D).

(F) Dynamics at a selected cluster exhibiting methylation changes preceding chromatin opening (cluster 14). Panels are similar as (D).

(Figure 4A). We next performed unsupervised clustering of the 130,236 regions, resulting in five major clusters that reflect the timing of the onset of DNA demethylation, and an additional group with high levels of 5mC in ESCs (Figure 4B). The enriched GO terms of these clusters broadly reflect their progression during reprogramming, starting with "immune-related processes," followed by "cell cycle," "chromatin," "mesenchymal-to-epithelial transition," and ending with "development and stem cells" (Figure 4C).

The 5 major clusters revealed three main scenarios between chromatin opening and active demethylation: (1) chromatin opening preceding demethylation (cluster 9), as exemplified by

the intragenic enhancer of *Lif* (Figure 4D); (2) simultaneous chromatin opening and demethylation (cluster 6), as exemplified by two GREs of *Krt8* (Figure 4E); and (3) demethylation precedes chromatin opening (cluster 14), which was quite unexpected. This is exemplified by the −15kb enhancer of *Tcf7*, which shows changes in 5hmC and 5mC at day 4, but an ATAC peak only in ESCs (Figure 4F).

In conclusion, comparing the kinetics of 5mC and ATAC-seq peaks, we identified genomic regions that are actively demethylated before displaying detectable chromatin accessibility. These will hereafter be designated DbAs (demethylation before accessibility).

Klf4 Induces Active Demethylation of Closed Chromatin Regions through Tet2 Recruitment

The discovery of DbAs raises the possibility that certain pioneer factors are able to target Tet2 to these regions. To test this hypothesis, we selected genomic regions that exhibited a >10% gain of 5hmC and a >10% loss of 5mC *before* becoming transposase accessible, resulting in 7,002 DbAs. These regions were grouped into 4 clusters (a–d) depending on the time point at which 5hmC and 5mC changes could first be detected (Figures 5A and 5B). To correlate the observed methylation changes with enhancer activity, we intersected the regions that show ATAC signals at day 4 and in ESCs, respectively, with the enhancer mark H3K4me2. Strikingly, a H3K4me2 gain at these regions was observed to occur concomitantly with active demethylation, *before* the emergence of chromatin accessibility (Figure 5C).

GO analysis of the 7,002 DbAs showed an enrichment for terms related to stem cells and development as well as for the Wnt, BMP, and transforming growth factor (TGF)-β signaling pathways (Figure S5A). TF motif analysis revealed a highly significant over-representation of Ap1, Klf, Oct/Pou, Esrr, and Snai motifs (Figure 5D). The Klf motif enrichment is of particular interest, as Klf4 has been shown to serve as a bridge between myeloid and pluripotent cells (Di Stefano et al., 2016). In addition, a recent report by the Plath laboratory described the binding of pluripotency factors during reprogramming of MEFs into iPSCs, in MEFs exposed for 48 hr to Klf4, Oct4, and Sox2 individually and in ESCs used as controls (Chronis et al., 2017). Analyzing this dataset in relationship with our DbAs in B cell reprogramming, we observed that Klf4, Oct4, and Sox2 bind to many regions in the absence of detectable ATAC-seq signals (Figure S5B). Moreover, Klf4 binds these DbAs even when individually expressed in MEFs. This is exemplified by Klf4 bound to a 100Kb GRE of the chromatin-related gene *Rybp* (Figure S5C). In contrast, no binding of chromatin inaccessible regions was observed for Oct4 and Sox2 in cells individually expressing these factors (Figure S5B).

The above data suggested that Klf4 can recruit Tet2 in the absence of nucleosome repositioning. To further explore this possibility, we intersected the Klf4 and Tet2 binding sites in ESCs and found that 51% of the Tet2 binding sites are also bound by Klf4, while 31% of the Klf4 sites overlap with Tet2 (Figure 5E). We next performed immunoprecipitation experiments and observed that precipitates obtained with an antibody against Tet2 contained readily detectable Klf4 protein. Conversely, precipitates prepared with an antibody to Klf4 were enriched for Tet2 (Figures 5F and S5D). In contrast, such an interaction was not observed between Tet2 and Esrrb (Figure S5E).

In conclusion, our data show that among the pluripotency factors Klf4, Oct4, and Sox2, described to have pioneer activity (Soufi et al., 2015), Klf4 is unique in its ability to bind transposase inaccessible chromatin that becomes demethylated and decorated with the enhancer mark H3K4me2. This property is mediated by recruitment of Tet2, endowing Klf4 with a novel type of pioneer activity.

A Subset of DbAs Identified during B Cell Reprogramming Can Also Be Detected during MEF to iPSC Reprogramming

To determine whether DbAs can also be observed with other reprogramming systems, we tested MEFs in a compound-enhanced reprogramming protocol (Bar-Nur et al., 2014) (Figure S5F). For this purpose, we established cultures of MEFs from reprogrammable mice grown in the presence of ascorbic acid and GSK3bi (hereafter referred as MEF^{AGi}) and performed ATAC-seq as well as hMeDIP-seq at 0, 2, and 4 days after OSKM induction. Focusing on ATAC positive regions, we observed a sequential gain of 5hmC after OSKM induction (Figure S5G), similar to B to iPSC reprogramming (Figure 1E). These major changes in 5hmC are consistent with the rapid upregulation of Tet2 observed in this system (Figure S5H). We next determined (Figure S5I) whether there is an overlap between the 3,827 dynamic 5hmC DbA regions detected during MEF^{AGi} reprogramming (Figure 5H) and the 7,002 DbAs identified during B cell reprogramming (Figure 5A). As shown in Figure S5J, 463 regions are shared between the two datasets (Figure S5J), as exemplified by the intragenic regulatory element of the chromatin-related factor Arid1b (Figure 5G). To determine whether the DbAs identified during MEF^{AGi} reprogramming contain regions bound by Klf4 before chromatin accessibility, we intersected the 5hmC data with the above-described conventional MEF to iPSC data from the Plath laboratory (Chronis et al., 2017). Briefly, during MEF^{AGi} reprogramming we identified 3,827 regions that exhibited a >2-fold gain of 5hmC *before* becoming transposase accessible (1,957 at day 2 and 1,870 at day 4; Figure 5H). These regions were enriched for the binding of Klf4, Oct4, and Sox2 in the absence of ATAC signals after OSKM induction, using Plath's MEF to iPSC dataset. Interestingly, again Klf4 but not Oct4 exhibited pioneer activity when expressed individually (Figure 5I). Examples of MEF^{AGi}-specific DbAs are shown for regulatory regions of *Smarcd2*, *Ebf3*, *Sall4*, or *Tet2* (Figure S5K).

Our results thus show that a subset of DbA regions identified in B cell to iPSC reprogramming can also be detected during accelerated reprogramming of MEF^{AGi} cells to iPSCs, suggesting that DbAs are a general feature of diverse somatic cell types reprogrammed into iPSCs. In addition, they support the proposal that Klf4 has a unique pioneer factor activity.

Demethylation of Core Pluripotency Factor Enhancers Occurs within 24 hr and Coincides with Their Activation

The demethylation of GREs controlling the expression of key pluripotency TFs, including Oct4 and Nanog, is a hallmark of successful iPSC reprogramming. We therefore examined demethylation in our accelerated reprogramming system, where key

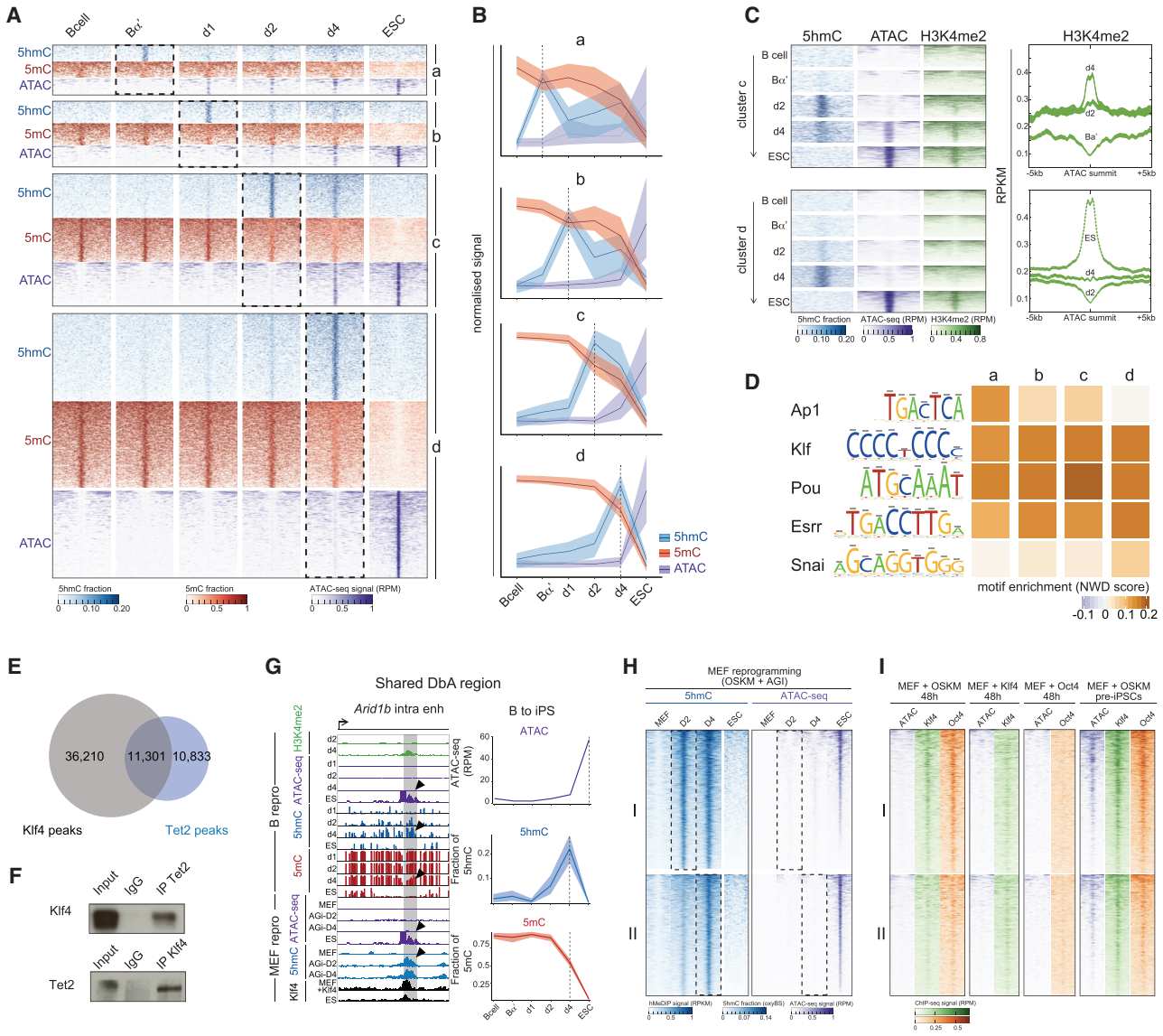


Figure 5. Klf4 Induces Active Demethylation of ATAC-Insensitive Regions

(A) Heatmaps of 5hmC, 5mC, and ATAC at regions exhibiting active demethylation before chromatin accessibility (DbAs). Regions are classified by the time point when they lose methylation (a) from B cells to Bα′; (b) from Bα′ to day 1 (d1); (c) from d1 to d2, and (d) from d2 to d4. Dashed rectangles indicate the step change for 5hmC and 5mC.

(B) Quantification of ATAC (purple), 5hmC (blue), and 5mC (red) signals at each groups of regions in (A). Plots represent the mean (color line) and inter-quartile range (shaded region). Dashed lines indicate the step change for 5mC and 5hmC at each cluster.

(C) H3K4me2 signal at clusters c (top) and d (bottom) from (A). Left: heatmaps of 5hmC, ATAC, and H3K4me2. Right: average plots of H3K4me2 ChIP-seq signal around ATAC peak summits at the time points before, during, and after the change of 5hmC.

(D) Heatmap displaying enrichment for transcription factors' motifs (maxNWD score, see STAR Methods) at selected clusters from (A).

(E) Venn diagram showing the overlap between Klf4 (GEO: GSE90895) (Chronis et al., 2017) and Tet2 (GEO: GSM2065691) (Xiong et al., 2016) ChIP-seq peaks in ESCs.

(F) Klf4-Tet2 co-immunoprecipitation. Interaction was assessed in ESCs using specific antibodies against Klf4 and Tet2.

(G) Shared DbA region at an intragenic regulatory element of Arid1b (shaded area). Snapshot showing signal for H3K4me2, ATAC-seq, 5hmC, and 5mC during reprogramming of B cells (tracks 1–14); signal for for ATAC-seq and 5hmC during accelerated reprogramming of MEF to iPSCs (tracks 15–21) and ChIP-seq signal for Klf4 in ESCs and in MEFs after ectopic expression of Klf4 alone during 48 hr (tracks 22–23). Right: quantification of ATAC, 5hmC, and 5mC at this region during reprogramming of B cells. Plots represent the mean (color line) and the inter-quartile range for 5mC/5hmC (shaded region).

(H) Heatmaps showing signal of 5hmC (blue) and ATAC (purple) at regions exhibiting active demethylation before chromatin accessibility (DbAs) during accelerated MEF reprogramming (OSKM+AGi). Regions are classified by the time point when they gain 5hmC: (I) from MEFs to D2 and (II) from D2 to D4. Dashed rectangles indicate the step change for 5hmC.

(I) Heatmaps showing signal for ATAC-seq (purple), Klf4 ChIP-seq (green), and Oct4 ChIP-seq (orange) during conventional MEF to iPSC reprogramming (GEO: GSE90895) (Chronis et al., 2017), at the clusters from (H). First panel: ATAC, Klf4, and Oct4 signals in MEFs OSKM 48 hr. Second panel: ATAC and Klf4 signals in MEFs overexpressing Klf4 during 48 hr. Third panel: ATAC and Oct4 signals in MEFs overexpressing Oct4 during 48 hr. Forth panel: ATAC, Klf4, and Oct4 signals in pre-iPSCs.

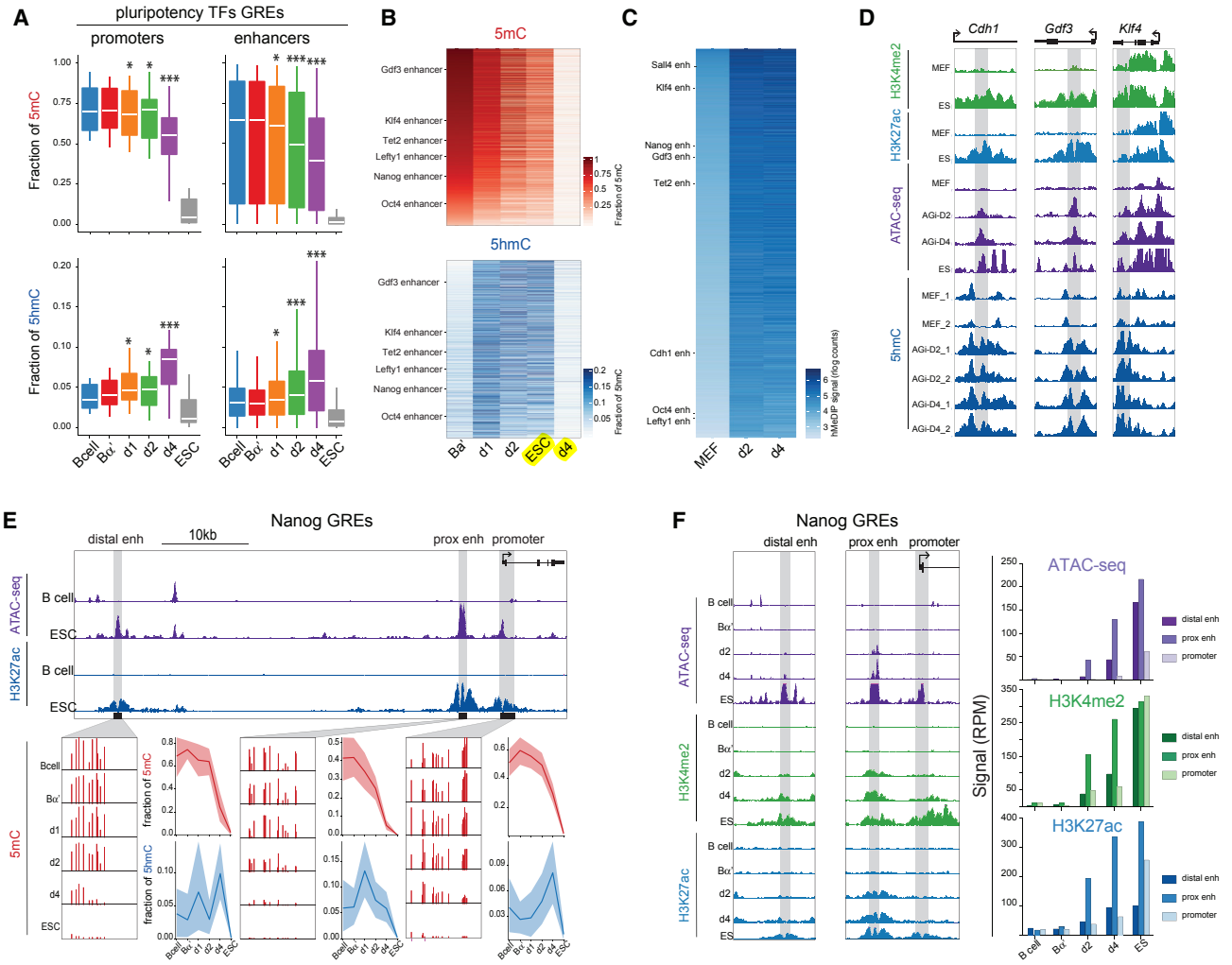


Figure 6. 5hmC-Mediated Early Demethylation of Regulatory Regions of Pluripotency Genes during Accelerated Reprogramming Systems

(A) CpG-methylation levels at regulatory regions of pluripotency genes during B cell reprogramming. Top: 5mC. Bottom: 5hmC. Plots display the median (bar), inter-quartile range (box), and 95% confidence interval (vertical bar). Statistical significance was determined using a two-tailed paired Wilcoxon's test (***$p < 0.001$; **$p < 0.01$; *$p < 0.05$).

(B) 5mC and 5hmC heatmaps at ESC-related regulatory regions during B cell reprogramming. Heatmaps display regions showing ≥10% 5mC loss by d1 compared to B cells. Selected GREs and their associated genes are indicated on the left.

(C) 5hmC heatmap at ESC-related regulatory regions during accelerated MEF reprogramming. Heatmap displays regions showing >2-fold 5hmC gain at day 2 (D2) compared to B cells. Selected GREs and their associated genes are indicated on the left.

(D) Examples of regulatory regions of *Cdh1*, *Gdf3*, and *Klf4* showing early gain of 5hmC during accelerated MEF to iPS reprogramming. Snapshots showing ChIP-seq signal for H3K4me2 and H3K27ac in MEFs and ESCs and ATAC-seq and 5hmC signal during reprogramming. Shaded regions indicate the ESC-related regulatory elements identified in (C).

(E) Snapshot showing early demethylation at the GREs of *Nanog* (gray areas) during B to iPS reprogramming. Top: ATAC-seq and H3K27ac ChIP-seq signal in B cells and ESCs. Bottom: zoom showing 5mC dynamics at *Nanog* distal enhancer (left) proximal enhancer (middle) and promoter (right). Plots on the right of 5mC tracks represent the mean (color line) and inter-quartile range (shaded region) for 5mC and 5hmC.

(F) Dynamics of chromatin accessibility and enhancer activation at *Nanog*'s GREs during B to iPS reprogramming. Left: snapshot showing ATAC-seq, H3K4me2 and H3K27ac ChIP-seq signals. Right: quantification of the ATAC-seq and ChIP-seq signals at Nanog's GREs.

pluripotency genes are activated 1–4 days after OSKM activation (Figure S6A and (Di Stefano et al., 2016)). For this purpose, we first identified regulatory regions of core pluripotency TFs and divided them into promoters and enhancers. Interestingly, enhancers displayed more pronounced changes than promoters (15% versus 25% at day 4) and exhibited an incipient 5mC decrease at day 1 (Figure 6A). Concomitantly, we observed an increase of 5hmC (Figure 6A), suggesting active demethylation. Next, to identify all pluripotency-related GREs that become demethylated early, we compiled H3K27ac-marked regions in ESCs that show ≥10% loss in 5mC by day 1, resulting in 2,700 sites. These included enhancers of the key pluripotency genes *Oct4*, *Nanog*, *Gdf3*, *Klf4*, *Tet2*, and *Lefty1/2* (Figure 6B). Similar results were obtained for MEF^{AGi} to iPSC

reprogramming, where 1,830 pluripotency-related GREs showed a significant gain of 5hmC by day 2 (Figure 6C), including GREs of the pluripotency genes *Klf4*, *Nanog*, *Gdf3*, *Oct4*, *Lefty1*, *Tet2*, and *Cdh1* (Figures 6D and S6B). At the GREs of *Oct4*, *Nanog*, *Tet2*, and *Lefty1* the most rapid changes during B cell reprogramming occurred at the proximal enhancers (Figures 6E and S6C–S6E). Of note, we observed a remarkable correlation between the hydroxymethylation kinetics of the *Nanog* GREs on the one hand with that of chromatin accessibility and H3K4me2 decoration on the other, in the order of proximal enhancer, distal enhancer, and promoter (Figures 6E and 6F).

Together, our data show that the GREs of key pluripotency TF genes exhibit rapid changes during iPSC reprogramming, with activation of enhancers preceding that of promoters.

Tfcp2l1 Binds to Early Demethylated GREs of Pluripotency Factors and Interacts with Tet2

In an attempt to identify the TF that binds to the most rapidly demethylated pluripotency factors' GREs, we looked for TFs predicted to target the proximal enhancer of Nanog, revealing a strong enrichment of the motif for the naive pluripotency factor Tfcp2l1 (Martello et al., 2013) (Figure 7A). To validate this prediction, we intersected the genomic distribution of Tfcp2l1 in ESCs (Chen et al., 2008) with the dynamic 5mC regions (Figure 4B). Surprisingly, we found that Tfcp2l1 was enriched not only at regions that get demethylated late (5mC loss in ESCs), but also at regions demethylated early (5mC loss at day 1) (Figure 7B). A reverse pattern was observed for the B cell factor Ebf1 at the same clusters, confirming the specificity of Tfcp2l1 enrichment (Figure S7A). Supporting this finding, 666 of the 2,700 pluripotency-related GREs that lose methylation by day 1 (Figure 6B) are bound by Tfcp2l1 in ESCs showing a highly significant enrichment over random regions (permutation test, p value = 9×10^{-5}). Similarly, Tfpc2l1 is enriched at GREs showing gain of 5hmC at day 2 during MEF^{AGi} reprogramming (Figure 7C), suggesting a possible role in both systems.

Consistent with the possibility that Tfcp2l1 recruits Tet2 to these regions, the GREs of *Nanog* bound by Tfcp2l1 (Figure 7D) largely overlapped with Tet2; similar findings were made for *Oct4* GREs (Figure S7B). This is also supported by a substantial genome-wide overlap between the binding sites in ESCs of the two factors (Figure 7E). Finally, we performed immunoprecipitation experiments and detected Tet2 in ESC extracts precipitated with a Tfcp2l1 antibody and Tfcp2l1 in extracts precipitated with a Tet2 antibody (Figures 7F and S7C), showing that the two proteins interact. Finally, chromatin immunoprecipitation (ChIP)-qPCR experiments confirmed that Tfcp2l1 is bound to the enhancers of *Oct4*, *Nanog*, and *Tet2* already in Bα' cells and even further bound in ESCs (Figure 7G). In line with this finding, we detected *Tfcp2l1* to be expressed at low levels in preB cells, not expressed in MEFs and highly expressed in ESCs (Figure 7H). However, in MEF^{AGi} cells, the gene is activated as early as 2 days upon OSKM induction (Figures S7D and S7E). This could explain, in part, the higher reprogramming efficiency of MEF^{AGi} cells compared to MEFs without the inhibitors, predicting that overexpression of Tfcp2l1 should enhance MEF reprogramming efficiency. Indeed, expression of Tfcp2l1 in MEFs significantly increased the cells' reprogramming efficiency (Figures 7I, 7J, and S7F).

Together, our data suggest that Tfcp2l1 drives the rapid demethylation and activation of GREs of pluripotency TF genes. In addition, the observed partial overlap between the binding sites of Tfcp2l1 and Tet2 on the one hand and the interaction between the two proteins on the other indicates that Tfcp2l1 can recruit Tet2 to these pluripotency-related GREs.

DISCUSSION

Here, we report the genome-wide dynamics of 5hmC and 5mC at single-nucleotide resolution during TF-induced cell reprogramming. Our high-resolution data allowed us to dissect the temporal relationships between DNA methylation, hydroxymethylation, and TF binding at key regulatory regions that govern cell identity. Selective temporal ablation of the Tet2 dioxygenase in B cells indicates that active demethylation is required both early and late during the cell reprogramming process. This can be explained by the finding that at least three different TFs, namely, C/EBPα, Klf4, and Tfcp2l1, recruit the enzyme to DNA where it facilitates regulation of enhancer activity throughout reprogramming (summarized in Figures 3L and 7K). In addition, our dynamic analyses permitted us to uncover enhancers exhibiting active demethylation before detectable chromatin accessibility, mediated in part by a unique pioneering activity of Klf4.

Surprisingly, among the regions robustly marked by 5hmC a substantial proportion maintained the 5hmC mark for 2 or more consecutive days, suggesting that here Tet2 activity is stalled or delayed. It is possible that at these regions 5hmC acts as an epigenetic mark, such as by recruiting Mbd3, a member of the repressive NURD chromatin remodeling complex, known to be involved in reprogramming (dos Santos et al., 2014; Rais et al., 2013). In line with this idea, chromatin regions of low Tet2 processivity have been shown to correspond to less accessible, repressed chromatin (Wu et al., 2014). The regions semi-stably marked by 5hmC in turn might represent transiently compacted chromatin areas.

Our data using two rapid and highly efficient iPSC reprogramming systems showed that GREs of key pluripotency factors such as *Oct4* and *Nanog* become actively demethylated as early as 1–2 days after induction of the Yamanaka factors, compared to around 21 days, the first time point previously described (Lee et al., 2014; Milagre et al., 2017; Polo et al., 2012). We also found an excellent correlation between DNA demethylation, chromatin opening, and enhancer activation at the *Nanog* GREs. However, *Nanog* only becomes expressed 2–4 days after OSKM induction, while *Oct4* becomes upregulated immediately. These differences could be explained by the finding that the activation of *Nanog*, but not *Oct4*, is preceded by topological chromatin changes of the locus (Stadhouders et al., 2018), suggesting that these are rate limiting. Our observations therefore show that active DNA demethylation is required—but not always sufficient—for gene activation, and that this depends on the chromatin context.

Based on the premise that Tet2 needs to be recruited by TFs to specific sites, we expected chromatin accessibility to emerge at enhancers either before or concomitantly with changes in 5hmC and 5mC. Although this was often the

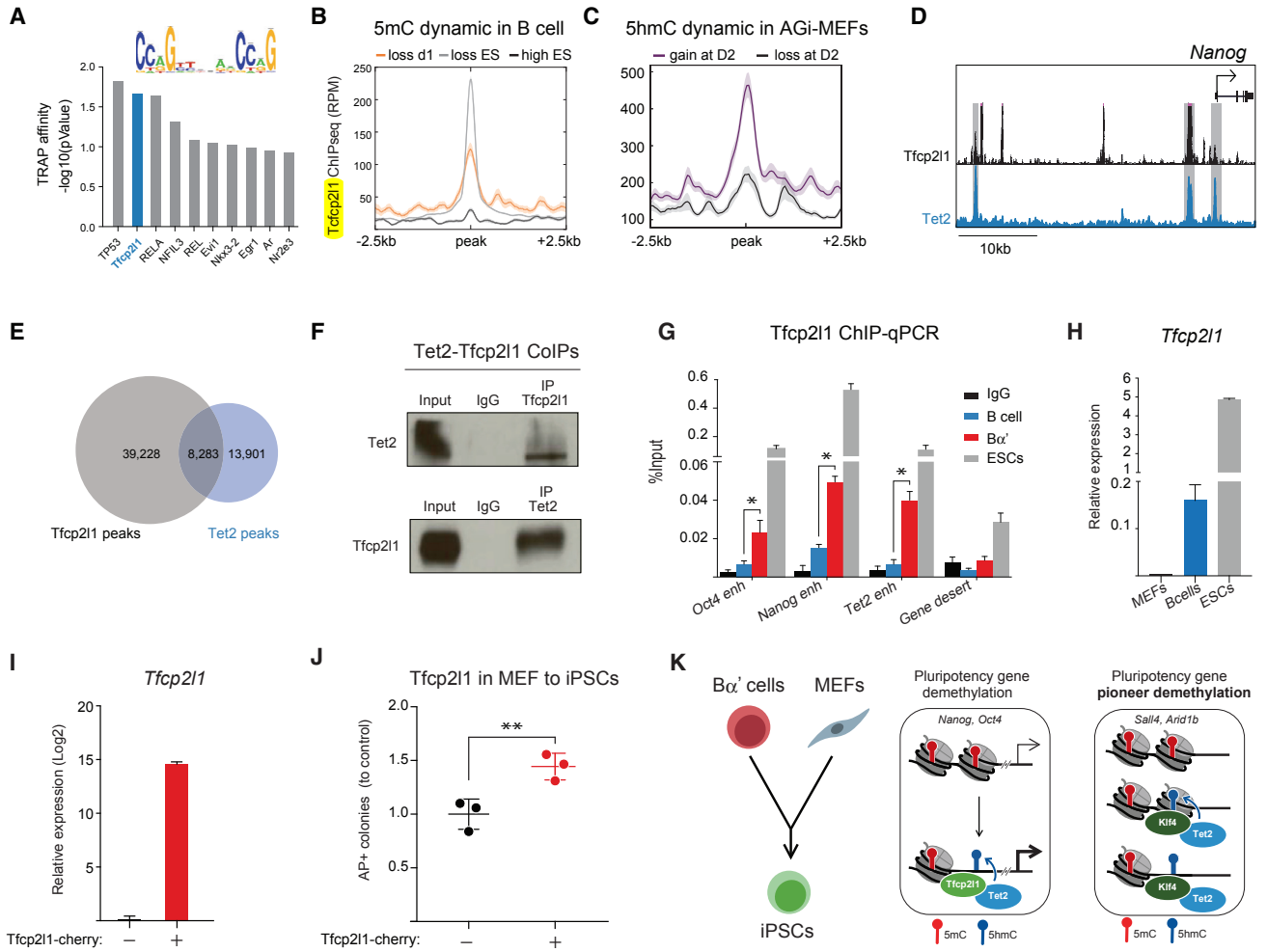


Figure 7. Tfcp2l1 Interacts with Tet2 to Induce Active Early Demethylation at Oct4 and Nanog Proximal Enhancers

(A) Putative TFs binding to *Nanog* proximal enhancer based on predicted affinity (TRAP using JASPAR database) (Thomas-Chollier et al., 2011). Tfcp2l1 motif (JASPAR MA0145.2) is shown as an inset in the graph and highlighted in blue in the plot.

(B) Average plot of Tfcp2l1 ChIP-seq signal in ESCs (GEO: GSM288350) (Chen et al., 2008), at regions losing methylation during B to iPS reprogramming at d1 (orange), in ESC (gray) or showing high methylation in ESC (black) based on the clustering in Figure 4B. Average signal is centered at ATAC peak summits.

(C) Average plot of Tfcp2l1 ChIP-seq signal in ESCs (GEO: GSM288350) (Chen et al., 2008), at regions gaining 5hmC at D2 (purple) or losing 5hmC at D2 (gray) during accelerated MEF to iPSC reprogramming based on the clustering in Figure S5D. Average signal is centered at ATAC peak summits.

(D) Representative snapshot showing Tfcp2l1-Tet2 co-binding at Nanog's GREs in ESCs. Regions analyzed are as in Figure 6E.

(E) Venn diagram showing the overlap between Tfcp2l1 (GEO: GSM288350) (Chen et al., 2008) and Tet2 (GEO: GSM2065691) (Xiong et al., 2016) ChIP-seq peaks in ESCs.

(F) Tfcp2l1-Tet2 co-immunoprecipitation. Interaction was assessed in ESCs using specific antibodies against Tet2 and Tfcp2l1.

(G) Tfcp2l1 ChIP-qPCR at selected enhancers of pluripotency genes in B, Bα′, and ESCs. Intergenic region (gene desert) is shown as a negative control region. Error bars indicate SD (n = 3 biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's t test (*$p < 0.05$).

(H) *Tfcp2l1* levels (qRT-PCR) in MEF, B, and ESCs. Values are expressed as the percentage of *Pgk* expression. Error bars indicate SD (n = 3 biologically independent samples).

(I) *Tfcp2l1* levels (qRT-PCR) in MEFs infected with a *Tfcp2l1*-Cherry construct. Values were normalized against *Pgk* expression. Error bars indicate SD (n = 3 biologically independent samples).

(J) Alkaline phosphatase positive (AP^+) iPSC colony counts after reprogramming MEFs ectopically expressing T*fcp2l1*. Error bars indicate SD (n = 3 biological independent experiments). Statistical significance was determined using a two-tailed unpaired Student's t test (**$p < 0.01$).

(K) Model of Tet2 interaction with Tfcp2l1 (left) and Klf4 (right) and induced hydroxymethylation at pluripotency enhancers

case, we detected many regions in which methylation changes and decoration of histones with active enhancer marks preceded chromatin opening. Further analyses revealed that Klf4 is capable of binding to such "demethylation-before-chromatin-accessibility" regions, or DbAs, in line with its known ability to bind to nucleosome dense regions (Soufi et al., 2015). That it can also do so without inducing nucleosome displacement suggests that it exhibits a novel type of pioneer factor activity. The detection of DbAs in the context of both B cell and fibroblast reprogramming raises the possibility that they represent a novel feature of cell-fate decisions during development and cancer.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- CONTACT FOR REAGENT AND RESOURCE SHARING
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
  - Cells and cell cultures
- METHOD DETAILS
  - B cell to iPSC reprogramming experiments
  - MEFs to iPSCs reprogramming experiments
  - Vectors and virus production and infection
  - Inhibition of Tet activity by DMOG
  - FACS analyses
  - Colony counting
  - Cell viability
  - RNA extraction
  - qRT-PCR analyses
  - ATAC seq
  - ChIP-seq
  - Chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR)
  - MeDIP/hMeDIP-qPCR
  - Western blots (WBs) and immunoprecipitation
  - Production of recombinant forms of C/EBPα and Tet2
  - C/EBPα-Tet2 pulldown assay
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Bisulfite and ox-bisulfite sequencing data processing
  - ATAC-seq analyses
  - ChIP-seq and hMeDIP-seq analyses
  - Transcription factor binding motif analyses
- DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and one table and can be found with this article online at https://doi.org/10.1016/j.stem.2018.08.016.

ACKNOWLEDGMENTS

We thank Dr. Xu Guoliang for Tet plasmids, the Sequencing Unit of CNAG-CRG, the Genomics Unit of the CRG, the Biomolecular Screening & Protein Technologies Facility of the CRG, and the Flow Citometry Unit of UPF-CRG for technical assistance; Maria Vila for bioinformatic support; and Graf lab members for critical reading of the manuscript. This work was supported by Spanish Ministry of Economy, Industry and Competitiveness (MEIC) (Plan Estatal 2015. SAF2015-68740-P) and the European Research Council Synergy grant (4D-Genome). J.L.S. was supported by MEIC (IJCI-2014-21872), and T.V.T. was supported by MEIC (FJCI-2014-22946). We also acknowledge support of the Spanish Ministry of Economy, Industry and Competitiveness to the EMBL partnership, the Centro de Excelencia Severo Ochoa, and the CERCA Programme-Generalitat de Catalunya.

AUTHOR CONTRIBUTIONS

J.L.S., S.C., and T.G. conceived the study and wrote the manuscript with input from all co-authors; J.L.S. designed and performed the cell-culture and animal and molecular biology experiments and analyzed the data; S.C. designed and performed bioinformatics analyses; T.V.T. performed the protein IPs and molecular biology experiments; A.G., S.H., and D.T. performed bioinformatics analyses; B.D.S. and C.B. performed cell-culture and animal work; J.B. designed and performed MEF^{AGi} reprogramming experiments; R.S. performed molecular biology experiments; C.S.-M. generated the Tet2KO reprogrammable mouse strain; M.G. and I.G.G. performed WGBS library preparation and sequencing; S.A. and L.D.C. designed and performed 5hmC immunoprecipitation on MEF^{AGi} reprogramming; and K.H. edited the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: December 19, 2017
Revised: July 7, 2018
Accepted: August 23, 2018
Published: September 13, 2018

REFERENCES

Amouroux, R., Nashun, B., Shirane, K., Nakagawa, S., Hill, P.W., D'Souza, Z., Nakayama, M., Matsuda, M., Turp, A., Ndjetehe, E., et al. (2016). De novo DNA methylation drives 5hmC accumulation in mouse zygotes. Nat. Cell Biol. *18*, 225–233.

Apostolou, E., and Hochedlinger, K. (2013). Chromatin dynamics during cellular reprogramming. Nature *502*, 462–471.

Bar-Nur, O., Brumbaugh, J., Verheul, C., Apostolou, E., Pruteanu-Malinici, I., Walsh, R.M., Ramaswamy, S., and Hochedlinger, K. (2014). Small molecules facilitate rapid and synchronous iPSC generation. Nat. Methods *11*, 1170–1176.

Bird, A. (2002). DNA methylation patterns and epigenetic memory. Genes Dev. *16*, 6–21.

Boiani, M., Eckardt, S., Schöler, H.R., and McLaughlin, K.J. (2002). Oct4 distribution and level in mouse clones: consequences for pluripotency. Genes Dev. *16*, 1209–1219.

Booth, M.J., Branco, M.R., Ficz, G., Oxley, D., Krueger, F., Reik, W., and Balasubramanian, S. (2012). Quantitative sequencing of 5-methylcytosine and 5-hydroxymethylcytosine at single-base resolution. Science *336*, 934–937.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Bussmann, L.H., Schubert, A., Vu Manh, T.P., De Andres, L., Desbordes, S.C., Parra, M., Zimmermann, T., Rapino, F., Rodriguez-Ubreva, J., Ballestar, E., and Graf, T. (2009). A robust and highly efficient immune cell reprogramming system. Cell Stem Cell *5*, 554–566.

Carey, B.W., Markoulaki, S., Beard, C., Hanna, J., and Jaenisch, R. (2010). Single-gene transgenic mouse strains for reprogramming adult somatic cells. Nat. Methods *7*, 56–59.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106–1117.

Chen, J., Liu, H., Liu, J., Qi, J., Wei, B., Yang, J., Liang, H., Chen, Y., Chen, J., Wu, Y., et al. (2013). H3K9 methylation is a barrier during somatic cell reprogramming into iPSCs. Nat. Genet. *45*, 34–42.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative binding of transcription factors orchestrates reprogramming. Cell *168*, 442–459.

Collins, L.S., and Dorshkind, K. (1987). A stromal cell line from myeloid long-term bone marrow cultures can support myelopoiesis and B lymphopoiesis. J. Immunol. *138*, 1082–1087.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Costa, Y., Ding, J., Theunissen, T.W., Faiola, F., Hore, T.A., Shliaha, P.V., Fidalgo, M., Saunders, A., Lawrence, M., Dietmann, S., et al. (2013).

NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. Nature *495*, 370–374.

Croft, D., O'Kelly, G., Wu, G., Haw, R., Gillespie, M., Matthews, L., Caudy, M., Garapati, P., Gopinath, G., Jassal, B., et al. (2011). Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. *39*, D691–D697.

de la Rica, L., Rodríguez-Ubreva, J., García, M., Islam, A.B., Urquiza, J.M., Hernando, H., Christensen, J., Helin, K., Gómez-Vaquero, C., and Ballestar, E. (2013). PU.1 target genes undergo Tet2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. Genome Biol. *14*, R99.

Di Stefano, B., Sardina, J.L., van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., and Graf, T. (2014). C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235–239.

Di Stefano, B., Collombet, S., Jakobsen, J.S., Wierer, M., Sardina, J.L., Lackner, A., Stadhouders, R., Segura-Morales, C., Francesconi, M., Limone, F., et al. (2016). C/EBPα creates elite cells for iPSC reprogramming by upregulating Klf4 and increasing the levels of Lsd1 and Brd4. Nat. Cell Biol. *18*, 371–381.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21.

Doege, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A., et al. (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature *488*, 652–655.

Domcke, S., Bardet, A.F., Adrian Ginno, P., Hartl, D., Burger, L., and Schübeler, D. (2015). Competition between DNA methylation and transcription factors determines binding of NRF1. Nature *528*, 575–579.

dos Santos, R.L., Tosti, L., Radzisheuskaya, A., Caballero, I.M., Kaji, K., Hendrich, B., and Silva, J.C. (2014). MBD3/NuRD facilitates induction of pluripotency in a context-dependent manner. Cell Stem Cell *15*, 102–110.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison, C.A., 3rd, and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Habibi, E., Brinkman, A.B., Arand, J., Kroeze, L.I., Kerstens, H.H., Matarese, F., Lepikhov, K., Gut, M., Brun-Heath, I., Hubner, N.C., et al. (2013). Whole-genome bisulfite sequencing of two distinct interconvertible DNA methylomes of mouse embryonic stem cells. Cell Stem Cell *13*, 360–369.

Hackett, J.A., and Surani, M.A. (2014). Regulatory principles of pluripotency: from the ground state up. Cell Stem Cell *15*, 416–430.

Han, D., Lu, X., Shih, A.H., Nie, J., You, Q., Xu, M.M., Melnick, A.M., Levine, R.L., and He, C. (2016). A Highly Sensitive and Robust Method for Genome-wide 5hmC Profiling of Rare Cell Populations. Mol. Cell *63*, 711–719.

Hasemann, M.S., Lauridsen, F.K., Waage, J., Jakobsen, J.S., Frank, A.K., Schuster, M.B., Rapin, N., Bagger, F.O., Hoppe, P.S., Schroeder, T., and Porse, B.T. (2014). C/EBPα is required for long-term self-renewal and lineage priming of hematopoietic stem cells and for the maintenance of epigenetic configurations in multipotent progenitors. PLoS Genet. *10*, e1004079.

Hu, X., Zhang, L., Mao, S.Q., Li, Z., Chen, J., Zhang, R.R., Wu, H.P., Gao, J., Guo, F., Liu, W., et al. (2014). Tet and TDG mediate DNA demethylation essential for mesenchymal-to-epithelial transition in somatic cell reprogramming. Cell Stem Cell *14*, 512–522.

Huber, W., Carey, V.J., Gentleman, R., Anders, S., Carlson, M., Carvalho, B.S., Bravo, H.C., Davis, S., Gatto, L., Girke, T., et al. (2015). Orchestrating high-throughput genomic analysis with Bioconductor. Nat. Methods *12*, 115–121.

Kallin, E.M., Rodríguez-Ubreva, J., Christensen, J., Cimmino, L., Aifantis, I., Helin, K., Ballestar, E., and Graf, T. (2012). Tet2 facilitates the derepression of myeloid target genes during CEBPα-induced transdifferentiation of pre-B cells. Mol. Cell *48*, 266–276.

Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839–843.

Kumar, L., and E Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. Bioinformation *2*, 5–7.

Lee, D.S., Shin, J.Y., Tonge, P.D., Puri, M.C., Lee, S., Park, H., Lee, W.C., Hussein, S.M., Bleazard, T., Yun, J.Y., et al. (2014). An epigenomic roadmap to induced pluripotency reveals DNA methylation as a reprogramming modulator. Nat. Commun. *5*, 5619.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550.

Marco-Sola, S., Sammeth, M., Guigó, R., and Ribeca, P. (2012). The GEM mapper: fast, accurate and versatile alignment by filtration. Nat. Methods *9*, 1185–1188.

Martello, G., Bertone, P., and Smith, A. (2013). Identification of the missing pluripotency mediator downstream of leukaemia inhibitory factor. EMBO J. *32*, 2561–2574.

Medina-Rivera, A., Abreu-Goodger, C., Thomas-Chollier, M., Salgado, H., Collado-Vides, J., and van Helden, J. (2011). Theoretical and empirical quality assessment of transcription factor-binding motifs. Nucleic Acids Res. *39*, 808–824.

Medina-Rivera, A., Defrance, M., Sand, O., Herrmann, C., Castro-Mondragon, J.A., Delerce, J., Jaeger, S., Blanchet, C., Vincens, P., Caron, C., et al. (2015). RSAT 2015: Regulatory Sequence Analysis Tools. Nucleic Acids Res. *43*, W50–W56.

Merkel, A., Fernandez-Callejo, M., Casals, E., Marco-Sola, S., Schuyler, R., Gut, I.G., and Heath, S.C. (2017). GEMBS: high through-put processing pipeline for DNA methylation data from Whole Genome Bisulfite Sequencing (WGBS). Bioinformatics. https://doi.org/10.1093/bioinformatics/bty690.

Milagre, I., Stubbs, T.M., King, M.R., Spindel, J., Santos, F., Krueger, F., Bachman, M., Segonds-Pichon, A., Balasubramanian, S., Andrews, S.R., et al. (2017). Gender differences in global but not targeted demethylation in iPSC reprogramming. Cell Rep. *18*, 1079–1089.

Neph, S., Kuehn, M.S., Reynolds, A.P., Haugen, E., Thurman, R.E., Johnson, A.K., Rynes, E., Maurano, M.T., Vierstra, J., Thomas, S., et al. (2012). BEDOPS: High-performance genomic feature operations. Bioinformatics *28*, 1919–1920.

Pastor, W.A., Aravind, L., and Rao, A. (2013). TETonic shift: Biological roles of TET proteins in DNA demethylation and transcription. Nat. Rev. Mol. Cell Biol. *14*, 341–356.

Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, C.M., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., Cloutier, J., et al. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell *151*, 1617–1632.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65–70.

Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. *44* (W1), W160-5.

Sardina, J.L., López-Ruano, G., Sánchez-Abarca, L.I., Pérez-Simón, J.A., Gaztelumendi, A., Trigueros, C., Llanillo, M., Sánchez-Yagüe, J., and Hernández-Hernández, A. (2010). p22phox-dependent NADPH oxidase activity is required for megakaryocytic differentiation. Cell Death Differ. *17*, 1842–1854.

Schmidl, C., Rendeiro, A.F., Sheffield, N.C., and Bock, C. (2015). ChIPmentation: Fast, robust, low-input ChIP-seq for histones and transcription factors. Nat. Methods *12*, 963–965.

Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. Nat. Rev. Genet. *14*, 204–220.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555–568.

Stadhouders, R., Vidal, E., Serra, F., Di Stefano, B., Le Dily, F., Quilez, J., Gomez, A., Collombet, S., Berenguer, C., Cuartero, Y., et al. (2018). Transcription factors orchestrate dynamic interplay between genome topology and gene regulation during cell reprogramming. Nat. Genet. *50*, 238–249.

Stadtfeld, M., Maherali, N., Borkent, M., and Hochedlinger, K. (2010). A reprogrammable mouse strain from gene-targeted embryonic stem cells. Nat. Methods *7*, 53–55.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Thomas-Chollier, M., Hufton, A., Heinig, M., O'Keeffe, S., Masri, N.E., Roider, H.G., Manke, T., and Vingron, M. (2011). Transcription factor binding predictions using TRAP for the analysis of ChIP-seq data and regulatory SNPs. Nat. Protoc. *6*, 1860–1869.

van Oevelen, C., Collombet, S., Vicent, G., Hoogenkamp, M., Lepoivre, C., Badeaux, A., Bussmann, L., Sardina, J.L., Thieffry, D., Beato, M., et al. (2015). C/EBPα activates pre-existing and de novo macrophage enhancers during induced pre-B cell transdifferentiation and myelopoiesis. Stem Cell Reports *5*, 232–247.

Vidal, S.E., Amlani, B., Chen, T., Tsirigos, A., and Stadtfeld, M. (2014). Combinatorial modulation of signaling pathways reveals cell-type-specific requirements for highly efficient and synchronous iPSC reprogramming. Stem Cell Reports *3*, 574–584.

Wang, Y., Xiao, M., Chen, X., Chen, L., Xu, Y., Lv, L., Wang, P., Yang, H., Ma, S., Lin, H., et al. (2015). WT1 recruits TET2 to regulate its target gene expression and suppress leukemia cell proliferation. Mol. Cell *57*, 662–673.

Wu, X., and Zhang, Y. (2017). TET-mediated active DNA demethylation: mechanism, function and beyond. Nat. Rev. Genet. *18*, 517–534.

Wu, H., Wu, X., Shen, L., and Zhang, Y. (2014). Single-base resolution analysis of active DNA demethylation using methylase-assisted bisulfite sequencing. Nat. Biotechnol. *32*, 1231–1240.

Xiong, J., Zhang, Z., Chen, J., Huang, H., Xu, Y., Ding, X., Zheng, Y., Nishinakamura, R., Xu, G.L., Wang, H., et al. (2016). Cooperative action between SALL4A and TET proteins in stepwise oxidation of 5-methylcytosine. Mol. Cell *64*, 913–925.

Yildirim, O., Li, R., Hung, J.H., Chen, P.B., Dong, X., Ee, L.S., Weng, Z., Rando, O.J., and Fazzio, T.G. (2011). Mbd3/NURD complex regulates expression of 5-hydroxymethylcytosine marked genes in embryonic stem cells. Cell *147*, 1498–1510.

Yin, Y., Morgunova, E., Jolma, A., Kaasinen, E., Sahu, B., Khund-Sayeed, S., Das, P.K., Kivioja, T., Dave, K., Zhong, F., et al. (2017). Impact of cytosine methylation on DNA binding specificities of human transcription factors. Science. Published online May 5, 2017. https://doi.org/10.1126/science.aaj2239.

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

Yu, M., Hon, G.C., Szulwach, K.E., Song, C.X., Zhang, L., Kim, A., Li, X., Dai, Q., Shen, Y., Park, B., et al. (2012). Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. Cell *149*, 1368–1380.

Zhang, D.E., Zhang, P., Wang, N.D., Hetherington, C.J., Darlington, G.J., and Tenen, D.G. (1997). Absence of granulocyte colony-stimulating factor signaling and neutrophil development in CCAAT enhancer binding protein alpha-deficient mice. Proc. Natl. Acad. Sci. USA *94*, 569–574.

Zhang, Q., Zhao, K., Shen, Q., Han, Y., Gu, Y., Li, X., Zhao, D., Liu, Y., Wang, C., Zhang, X., et al. (2015). Tet2 is required to resolve inflammation by recruiting Hdac2 to specifically repress IL-6. Nature *525*, 389–393.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Biotin Anti-mouse CD19 (1D3)	BD Biosciences	Cat# 553784; RRID: AB_395048
APC Anti-human CD4 (RPA-T4)	BD Biosciences	Cat# 555349; RRID: AB_398593
Biotin anti-human CD4 (RPA-T4)	eBioscience	Cat# 13-0049; RRID:AB_466337
Purified Rat Anti-Mouse CD16/CD32 (Mouse BD Fc Block)	BD Biosciences	Cat# 553142; RRID: AB_394654
Anti-H3K4me2 antibody	Abcam	Cat# ab7766; RRID: AB_2560996
Anti-5hmC	Active Motif	Cat# 39769; RRID: AB_10013602
Anti-5mC	Diagenode	Cat# C15200081; RRID: AB_2572207
Anti-Tet2 antibody	Abcam	Cat# ab124297; RRID: AB_2722695
C/EBPα (14AA) antibody	Santa Cruz Biotechnology	Cat# sc-61; RRID: AB_631233
Klf4 antibody	R&D systems	Cat# AF3158; RRID: AB_2130245
β-Tubulin (SAP.4G5) antibody	Sigma-Aldrich	Cat# T7816; RRID: AB_261770
Tfcp2l1 antibody	R&D systems	Cat# AF5726; RRID: AB_2202564
Esrrb antibody	R&D systems	Cat# PP-H6705-00; RRID: AB_2100412
Mouse IgG - Isotype Control antibody	Abcam	Cat# ab37355; RRID: AB_2665484
Rabbit IgG, polyclonal - Isotype Control	Abcam	Cat# ab171870; RRID: AB_2687657
Goat IgG, polyclonal - Isotype Control	Abcam	Cat# ab37373
Anti-Flag M2 antibody	Sigma-Aldrich	Cat# F1804; RRID: AB_262044
GAPDH (6C5) antibody	Santa Cruz Biotechnology	Cat# sc-32233; RRID: AB_627679
Bacterial and Virus Strains		
E. coli: Subcloning Efficiency DH5α Competent Cells	Invitrogen	Cat# 18265017
E. coli: BL21(DE3) Competent	New England Biolabs	Cat# C2527I
Chemicals, Peptides, and Recombinant Proteins		
Recombinant murine IL-7	Peprotech	Cat# 217-17
Recombinant murine IL-4	Peprotech	Cat# 214-14
Recombinant murine IL-15	Peprotech	Cat# 210-15
ESGRO Recombinant mouse LIF protein	Merk Millipore	Cat# ESG1106
Human Interferon-Alpha1	Pbl Assay Science	Cat# 11175-1
DMEM Medium	GIBCO	Cat# 12491015
RPMI 1640 Medium	GIBCO	Cat# 12633012
Knockout-DMEM	GIBCO	Cat# 10829018
Neurobasal Medium	GIBCO	Cat# 21103049
DMEM-F12 Medium	GIBCO	Cat# 12634010
Fetal Bovine Serum, E.U.-approved, South America origin	GIBCO	Cat# 10270-106
Embryonic stem-cell FBS, qualified, US origin	GIBCO	Cat# 10270-106
KnockOut Serum Replacement	GIBCO	Cat# A3181502
Pen Strep	GIBCO	Cat# 15140122
L-Glutamine (200mM)	GIBCO	Cat# 25030081
Sodium Pyruvate (100mM)	GIBCO	Cat# 11360070
MEM Non-Essential Amino Acids Solution (100X)	GIBCO	Cat# 11140068
2-Mercaptoethanol	Invitrogen	Cat# 31350010
N-2 Supplement (100X)	GIBCO	Cat# 17502048
B-27 Serum-Free Supplement (50X)	GIBCO	Cat# 17504044
MEK inhibitor (PD0325901)	Selleckchem	Cat# S1036
GSK3b inhibitor (CHIR-99021)	Selleckchem	Cat# S1263

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
L-Ascorbic Acid	Sigma-Aldrich	Cat# A92902
17β-estradiol	Merck Millipore	Cat# 3301
Doxycycline hyclate	Sigma-Aldrich	Cat# D9891
Dimethyloxalylglycine (DMOG)	Sigma-Aldrich	Cat# D3695
TrypLE Express Enzyme (1X)	GIBCO	Cat# 12605010
Accutase cell detachment solution	Merck Millipore	Cat# SCR005
Trypsin-EDTA (0.05%)	GIBCO	Cat# 25300054
Phusion High-Fidelity DNA polymerase	Thermo Scientific	Cat# F530L
Gibson Cloning Master Mix	Produced in-house	N/A
MACS Streptavidin MicroBeads	Miltenyi Biotec	Cat# 130-048-101
MACS LS magnetic columns	Miltenyi Biotec	Cat# 130-042-401
Dynabeads Protein A for Immunoprecipitation	Life Technologies	Cat# 10002D
Dynabeads Protein G for Immunoprecipitation	Life Technologies	Cat# 10009D
Proteinase K	New England Biolabs	Cat# P8107S
Complete mini protease inhibitors	Roche	Cat# 11836153001
Unmethylated bacteriophage λ DNA	Promega	Cat# D1521
3C-protease	Produced in-house	N/A
Sf-900 II SFM	Thermo Scientific	Cat# 10902096
HisTrap FF 5ml columns	GE Healthcare	Cat# 17525501
Protino® 96 Ni-IDA	Macherey-Nagel	Cat# 745300
Critical Commercial Assays		
Alkaline Phosphatase Staining Kit II	Stemgent	Cat# 00-0055
Pacific Blue Annexin V/SYTOX AADvanced Apoptosis Kit	Thermo Scientific	Cat# A35136
Blood & Cell Culture DNA Mini kit	QIAGEN	Cat# 13323
TrueMethyl® Whole Genome kit	Cambridge Epigenetix	Cat# OP-06-001
KAPA Library Quantification Kit for Illumina Platforms, v1.14	Kapa Biosystems	Cat# KR0405
TruSeq SBS Kit v3-HS	Illumina	Cat# FC-401-3001
miRNeasy mini kit	QIAGEN	Cat# 217004
High Capacity RNA-to-cDNA kit	Applied Biosystems	Cat# 4387406
SYBR Green QPCR Master Mix	Applied Biosystems	Cat# 4309155
Nextera DNA sample preparation kit (ATAC-transposase and buffer)	Illumina	Cat# FC-121-1030
NEBNext High-Fidelity 2X PCR Master Mix (ATAC-library amplification)	New England BioLabs	Cat# M0541S
MinElute PCR Purification Kit	QIAGEN	Cat# 28004
QIAquick Gel Extraction Kit	QIAGEN	Cat# 28704
Agencourt AMPure XP-PCR purification kit	Beckman Coulter	Cat# A63882
High Sensitivity DNA Analysis Kits	Agilent	Cat# 5067-4626
NEBNext Ultra library prep kit	New England Biolabs	Cat# E7370
HiSeq SBS Kit v4	Illumina	Cat# FC-401-4002
X-tremeGENE HP DNA Transfection Reagent	Sigma-Aldrich	Cat# 06366244001
Deposited Data		
BS-seq/oxBS-seq data during B to iPS reprogramming	This paper	GEO:GSE103469
hMeDIP-seq data during MEF^{AGi} reprogramming	This paper	GEO: GSE117919
ATAC-seq data during MEF^{AGi} reprogramming	This paper	GEO: GSE117920
ATAC-seq and H3K4me2 ChIP-seq data during B to iPS reprogramming	This paper and (Stadhouders et al., 2018)	GEO: GSE103470
ChIP-seq data in Bcell and Bα'cell for: C/EBPα, PU.1 and H3K27ac	(Di Stefano et al., 2016)	GEO: GSE71218

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
C/EBPα ChIP-seq data in GMPs	(Hasemann et al., 2014)	GEO: GSE43007
hMeDIP-seq data in hematopoietic progenitors	(Han et al., 2016)	GEO: GSE77967
Tet2 ChIP-seq data in ESCs	(Xiong et al., 2016)	GEO: GSM2065691
ATAC-seq and ChIP-seq data for TFs during MEF to iPS reprogramming	(Chronis et al., 2017)	GEO: GSE90895
Tfcp2l1 ChIP-seq data in ESCs	(Chen et al., 2008)	GEO: GSM288350
Experimental Models: Cell Lines		
ESCs (E14TG2a)	ATCC	Cat# CRL-1821; RRID: CVCL_9108
Platinum-E retroviral packaging cells	Cell Biolabs, INC	Cat# RV-101; RRID: CVCL_B488
S17 stromal cell line	From Dr. Dorshkind, UCLA. (Collins and Dorshkind, 1987)	RRID: CVCL_E226
C10 pre-B cell line	Produced in-house (Bussmann et al., 2009)	N/A
Mouse Embryonic Fibroblasts, Irradiated	MTI-GlobalStem	Cat# GSC-6001G
Sf21 insect cell line	Thermo Scientific	Cat# 11497013
Experimental Models: Organisms/Strains		
Mouse: Oct4-GFP OSKM-reprogrammable	(Di Stefano et al., 2014)	N/A
Mouse: $Tet2^{f.}$ B6;129S-Tet2tm1.1Iaai/J	The Jackson Laboratory	Cat# JAX:017573; RRID:IMSR_JAX:017573
Mouse: Mx-Cre. B6; Cg-Tg(Mx1-cre)1Cgn/J	The Jackson Laboratory	Cat# JAX:003556, RRID:IMSR_JAX:003556
Mouse: $Tet2^{f}$ /Mx-Cre	This paper	N/A
Mouse: $Tet2^{f}$ /Mx-Cre Oct4-GFP OSKM-reprogrammable	This paper	N/A
Mouse: OKSM-reprogrammable	(Stadtfeld et al., 2010)	N/A
Oligonucleotides		
See Table S1	N/A	N/A
Recombinant DNA		
pMMLV-Tfcp2l1-IRES-mCherry	Vector Builder	N/A
pMSCV-Cebpa-IRES-hCD4	Produced in-house, (Bussmann et al., 2009)	N/A
pMX-Flag-Tet1CD	From Dr. Guo-Liang Xu, (Hu et al., 2014)	N/A
pMX-Flag-Tet2CD	From Dr. Guo-Liang Xu, (Hu et al., 2014)	N/A
pMX-Flag-Tet2HD	From Dr. Guo-Liang Xu, (Hu et al., 2014)	N/A
pMX-Flag-Tet3CD	From Dr. Guo-Liang Xu, (Hu et al., 2014)	N/A
pMMLV-Flag-Tet1CD-IRES-mCherry	This paper	N/A
pMMLV-Flag-Tet2CD-IRES-mCherry	This paper	N/A
pMMLV-Flag-Tet2HD-IRES-mCherry	This paper	N/A
pMMLV-Flag-Tet3CD-IRES-mCherry	This paper	N/A
C/EBPαDBD-pCoofy1 (his Tag)	This paper	N/A
C/EBPαTD-pCoofy2 (trx-histag)	This paper	N/A
pcoofy27-His-Tet2CD	This paper	N/A
Software and Algorithms		
R v3.1.0	N/A	https://www.r-project.org/
Bioconductor v3.0	(Huber et al., 2015)	https://bioconductor.org/
GEM aligner (v3)	(Marco-Sola et al., 2012)	https://biocore.crg.eu/wiki/GEM_Mapper
GEMBS pipeline	(Merkel et al., 2017)	N/A
STAR	(Dobin et al., 2013)	https://github.com/alexdobin/STAR
DeepTools 1.5.9.1	(Ramírez et al., 2016)	https://github.com/deeptools/deepTools/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
MACS2 v2.1.0.20140616	(Zhang et al., 2008)	https://github.com/taoliu/MACS
Bedops	(Neph et al., 2012)	https://bedops.readthedocs.io/en/latest/
RSAT suite	(Medina-Rivera et al., 2015)	http://rsat.eu
TRAP	(Thomas-Chollier et al., 2011)	http://trap.molgen.mpg.de/cgi-bin/home.cgi
Rpackage mFuzz v2.26.0	(Kumar and E Futschik, 2007)	https://bioconductor.org/packages/release/bioc/html/Mfuzz.html
Rpackage ReactomePA v1.10.1	(Croft et al., 2011)	https://bioconductor.riken.jp/packages/3.0/bioc/html/ReactomePA.html
Rpackage DESeq2 v1.6.3	(Love et al., 2014)	https://bioconductor.org/packages/release/bioc/html/DESeq2.html

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Thomas Graf (thomas.graf@crg.eu).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Mice

As a source for the B cells used in our experiments we crossed "OSKM-reprogrammable mice" containing a doxycycline-inducible OSKM cassette and the tetracycline transactivator (Carey et al., 2010) with an *Oct4*-GFP reporter strain (Boiani et al., 2002), as previously described (Di Stefano et al., 2016; Di Stefano et al., 2014). In addition, we also used as a source of B cells a newly established line with an interferon inducible *Tet2* knockout by crossing "*Oct4*-GFP/OSKM-reprogrammable mice" with *Tet2*^{fl} mice and MxCre mice (Figure S2B) that were obtained from The Jackson Laboratories. Reprogrammable MEFs were isolated at E13.5 from embryos containing a *Oct4*, *Klf4*, *Sox2* and *c-Myc* (OKSM) polycistronic cassette in the *Col1a1* locus and harboring homozygous *Rosa26-M2rtTA* alleles (Stadtfeld et al., 2010) "OKSM-reprogrammable mice." During experiments the number of female and male mice was balanced. Mice were housed in standard cages under 12h light–dark cycles and fed *ad libitum* with a standard chow diet. All experiments were approved by the Ethics Committee of the Barcelona Biomedical Research Park (PRBB) and performed according to Spanish and European legislation.

Cells and cell cultures

ESCs (E14TG2) were cultured on gelatinized plates and grown either in FBS+LIF conditions [Knockout-DMEM (GIBCO Cat#10829018) containing 15% ES-FBS (GIBCO Cat# 10270-106), Pen Strep (100X) (GIBCO Cat#15140122), L-Glutamine (100X) (GIBCO Cat#25030081), MEM Non-Essential Amino Acids Solution (100X) (GIBCO Cat#11140068), Sodium Pyruvate (100X), 2-Mercaptoethanol (Invitrogen, Cat#31350010) and 1000U/ml Leukemia Inhibitory Factor (LIF) (Merk Millipore Cat# ESG1106)] or in 2i-serum free conditions [50% DMEM-F12 (GIBCO Cat#12634010), 50% Neurobasal (GIBCO Cat#21103049) medium containing N2 supplement (100X) (GIBCO Cat#17502048), B27 supplement (50X) (GIBCO Cat#17504044), Pen Strep (100X) (GIBCO Cat#15140122), L-Glutamine (100X) (GIBCO Cat#25030081), MEM Non-Essential Amino Acids Solution (100X) (GIBCO Cat#11140068), Sodium Pyruvate (100X), 2-Mercaptoethanol (Invitrogen, Cat#31350010) 1 μM PD0325901 (Selleckchem Cat#S1036), 3 μM CHIR99021 (Selleckchem Cat#S1263) and 1000U/ml Leukemia Inhibitory Factor (LIF) (Merk Millipore Cat# ESG1106)]. CD19^{+} B cells were isolated from the bone marrow with a monoclonal antibody to CD19 (BD Biosciences Cat#553784), using MACS sorting technology (Miltenyi Biotech) as previously described (Di Stefano et al., 2016). The purity of the sorted cell fractions was assessed by FACS (using an LSR2 machine from BD Biosciences) and was found to be typically > 98%. B cells were grown in 20%FBS-RPMI medium [RPMI 1640 medium (GIBCO Cat#12633012) supplemented with 20% Fetal Bovine Serum (GIBCO Cat# 10270-106), Pen Strep (100X) (GIBCO Cat#15140122), L-Glutamine (100X) (GIBCO Cat#25030081), 2-Mercaptoethanol (1000X) (Invitrogen, Cat#31350010) and 10ng/ml IL-7 (Peprotech Cat# 217-17)].

METHOD DETAILS

B cell to iPSC reprogramming experiments

B cells isolated from 8-16 weeks *Oct4*-GFP/OSKM-reprogrammable mice were infected with C/EBPαER-hCD4 retrovirus, plated at 500 cells/cm^2 in gelatinized plates (12 wells) on irradiated MEF feeders (MTI-GlobalStem Cat# GSC-6001G) in 20%FBS-RPMI medium. To activate C/EBPα, cultures were treated for 18h with 100nM β-Estradiol (E2) (Merck Millipore Cat# 3301), resulting in Bα' cells. After E2 washout, the cultures were switched to serum-free N2B27 medium [50% DMEM-F12 (GIBCO Cat#12634010),

50% Neurobasal (GIBCO Cat#21103049) medium containing N2 supplement (100X) (GIBCO Cat#17502048), B27 supplement (50X) (GIBCO Cat#17504044), Pen Strep (100X) (GIBCO Cat#15140122), L-Glutamine (100X) (GIBCO Cat#25030081), MEM Non-Essential Amino Acids Solution (100X) (GIBCO Cat#11140068), Sodium Pyruvate (100X), 2-Mercaptoethanol (Invitrogen, Cat#31350010) and 1000U/ml Leukemia Inhibitory Factor (LIF) (Merk Millipore Cat# ESG1106)] supplemented with IL-4 10ng/ml (Peprotech Cat#214-14), IL-7 10ng/ml (Peprotech Cat#217-17) and 2ng/ml IL-15 (Peprotech Cat#210-15). To activate OSKM the cultures were treated with 2 μg/ml of doxycycline (Sigma-Aldrich Cat#D9891). From day 2 onward the serum-free N2B27 medium was supplemented with 20% KnockOut™ Serum Replacement (KSR) (GIBCO Cat#A3181502).

MEFs to iPSCs reprogramming experiments

Reprogrammable MEFs were isolated from "OKSM-reprogrammable mice" (Stadtfeld et al., 2010). During expansion, cells were grown in 10%FBS-DMEM medium [DMEM Medium (GIBCO Cat#12633012) supplemented with 10% Fetal Bovine Serum (GIBCO Cat# 10270-106), Pen Strep (100X) (GIBCO Cat#15140122), L-Glutamine (100X) (GIBCO Cat#25030081), MEM Non-Essential Amino Acids Solution (100X) (GIBCO Cat#11140068) and 2-Mercaptoethanol (Invitrogen, Cat#31350010)]. To induce reprogramming, cells were seeded onto 0.1% gelatin-coated plates at a density of 3000 cells/cm^2 in reprogramming medium [Knockout-DMEM (GIBCO Cat#10829018) containing 15% ES-FBS (GIBCO Cat# 10270-106), Pen Strep (100X) (GIBCO Cat#15140122), L-Glutamine (100X) (GIBCO Cat#25030081), MEM Non-Essential Amino Acids Solution (100X) (GIBCO Cat#11140068), 2-Mercaptoethanol (Invitrogen, Cat#31350010), 1000U/ml Leukemia Inhibitory Factor (LIF) (Merk Millipore Cat# ESG1106) and 2 μg/ml doxycycline (Sigma-Aldrich Cat#D9891)]. Where indicated, 50 μg/ml ascorbic acid (Sigma-Aldrich Cat#A92902) and 3 μM GSK3 inhibitor CHIR-99021 (Selleckchem Cat#S1263) were added to the culture medium.

Vectors and virus production and infection

The C/EBPαER-hCD4 retroviral vector was generated as described before (Bussmann et al., 2009). The pMMLV-Tfcp2l1-mCherry retroviral vector was purchased from Vector Builder.

pMX plasmids coding for Flag-tagged Tet1CD, Tet2CD, Tet3CD and Tet2CD (HD) were kindly provided by Dr. Xu Guoliang. These plasmids were used to generate coding sequences for the Flag-tagged Tet1, Tet2, Tet3 or Tet2 mutant(HD) catalytic domains(CD), then sub-cloned into a retroviral backbone expressing mCherry by Gibson cloning (Gibson et al., 2009). Briefly, linearized recipient pMMLV plasmid and Flag-tagged Tet1CD, Tet2CD, Tet3CD and Tet2CD (HD) constructs were amplified by polymerase chain reaction (PCR) using Phusion High-Fidelity DNA polymerase (Thermo Scientific, Cat#F530L) according to manufacturer's instructions with primers listed in Table S1. The PCR fragments with expected size were excised from agarose gel and purified using QIAquick Gel Extraction Kit (QIAGEN Cat#28704) following manufacturer's instructions. Gibson assembly reactions were performed using an in-house produced Gibson Cloning Master Mix at 50°C for 1 hour and then transformed into DH5α competent cells (Invitrogen Cat#18265017). Positive colonies were selected and verified by Sanger sequencing. Production of the different Flag-Tet fusion proteins was assessed by western blot using a specific antibody against Flag (Sigma-Aldrich Cat# F1804).

Viral production with platinum E cells (Cell Biolabs, INC Cat# RV-101) and infection of B cells and MEFs were performed as described previously (Di Stefano et al., 2014).

Inhibition of Tet activity by DMOG

Dimethyloxalylglycine (DMOG) (Sigma-Aldrich Cat#D3695) was used to inhibit Tet activity as previously described (Amouroux et al., 2016). Briefly, B cells isolated from "*Oct4*-GFP reprogrammable mice" were incubated with 1mM of DMOG during different times (Figure S2G) preceding C/EBPα and OSKM induction.

FACS analyses

Oct4-GFP expression was monitored with an LSR II flow cytometer (BD Biosciences) using Diva v6.1.2 (BD Biosciences) and FlowJo software v10 (TreeStar).

Colony counting

Reprogramming efficiency at day12 was determined using the Alkaline Phosphatase Staining Kit II (Stemgent Cat#00-0055) following the instructions provided by the manufacturer.

Cell viability

Cell viability was assessed using the Pacific Blue Annexin V/SYTOX® AADvanced Apoptosis Kit for flow cytometry (Thermo Fisher Cat#A35136).

Whole genome oxidative-bisulfite and bisulfite sequencing (oxBS-seq/BS-seq)

DNA was extracted using the Blood & Cell Culture DNA Mini kit (QIAGEN Cat#13323). The preparation of the sequencing libraries for oxBS and BS followed the workflow of CEGX TrueMethyl® Whole Genome kit, v3.1 (Cambridge Epigenetix Cat# OP-06-001) in HydroxyMethyl mode, with minor modifications. Each sample was prepared in duplicate to improve the diversity and the evenness of the sequencing library coverage. Briefly, HMW DNA (400ng) was spiked with unmethylated bacteriophage λ DNA (5 ng of λ DNA per microgram of genomic DNA; Promega Cat#D1521), with methylated T7 phage DNA (5 ng of T7 DNA per microgram of genomic DNA) and also with the manufacturer's digestion and sequencing control (20 ng per each microgram of genomic DNA).

Genomic DNA with the spike-in controls was sheared using Covaris LE220 focused-ultrasonicator to a mean fragment size of 800bp, purified, denatured and the sample was evenly split for the oxidation reaction and the mock-oxidation reaction where the oxidant solution was replaced by water. Both aliquots were then processed in parallel for all stages of the protocol. After the oxidation reaction where 5-hydroxymethylcytosine is oxidized to 5-formylcytosine (5fC) and 5-methylcytosine (5mC) stays unchanged, the bisulfite treatment converts 5fC and all non-methylated cytosines to uracil, while 5mC is not altered.

Illumina compatible adapters and library specific indexes were incorporated through ten PCR cycles. The quality control of the library was monitored using the Agilent 2100 Bioanalyzer DNA 7500 assay, and the concentration was estimated using the KAPA Library Quantification Kit for Illumina® Platforms, v1.14 (Kapa Biosystems Cat#KR0405).

Paired-end DNA sequencing (2 × 101 bp) of the oxBS and BS libraries was performed using the HiSeq2000 (Illumina) with TruSeq SBS Kit v3-HS (Illumina Cat#FC-401-3001) following the manufacturer's protocol with HiSeq Control Software (HCS) v1.5.15.1. Image analysis, base calling and quality scoring of the run were processed using the manufacturer's software Real Time Analysis (RTA 1.13.48) and followed by generation of FASTQ sequence files by bcl2fastq v1.8.4.

RNA extraction

Cells were trypsinized and pre-plated for 30min on new dishes to remove the feeders, the floating cells harvested and RNA isolation performed with the miRNeasy mini kit (QIAGEN Cat#217004). RNA was eluted from the columns using RNase-free water and quantified by Nanodrop. cDNA was produced with the High Capacity RNA-to-cDNA kit (Applied Biosystems Cat#4387406).

qRT-PCR analyses

qRT-PCR reactions were set up in triplicate with the SYBR Green QPCR Master Mix (Applied Biosystems Cat#4309155) and primers as listed in Table S1. Reactions were run on an AB7900HT PCR machine with 40 cycles of 30 s at 95°C, 30 s at 60°C and 30 s at 72°C.

ATAC seq

ATAC-seq was performed as previously described (Buenrostro et al., 2013). The standard protocol was used with the following modifications. 100.000 cells were washed once with 100μl PBS and resuspended in 50μl lysis buffer (10mM Tris-HCl pH 7.4, 10mM NaCl, 3mM $MgCl_2$, 0.2% IGEPAL CA-630). The suspension of nuclei was then centrifuged for 10min at 500 g at 4°C, followed by the addition of 50μl transposition reaction mix (25μl TD buffer, 2.5μl Tn5 Transposase and 22.5μl Nuclease Free H_2O) (Illumina Cat#FC-121-1030) and incubation at 37°C for 45min. DNA was isolated using MinElute Kit (QIAGEN Cat#28004). Library amplification was done by two sequential PCR reactions (8 and 5 cycles, respectively) using NEBNext High-Fidelity 2X PCR Master Mix (New England BioLabs Cat#M0541S). Library quality was checked on a Bioanalyzer using High Sensitivity DNA Analysis Kits (Agilent Cat# 5067-4626), followed by paired-end sequencing (2x50bp) on an Illumina HiSeq2500.

ChIP-seq

ChIP-Seq using tagmentation (ChIPm-Seq) was performed as previously described (Schmidl et al., 2015) with 100,000 crosslinked cells using 1 μL of H3K4me2 antibody (Abcam, Cat#ab7766) per IP. Tagmentation of immobilized H3K4me2-enriched chromatin was performed for 2 min at 37°C in 25 μL transposition reaction mix (12.5 μL TD buffer, 1.0 μL Tn5 transposase and 11.5 μL nuclease free water) (Illumina Cat#FC-121-1030). Library amplification was performed as described for ATAC-Seq. Library quality was checked on a Bioanalyzer using High Sensitivity DNA Analysis Kits (Agilent Cat# 5067-4626), followed by sequencing (1x75bp) on an Illumina NextSeq500.

Chromatin immunoprecipitation followed by quantitative PCR (ChIP-qPCR)

ChIP-qPCR experiments were performed as described previously (Kallin et al., 2012). Briefly, cells were crosslinked with 1% paraformaldehyde and sonicated. Solubilized chromatin was immunoprecipitated with antibodies against Tet2 (Abcam Cat# ab124297) or rabbit IgG (Abcam Cat#ab171870) as a negative control. Antibody-chromatin complexes were pulled down using a 1:1 mixture of BSA-blocked Protein A and Protein G magnetic beads (Life Technologies Cat#10002D and Cat#10009D respectively), washed, and then eluted. After crosslink reversal and proteinase K treatment (New England Biolabs Cat#P8107S), immunoprecipitated DNA was extracted with phenol-chloroform and ethanol precipitated. The DNA fragments were further analyzed by qPCR.

MeDIP/hMeDIP-qPCR

DNA extraction and immunoprecipitation was carried out by a previously published protocol (http://www.epigenome-noe.net/, PROT33) modified by (Di Stefano et al., 2014). Briefly, 1 μg of genomic DNA was sonicated and incubated with 5 μg of either anti-5hmC antibody (Active Motif Cat#39769) or anti-5mC antibody (Diagenode Cat#C15200081) at 4C for 2hr. Antibody-DNA complexes were pulled down using a 1:1 mixture of BSA-blocked Protein A and Protein G magnetic beads (Life Technologies Cat#10002D and Cat#10009D respectively), washed, and then eluted. DNA was extracted with phenol-chloroform and ethanol precipitated. 5mC and 5hmC content of the samples was assessed by qPCR.

hMeDIP-Seq

DNA was extracted from MEF reprogramming samples using the Blood & Cell Culture DNA Mini kit (QIAGEN Cat#13323). Purified genomic DNA was sonicated using Covaris to obtain fragments of 200–600 bp. Library preparation was performed using the NEBNext Ultra library prep kit (New England Biolabs Cat#E7370) substituting the NEB hairpin adaptors by full-length barcoded

TruSeq adaptors from Illumina (Cat# FC-401-3001). Fragmented DNA with compatible adaptors were pooled together. Pooled DNA was heated for 10 minutes at 99°C, cold down on ice, and incubated overnight with 5hmC (1 μg Ab/μg DNA; Active Motif Cat#39769) antibody in IP buffer (10 mM Sodium Phosphate buffer, 140 mM NaCl, 0.05% Triton X-100). Immunocomplexes were recovered using pre-equilibrated Protein G/A Dynabeads (Life Technologies Cat#10002D and Cat#10009D respectively) for 2 hr at 4°C. After washing with IP buffer, immunocomplexes were eluted from columns by heating at 55°C for 30 minutes in proteinase digestion buffer [50 mM Tris-HCl pH = 8, 10 mM EDTA, 0.5% SDS, 35 mg Proteinase K (New England Biolabs Cat#P8107S)] and the DNA purified using MinElute PCR Purification kit (QIAGEN Cat#28004). Recovered DNA was amplified by PCR and sequenced in Illumina HiSeq2500 using 125-bp paired-end reads with HiSeq v4 chemistry (Illumina Cat#FC-401-4002).

Western blots (WBs) and immunoprecipitation

For WB analyses, proteins were extracted using NET-2 buffer [50mM Tris pH 7.4, 200mM NaCl, 0,1% triton and protease inhibitors (Roche Cat#11836153001)] loaded on SDS-PAGE and transferred onto PVDF membranes as previously described (Sardina et al., 2010).

For endogenous Co-IP experiments, lysates were prepared using lysis buffer [50mM Tris pH 7.5, 150mM NaCl, 1mM EDTA, 1mM EGTA, 5mM $MgCl_2$, 0,5% triton and protease inhibitors (Roche Cat#11836153001)] and centrifuged to eliminate insoluble material. The extract was pre-cleared by incubation with magnetic beads for 2h on a rotating wheel at 4°C. To crosslink the antibodies to Dynabeads® A or G (Life Technologies Cat#10002D and Cat#10009D respectively), 50 μL of beads were incubated with 3ug of the specific antibody or with 3ug the corresponding isotype control-IgG during 2 hr at room temperature on a rotating wheel. The mix was then washed once with PBS and twice with triethanolamine (pH8,2, 0,2M) and then incubated during 20min at room temperature with dimethyl pimelimidate (DMP, 20 m). To stop the reaction two 5 min washes with Tris-HCl (pH8, 50mM) were performed followed by 3 washes with PBS. Then beads were incubated with citric acid (0,1M, pH3) for 2 min. Finally, the mix was washed twice with lysis buffer before incubating it with the precleared protein extracts overnight at 4°C on a wheel. Afterward 1/20th of the mixture was kept as the unbound fraction and the rest washed six times with lysis buffer. To elute, the last wash was removed and the beads resuspended in 1X Laemmli buffer without β-mercaptoethanol and heated for 20 min at 60°C. Magnetic beads were then separated, the supernatant taken, complemented with 5% β-mercaptoethanol and boiled before loading on an SDS-PAGE acrylamide gel to analyze proteins by WB.

Production of recombinant forms of C/EBPα and Tet2

C/EBPα fragments were generated in the *E. coli* strain BL21(DE3) (New England Biolabs Cat#C2527I) by transforming them with C/EBPα DBD (DNA Binding and Dimerization Domain)-pCoofy1 (his tag) or C/EBPα TD (Transactivation Domain)-pCoofy2 (trx-his-tag). Cells growing in 1 l of 0.5mM IPTG induced-2XTY medium at OD = 0.6 were harvested at 4,000 g during 30min. Pellets were resuspended in a lysis buffer containing: 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 0.5% Triton X-100, 2 mM DTT 20 mM Imidazole, and protease inhibitors. Cells were then broken using a French Press and ultracentrifuge at 30,000 g during 30 min. The cell extract was loaded into a HisTrap FF 5ml column (GE Healtchare Cat#17525501) at 4°C. The column was washed with wash buffer [50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 2mM DTT, 20mM imidazole, and 10% glycerol)] and then eluted with 50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 2mM DTT, 500 mM imidazole. The eluate was pooled and buffer exchanged [50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 2mM DTT, 20mM imidazole, and 10% glycerol]. His-tags removal from recombinant C/EBPα peptides was performed using 3C protease (produced in-house) O/N at 4°C. The mixture was loaded again into HisTrap FF 5ml column (GE Healthcare Cat#17525501) in order to bind de 3C and the His-tag. The flow-through of the column was collected and buffer exchanged using an exchange column into 50mM Tris HCl pH = 7,4, 300mM NaCl, 10% glycerol, 2 mM DTT and flash frozen.

6His-Tet2CD (Catalytic Domain) peptide was generated in baculoviruses by transfecting Sf21 insect cells (Thermo Scientific Cat# 11497013) with pcoofy27-His-Tet2CD by X-tremeGENE HP DNA Transfection Reagent (Sigma Aldrich Cat# 06366244001) for 60 hours and amplified twice to get the P3 stock. Sf21 cells were cultured in Sf-900 II SFM (Thermo Scientific Cat# 10902096) at 27°C and infected with P3 stock of 6His-Tet2CD baculovirus for 72 hours; cells were harvested and centrifuged for 15 min at 3,000 g and 4°C. Cells were then lysed in a Triton lysis buffer [50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 1% Triton X-100, 2 mM DTT, 20 mM Imidazole, and protease inhibitors (Roche Cat#11836153001)] and broken using a French Press. Recombinant proteins were affinity-purified using HisTrap FF 5ml columns (GE Healthcare Cat#17525501) at 4°C. The column was then washed with wash buffer [50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 2mM DTT, 20mM imidazole, and 10% glycerol], eluted [50 mM Tris-HCl (pH 7.5), 500 mM NaCl, 2mM DTT, 500mM imidazole and 10% glycerol] and then stored at −80°C.

C/EBPα-Tet2 pulldown assay

Prior to the pulldown assay, recombinant peptides for C/EBPα and Tet2 were buffer exchanged to 50mM Tris HCl (pH 7.4), 150mM NaCl.

For the pulldown assay, the same amount of purified recombinant 6His-Tet2CD and C/EBPα-DBD or 6HisTet2CD and C/EBPα-TD were mixed and incubated for 30 min at RT, the mixture was loaded into a Protino® Ni_IDA 96 Ni-IDA (Macherey-Nagel Cat#745300). Each column of the Protino® was washed using the buffer mentioned above and the elution was performed in a buffer containing 50mM Tris HCl (pH 7.4) 150mM NaCl, 500mM Imidazole. Protein-protein interaction was examined by WB using specific antibodies against Tet2 (Abcam Cat#ab124297) and C/EBPα (Santa Cruz Biotechnology Cat#sc-61).

QUANTIFICATION AND STATISTICAL ANALYSIS

All sequencing data obtained were mapped onto the mouse genome assembly mm10 (Ensembl GRCm38.78) and analyzed with R (3.1.0) using packages from the bioconductor suite (v3.0) (Huber et al., 2015). For peak calling, regions overlapping the 'Encode blacklist' regions were removed (ENCODE Project Consortium, 2012). All the clusterings were performed using the Rpackage Mfuzz (2.26.0) (Kumar and E Futschik, 2007) on standardized data. All GO enrichment analyses were performed using the Rpackage ReactomePA (1.10.1)(Croft et al., 2011). The Principal Component Analysis (PCA) was performed using the prcomp function in the stats (v3.3.2) R package.

Bisulfite and ox-bisulfite sequencing data processing

Read mapping was carried out using the GEM aligner (v3) (Marco-Sola et al., 2012) against a composite reference containing two copies of the mouse mm10 Ensembl GRCm38.78 genome and two copies of the NCBI viral genome database (v35). For both the mouse and viral references, one copy had all cytosine bases replaced by thymine bases and the other had all guanine bases replaced by adenine bases. Before mapping was performed, the original sequence for each read was stored. The first read from each pair then had all cytosine bases replaced by thymine bases, and the second read had all guanine bases replaced by adenine bases. Read mapping with GEM was performed, allowing up to four mismatches per read with respect to the reference. After read mapping, the original sequence for each read was restored.

Estimation of cytosine levels was carried out on read pairs where both members of the pair mapped to the same contig with consistent orientation and there was no other such configuration at the same or a smaller edit distance from the reference. After mapping, we restored the original read data in preparation for the inference of genotype and methylation status. We estimated genotype and DNA methylation status simultaneously using the GEMBS pipeline (Merkel et al., 2017), taking into account the observed bases, base quality scores and the strand origin of each read pair. For each genome position, we produced estimates of the most likely genotype and the methylation proportion (for genotypes containing a cytosine base on either strand). After the initial calling step, the data were combined across samples and the genotypes were recalled assuming no sequence variability between samples. This increased the power to call genotypes and therefore also increased the number of sites that could be used for downstream analyses. Sites were selected where the combined genotype call across samples was for a homozygous CC followed by GG with a Phred score of at least 20, corresponding to an estimated genotype error level of $\leq$ 1%. Sites with > 500-coverage depth were excluded to avoid centromeric or telomeric repetitive regions. All subsequent analyses used this selected set.

ATAC-seq analyses

For the ATAC-seq analyses, reads were mapped using STAR (Dobin et al., 2013) (parameters: -outFilterMultimapNmax 1 -outFilterMismatchNmax 999 -outFilterMismatchNoverLmax 0.06 -alignIntronMax 1 -alignEndsType EndToEnd -alignMatesGapMax 2000). Duplicates reads were removed using Picard (http://broadinstitute.github.io/picard/) (function MarkDuplicates, parameter REMOVE_DUPLICATES = true). Bigwig tracks were made using DeepTools BamCoverage (1.5.9.1) (Ramírez et al., 2016) (parameters: -binSize 1 -normalizeUsingRPKM).

Peak calling was performed using macs2 (2.1.0.20140616) (Zhang et al., 2008) (parameters: -f BAMPE -g mm -bw 300 -p 0.001 -nolambda -keep-dup auto –call-summits). For quantitative analyses, peaks for all time points were merged as one set of non-overlapping regions using Bedops (Neph et al., 2012) and selecting from each individual peak calling the overlapping summit with the highest pValue (python script). Reads were counted on merged regions for each time point, using the Rpackage csaw (1.0.7) (function regionCounts, parameter: max.frag = 2000, pet = 'both'). DESeq2 (Love et al., 2014) was then used to perform quantitative analyses. Counts were scaled on genome wide fragments rather than on merged regions; to do so we counted reads in 10kb bins across the genome *(*using *csaw* function *WindowCounts),* used *DESeq2* to calculate the size factors for each sample and applied this to the counts on merged regions.

For the analysis of methylation at ATAC peaks, 500bp were selected around the peaks summits and only peaks with at least 3 CpG covered > 10x were kept. The fraction of 5mC/5hmC of all overlapping CpGs used to calculated 5mC/5hmC average score and inter-quartile range.

ChIP-seq and hMeDIP-seq analyses

ChIP-seq and hMeDIP-seq data were analyzed in the same way as ATAC-seq. Quantification of ChIPseq and hMeDIP-seq at ATAC peaks was performed the same way as for ATAC but fragments were quantified in a 2kb window around the summit of the peak.

Transcription factor binding motif analyses

Motif analyses were performed using the *RSAT suite* softwares (Medina-Rivera et al., 2015). Sequences were retrieved using the fetch-sequencing program, and repeats were masked to avoid artifactual motifs enrichment, using *bedtool maskfasta* and the mm10 Ensembl repeat annotation (GRCm38.75). Motif discovery was performed using *peaks-motifs*. Motif enrichment was performed with *matrixQuality*, using as background a Markov model of order 1 made on the purged sequences outputted by peak-motifs (i.e., overlapping regions appear only once and duplicates are masked, to avoid artifact motif enrichment), merged from all the

clusters. The Maximum-normalized-Weight-Difference Score (MNWD) was used as a motif enrichment score (Medina-Rivera et al., 2011). TFs binding prediction at enhancers of pluripotency genes was performed using TRAP (Thomas-Chollier et al., 2011) with JASPAR database.

DATA AND SOFTWARE AVAILABILITY

The accession number for all the sequencing data reported in this paper is NCBI GEO: GSE103470