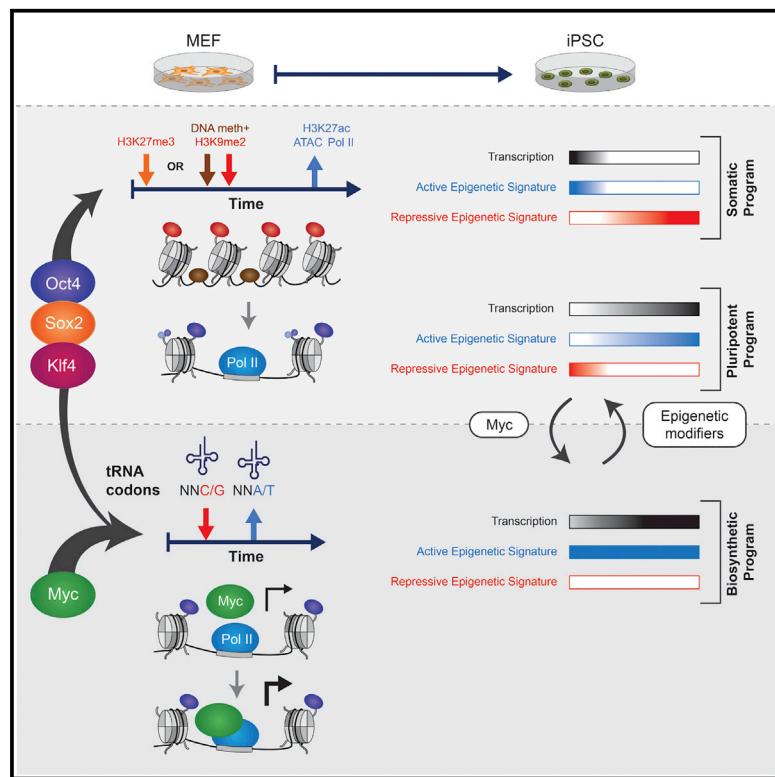


Deterministic Somatic Cell Reprogramming Involves Continuous Transcriptional Changes Governed by Myc and Epigenetic-Driven Modules

Graphical Abstract



Authors

Asaf Zviran, Nofar Mor, Yoach Rais, ..., Yitzhak Pilpel, Noa Novershtern, Jacob H. Hanna

Correspondence

yoach.rais@weizmann.ac.il (Y.R.), noa.novershtern@weizmann.ac.il (N.N.), jacob.hanna@weizmann.ac.il (J.H.H.)

In Brief

The epigenetic dynamics of iPSC reprogramming in correctly reprogrammed cells at high-resolution remain largely undefined. Hanna and colleagues now provide comprehensive characterization during the entire course of murine fibroblast reprogramming, by using Gata2a-Mbd3/NuRD-depleted and radically efficient reprogramming systems. These data provide insights into key questions underlying successful naive iPSC reprogramming.

Highlights

- Unbiased high-resolution profiling of mouse deterministic reprogramming to naive iPSCs
- Early DNA demethylation of pluripotency enhancers definitively marks future iPSCs
- Myc activity is indispensable for conducive iPS formation from somatic cells
- Optimized changes in tRNA codon usage amplify the output of Myc governed program



Deterministic Somatic Cell Reprogramming Involves Continuous Transcriptional Changes Governed by Myc and Epigenetic-Driven Modules

Asaf Zviran,^{1,2,15} Nofar Mor,^{1,15}* Yoach Rais,^{1,15,*} Hila Gingold,¹ Shani Peles,¹ Elad Chomsky,^{1,3,4} Sergey Viukov,¹ Jason D. Buenrostro,^{5,6} Roberta Scognamiglio,⁷ Leehee Weinberger,¹ Yair S. Manor,¹ Vladislav Krupalnik,¹ Mirie Zerbib,¹ Hadas Hezroni,³ Diego Adhemar Jaitin,⁸ David Larastiaso,⁸ Shlomit Gilad,⁹ Sima Benjamin,⁹ Ohad Gafni,¹ Awni Mousa,¹ Muneef Ayyash,¹ Daoud Sheban,^{1,8} Jonathan Bayerl,¹ Alejandro Aguilera-Castrejon,¹ Rada Massarwa,¹ Itay Maza,^{1,10} Suhair Hanna,^{1,11} Yonatan Stelzer,¹² Igor Ulitsky,³ William J. Greenleaf,^{13,14} Amos Tanay,^{3,4} Andreas Trumpp,⁷ Ido Amit,⁸ Yitzhak Pilpel,¹ Noa Novershtern,^{1,16,*} and Jacob H. Hanna^{1,16,17,*}

¹Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 761001, Israel

²New York Genome Center, New York, NY, USA

³Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel

⁴Department of Computer Science, Weizmann Institute of Science, Rehovot, Israel

⁵Broad Institute of MIT and Harvard, Cambridge, MA, USA

⁶Harvard Society of Fellows, Harvard University, Cambridge, MA, USA

⁷Heidelberg Institute for Stem Cell Technology and Experimental Medicine (HI-STEM gGmbH), Division of Stem Cells and Cancer, German Cancer Research Center (DKFZ), Heidelberg, Germany

⁸Department of Immunology, Weizmann Institute of Science, Rehovot, Israel

⁹Israel National Center for Personalized Medicine (G-INCPM), Weizmann Institute of Science, Rehovot, Israel

¹⁰Department of Gastroenterology, Rambam Hospital, Haifa, Israel

¹¹Department of Pediatrics, Rambam Hospital, Haifa, Israel

¹²Department of Molecular and Cell Biology, Weizmann Institute of Science, Rehovot, Israel

¹³Department of Genetics, Stanford University, Palo Alto, CA, USA

¹⁴Chan Zuckerberg Biohub, San Francisco, CA, USA

¹⁵These authors contributed equally

¹⁶Senior author

¹⁷Lead Contact

*Correspondence: yoach.rais@weizmann.ac.il (Y.R.), noa.novershtern@weizmann.ac.il (N.N.), jacob.hanna@weizmann.ac.il (J.H.H.)

<https://doi.org/10.1016/j.stem.2018.11.014>

SUMMARY

The epigenetic dynamics of induced pluripotent stem cell (iPSC) reprogramming in correctly reprogrammed cells at high resolution and throughout the entire process remain largely undefined. Here, we characterize conversion of mouse fibroblasts into iPSCs using Gatad2a-Mbd3/NuRD-depleted and highly efficient reprogramming systems. Unbiased high-resolution profiling of dynamic changes in levels of gene expression, chromatin engagement, DNA accessibility, and DNA methylation were obtained. We identified two distinct and synergistic transcriptional modules that dominate successful reprogramming, which are associated with cell identity and biosynthetic genes. The pluripotency module is governed by dynamic alterations in epigenetic modifications to promoters and binding by Oct4, Sox2, and Klf4, but not Myc. Early DNA demethylation at certain enhancers prospectively marks cells fated to reprogram. Myc activity drives expression of the essential biosynthetic module and is associated with optimized changes in tRNA codon usage. Our functional validations highlight interwoven

epigenetic- and Myc-governed essential reconfigurations that rapidly commission and propel deterministic reprogramming toward naive pluripotency.

INTRODUCTION

The ability to reprogram somatic cells into induced pluripotent stem cells (iPSCs) with Oct4, Sox2, Klf4, and Myc (abbreviated as OSKM) (Takahashi and Yamanaka, 2006) has provoked interest to define the molecular characteristics of this process. Previous epigenetic mapping studies on iPSC reprogramming were conducted on inefficient and asynchronous systems undergoing protracted reprogramming (Hussein et al., 2014) or by sorting pre-iPSCs, most of which do not progress to become iPSCs (Polo et al., 2012). However, this low efficiency and heterogeneity have limited genome-wide analysis of well-characterized, relatively homogeneous populations of cells that successfully complete this process.

Our group has demonstrated that optimized hypomorphic depletion of Mbd3 or Gatad2a, representing core members of the Gatad2a-Chd4-Mbd3/NuRD repressor complex, results in deterministic, up to 100% efficient, and more synchronized reprogramming in mouse cells within 8 days (Rais et al., 2013; Mor et al., 2018). Such systems enable high-resolution temporal



dissection of epigenetic dynamics underlying conducive naive iPSC formation, while simultaneously reducing “noise” from heterogeneous populations that fail to correctly complete the reprogramming course. With this opportunity to map epigenetics of reprogramming toward ground state naive pluripotency in highly efficient and homogeneous systems, without cell passaging or sorting for sub-populations, we provide comprehensive characterization during the entire 8-day course of mouse fibroblast reprogramming.

RESULTS AND DISCUSSION

Mapping Deterministic Reprogramming

We analyzed reprogramming of two independently generated Mbd3^{f/-} and Gatad2a^{-/-} secondary mouse embryonic fibroblast (MEF) clonal systems, carrying a doxycycline (DOX)-inducible human OKSM transgene. Cells were harvested every 24 hr until day 8, in which they are fully reprogrammed and processed for library preparation and sequencing. Two wild-type (WT) MEF secondary reprogramming systems (WT-1 and WT-2) were used as controls, where WT-2 is an isogenic genetically matched cell line to the Gatad2a^{-/-} cells (Mor et al., 2018; Rais et al., 2013). The use of independent Mbd3- and Gatad2a-depleted highly efficient systems with different single copy OKSM transgene integration pattern excludes cell line-specific signatures (Figure S1A). We sequenced 212 libraries from the NuRD-depleted systems and 21 libraries from WT1/2 systems (Table S1). The libraries span transcriptome (RNA sequencing [RNA-seq], small RNA-seq), chromatin modifications (chromatin immunoprecipitation sequencing [ChIP-seq] for H3K27ac, H3K27me3, H3K4me1, H3K4me3, H3K36me3, H3K9me2, and H3K9me3), DNA methylation, chromatin accessibility (ATAC-seq), and factor binding (Oct4, Sox2, Klf4, c-Myc, and RNA-Pol II ChIP-seq). Overall, we aligned 12.12 billion reads (Table S1). RNA-seq samples were reproducible with average correlation of $R = 0.93$ between consecutive samples (Figures 1B and 1C). TF and chromatin modification ChIP-seq samples showed high overlap between consecutive samples (Jaccard index > 0.3 ; Figures 1D and 1E). iPSC and embryonic stem cell (ESC) samples showed high consistency with previous measurements (Fisher exact test $p < 10^{-4}$; Figure 1F).

Deterministic Reprogramming Is Accompanied by Continuous Transcriptional Changes

Gene expression profiles in Mbd3^{f/-} and Gatad2a^{-/-} systems were highly similar (average $R = 0.88$ in Figures 1C and 1G and Tables S2 and S3). To further evaluate the kinetics of the two systems, we compared them to two WT secondary reprogramming systems (WT-2 series is isogenic to Gatad2a^{-/-} series), and to some of the previously published datasets that mapped iPSC reprogramming from WT fibroblasts. Principal component analysis (PCA) mapped samples in a trajectory that reflects the progression of reprogramming from MEF to iPS and/or ES (Figure 1H). MEF samples of all systems are clustered together in close proximity to WT samples measured in days 2, 4, 6, and 8, emphasizing the fast kinetics of NuRD-depleted systems compared to WT systems. Mbd3^{f/-} and Gatad2a^{-/-} samples from the same time points are closely mapped in both dimensions (Figure 1H). Importantly, although previously published data measured in WT MEF and iPS (Polo et al., 2012) are clustered together with our corre-

sponding samples, all other samples, which were measured from sorted cells undergoing reprogramming, are positioned in clusters based on the marker used for sorting and not according to the reprogramming day (Figure 1H). In addition, Thy1⁺, Thy1⁻, and SSEA1⁺ WT pre-iPSC samples do not cluster with any of the samples from Mbd3^{f/-} or Gatad2a^{-/-} systems beyond day 3 of reprogramming (PC1), possibly consistent with the notion that reprogramming measurements on inefficient stochastic reprogramming systems focus on early time points and infer data on populations most of which are not proceeding toward iPSCs.

Previous analysis on stochastic reprogramming systems has indicated only two waves of major transcriptional changes, one at the beginning and the other at the end of iPSC reprogramming, while in between there are very minor changes in transcriptional patterns (Hussein et al., 2014; Polo et al., 2012). We set out to test whether highly efficient reprogramming systems will show a similar pattern. A total of 8,705 genes (of which 8,042 are polyA⁺) were identified as differentially expressed along the Mbd3^{f/-} MEF to naive iPSC 8-day reprogramming course (Table S3). These genes show a sequential activity, and can be sorted according to their expression temporal pattern, showing a continuous dynamic transition from the somatic program to the pluripotent one. Three major expression shifts are observed during the continuous dynamic transition (Figure 1G). First, a large group of genes that are active in MEFs are downregulated as early as day 1. The second is a transient activation of genes between days 1 and 4. Finally, there is a gradual establishment of iPS and/or ES signature starting at day 5. Functional enrichment analysis in a single-day resolution (Figure 1I) characterized these changes: genes that are active in MEF and downregulated after DOX induction are enriched for somatic program processes (e.g., developmental process). Genes induced between days 1 and 6 are enriched for processes related to biosynthetic pathways (DNA and purine biosynthesis, translation). These processes are followed by induction of genes enriched for epigenetic remodeling and DNA repair processes. Finally, at day 6 there is a prominent induction of pluripotency maintenance master regulators including Nanog and Prdm14 (Figure S1B). In summary, at the transcriptional level, during conducive reprogramming trajectory, somatic cell repression- and pluripotency gene reactivation-associated changes do not occur simultaneously and are separated in time. However, many other changes related to cellular adaptation occur in between, thus rendering global transcriptional changes rather continuous and not confined only to early and late stages of iPSC reprogramming (Figures S1C and S1D; Table S4).

Dynamic OSK Binding Governs Conducive iPSC Formation

We identified 40,174 enhancers with a dramatic change in activity, which is consistent between the two NuRD depletion approaches used (Figure S2). We observed dynamic binding pattern for OSK in enhancers (Figure 1D), which is changing during reprogramming from an early pattern to late pluripotency-related pattern (Figures 2A–2C). cMyc has a strong preference to bind promoters over enhancers during reprogramming (Figures 2B and 2D). Inspecting the binding co-localization of OSKM shows a clear difference between promoters and enhancers. Oct4-binding enhancers overlap with Sox2 and to a lesser extent with Klf4 targets throughout the process

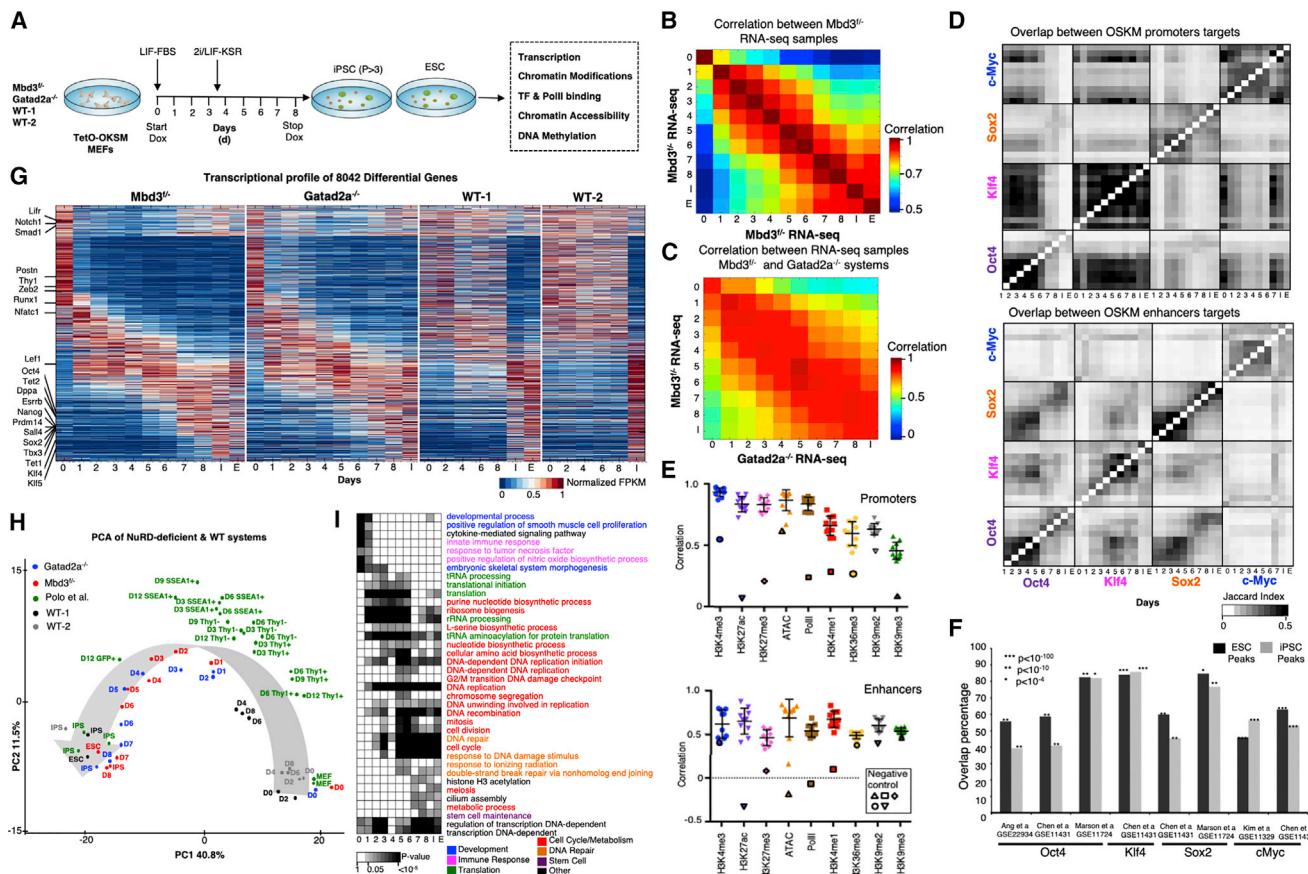


Figure 1. Continuous and Coordinated Progression of Conducive Reprogramming in Two Independent NuRD-Deficient Systems

- (A) Experimental scheme.
- (B) Spearman correlation between expression profiles of *Mbd3^{-/-}* system. Calculated over all differential genes (n = 8,042) showing an average correlation of R = 0.93 between consecutive samples.
- (C) As in (B), but between *Mbd3^{-/-}* and *Gatad2a^{-/-}* systems.
- (D) Overlap between targets of OSKM in promoters and enhancers. Pixel shade indicates Jaccard index.
- (E) Correlation between consecutive samples in *Mbd3^{-/-}* system (MEF-day 1, day 1-day 2, ...day 8-iPS), measured over all ESPGs promoters (promoters with differential chromatin pattern, n = 3,593, top), or all differential enhancers (n = 40,174, bottom), for each chromatin mark. Negative controls were calculated between MEF and iPS and are marked with solid border.
- (F) Overlap between binding targets of Oct4, Sox2, Klf4, or Myc, and previously published binding data of the same factors calculated in ES and iPS samples. Percentage out of our measured binding targets is presented, along Fisher exact test p values.
- (G) Global transcriptional pattern of 8,042 differential genes (FC > 4 and maximal fragments per kilobase million [FPKM] value > 1), sorted by their temporal pattern in *Mbd3^{-/-}* system (the same gene order was applied for the other reprogramming systems). Heatmap represents unit-transformation of FPKM values.
- (H) PCA analysis of all samples alongside samples from a previous publication (Polo et al., 2012). PCA was calculated on the same set of genes and normalization as in (G).
- (I) GO categories enriched among the genes that are active in each day. Gene is defined to be active in samples where FPKM is above 0.5 of the gene max value. p values were calculated with Fisher exact test and FDR-corrected. Categories with corrected p value <0.01 in at least two time points are presented. Gray shades represent FDR-corrected p values.

(Figure 2D). The average probability to see co-localization of Oct4 and Sox2 in enhancers is 0.61, while in promoters, it decreases to 0.39. However, the probability of Oct4 and Klf4 co-localization in promoters is higher by ~20% compared to enhancers. Differences between binding of enhancers and promoters are also apparent at the DNA motif level (Figure 2E). While OSK-binding promoters are enriched mainly for temporally stable OSK-binding motifs, OSK-binding enhancers are enriched for many additional and temporally varying binding motifs (Figure 2E). This change in motif preference may indicate a change in the collaborative binding of OSK during reprogramming

(Chronis et al., 2017), and may be responsible for the more dramatic changes in enhancer activity: 5-fold increase in total differential enhancers (40,174 in comparison to differential gene promoters [8,042]) underscores the magnitude of enhancer reprogramming.

We next asked if OSK are directly responsible for the repression of the somatic program. When inspecting enhancers that are active in MEF and repressed already at day 1, we observed that they are not significantly bound by OSK at any stage (not even when OSK are already expressed at day 1) (Figure 2B). This observation is different from that previously reported

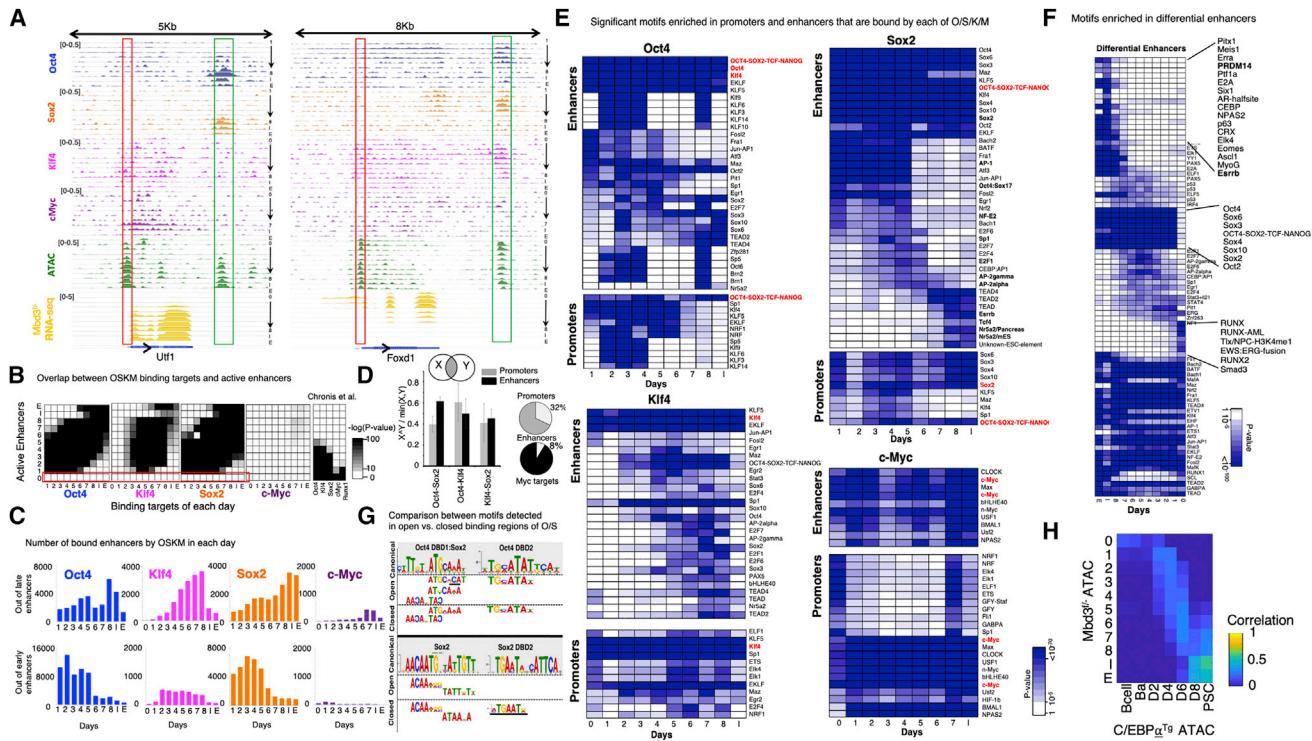


Figure 2. Stage-Specific Binding Preference and Collaboration of OSKM

(A) ChIP-seq landscape of two examples. Promoters are marked in red, enhancers are marked in green. Signals are normalized to sample size (RPM).

(B) Overlap between binding of OSKM and active enhancers in each day of reprogramming. Enhancer is defined as active in a specific day if its ATAC-seq Z score is above 1.5 SD in that day. Gray shades indicate Fisher exact test p value for overlap between compared samples. Note that OSKM do not bind the enhancers that are active in MEF (D0, marked in red); these enhancers are not significantly bound by OSKM at any day during reprogramming. Rightmost square indicates binding of the same enhancers by OSKM as measured previously in a highly inefficient WT system (Chronis et al., 2017).

(C) Number of enhancers bound by each of OSKM factors in each day of reprogramming. Top: out of enhancers that are bound by the factor in late stages (day 8, iPS, ESC). Bottom: out of enhancers that are bound by the factor in early stages (day 1–day 3).

(D) Probability to observe co-localized binding of transcription factors in promoters (gray) and enhancers (black). Calculated in days 1, 8, and iPS (error bars indicate SEM). Right: Myc binds 32% of promoters and 8% of active enhancers.

(E) Significant motifs enriched in promoters and enhancers that are bound by each of Oct4, Sox2, Klf4, and c-Myc at different days of reprogramming, as detected by Homer/4.7 software. p values, indicated by color shade, were reported by Homer and are FDR-corrected. Motifs that are significantly enriched (corrected $p < 10^{-30}$) in at least one time point are presented.

(F) Motifs enriched in differential enhancers that are active in each day of reprogramming (ATAC-seq Z score > 1.5). p values, indicated by color shade, were reported by Homer and are FDR corrected. Motifs that are significantly enriched (corrected $p < 10^{-50}$) in at least one time point are presented.

(G) Motifs found in “closed” versus “open” binding targets of the indicated transcription factor. Accessibility of targets was calculated based on ATAC-seq. Motifs found in OSK binding targets calculated in Mbd3^{flox/-} day1. Motifs that are different between open and closed binding targets are marked in black line. Complementary motifs to canonical motif appear in reverse order.

(H) Spearman correlation between ATAC-seq profiles of the two efficient reprogramming systems: Mbd3^{flox/-} MEF and C/EBPα^{Tg} B cell systems calculated over 40,174 differential enhancers.

(Chronis et al., 2017), who observed predominant OSK binding on MEF open enhancers at day 2 of the process (Figure 2B). This difference is likely due to their usage of a system with <1% iPSC efficiency (Carey et al., 2010). Our observations and those of others who used highly efficient systems (Li et al., 2017) suggest that different regulators may mediate the repression of MEF enhancers during successful reprogramming. Indeed, MEF enhancers are enriched for binding motif of Runx1, Tead, Nf1, and Erg ($p < 10^{-70}$; Figure 2F), and to a much lower extent to Oct4 or Sox2 ($p = 10^{-6}$). This begins to change from day 1, when active enhancers are significantly enriched ($p < 10^{-250}$) for Oct4 and Sox2 binding motif (Figure 2F). This includes late stage enhancers that are enriched ($p < 10^{-50}$) for other pluripotent transcription factors such as Prdm14 that are upregulated during the process.

It has been previously shown that Oct4 and Sox2 are pioneer factors that have the ability to bind closed chromatin and to activate new regulatory elements. Specifically, 70% of the enhancers bound by OSK after 48 hr of reprogramming are in a closed chromatin state in human fibroblasts (Soufi et al., 2015). In the highly efficient mouse system used here, out of the 4,858 enhancers that are bound by OSK on day 1, 74% were in a closed state (no mark) in MEF (fully consistent with findings reported [Soufi et al., 2015] in human reprogramming), and 4% were repressed by either H3K9me2 or H3K27me3. When we examined the binding motifs abundant in closed versus opened binding sites of OSK in day1 of Mbd3^{flox/-} system, the canonical TF binding motifs of OSK were detected after 1 day of DOX induction in Mbd3^{flox/-} cells in both closed and open regions

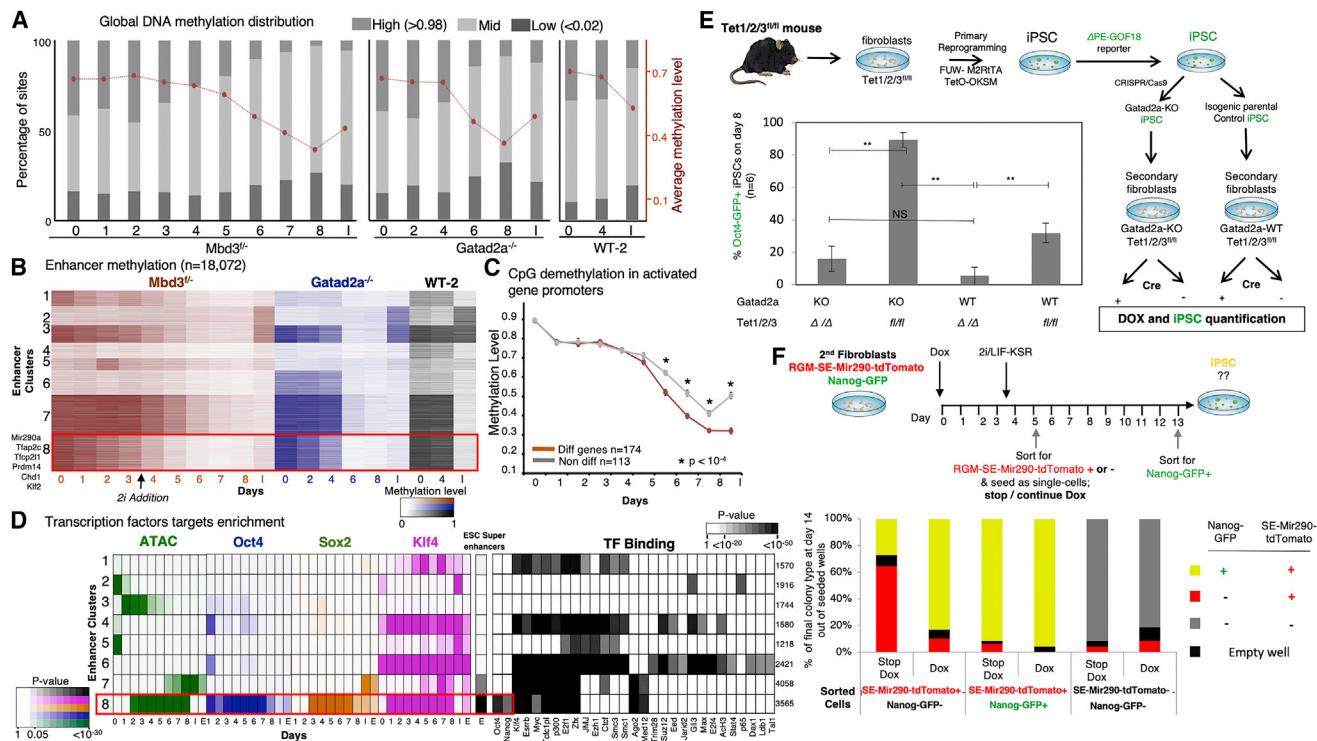


Figure 3. Rapid DNA Demethylation of Naive ESC Super-Enhancers during Conducive iPSC Reprogramming

- (A) Distribution of low (<0.02), mid (0.02–0.98), and high (>0.98) methylated CpG sites along reprogramming. Average and SEM are indicated in red plot.
- (B) Methylation level measured in covered enhancers ($n = 18,072$), in Mbd3^{fl/fl}, Gatad2a^{-/-}, and WT-2 systems. Enhancers are clustered into eight clusters using k-means. Cluster 8 consists of enhancers that undergo fast demethylation, compared to clusters 3 and 7.
- (C) Average methylation measured in promoters of genes that were highly methylated (>80%) in day 0. Genes that change their expression level (red) are compared to genes that do not change their expression level (gray). Wilcoxon p value indicates places where methylation of differential genes is significantly lower than methylation of non-differential genes.
- (D) Left: enrichment of enhancer clusters, as shown in (B), for OSK binding, DNA accessibility, and super-enhancers showing that cluster 8 is highly enriched for OSK binding and overlaps with super-enhancers. Color shades represent FDR-corrected enrichment p value. Right: enrichment of the same enhancer clusters to transcription factor binding taken from hmChIP database. Cluster size is indicated on the right.
- (E) Experimental scheme summary. Reprogramming efficiency was measured by Oct4-GFP⁺ cells percentage in Tet1/2/3 null(Δ) and Tet1/2/3^{fl/fl} with and without Gatad2a expression, after 8 days. * $p < 0.01$, ** $p < 0.001$ (Student's t test), $n = 6$, error bars indicate SD.
- (F) Secondary MEF harboring Mir290-RGM and Nanog GFP-reporter were sorted after reprogramming to 3 different populations: RGM-SE-Mir290-tdTomato-positive cells (sorted at day 5), Nanog-GFP- and Mir290-RGM-positive cells (sorted at days 10–14), and “double negative” cells (sorted at day 5). The cells were seeded as single cell-per-well and were treated with medium either supplemented with Dox or lacking Dox. On day 14 following reseeding, colonies were inspected for GFP and mCherry (RGM) markers.

(Figure 2G). The latter is entirely consistent with a pioneer TF activity for OSK in both mouse and human (Soufi et al., 2015) and contradicts the previous claim for such differences between these two species (Chronis et al., 2017).

Early Enhancer Demethylation Marks Commissioning of Conducive Reprogramming

We observed a global reduction in DNA methylation (Figures 3A and S3A), which reaches its lowest level at day 8. We clustered enhancers based on their methylation levels (Figure 3B). All 8 clusters showed different variations of progression exclusively entailing loss of DNA methylation, and none of the clusters showed continuous increase in methylation levels during the 8 days of reprogramming. The latter indicate that *de novo* DNA methylation is neither required for highly efficient and conducive iPSC reprogramming nor for repression of somatic lineage genes in naive pluripotency reprogramming conditions (Lee et al., 2014; Polo et al., 2012).

We wanted to test whether different rates of demethylation exist for certain gene groups. We considered genes that are methylated in MEF (>80% methylation) and compared those that change their expression (fold change [FC] > 4) to those that do not change their expression (FC < 1.5; Figure 3C). We found that genes that are upregulated at some point during the process undergo significantly faster demethylation compared to the non-changing genes, starting from day 6 following OSKM induction. When examining the methylation of enhancers (Figure 3B), we identified one cluster (number 8), which is 68% methylated in MEF, and then undergoes fast demethylation (with average 43% methylation level on day 3), even before introducing 2i at day 3.5. The enhancers in this cluster are accessible between days 2 and 8, are highly enriched for the binding of OSK (Figure 3D), and highly overlap ($p < 10^{-43}$) with ESC super-enhancers including Mir290, Tffap2c, and Prdm14, which are known to boost iPS efficiency (Table S5). Another cluster of

enhancers (cluster 7) is enriched for SK binding and for ESC super-enhancers, but it undergoes a slightly slower demethylation than cluster 8 (with average methylation of 67% in day 3), and its enhancers are accessible as measured by ATAC-seq only from day 6 until iPS and/or ES (Figure 3D). Both clusters are enriched for binding by Esrrb, E2f1, and Klf4, but only cluster number 8 is enriched for Nanog and Oct4 binding.

We next aimed to unravel the mechanism underlying these different demethylation rates in our system and whether this early demethylation is important for achieving efficient reprogramming. Given that this demethylation occurs before introducing 2i, we suspected that Tet enzymes, known to target and demethylate key pluripotency genes in ESC, might regulate this change. To test this, we established Tet1/2/3 triple-floxed conditional knockout mouse model, from which we derived secondary iPSCs, generated isogenic *Gata2a*^{-/-} iPSC lines with CRISPR and Cas9, and subsequently re-isolated DOX inducible reprogrammable MEFs (Figure 3E). Depletion of Tet enzymes in *Gata2a*-WT decreased iPSC efficiency (from 32% to 6%). However, upon ablation of Tet enzymes in the *Gata2a*^{-/-} deterministic reprogramming system, reprogramming efficiency dropped from 93% down to 6%–18%, similar to that in WT system, thus abolishing the beneficial effect of *Gata2a* depletion (Figure 3E). The latter indicates that Tet activity early in reprogramming is essential for highly efficient conducive iPSC reprogramming in NuRD-depleted systems. Knockdown of *Tfap2c* or *Tfcp2l1* reduced reprogramming efficiency by 15% (Figure S3C). The latter suggests that early demethylation of selected enhancers by Tet enzymes promotes the commissioning of several pro-reprogramming factors that synergistically contribute to highly efficient reprogramming.

We tested whether the rapid demethylation of cluster 8 super-enhancers specifically detected during deterministic iPSC reprogramming, but not in bulk WT reprogramming samples (Figure 3B), can be used as an early marker to prospectively enrich for the rare correctly commissioned WT cells to become iPSCs. To isolate cells in real time during reprogramming based on their DNA methylation status of a certain locus and at the single-cell level, we utilized a recently generated reporter system for endogenous genomic DNA methylation (RGM) (Stelzer et al., 2015). We chose a validated RGM construct for Mir290 enhancer encoding tdTomato (RGM-SE-miR290-tdTomato), which was enriched in cluster 8, and introduced it in two OKSM DOX-inducible reprogramming systems carrying Nanog-GFP reporter (Figure 3F). In these inefficient WT systems, the first Nanog-GFP⁺ cells appeared at days 10–14 following DOX induction, which were sorted and plated as single cells in naive ESC media with or without continued DOX. As expected, over 90% iPSC efficiency was obtained following sorting Nanog-GFP⁺ cells irrespective to the continued use of DOX to induce transgenes after sorting (Figure 3F), confirming that Nanog-GFP⁺ cells are already bona fide committed iPSCs that no longer need OSKM transgene expression. On the contrary, SE-miR290-tdTomato⁺ cells appeared at very low frequency already at day 4 during reprogramming of Mbd3/*Gata2a*-WT cells as single-positive tdTomato⁺ cells (Figure S3B). tdTomato^{+/GFP-} cells at day 5 were sorted and plated as single cells in naive ESC media with or without continued DOX treatment. Remarkably, >85% iPSC efficacy, as measured by Nanog-GFP, was obtained from day 5 sorted tdTomato^{+/GFP-}

cells only upon continued DOX supplementation (Figure 3F), compared to <10% iPSC efficiency without sorting. In the absence of continued DOX, 26% efficiency was obtained from early tdTomato^{+/GFP-}-sorted cells, suggesting that the sorted tdTomato^{+/GFP-} cells are not bona fide iPSCs, however, they were correctly “commissioned” and become committed to becoming iPSCs if OSKM expression is continuously delivered to drive the process toward completion. Day 5 double-negative-sorted cells did not yield any iPSCs after 10 days of DOX induction, indicating that this fraction marks somatic cells that did not optimally embark on a conducive trajectory toward becoming iPSCs. These results indicate that early demethylation of Mir290 super-enhancer marks correctly commissioned NuRD-WT somatic cells following DOX induction, that rapidly assume a conducive trajectory to becoming iPSCs if OKSM induction is continued. This also provides a means for early prospective isolation of adequately commissioned somatic cells for a successful reprogramming trajectory based on endogenous epigenetic feature.

Two Synergistic and Distinctly Regulated Gene Programs Ignite Deterministic Reprogramming

We next wanted to characterize the epigenetic changes and examine their connection to the changes in gene expression. For each differentially expressed gene that showed a significant epigenetic modification in its promoter ($n = 7,801$), we calculated the correlation between its transcriptional temporal pattern and chromatin modification patterns, measured around the transcription start site (TSS) or transcription end site (TES) across all time points. When we cluster these genes and chromatin marks (Figure 4A), we observed that chromatin marks separate into two clusters. One consists of marks that are positively correlated to gene expression, and are indeed known to be associated with active transcription, such as H3K4me3, H3K27Ac, H3K36me3 (in TES), and chromatin accessibility. The other consists of marks that negatively correlate with gene expression and are known repression-associated marks, such as H3K27me3 and H3K9me3. Interestingly, the genes also separate into two main clusters. One consists of genes that display high correlation (positive or negative) between expression and chromatin modifications and the other consists of genes that are not correlated, despite the fact that the genes are differentially expressed. Notably, each of these two gene groups contains both induced and repressed genes (Figures S4A and S4B). We inspected the actual transcriptional and epigenetic patterns for these two gene clusters, focusing on H3K27ac and H3K27me3 marks, which showed the highest positive and negative correlation to transcription (Figure 4B). The genes in the first group showed a clear switch-like on-off behavior between the epigenetic marks (Figures 4B and S4A), correlated with the activation or repression. We therefore concluded that these are genes with epigenetically switched promoters (abbreviated as ESPGs) (Table S6). In the second group, the majority of the genes ($N = 3,049$, 72%) had differential transcription (above 4-fold change), but with consistently high levels of H3K27ac and low levels of H3K27me3 (Z score < 0.7; Figures 4B and S4B). The promoters of these genes show a constitutive active chromatin signature, suggesting that these genes are regulated by distinct mechanisms. We refer to this group as CAPGs (constitutively active

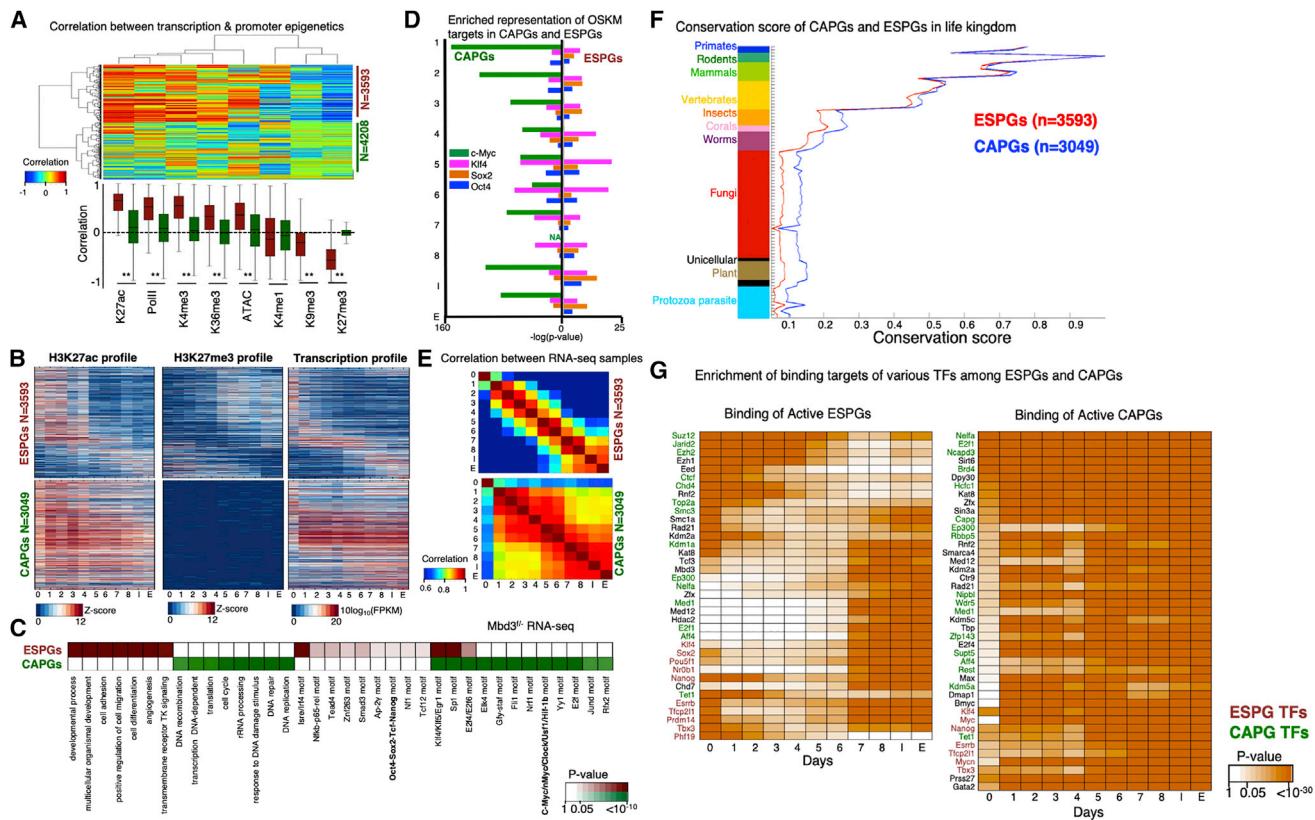


Figure 4. Distinct Regulation of Cell Fate Genes and of Biosynthetic Processes by MYC

(A) Spearman correlation between the expression change of each differential gene (row) and change in promoter chromatin modification of each indicated mark (column). Analysis was done on all differential genes that have at least two marks in their promoter (i.e., with Z score >1 SD), resulting in 7,801 genes. Hierarchical clustering clustered the genes into two distinct groups: genes with correlation between gene expression and promoter epigenetic modifications ($n = 3,593$) and genes with no trend of correlation ($n = 4,208$). Top: clustered correlation matrix. Bottom: correlation distribution of each mark in the two gene groups, where red denote correlated group, and green denote the non-correlated group. **Wilcoxon p value $< 10^{-170}$.

(B) Pattern of H3K27ac, H3K27me3, and expression in epigenetically switched genes (ESPGs, top), and in genes with consistency-active-promoters (CAPGs, bottom). Each row corresponds to a single gene, genes are sorted according to their expression pattern and the same sorting was applied to the epigenetic marks.

(C) Enrichment of GO categories and binding motifs in ESPGs (red) and CAPGs ($n = 3,049$, green). Color shades indicate FDR corrected Fisher exact test p value or motif enrichment p value.

(D) Enrichment of OSKM binding targets in promoters of ESPGs compared to CAPGs. Minus \log_{10} of Fisher exact test p values are indicated.

(E) Spearman correlation matrix between $Mbd3^{fl}$ RNA-seq samples calculated over ESPGs (top), showing a gradual change along reprogramming and over CAPGs (bottom), showing two waves of change, on day 1 and on days 5–6.

(F) Conservation score of ESPGs and CAPGs calculated with PhylogeNE software. Graph includes the mean and SEM values for each gene set (one-tailed Wilcoxon test, p value of non-vertebrate organisms $< 10^{-30}$).

(G) Scheme showing transcription factors that significantly bind ESPGs or CAPGs promoters that are active in each day of reprogramming, based on ChIP-seq databases (FPKM > 0.5 of the gene max value). Red, TFs with ESPG pattern; green, TFs with CAPG pattern; orange shades, FDR-corrected p values (Fisher exact test).

promoter genes) (Table S6). In accordance with chromatin modifications, DNA methylation in the promoters of the two groups is different (Figure S4C): CAPGs show a consistent hypomethylation, regardless of their transcriptional pattern, whereas ESPGs, which are regulated on the chromatin level, are also regulated by DNA methylation. Chromatin state in enhancers is reminiscent of that on promoters (Figure S4D).

Inspecting the functional enrichment of the two groups, we found a specific association of ESPGs to cell fate determination processes, indicating that epigenetic regulation is highly specific for cell fate genes (Figure 4C). CAPGs are enriched for biosynthetic pathways, including DNA synthesis, proliferation, DNA

repair, and chromatin reorganization (Figure 4C). The two programs show a distinct conservation pattern during the evolution of vertebrate organisms: while in vertebrates CAPGs and ESPGs are conserved in a similar degree (Figure 4F), in fungi and other non-vertebrates CAPGs are more conserved than ESPGs ($p < 10^{-30}$), emphasizing their basic role in cellular maintenance. The two groups also show distinct regulation by c-Myc: CAPGs, but not ESPGs, are significantly bound by c-Myc (sample median $p < 10^{-75}$ for CAPGs, sample median $p > 0.9$ for ESPGs) (Figure 4D). This is supported also by over-representation of c-Myc motif only in CAPG promoters (Figure 4C). Additional TF binding motifs show enrichment specific to one group and not

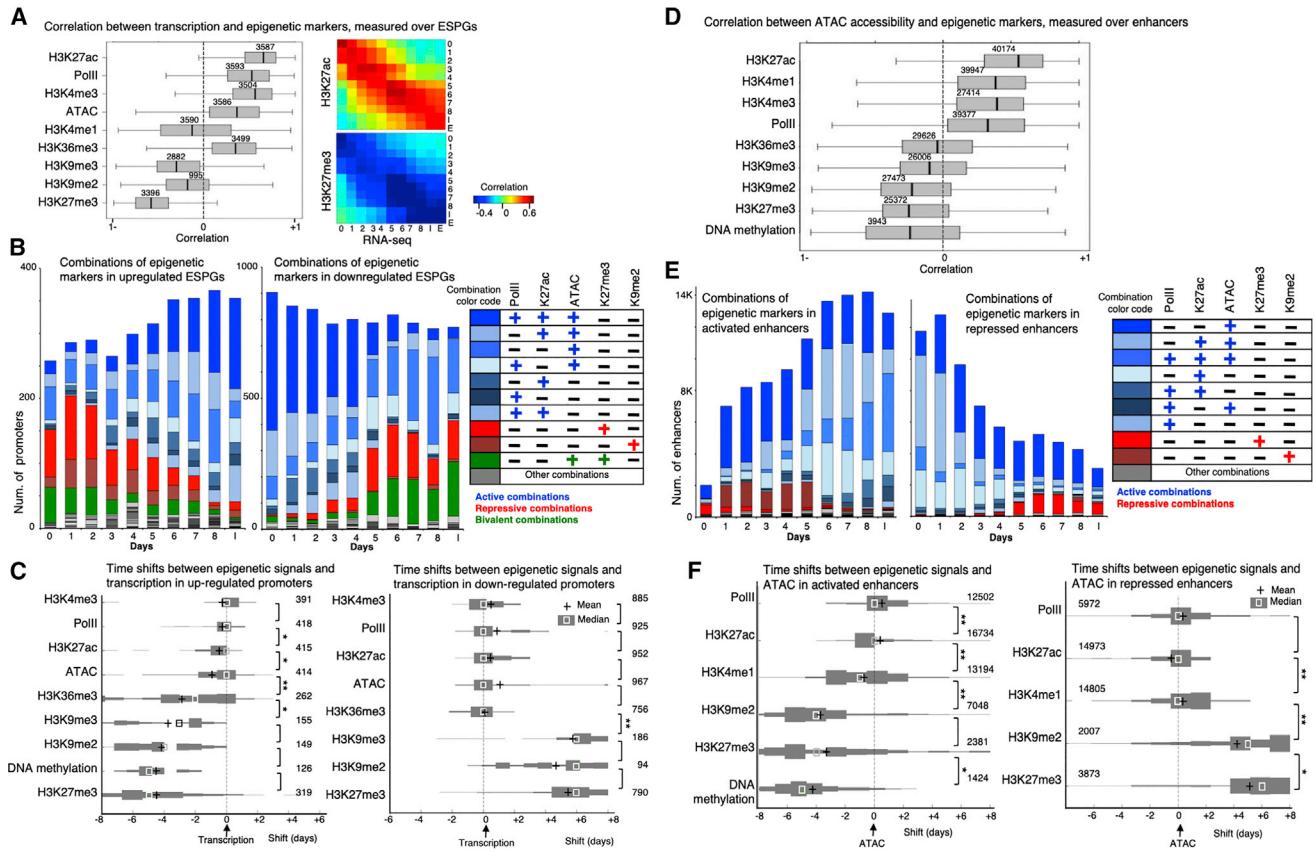


Figure 5. Mapping the Order of Epigenetic Events that Drive Transcription Initiation and Repression

- (A) Left: correlation between each of the indicated chromatin modifications and gene expression patterns that were measured in the promoters of ESPGs that have the modifications (numbers are indicated). Correlations were calculated for each gene over 11 time points. H3K27ac, Pol II, and H3K4me3 show positive correlation with gene expression, while H3K27me3 and, to lesser extent, H3K9me2,3 show negative correlation with expression. Right: Spearman correlation matrix over all ESPGs, between H3K27ac or H3K27me3 and gene expression (RPKM) for each sample separately.
- (B) Stacked bar chart of all combinations of the indicated chromatin modifications, as measured in promoters of upregulated ESPGs (left) and downregulated ESPGs (right). Right, color code of frequent combinations (>3% of the promoters across samples).
- (C) Distribution of time shifts between each of the indicated epigenetic modification and gene expression profile, measured using cross-correlation. The distribution, presented as histogram, was measured over upregulated ESPGs (left) and downregulated ESPGs (right) that have a changing epigenetic modification (max-min Z score > 0.5). The number of promoters tested is indicated. x axis indicates the temporal time shift in days. Plus indicates mean, and square indicates median. * $p < 10^{-5}$, ** $p < 10^{-25}$ (Wilcoxon test). p values measure significant time difference between consecutive epigenetic events. Negative shift, preceding event; positive shift, lagging event.
- (D) Correlation between each of the indicated chromatin modification and DNA accessibility patterns that were measured in all differential enhancers that have the modifications (numbers are indicated). Correlations were calculated over 11 time points.
- (E) Stacked bar chart of all combinations of the indicated chromatin modifications, as measured in activated enhancers (left) and in repressed enhancers (right). Right, color code of frequent combinations (>3% of the enhancers across samples).
- (F) Distribution of time shifts between each of the indicated epigenetic modification and accessibility profile, measured using cross-correlation. The distribution, presented as histogram, was measured over activated enhancers (left) and repressed enhancers (right) that have a changing epigenetic modification (max-min Z score > 0.5). The number of enhancers tested is indicated. x axis indicates the temporal time shift in days. Plus indicates mean, and square indicates median. * $p < 10^{-4}$, ** $p < 10^{-50}$ (Wilcoxon test).

the other (Figure 4C), further supporting a model of separate regulation. Finally, the two groups have different temporal behavior: while ESPGs have a gradual change in activity along reprogramming, CAPGs converge to their final activity pattern as early as day 1 (Figures 4B and 4E). Importantly, however, these two programs retain a coupled and cross-coordinated regulation. Protein binding enrichment in ESPGs and CAPGs using public protein-DNA databases shows a number of proteins that are associated with one of the groups but bind the opposite group. Several epigenetic modifying components such as Poly-

comb and Wdr5 show a constitutively active promoter configuration, but regulate ESPGs (Figure 4G).

Two Divergent Modes of Epigenetic Repression of ESPGs

We next sought to discern epigenetic regulation during iPSC formation, and from the differentially expressed genes, we focused on ESPGs as they are the ones that undergo a repressive to active switch or vice versa. We used the chromatin modification coverage in promoters and the RNA expression level and

calculated the temporal correlation distribution for all ESPGs (Figure 5A), i.e., correlation that is calculated for each gene over all time points. RNA-Pol II, H3K27ac, and H3K4me3 in promoters are highly correlated to gene expression of the genes they decorate (median $r = 0.55$, 0.7, and 0.6, respectively). H3K27me3, H3K9me2, and H3K9me3 show negative correlation to gene expression. Examining the frequencies of combinations of chromatin modifications (Figure 5B) on gene promoters, we observed that in upregulated ESPGs ($n = 431$), there is a rapid reduction of H3K9me2 and H3K27me3. In addition, there is a substantial increase in H3K27ac and binding of Pol II, such that by day 8 and iPSC, 45% of the promoters are decorated by the combination of H3K27ac and Pol II. In the downregulated ESPGs ($n = 974$), we observe the opposite pattern with loss of H3K27ac and Pol II binding and gain of H3K27me3 starting from day 5 (Figure 5B).

The analysis above also highlights the most frequent combinations and the combinations that are not apparent in the data and are mutually exclusive. The latter allowed us to ask whether mutually exclusive modes of repression exist in iPSC reprogramming. Active marks (H3K27ac, RNA-Pol II, and ATAC) tend to appear together on promoters (Figure 5B), and we did not discern distinct mutually exclusive modes of acquiring activation marks. On the contrary, repressive marks (H3K27me3 and H3K9me2) work separately from one another. We observed that <1% of the promoters are marked by both H3K9me2 and H3K27me3, suggesting these are mutually exclusive marks. Indeed, our data show a clear association between H3K9me2 and DNA methylation (Figures S5A and S5E), and this may explain why in our system, which undergoes substantial DNA demethylation, there is a limited gain of H3K9me2 on downregulated ESPGs. Furthermore, H3K27me3 decorates genes that are enriched for functions in development, while H3K9me2 decorates genes related to signaling pathways (Figure S5B). H3K27me3 genes are naturally highly enriched for Polycomb targets (Figure S5B), and an induction in the expression of Polycomb members is observed, which overlaps with the increase in H3K27me3 peaks starting from day 5 (Figure 5B). Altogether, this analysis uncovers two divergent modes of epigenetic repression by H3K27me3 and H3K9me2 during iPSC reprogramming with opposing association with DNA methylation and distinct associated regulatory functions.

Bivalent promoters, which carry both H3K27me3 repressive mark and H3K4me3 active mark, constituted 38% of ESPGs and were found to constitute a third distinct mode of repression (Figures 5B and S5F). Bivalent promoters are highly enriched for developmental regulators (Fisher exact test, $p < 10^{-90}$, false discovery rate [FDR] corrected), and overlap with bivalent promoters that were detected previously in MEFs and in ESCs (Figure S5B). When comparing bivalent promoters to H3K27me3-only promoters, we observe that the repression of transcription is stronger in H3K27me3-only promoters than in bivalent ones (one-tailed Wilcoxon test, $p < 10^{-98}$). Moreover, the chromatin of bivalent promoters is much more accessible compared to H3K27me3 only or to H3K9me2 promoters, which decorate closed chromatin (Figures S5C–S5F). To rule out the possibility that bivalent signature is a mere result of a residual mixed cell population in the highly efficient system, we note that other combinations that are mutually exclusive in promoters,

such as H3K27me3 and H3K27ac, appear in much lower frequency (<3%) compared to the bivalent combinations (>25%).

Repressive and Active Chromatin Mark Switching Are Temporally Separated over ESPGs

The temporal interplay between the different chromatin marks remains to be defined at high-resolution, and it is unclear whether during transitions from repressed to activated state changes in repressive and activating epigenetic modes co-occur simultaneously or are well separated. Because our data consist of time series, we used cross-correlation, a signal-processing algorithm widely used to detect and quantify the temporal offset between signals (Kiviet et al., 2014) (Figure S6), to test whether the deposition of these modifications has a temporal order. We estimated the distribution of offsets across ESPGs (Figure 5C) (i.e., for each ESPG we calculate the cross-correlation between its temporal expression and the temporal pattern of each of the epigenetic marks) across MEF to iPSC samples. The analysis clearly highlighted separation between accumulation of repressive and activation marks at gene promoters. In induced genes, first, DNA is demethylated and H3K27me3 is removed, and only then chromatin becomes accessible. Finally, H3K27ac, RNA-Pol II, and H3K4me3 accumulate in the promoter in a close proximity to transcriptional activation (Figure 5C). The latter also excludes an alternative scenario wherein gene activation, removal of repressive marks follows epigenetic activation and transcription initiation (Figure 5C). In repressed genes, PolII dissociates from its bound promoters in close proximity to the eviction of H3K27ac and H3K4me3 and chromatin closure (Figure 5C). Only afterward, repressive marks like H3K27me3 and H3K9me3 gradually accumulate during the following days.

Our next aim was to elucidate the temporal order of epigenetic changes that occur in differential enhancers and how they compare to those observed in promoters. 43% of all annotated enhancers ($n = 40,174$) showed differential ATAC-seq and H3K27ac signals in both *Mbd3^{f/-}* and *Gata2a^{-/-}* systems and were identified as differential enhancers (Figure S2A). The enhancer activation kinetics in the two NuRD-depleted systems were highly consistent and faster than the WT systems (Figures S2B–S2D). We calculated the correlation between chromatin accessibility and chromatin modification in each of the differential enhancers (Figure 5D), and observed positive correlation with H3K27ac modification (median $r = 0.55$). Interestingly, positive correlation was evident between enhancer accessibility and RNA-Pol II binding. Furthermore, we observed negative correlation between enhancer accessibility and DNA-methylation, H3K27me3 and H3K9me2, but to less extent with H3K9me3 modification. H3K27me3 does not always decorate repressed enhancers. In fact, when all possible combinations of chromatin marks are inspected in differential enhancers (Figure 5E), 85% of the enhancers that are active in day 8 are in a closed chromatin state on day 0 (MEF) but are not marked by any of the histone marks measured herein. Like in promoters, H3K9me2 repression can be observed in the first days of reprogramming, is later depleted, and is mutually exclusive to H3K27me3 (Figure 5E). Unlike its abundance on promoters during reprogramming, bivalency at enhancers (H3K27me3 with H3K4me1) is rare, and H3K27me3 is rarely deposited on accessible enhancers (<4%; Figures 5E and S5E).

To examine the sequence of epigenetic events during enhancer activation and suppression, we used cross-correlation and quantified the temporal offset between chromatin changes and DNA accessibility in each differential enhancer. We found that in activated enhancers ($n = 17,174$), H3K27me3 is first removed, then H3K4me1 is deposited, followed by chromatin accessibility and deposition of H3K27ac and finally, by binding of RNA-Pol II (Figure 5F). In repressed enhancers, Pol II release and the removal of H3K27ac and H3K4me1 happen all in close proximity to chromatin closure timing, followed by gradual deposition of H3K27me3 or H3K9me2. Thus, overall, the orderly switches from activation to repression (or vice versa) over enhancers are similar to those seen over promoters (Figures 5C and 5F). Cross-correlation was used to quantify the temporal order of epigenetic changes in enhancers and promoters in relation to measured transcription changes (Figure S7A). No significant temporal differences were observed in deposition or removal of repressive chromatin marks between enhancers and promoters during repression or activation of ESPGs, respectively (Figure S7A). However, we could see that active modifications are deposited on enhancers before they are deposited on the associated promoters during gene activation (paired sample t test, ATAC-seq, $p < 10^{-7}$; H3K27ac, $p < 10^{-11}$; H3K4me3, $p < 10^{-2}$, respectively). In contrast, during ESPG repression, eviction of activation marks on enhancers was significantly lagging in comparison to promoters (Figure S7A). Unexpectedly, RNA-Pol II binds enhancers and showed similar behavior to the activating epigenetic marks (Pol II binds enhancers slightly before it binds to promoters [$p < 10^{-3}$; Figures S7A–S7C] and leaves the enhancers slightly after it leaves the promoters [$p < 10^{-23}$]). RNA-Pol II binding in enhancers is highly correlated both to gene transcription (Figure S7D) and to enhancer activity (Figure 5D). Independent RNA-Pol II binding data, measured in mouse ESC (Rahl et al., 2010), was also highly enriched among enhancers that are active in late reprogramming stage ($p \leq 10^{-200}$; Figure S7B). These results indicate that the phenomenon of Pol II recruitment to enhancers as an early event of enhancer commissioning is widely abundant during iPSC reprogramming.

Myc Activity Is Essential for iPSC Reprogramming

CAPGs are predominantly regulated by Myc and drive cellular biosynthetic processes. As exogenous Myc is dispensable for iPSC formation from WT and NuRD-depleted somatic cells (Nakagawa et al., 2008), this raised the possibility that the observed CAPGs induction is a side effect of c-Myc overexpression and is not essential for the reprogramming process. To test this, we introduced perturbations to the highly efficient optimally NuRD-depleted reprogramming protocols. First, we tested reprogramming with a viral induction of only 3 factors OSK (Figure 6Ai). Notably, CAPGs that were upregulated in the original protocol, were still significantly upregulated compared to MEF ($p < 10^{-12}$; Figure 6B). However, we noticed that in OSK reprogramming, endogenous c-Myc continues to be highly expressed and endogenous n-Myc is induced after OSK induction ($FC > 1.8$, for both c-Myc and n-Myc). We tested OSK reprogramming under inhibition of endogenous Myc family members by treating MEFs that carry OSK cassette with siRNAs for c-Myc, n-Myc, and l-Myc starting on day –3 prior to DOX induction (Figure 6Aii). Myc inhibition resulted in dramatic reduction in reprog-

ammed colonies (Figures 6C and 6D). The downregulation and upregulation of ESPGs was also diminished by Myc inhibition (Figure 6E), although Myc does not bind them directly (Figure 4), suggesting that this change is caused indirectly. We used conditional knockout fibroblasts for both c-Myc and n-Myc genes and carrying Lox-stop-Lox-YFP reporter in the Rosa26 locus that can mark floxed cells upon Cre treatment. Fibroblasts were treated with CAGGS-Cre plasmid, sorted for YFP, and subjected to either OSK or OSKM transduction (Figure 6F). Remarkably, we could not obtain any YFP⁺ iPSC colonies following OSK induction and follow-up of over 30 days of reprogramming from Cre-treated cells (Figure 6G). Following applying MYC inhibition during the first 4 days of reprogramming by in secondary Mbd3^{fl/fl} cells, we noted that downregulation of somatic marker Thy1, one of the earliest events in MEF reprogramming, is abrogated (Figure 6H). These findings show that there is no initiation of reprogramming process in the absence of Myc activity and reveals an early critical role for MYC in conducive iPSC formation. Inhibition of MYC activity abolished highly efficient B cell reprogramming by C/EBPα+OSKM, mouse common myeloid progenitors, and human iPS reprogramming by OSK (Figure 6J). The latter findings are consistent with the high similarity and convergence in gene expression and accessibility changes found in our NuRD-depleted MEF systems and that in highly efficient B cell to iPSC reprogramming (Figures 2H and S2E–S2G).

Molecularly, c-Myc overexpression (OE) in MEF, without the induction of other reprogramming factors, induced CAPGs expression changes in the same way it changes during reprogramming by OSKM (Figure 6B), also causing significant repression of downregulated ESPGs (somatic genes), but did not lead to the induction of upregulated ESPGs (pluripotency genes). We further validated Myc-induced CAPG changes by looking at specific functional groups of genes: genes related to cell biosynthesis, which are bound by c-Myc (Figures 6K and 6L), are induced upon overexpression of c-Myc. These expression changes are consistent with previously published data (Scognamiglio et al., 2016) of Myc inhibition and reconstitution measured independently during naive mouse ESC maintenance (Figure 6M). Interestingly, we observed that reprogramming related chromatin modifiers such as Prc2 members, Tet1, and Wdr5 are induced by the mere OE of c-Myc and fail to be induced upon its inhibition (Figure 6N). This indicates that Myc has a critical role in igniting the biosynthetic pathways that are dispensable for pluripotency maintenance (Scognamiglio et al., 2016), yet essential for reestablishing pluripotency in somatic cells and must be provided either endogenously or exogenously.

Rapid Rewiring of tRNA Pool Boosts Myc-Dominated CAPG

The rapid change in CAPGs expression, without associated changes in their epigenetic signature, raised the possibility that CAPGs may be differently regulated. A recent study (Gingold et al., 2014) documented a cancer-promoting mechanism that supports loss of somatic identity and acquisition of a highly active metabolic state during cancer transformation involving coordinated changes in the tRNA pool and the codon usage preference of tRNA. We thus examined if such shifts occur at the codon usage level of the transcriptome and at tRNA transcription status when somatic cells undergo reprogramming toward

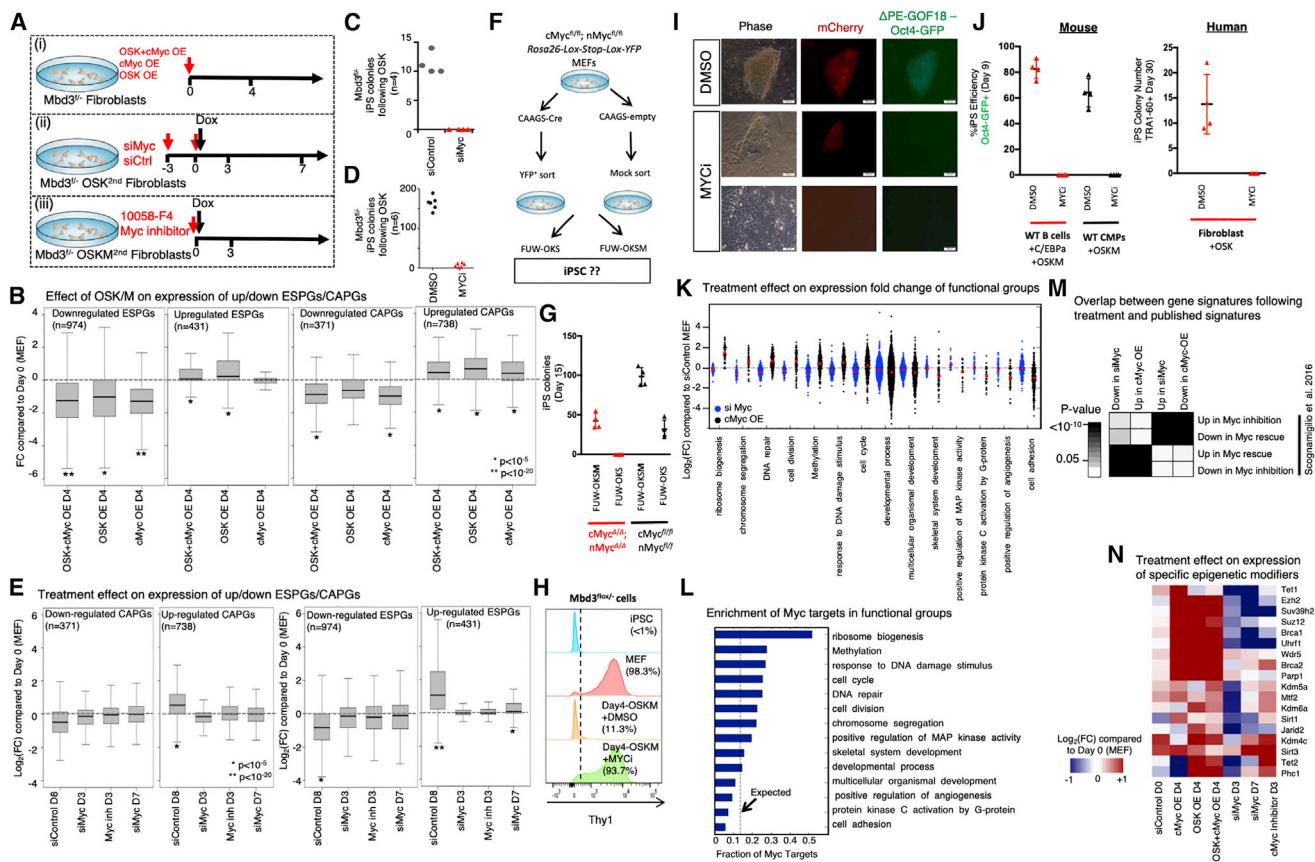


Figure 6. Biosynthetic Processes Are Regulated by Myc Activity and Are Essential for Reprogramming

(A) Experimental flow describing three experimental perturbation settings: (i) Mbd3^{f/f} MEFs were virally infected with cMyc overexpression (OE) cassette, OSK-OE cassette, or both cassettes. Gene expression was measured on day 4 following infection. (ii) Mbd3^{f/f} MEFs carrying OSK Dox-dependent cassette were treated for knockdown of c-Myc, n-Myc, and l-Myc. Gene expression was measured on days 3 and 7, and colony formation was measured on day 11. (iii) Mbd3^{f/f} MEFs carrying OSKM Dox-dependent cassette were treated with inhibitor of cMyc (10058-F4) and with Dox. Gene expression and colony formation were measured on day 3.

(B) Distribution of expression fold change (FC) compared to WT MEF of up or downregulated ESPGs (downregulated ESPGs are enriched for somatic genes) and CAPGs. Presented perturbations are overexpression of OSK cassette, overexpression of c-Myc cassette, or overexpression of the two cassettes together. *p < 10⁻⁵, **p < 10⁻²⁰, Wilcoxon test.

(C and D) Reprogrammed colony formation in Myc knockdown (C) or small molecule inhibition (D), measured 11–14 days after Dox.

(E) Distribution of expression fold change (FC, in log₂ scale) compared to MEF of up- and/or downregulated ESPGs and CAPGs. Presented perturbations are Myc knockdown, inhibition of Myc activity with small molecular inhibitor (10058-F4). *p < 10⁻⁵, **p < 10⁻²⁰, Wilcoxon test.

(F) Experimental scheme.

(G) iPSC reprogramming efficiency in different cells expressing both endogenous and/or exogenous cMyc and nMyc.

(H) FACS analysis for surface expression of fibroblast surface marker Thy1 on the indicated Mbd3^{f/f} cell types. Dotted line indicates positive threshold for detection.

(I) Representative pictures of Mbd3^{f/f} cells harboring mCherry-NLS and ΔPE-GOF18 Oct4-GFP cassettes after 13 days of reprogramming in the presence of MYCi. Scale, 100 μM.

(J) Left: iPSC reprogramming efficiency by applying highly efficient mouse B cell and WT CMP reprogramming protocols by OSKM in the presence or absence of MYC small molecule inhibitor (MYCi). Right: human iPSC reprogramming efficiency by applying OKS lentiviral transduction in the presence or absence of MYCi.

(K) Expression fold-change distribution (log₂ scale) of selected GO categories in Myc overexpression or Myc knockdown showing that upon overexpression of Myc, processes such as ribosomal biogenesis and chromosome segregation are induced.

(L) Fraction of Myc targets in significantly induced and repressed GO categories, compared to what is expected by random (dashed line).

(M) Overlap between differential genes detected in Myc perturbation experiments, and differential genes detected in previous published perturbations (Scognamiglio et al., 2016). Fisher exact test p values are presented.

(N) Expression fold change of selected chromatin modifiers.

pluripotency. To characterize putative changes in the codon usage of the transcriptome, we calculated the average codon usage distribution of all differential genes in the four reprogramming systems. Using PCA, we characterized the codon combination that shows the highest variability during reprogramming (Figure 7A)

and noticed a change in codon combination that separates between early and late stages of reprogramming. The observed change in codon usage corresponds to a shift from G/C-ending codons to A/T-ending codons (Figure 7B), with the most prominent change occurring already at the first day of reprogramming

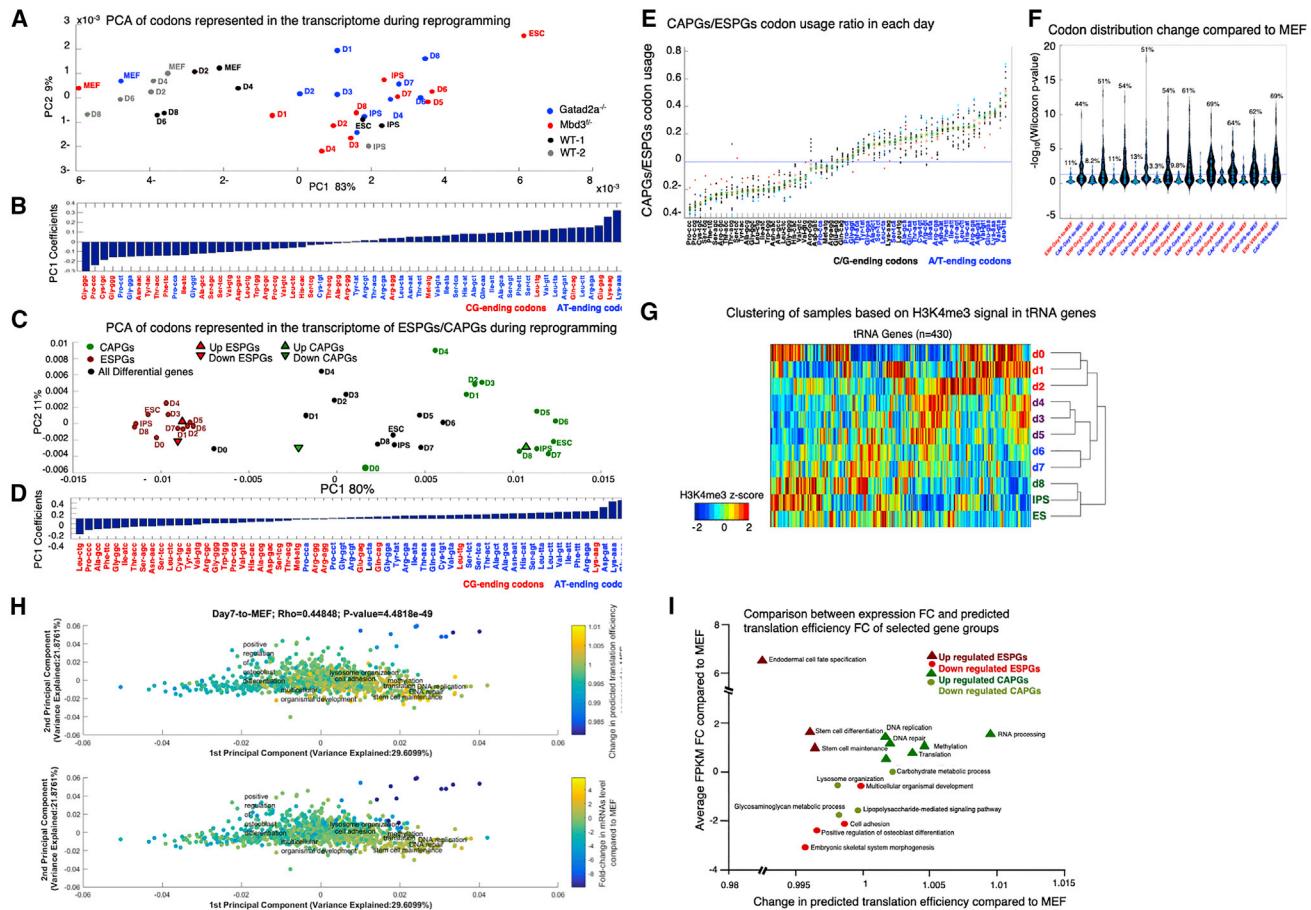


Figure 7. Biosynthetic Processes Are Coordinated by Optimization of Available tRNA Pool

(A) A PCA projection of codons' representation in the transcriptome along reprogramming in *Mbd3^{fl/fl}*, *Gata2a^{-/-}*, and the two WT systems. The representation of the codons in the transcriptome was determined by multiplying the number of occurrences of each codon in each gene by the scaled expression level of each gene in each time point per cell type. The variance percentage, out of the total original variance in the high-dimensional space, spanned by the first and second PCs is indicated on the x and y axis, respectively.

(B) Coefficients associated with first principle component of (A). Blue, A/T ending codons; red, G/C ending codons.

(C) A PCA projection of the codon usage of all differential genes (black), ESPGs (red), and CAPGs (green) show a striking separation between ESPGs and CAPGs.

(D) Coefficients associated with first principle component of (C). Blue, A/T ending codons; red, G/C ending codons.

(E) ESPGs prefer G/C-ending codons, while CAPGs use A/T-ending ones. Scatterplot analysis of the CAPGs-to-ESPGs ratio of codon usage. For each codon, shown are 11 calculated CAPGs-to-ESPGs ratios of codon usage along reprogramming and in ESCs. The mean and median values of each codon are shown as red crosses and green squares, respectively. Three time points are color-coded: red, CAPGs-to-ESPGs ratio in MEF; blue, CAPGs-to-ESPGs ratio in iPS; light blue, CAPGs-to-ESPGs ratio in ES.

(F) CAPGs change significantly their codon usage compared to MEF already in day 1 of reprogramming, where ESPGs are much less variable. Violin plot of the significance ($-\log_{10}[\text{Wilcoxon p value}]$) of the differences between the usage of each of the 61 sense codons in each day compared to MEF, calculated for ESPGs (red) and CAPGs (blue) sets (for each day separately).

(G) Hierarchical clustering of various time points and/or cell types based on the H3K4me3 signature in the vicinity of individual tRNA genes. ($n = 430$, Spearman correlation metric and average linkage were used).

(H) Projection of the tRNA translation efficiency and mRNA expression changes on the codon-usage map. PCA projection of GO category gene sets. The location of each gene set in this space is determined by the average codon usage of all the genes that belong to it. The % variance spanned by the first and second PCs is indicated on the x and y axis, respectively. The color code represents the predicted translation efficiency of tRNAs and expression change of mRNAs (top and bottom, respectively) in day 7 compared to MEF. Top: each gene category is color coded according to the relative change in the availability of the tRNAs that correspond to the codon usage of its constituent genes, averaged over all genes in the category; the tRNA availability of each individual gene was calculated similarly to the TAI measure of translation efficiency, where the expression of individual tRNA genes was evaluated by the H3K4me3 reads in its vicinity. A yellow color for a given gene category indicates that on average the genes in that category have codons that mainly correspond to the tRNAs that are induced in day 7, whereas a blue color indicates that the codon usage in the categories is biased toward the tRNAs that were repressed in day 7. Bottom: changes at the mRNA level, averaged over all the genes in each gene category as in the top, red means that the genes were induced in day 7.

(I) Comparison between predicted changes in translation and changes in transcription for various GO categories that show the highest association with up or downregulated CAPGs/ESPGs.

of NuRD-depleted, but not WT cells. We characterized the codon combination that shows the highest variability, for each subset of ESPGs, CAPGs, or total differential genes (Figures 7C and 7D). Surprisingly, the codon usage in ESPGs (red) and CAPGs (green) clustered at the lower and upper margins of the first principle component, respectively. The latter showed divergence in codon usage programs between the CAPGs and ESPGs: while ESPGs mainly tend to use codons that end with a G/C at the third codon position, CAPG genes split into two programs: the genes that are induced during reprogramming are encoded with A/T ending codons, while those that are repressed in the process mainly use G/C-ending codons (Figures 7C–7F). Interestingly, we did not see any significant change in codon usage when comparing different time points of ESPGs, but we do see a rapid and significant change in codon usage of CAPGs emerging already between day 0 and day 1 (Figure 7C) that underlies the global change observed during reprogramming.

The efficiency of translation elongation is determined by the relation between the supply of tRNAs and the demand for specific tRNA types, governed by the representation of the 61 sense codons in the transcriptome. We asked whether the changes in codon usage along reprogramming are accompanied with a co-ordinated change in the tRNA pool. We measured the chromatin mark H3K4me3 in the vicinity of the tRNA genes and observed a change in tRNA activation throughout reprogramming (Figure 7G). We next asked whether the change in the tRNA pool along reprogramming correspond to the observed change in the codon usage of the translated transcriptome and calculated the expected translational efficiency for genes belonging to the most highly enriched Gene Ontology (GO) categories corresponding to up or downregulated ESPGs and CAPGs based on their codon sequence and the tRNA epigenetic status. We observed a global significant positive correlation between the changes in transcription and translation, suggesting that the anticodons whose expression is elevated along reprogramming correspond to the codons that are enriched in the transcriptome of the respective cell state (Spearman $r = 0.45$, $p < 4.5e-49$; Figure 7H). However, while GO annotations that are associated with upregulated CAPGs showed an increase in predicted translation efficiency (Figure 7I), GO annotations associated with upregulated ESPGs show an opposite trend: a decrease in predicted translation efficiency, corresponding to their G/C-ending codon preference (Figures 7C and 7D). Thus, the CAPG program is responsible for biosynthetic processes and is optimally boosted by Myc and tRNA codon usage.

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
 - Cell Culture
- METHOD DETAILS
 - Generation of Gata2a-knockout Reprogrammable secondary MEF lines

- Generation of reprogrammable *Mbd3*^{flox/-} secondary MEF lines
- Mouse embryo micromanipulation
- Reprogramming of MEF to naive ground state naive iPSC
- Primary and secondary reprogrammable lines by viral infection
- Knockdown endogenous Myc during reprogramming
- Generation of triple *Tet1,2,3*^{flox/flox} mice and cell lines
- RT-PCR analysis
- AP Staining
- Imaging, quantifications, and statistical analysis
- ChIP-seq library preparation
- PolyA-RNA-seq library preparation
- Small RNA-seq library preparation
- ATAC-seq library preparation
- Whole-Genome Bisulfite Sequencing (WGBS) library preparation
- Reduced-Representation Bisulfite (RRBS) library preparation

● QUANTIFICATION AND STATISTICAL ANALYSIS

- ChIP-seq analysis
- RNA-seq analysis
- Analysis and integration of previously published datasets
- Extended Differential lncRNA analysis
- Phylogenetic analysis
- Functional Enrichment
- Protein-DNA binding enrichment analysis
- ATAC-seq analysis
- Identifying accessible chromatin regions
- Enhancer Identification
- Generating ATAC-seq normalized profiles in TSS and in enhancers
- Methylation Analysis of WGBS and RRBS
- Methylation estimation
- Combinatorial analysis for histone marks localization
- Motif analysis
- Motif analysis in open versus closed binding targets
- Translation Analysis

● DATA AND SOFTWARE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and six tables and can be found with this article online at <https://doi.org/10.1016/j.stem.2018.11.014>.

ACKNOWLEDGMENTS

J.H.H. is supported by a gift from Ilana and Pascal Mantoux and grants from European Research Council (ERC-2016-COG-726497; CellNaivety and ERC-2015-POC-692945; FORMAT), Flight Attendant Medical Research Institute (FAMRI), Israel Ministry of Science (IMOS), Israel Ministry of Health (IMO), Israel Science Foundation (ISF-ICORE, ISF-NFSC, and ISF-Morasha [also to N.N.]), Kamin-Yeda (IMOST), Minerva Stiftung, ICRF Research Professorship, BSF, Human Frontiers Science Program (HFSP), the Benoziyo Endowment Fund, Weizmann-U. Michigan research program, Weizmann-UK research program, Weizmann-EPFL research program, Benoziyo Endowment Fund, New York Stem Cell Foundation (NYSCF), Kekst Research Center, Shapell Foundation, Kimmel Innovator Research Award, and the Helen and Martin Kimmel Institute for Stem Cell Research. A.Z. is supported by the EMBO long-term

fellowship (ALTF 140-2016). Y.P. is supported by ISF and ERC-CoG research grants. J.H.H. is a NYSCF-Robertson Investigator. A. Trumpp is supported by the SFB873 funded by the Deutsche Forschungsgemeinschaft (DFG) and the Dietmar Hopp Foundation.

AUTHOR CONTRIBUTIONS

A.Z., N.M., Y.R., N.N., and J.H.H. conceived the idea for this project, conducted experiments, and wrote the manuscript. M.Z., R.M., and Y.R. conducted micro-injections. S.B. and S.G. conducted and supervised high-throughput sequencing. A.M. and Y.S.M. assisted in RNA-seq analysis. I.U. and H.H. conducted lncRNA analysis. W.J.G. and J.B. assisted in analyzing ATAC-seq. E.C. and A. Tanay assisted in WGBS. H.G. and A.Z. performed tRNA analysis under the supervision of Y.P. I.A., D.L., and D.A.J. assisted in ChIP-seq experiments. N.N. supervised bioinformatics and analyzed ChIP-seq data. S.P., M.A., I.M., S.H., A.A.-C., J.D.B., O.G., D.S., and V.K. assisted in tissue culture. Y.S. assisted in RGM reporter experiments. Y.R., L.W., and N.M. engineered cell lines under S.V. supervision. A. Trumpp and R.S. provided Myc mutant lines and interpretation. N.N. and J.H.H. supervised executions of experiments and adequate analysis of data.

DECLARATION OF INTERESTS

J.H.H. is an advisor to Biological Industries, Ltd. J.H.H., N.N., and Y.R. filed related patents.

Received: February 21, 2018

Revised: September 21, 2018

Accepted: November 9, 2018

Published: December 13, 2018

REFERENCES

- Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. (2006). Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* 44, 23–28.
- 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.
- 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.e20.
- Dos Reis, M., Savva, R., and Wernisch, L. (2004). Solving the riddle of codon usage preferences: A test for translational selection. *Nucleic Acids Res.* 43, 154–159.
- Gingold, H., Tehler, D., Christoffersen, N.R., Nielsen, M.M., Asmar, F., Kooistra, S.M., Christoffersen, N.S., Christensen, L.L., Borre, M., Sørensen, K.D., et al. (2014). A dual program for translation regulation in cellular proliferation and differentiation. *Cell* 158, 1281–1292.
- Hussein, S.M.I., Puri, M.C., Tonge, P.D., Benevento, M., Corso, A.J., Clancy, J.L., Mosbergen, R., Li, M., Lee, D.-S., Cloonan, N., et al. (2014). Genome-wide characterization of the routes to pluripotency. *Nature* 516, 198–206.
- Kiviet, D.J., Nghe, P., Walker, N., Boulneau, S., Sunderlikova, V., and Tans, S.J. (2014). Stochasticity of metabolism and growth at the single-cell level. *Nature* 514, 376–379.
- Lee, D.S., Shin, J.Y., Tonge, P.D., Puri, M.C., Lee, S., Park, H., Lee, W.C., Hussein, S.M.I., 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.
- Li, D., Liu, J., Yang, X., Zhou, C., Guo, J., Wu, C., Qin, Y., Guo, L., He, J., Yu, S., et al. (2017). Chromatin accessibility dynamics during iPSC reprogramming. *Cell Stem Cell* 21, 819–833.e6.
- Mor, N., Rais, Y., Sheban, D., Peles, S., Aguilera-Castrejon, A., Zviran, A., Elinger, D., Viukov, S., Geula, S., Krupalnik, V., et al. (2018). Neutralizing Gatad2a-Chd4-Mbd3/NuRD complex facilitates deterministic induction of naive pluripotency. *Cell Stem Cell* 23, 412–425.e10.
- Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat. Biotechnol.* 26, 101–106.
- 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.
- Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. *Cell* 141, 432–445.
- 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.
- Scognamiglio, R., Cabezas-Wallscheid, N., Thier, M.C., Altamura, S., Reyes, A., Prendergast, Á.M., Baumgärtner, D., Carnevali, L.S., Atzberger, A., Haas, S., et al. (2016). Myc depletion induces a pluripotent dormant state mimicking diapause. *Cell* 164, 668–680.
- 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.
- Stelzer, Y., Shivalila, C.S., Soldner, F., Markoulaki, S., and Jaenisch, R. (2015). Tracing dynamic changes of DNA methylation at single-cell resolution. *Cell* 163, 218–229.
- 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.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
H3K4Me1	Abcam	Cat# ab8895; RRID: AB_306847
H3K4Me3	Abcam	Cat# Ab8580, RRID: AB_306649
H3K27ac	Abcam	Cat# ab4729, RRID: AB_2118291
H3K27Me3	Millipore	Cat# 07-449, RRID: AB_310624
H3K9Me3	Abcam	Cat# ab8898, RRID: AB_306848
H3K36Me3	Abcam	Cat# ab9050, RRID: AB_306966
H3K9Me2	MBL	Cat# MABI0317, RRID: N/A
Oct4	Santa Cruz	Cat# SC8628, RRID: AB_653551
Klf4	R&D	Cat# AF3158, RRID: AB_2130245
Sox2	Millipore	Cat# AB5603, RRID: AB_2286686
C-Myc	Santa Cruz	Cat# sc764, RRID: AB_631276
Pol II (N20)	Santa Cruz	Cat# sc899, RRID: AB_632359
Chemicals, Peptides, and Recombinant Proteins		
PD0325901	Axon Medchem	1408
CHIR99021	Axon Medchem	#1386
Recombinant human LIF	Peprotech	300-05
c-Myc inhibitor 10058-F4	Axon Medchem	#2222
cOmplete, Protease inhibitor	Roche	04693159001
Protease inhibitor cocktail (Sigma)	Sigma-Aldrich	P8340
Critical Commercial Assays		
Alkaline Phosphatase Kit	Millipore	SCR004
Lipofectamine RNAiMAX	ThermoScientific	#13778075
TruSeq RNA Sample Preparation Kit v2	Illumina	RS-122-2001
EZ DNA Methylation-Gold kit	Zymo	D5005
EpiGenome Methyl-Seq	Illumina	EGMK81312
Truseq small RNA sample preparation kit	Illumina	RS-200-0012
Deposited Data		
ATAC-Seq, ChIP-Seq, RNA-Seq, WGBS	This Paper	GEO: GSE102518
Experimental Models: Cell Lines		
<i>Mbd3</i> ^{flox/-} cell lines that carries the GOF18-Oct4-GFP transgenic reporter, FUW-M2RtTA; FUW-TetO-STEMCCA-humanOKSM – ES and Secondary MEF	Rais et al., 2013	N/A
RGM-miR290-SE-tdTomato Nanog-GFP	Stelzer et al., 2015	N/A
FUW-M2RtTA; FUW-TetO-STEMCCA-humanOKSM, ΔPE-GOF18-Oct4-GFP cells (Both Gatad2a WT and KO)	Mor et al., 2018	N/A
Tet1,2,3 ^{t/t} MEF and iPSC cell line	This paper	N/A
c-Myc ^{t/t} , n-Myc ^{t/t} Rosa26-Lox-stop-Lox-YFP	Scagnamiglio et al., 2016	N/A
V6.5 murine ESC line	Beard et al., 2006	N/A
Experimental Models: Organisms/Strains		
Tet2 ^{flox/flox} (B6;129S-Tet2tm1.1laai/J) mouse strain	Jackson	# 017573, RRID: IMSR_JAX:017573
Tet1,2,3 ^{flox/flox}	This paper	N/A
Oligonucleotides		
siRNA targeting Mouse-cMYC	Invitrogen	MSS-237326, MSS-237327, MSS-237328
siRNA targeting Mouse-IMYC	Invitrogen	MSS-275360, MSS-275361, MSS-275362

(Continued on next page)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
siRNA targeting Mouse-nMYC	Invitrogen	MSS-207081, MSS-207082, MSS-276042
Stealth RNAi siRNA Negative Control, Med GC	Invitrogen	12935300
siRNA targeting Mouse Tfap2c	Invitrogen	MSS-210701, MSS-277866, MSS-277867
siRNA targeting Mouse Bend3	Invitrogen	MSS-221180, MSS-221181, MSS-221182
siRNA targeting Mouse Tfcp2l1	Invitrogen	MSS-294469, MSS-294470, MSS-294471
Recombinant DNA		
pLM-mCerulean-cMyc	Addgene	Addgene #23244
FUW-M2rtTA	Addgene	Addgene #20342
FUW-TetO-STEMCCA-humanOKS-mCherry	Mor et al., 2018	N/A
FUW-TetO-STEMCCA-humanOKSM	Mor et al., 2018	N/A
Software and Algorithms		
Tophat 2.0.10	https://ccb.jhu.edu/software/tophat/index.shtml	https://ccb.jhu.edu/software/tophat/index.shtml
Cufflinks 2.2.1	http://cole-trapnell-lab.github.io/cufflinks/	http://cole-trapnell-lab.github.io/cufflinks/
R pheatmap package	https://cran.r-project.org/web/packages/pheatmap/index.html	https://cran.r-project.org/web/packages/pheatmap/index.html
R prcomp package	https://stat.ethz.ch/R-manual/R-devel/library/stats/html/prcomp.html	https://stat.ethz.ch/R-manual/R-devel/library/stats/html/prcomp.html
Bowtie 2	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
Picard tools	http://broadinstitute.github.io/picard/	http://broadinstitute.github.io/picard/
MACS 1.4.2	http://liulab.dfci.harvard.edu/MACS/	http://liulab.dfci.harvard.edu/MACS/
bedtools	https://bedtools.readthedocs.io/en/latest/	https://bedtools.readthedocs.io/en/latest/
samtools	http://www.htslib.org/doc/samtools.html	http://www.htslib.org/doc/samtools.html
IGV	https://software.broadinstitute.org/software/igv/	https://software.broadinstitute.org/software/igv/
Python misha package	https://bitbucket.org/tanaylab/misha-package	https://bitbucket.org/tanaylab/misha-package
MATLAB	MathWorks	https://www.mathworks.com/products/matlab.html
Prism	GraphPad software	https://www.graphpad.com/scientific-software/prism/
FlowJo	FlowJo	https://www.flowjo.com/
ZEN Software	Zeiss	https://www.zeiss.com/microscopy/int/products/microscope-software/zen-lite.html

CONTACT FOR REAGENT AND RESOURCE SHARING

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

EXPERIMENTAL MODEL AND SUBJECT DETAILS**Mice**

Tet2^{flox/flox} mice were obtained from Jackson Laboratories (Stock number 017573). All animal experiments were performed according to the Animal Protection Guidelines of Weizmann Institute of Science, Rehovot, Israel. All animal experiments described herein were approved by relevant Weizmann Institute IACUC (#00330111-Hanna). All efforts were made to minimize animal discomfort.

Cell Culture

WT or Mutant mouse ESC/iPSC lines and sub-clones were routinely expanded in mouse ES medium (mESM) consisting of: 500ml DMEM-high glucose (ThermoScientific), 15% USDA certified Fetal Bovine Serum (Biological Industries), 1mM L-Glutamine (Biological Industries), 1% nonessential amino acids (Biological Industries), 0.1mM β - mercaptoethanol (Sigma), penicillin-streptomycin (Biological Industries), 10 μ g recombinant human LIF (Peprotech). For ground state naive conditions (N2B27 2i/LIF), murine naive pluripotent cells (iPSCs and ESCs) were conducted in serum-free chemically defined N2B27-based media: N2B27-based media: 250ml Neurobasal (ThermoScientific), 250ml DMEM:F12 (ThermoScientific) 5ml N2 supplement (Invitrogen; 17502048), 5ml B27 supplement (Invitrogen; 17504044), 1mM glutamine (Invitrogen), 1% nonessential amino acids (Invitrogen), 0.1mM β -mercaptopethanol (Sigma), penicillin-streptomycin (Invitrogen), 5mg/ml BSA (Sigma), small-molecule inhibitors CHIR99021 (CH, 3 μ M - Axon Medchem) and PD0325901 (PD, 0.3-1 μ M - Axon Medchem). Mycoplasma detection tests were conducted routinely every month with MycoALERT ELISA based kit (Lonza) to exclude mycoplasma free conditions and cells throughout the study.

METHOD DETAILS

Generation of *Gatad2a*-knockout Reprogrammable secondary MEF lines

Secondary MEF for *Gatad2a*^{-/-} cell line and WT-2 were obtained as described in [Mor et al. \(2018\)](#). Shortly, iPSCs were established following primary reprogramming of cells using M2rtTA and TetO-OKSM-STEMCCA (human OSKM cDNA inserts were used). The iPSC, harboring mCherry constitutive expression (to label viable cells) and Δ PE-GOF18-Oct4-GFP cassette (Addgene plasmid# 52382), were then subjected to CRISPR and Cas9 targeting *Gatad2a* (sgRNA- cgccgtatgtattgtct), resulting in *Gatad2a*-knockout cells ([Mor et al., 2018](#)). Both *Gatad2a*-KO and its isogenic wild-type line (WT-2) were then injected into blastocysts, and MEF were harvested at E13.5. MEFs were harvested at E13.5 and grown in MEF medium, which contained 500 mL DMEM (Invitrogen), 10% fetal calf serum (Biological Industries), 1 mM glutamine (Invitrogen), 1% non-essential amino acids (Invitrogen), 1% penicillin-streptomycin (Invitrogen), 1% sodium pyruvate (Invitrogen). All animal studies were conducted according to the guideline and following approval by the Weizmann Institute IACUC (approval # 33550117-2 and 33520117-3). Cell sorting and FACS analysis were conducted on 4 lasers equipped FACS Aria III cells sorter (BD). Analysis was conducted with either DIVA software or Flowjo. Throughout this study, all cell lines were monthly checked for Mycoplasma contaminations (LONZA – MYCOALERT KIT), and all samples analyzed in this study were never tested positive or contaminated.

Generation of reprogrammable *Mbd3*^{flx/-} secondary MEF lines

All secondary reprogrammable lines harbor constitutive expression of the M2rtTA from the Rosa26 locus and TetO-OKSM cassette (human OKSM cDNA inserts were used) introduced either by viral transduction of knock-in in the Col1a1 locus. Secondary mouse embryonic fibroblast (MEF) from *Mbd3*^{flx/-} cell line (A12 clone: *Mbd3*^{flx/-} cell lines that carries the GOF18-Oct4-GFP transgenic reporter (complete Oct4 enhancer region with distal and proximal enhancer elements) (Addgene plasmid #60527)) and WT-1 cell line (WT-1 clone that carries the deltaPE-GOF18-Oct4-GFP reporter (Addgene plasmid#52382) were previously described ([Rais et al., 2013](#)). Note that we do not use Oct4-GFP or any other selection for cells before harvesting samples for conducting genomic experiments.

Mouse embryo micromanipulation

Pluripotent mouse ESCs and iPSCs were injected into BDF2 diploid blastocysts, harvested from hormone primed BDF1 6-week-old females. Microinjection into E3.5 blastocysts placed in M16 medium under mineral oil was done by a flat-tip microinjection pipette. A controlled number of 10-12 cells were injected into the blastocyst cavity. After injection, blastocysts were returned to KSOM media (Invitrogen) and placed at 37°C until transferred to recipient females. Ten to fifteen injected blastocysts were transferred to each uterine horn of 2.5 days post coitum pseudo-pregnant females.

Reprogramming of MEF to naive ground state naive iPSC

Reprogramming of the optimally NuRD depleted and WT platform cell lines to iPSC was performed for the first 3 days with MES medium, which contained 500 mL DMEM (Invitrogen), 15% fetal calf serum, 1 mM glutamine (Invitrogen), 1% non-essential amino acids (Invitrogen), 1% penicillin-streptomycin (Invitrogen), 1% sodium pyruvate (Invitrogen), 0.1 mM β -mercaptopethanol (Sigma), 20 ng/ml human LIF (in house prepared). MES medium for reprogramming was supplemented with Doxycycline (DOX) (2 μ g ml⁻¹), which activated the OKSM cassette and the reprogramming process. On day 3.5, medium was replaced to FBS-free media composed of: 500 ml DMEM (Invitrogen), 15% knockout serum replacement (Invitrogen; 10828), 1 mM glutamine (Invitrogen), 1% non-essential amino acids (Invitrogen), 0.1 mM β -mercaptopethanol (Sigma), 1% penicillin-streptomycin (Invitrogen), 1% sodium pyruvate (Invitrogen), 20 ng/ml recombinant human LIF (Peprotech or in house-prepared), CHIR99021 (3 μ M; Axon Medchem), PD0325901 (PD, 0.3-1 μ M; Axon Medchem). After DOX treatment medium was replaced to KSR-based with the addition of MEK and GSK3 inhibitors (2i), supplemented with Doxycycline (DOX) (1 μ g ml⁻¹), until the end of the reprogramming regimen (i.e., day 8). Cells were harvested at first time point (MEF) and every 24 hours until day 8 and were used for library preparation followed by sequencing. *Mbd3*^{f/-}, *Gatad2a*^{-/-} and WT established iPSC line (after 3 passages or more), and *Mbd3*^{f/-} or WT V6.5 mouse ESCs were used as controls. Reprogramming was conducted in 5% O₂ conditions during the first 3.5 days, and then cells were moved to 20% O₂ conditions. For all mouse iPSC reprogramming experiments, irradiated human foreskin fibroblasts were used as feeder cells, as any sequencing

input originating from the use of human feeder cells cannot be aligned to the mouse genome and is therefore omitted from the analysis. All cell undergoing reprogramming were harvested without any prior passaging or sorting for any subpopulations during the reprogramming process. No blinding was conducted when testing outcome of reprogramming experiments.

Primary and secondary reprogrammable lines by viral infection

For primary cell reprogramming, ~3x106 293T cells in a 10cm culture dish were transfected with JetPEI (Polyplus) 20ul reagent for 10ug DNA as follow: pPAX (3.5 µg), pMDG (1.5 µg) and 5 µg of the lentiviral target plasmid (pLM-mCerulean-cMyc (Plasmid #23244), FUW-STEMCCA-OKS-mCherry or FUW-M2rtTA, FUW-TetO-STEMCCA-OKS-mCherry (a kind gift from Gustavo Mostoslavsky). Viral supernatant was harvest 48 and 72 hours post transfection, filtered through 0.45micron sterile filters (Nalgene) and added freshly to the primary MEF that was isolated from Mbd3^{flox/-} chimeric mice (unless indicated otherwise). At day 4 cells was sorted by the relevant florescent filter (mCerulean (cMyc OE), mCherry (OSK OE) or double positive (OSK+M OE) cell was collected for RNA extraction or seeded for farther growth.

Knockdown endogenous Myc during reprogramming

For secondary Mbd3^{f/-} OSK^{2nd} production, primary MEFs from Mbd3^{flox/-} chimeric mice were infected with FUW-TetO-STEMCCA-OKS-mCherry and FUW-M2rtTA. iPS cells were isolated and injected into BDF2 blastocysts for the isolation of secondary MEFs. Secondary MEFs were transfected at day -3 and again at day 0 (starting reprogramming by adding DOX) with siRNA for cMyc, nMyc or control (Stealth siRNA- mix of 3 as indicated in the table below) with RNAiMAX (Invitrogen). For molecular analysis, cells were collected at day 3 and day 7 or day 8 as indicated.

Generation of triple Tet1,2,3^{flox/flox} mice and cell lines

Tet2^{flox/flox} mice were obtained from Jackson Laboratories (Stock number 017573). Tet1^{flox/flox} mice were generated by using conditional knockout targeting vector against Exon 4 in V6.5 ESC. After removal of Neomycin selection cassettes by Flippase in correctly targeted ESCs (validated both by Southern Blot and PCR analysis), chimeric blastocyst injections followed by successful germline transmission allowed us to establish Tet1^{flox/flox} mouse colony. Tet3^{flox/flox} mice were generated by gene targeting of the endogenous Exon 7 (contains Fe(ii) catalytic domain) Tet3 locus. After removal of Neomycin selection cassettes by Flippase in correctly targeted ESCs (validated both by Southern Blot and PCR analysis), chimeric blastocyst injections followed by successful germline transmission allowed us to establish Tet3^{flox/flox} mouse colony. Triple floxed homozygous mice were generated by interbreeding, after which Tet1^{flox/flox} Tet2^{flox/flox} Tet3^{flox/flox} mouse strain was obtained. Genotyping primers: Tet1_gen1_F: AGGAGTGT CAGGGTCAAGGCCATC; Tet1_gen1_R:TCCCTGACAGCAGCCACACTTG; Tet2_lox_F: AAGAATTGCTACAGGCCTGC; Tet2_lox_R: TTCTTTAGCCCTTGCTGAGC; Tet3_lox_f: agttccctgacgttggagagtgg; Tet3_lox_r: ggaactcaagctcctcagaggaagc. The Tet1 floxed allele gives a band of 500bp, compared to the 450bp WT. The Tet2 flox allele gives a band of 427bp, compared to the 249bp WT. The Tet3 floxed allele gives a band of 300bp, compared to the 200bp WT. MEFs, ESCs and iPSCs were derived from triple Tet1/2/3^{flox/flox} mice and were used as indicated in the figures. Deleting Gatad2a in Tet1/2/3^{flox/flox} iPSCs was done with CRISPR and Cas9 as indicated in methods above.

RT-PCR analysis

Total RNA was isolated using Trizol (ThermoFisher). 1 µg of DNase-I-treated RNA was reverse transcribed using a First Strand Synthesis kit (Invitrogen) and ultimately re-suspended in 100 µl of water. Quantitative PCR analysis was performed in triplicate using 1/50 of the reverse transcription reaction on ViiA7 platform (Applied Biosystems). Error bars indicate standard deviation of triplicate measurements for each measurement.

AP Staining

Alkaline phosphatase (AP) staining was performed with AP kit (Millipore SCR004) according to manufacturer's instructions.

Imaging, quantifications, and statistical analysis

Imaged were acquired with D1 inverted microscope (Carl Zeiss, Germany) equipped with DP73 camera (Olympus, Japan) or with Zeiss LSM 700 inverted confocal microscope (Carl Zeiss, Germany) equipped with 405nm, 488nm, 555nm and 635 solid state lasers, using a 20x Plan-Apochromat objective (NA 0.8). All images were acquired in sequential mode. For comparative analysis, all parameters during image acquisition were kept constant throughout each experiment. Images were processed with Zen blue 2011 software (Carl Zeiss, Germany), and Adobe Photoshop.

ChIP-seq library preparation

Cells were crosslinked in formaldehyde (1% final concentration, 10 min at room temperature), and then quenched with glycine (5 min at room temperature). Fixed cells were lysed in 50 mM HEPES KOH pH 7.5, 140 mM NaCl, 1 mM EDTA, 10% glycerol, 0.5% NP-40 alternative, 0.25% Triton supplemented with protease inhibitor at 4°C (Roche, 04693159001), centrifuged at 950g for 10 min and re-suspended in 0.2% SDS, 10 mM EDTA, 140 mM NaCl and 10 mM Tris-HCl. Cells were then fragmented with a Branson Sonifier (model S-450D) at -4°C to size ranges between 200 and 800 bp and precipitated by centrifugation. Antibody was pre-bound by incubating with Protein-G Dynabeads (Invitrogen 100-07D) in blocking buffer (PBS supplemented with 0.5% TWEEN and 0.5%

BSA) for 1 h at room temperature. Washed beads were added to the chromatin lysate for an incubation periods of either 6 or 18 hours. Samples were washed five times with RIPA buffer, twice with RIPA buffer supplemented with 500 mM NaCl, twice with LiCl buffer (10 mM TE, 250mM LiCl, 0.5% NP-40, 0.5% DOC), once with TE (10Mm Tris-HCl pH 8.0, 1mM EDTA), and then eluted in 0.5% SDS, 300 mM NaCl, 5 mM EDTA, 10 mM Tris HCl pH 8.0. Eluate was incubated treated sequentially with RNaseA (Roche, 11119915001) for 30 min and proteinase K (NEB, P8102S) for 2 h in 65°C for 8 h, and then. DNA was purified with The Agencourt AMPure XP system (Beckman Coulter Genomics, A63881). Libraries of cross-reversed ChIP DNA samples were prepared according to a modified version of the Illumina Genomic DNA protocol. All chromatin immunoprecipitation data are available at the National Center for Biotechnology Information Gene Expression Omnibus database under the series accession GEO no. GSE102518. Samples were run with various protocols and machines ([Table S1](#)). Please note that while it seems that Klf4 is starting to significantly bind enhancers only on day 2, we note that this is actually a result of the specific Klf4 antibody used for ChIP-seq, that is known to have better affinity for endogenous form of mouse Klf4 which becomes highly upregulated later in reprogramming, than the exogenous transgene derived human KLF4 version that is induced from early stage upon DOX addition.

PolyA-RNA-seq library preparation

Total RNA was isolated from indicated cell lines, RNA was extracted from Trizol pellets by Direct-zol RNA MiniPrep kit (Zymo) and utilized for RNA-Seq by TruSeq RNA Sample Preparation Kit v2 (Illumina) according to manufacturer's instruction. See [Table S1](#) for details of protocol and sequencing machine used.

Small RNA-seq library preparation

1ug of total RNA from each sample was processed using the TruSeq small RNA sample preparation kit (RS-200-0012 Illumina) followed by 12 cycles of PCR amplification. Libraries were evaluated by Qubit and TapeStation. For purification of the small RNA fragments, they were size selected using Blupippne machine (Sage Science) with 3% gel cassette followed by clean-up with mini-elite PCR purification kit (QIAGEN). The libraries were constructed with different barcodes to allow multiplexing of 11 samples. See [Table S1](#) for details of protocol and sequencing machine used.

ATAC-seq library preparation

Cells were trypsinized and counted, 50,000 cells were centrifuged at 500g for 3 min, followed by a wash using 50 µL of cold PBS and centrifugation at 500g for 3 min. Cells were lysed using cold lysis buffer (10 mM Tris-HCl, pH 7.4, 10 mM NaCl, 3 mM MgCl₂ and 0.1% IGEPAL CA-630). Immediately after lysis, nuclei were spun at 500g for 10 min using a refrigerated centrifuge. Next, the pellet was resuspended in the transposase reaction mix (25 µL 2 × TD buffer, 2.5 µL transposase (Illumina) and 22.5 µL nuclease-free water). The transposition reaction was carried out for 30 min at 37°C and immediately put on ice. Directly afterward, the sample was purified using a QIAGEN MinElute kit. Following purification, the library fragments were amplified using custom Nextera PCR primers 1 and 2 for a total of 12 cycles. Following PCR amplification, the libraries were purified using a QIAGENMinElute Kit and sequenced as indicated in [Table S1](#).

Whole-Genome Bisulfite Sequencing (WGBS) library preparation

DNA was isolated from snap-frozen cells using the Quick-gDNA mini prep kit (Zymo). DNA was then converted by bisulfite using the EZ DNA Methylation-Gold kit (Zymo). Sequencing libraries were created using the EpiGnome Methyl-Seq (Epicenter) and sequenced as indicated in [Table S1](#)

Reduced-Representation Bisulfite (RRBS) library preparation

RRBS libraries were generated as described previously with slight modifications⁴⁰. Briefly, DNA was isolated from snap-frozen cell pellets using the Quick-gDNA mini prep kit (Zymo). Isolated DNA was then subjected to Mspl digestion (NEB), followed by end repair using T4 PNK/T4 DNA polymerase mix (NEB), A-tailing using Klenow fragment (3'5' exo-) (NEB), size selection for fragments shorter than 500 bp using SPRI beads (Beckman Coulter) and ligation into a plasmid using quick T4 DNA ligase (NEB). Plasmids were treated with sodium bisulphite using the EZ DNA Methylation-Gold kit (Zymo) and the product was PCR amplified using GoTaq Hot Start DNA polymerase (Promega). The PCR products were A-tailed using Klenow fragment, ligated to indexed Illumina adapters using quick T4 DNA ligase and PCR amplified using GoTaq DNA polymerase. The libraries were then size-selected to 200–500 bp by extended gel electrophoresis using NuSieve 3:1 agarose (Lonza) and gel extraction (QIAGEN). See [Table S1](#) for sequencing protocol used.

QUANTIFICATION AND STATISTICAL ANALYSIS

ChIP-seq analysis

Alignment and peak detection

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

TSS, TES and Enhancer definition

Transcription start sites (TSS) and transcription end sites (TES) were taken from mm10 assembly (UCSC, December 2011). Promoters/TES intervals were defined as 1000bp around each TSS/TES, and enhancers were defined as 300bp around enhancer detection summit point (see enhancer identification below).

Chromatin modification profile estimation in TSS, TES and in enhancers

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

Transcription factor binding in promoter and enhancer

Promoter or enhancer was defined as bound by a TF if it overlapped a binding peak of the TF, as detected by MACS. Of note, 94.3% of the identified peaks of OSK overlap with either promoter or enhancer.

Transcription factor binding taken from previously published data (Chronis et al., 2017)

OSKM/Runx1 Binding data were downloaded from NCBI GEO: GSE90893, and were analyzed using the same pipeline as described above.

RNA-seq analysis

Read Alignment for PolyA-RNA-seq

Tophat software version 2.0.10 was used to align reads to mouse mm10 reference genome (UCSC, December 2011). FPKM values were calculated over all genes in mm10 assembly GTF (UCSC, December 2011), using cufflinks (version 2.2.1). Genes annotated as protein coding, pseudogene or lncRNA (n = 24,439) were selected for further analysis.

Read Alignment for Small RNA-seq

Bowtie software version 2 was used to align reads to mouse mm10 reference genome (UCSC, December 2011). FPKM values were calculated over all genes in mm10 assembly GTF (UCSC, December 2011), using cufflinks (version 2.2.1). Genes annotated as rRNA, miRNA, snoRNA were selected for further analysis.

Subsequently, PolyA and small RNA-seq FPKM were combined and processed together.

Active and Differential genes

Gene was defined to be active in samples where FPKM is above 0.5 of the gene max value. Differential genes were defined by (FC > 4) & (maximum value > 1). Subsequent filtering was done to reject oscillatory or non-continuous time series by comparing the sum of derivatives to the total span. Specifically, the filtering scheme is $[\sum_{j=1}^{t-1} (R_j^i - R_{j-1}^i) / (\max_j(R_j^i) - \min_j(R_j^i))] < 2.5$, where j is the sample index, i is the gene index, and R_j^i is the level of gene i in sample j.

Expression HeatMap

Gene sorting in expression heat-maps (e.g. Figure 1G) was done according to the average position of gene active samples, i.e., calculating the average of sample indexes (j) where the gene is active. Unit normalized FPKM was calculated using the following formula $R_j^i * = R_j^i / [\max_j(R_j^i) + 1]$ where j is the sample index, i is the gene index and FPKM = 1 is the transcription noise threshold, and $\max_j(R_j^i)$ is the maximal level in each dataset. This normalization scheme allowed easy comparison of gene temporal pattern with normalized dynamic range.

Correlations

All correlation tests were done using Spearman correlation.

PCA

PCA analysis (e.g. Figure 1H) was carried out over all differential genes in unit normalization by MATLAB (version R2011b) princomp command.

Analysis and integration of previously published datasets

C/EBP α +OSKM mouse B cell reprogramming RNA-seq data was downloaded from NCBI GEO: GSE96611, and was analyzed using the same pipeline as described above.

Extended Differential lncRNA analysis

lncRNA dataset was annotated using PLAR. FPKM values were calculated for all lncRNAs in PLAR mm9 dataset using cuffdiff (version 2.2.1). lncRNA coordinates were then converted to mm10 using liftOver utility. Differential lncRNAs were defined by (FC > 4) & (maximum value > 1). Subsequent filtering removed lncRNAs that were suspected to be expressed due to B1/B2-repeats by removing all sequence reads overlapping B1 or B2 repeats, resulting in 560 differential lncRNAs, out of them 221 differential lncRNA not previously annotated by Ensembl. Hierarchical clustering was performed over all differential lncRNAs with Spearman correlation metric and average linkage, further separating the differential lncRNAs to up, downregulated and intermediate induced lncRNAs (Table S4).

RNA-seq data are available at the National Center for Biotechnology Information Gene Expression Omnibus database under the series accession GEO: GSE102518.

Phylogenetic analysis

Conservation scores of CAPGs and ESPGs were extracted from PhyloGene database <http://genetics.mgh.harvard.edu/phylogene/>.

Functional Enrichment

Active genes at each sample (day) are tested for enrichment of functional gene sets taken from Gene Ontology (GO, <http://www.geneontology.org>), using Fisher exact test. Gene is defined to be active in samples where FPKM is above 0.5 of the gene max value. All enrichment values for each day were FDR corrected using Benjamini and Hochberg method. GO annotations were filtered to include only annotations with FDR-corrected p-val < 0.01 in at least two samples, annotations are sorted according to average position of enrichment pattern.

Protein-DNA binding enrichment analysis

Active genes at each sample (day) are tested for enrichment (fisher exact test) to previously published protein-DNA binding ChIP-seq obtained from the MRC mouse ES ChIP-seq compendium (Figure 4G), hmChIP and BindDB databases. Gene is defined to be active in samples where RPKM is above 0.5 of the gene max value. All enrichment values for each day were passed through FDR test, using the Benjamini and Hochberg method. Subsequently, TF annotations per day were filtered to include only annotations with FDR-corrected Pval < 10⁻³⁰ in at least one sample. Further filtering the predicted TF to include only TF that are also differentially expressed during reprogramming according to our collected RNA-seq.

ATAC-seq analysis

Reads were aligned to mm10 mouse genome using Bowtie2 with the parameter -X2000 (allowing fragments up to 2 kb to align). Duplicated aligned reads were removed using Picard MarkDuplicates tool with the command REMOVE_DUPLICATES = true. To identify chromatin accessibility signal we considered only short reads ($\leq 100\text{bp}$) that correspond to nucleosome free region. C/EBPa+OSKM B cell reprogramming ATAC-seq data was downloaded from NCBI GEO: GSE96611, and was analyzed using the same pipeline as described above.

Identifying accessible chromatin regions

To detect and separate accessible loci in each sample, we used MACS version 1.4.2-1 with--call-subpeaks flag (PeakSplitter version 1.0). Next, summits in previously annotated spurious regions were filtered out using a custom blacklist targeted at mitochondrial homologs. To develop this blacklist, we generated 10,000,000 synthetic 34-mer reads derived from the mitochondrial genome. After mapping and peak calling of these synthetic reads we found 28 high-signal peaks for the mm10 genome. For all subsequent analysis, we discarded peaks falling within these regions.

Enhancer Identification

Each ATAC-seq peak in each sample was represented by a 300bp region around the summit center. H3K27ac peaks were detected in a similar manner, using MACS version 1.4.2-1, and merged for all time points using bedtools merge command. All ATAC peaks were filtered to include only peaks which co-localized with the merged H3K27ac peaks, meaning only ATAC peaks that have H3K27ac mark on at least one of the time points were passed to further processing. Finally, the peaks from all samples were unified and merged (using bedtools unionbedg and merge commands), further filtered to reject peaks that co-localized with promoter or exon regions based on mm10 assembly (UCSC, December 2011). Finally we were left with 93,137 genomic intervals which we annotated as active enhancers, of which 78% of overlap with H3K4me1 modification, and 69.9% are bound by at least one of the transcription factors mapped (RNA Pol II/O/S/K/M), and 54% are bound by at least one of O/S/K. MEF-enhancers significantly overlap with ENCODE Spleen and Heart enhancers (p value < 1e-13, Enrichment fold-change > 1.4), and Day7 & Day8 enhancers significantly overlap with ENCODE mESC enhancers (p value < 3.2e-11, Enrichment fold-change > 1.3). All enhancers were then annotated by their most proximal gene using annotatePeaks function (homer/4.7 package). Enhancers were considered as differential if both their ATAC-seq and H3K27ac signals show significant change during reprogramming (min zscore < 0.5, max zscore > 1.5, for both chromatin marks). ATAC-seq data are deposited under GEO no. GSE102518.

Generating ATAC-seq normalized profiles in TSS and in enhancers

ATAC-seq profiles were calculated using in-house script over all genomic intervals defined for TSS and enhancers. Shortly, the genomic interval is divided to 50bp size bins, and the coverage in each bin is estimated. Each bin is then converted to z-score by normalizing each position by the mean and standard deviation of the sample noise ($X^j = (X_j - \mu_{\text{noise}})/\sigma_{\text{noise}}$). Noise parameters were estimated for each sample from 6×10^7 random bp across the genome. Finally, the 3rd highest bin z-score of each interval is set to represent the coverage of that interval.

Methylation Analysis of WGBS and RRBS

Alignment of RRBS data

The sequencing reads were aligned to the mouse mm10 reference genome (UCSC, December 2011), using Bismark aligner (parameters -n 1 -l 20). Mapping was done independently for the two ends of each pair. Read pairs that mapped uniquely to two different fragments were discarded. In cases where one read uniquely mapped on a restriction site but its pair could not be mapped uniquely

or could not be mapped at all, we attempted to re-align the entire read pair to the fragment. Read pairs showing more than one unconverted non-CpG cytosine, which occur at very low frequency were filtered out.

Alignment of WGBS data

The sequencing reads were aligned to the mouse mm10 reference genome (UCSC, December 2011), using a proprietary script based on Bowtie2. In cases where the two reads were not aligned in a concordant manner, the reads were discarded.

Methylation estimation

Methylation levels of CpGs calculated by RRBS and WGBS were unified. Mean methylation was calculated for each CpG that was covered by at least 5 distinct reads (X5). Average methylation level in various genomic intervals was calculated by taking the average over all covered X5 covered CpG sites in that interval. Please note that in both systems higher global DNA methylation levels were observed in iPSC that were cultured over prolonged time in 2i/LIF conditions, compared to newly generated iPSCs on day 8 possibly because OSKM transgene can boost hypomethylation independent of 2i.

Correlation of chromatin modifications

Correlation between chromatin modification to gene expression and to accessibility signal were estimated using Spearman correlation (Figures 4A, 5A, S2C, and S2D). Promoters or enhancers with z-score above zero were included in the analysis, resulting in different number of promoter or enhancers for each chromatin marks (which are indicated in the figures).

Cross-correlation of chromatin modifications

Cross-correlation method^{50,51} measures the overlap between two signals, while shifting the signals in their x axis (convolution). In our case, the x axis is time. Cross-correlation score was calculated using MATLAB R2013b xcorr command. The offset showing the highest xcorr coefficient was defined as the optimal offset between the two signals. Cross-correlation was calculated in three systems: (i) Between chromatin modifications in promoters and gene expression pattern of ESPGs (Figure 5C). (ii) Between chromatin modifications and accessibility signal in differential enhancers (Figure 5F). (iii) Between chromatin modifications in promoters and enhancers that are associated with these promoters, and gene expression (Figure S7A).

In all these cases promoters/enhancers were included only if the modification z-score was changing ($\text{max-min} > 0.5$), resulting in different number of promoters/enhancers as indicated in the graphs. Please note that we could not present cross-correlation with OSKM binding, because the method requires quantitative information (z-score), and we used OSKM binding only as binary data (MACS binding peaks).

Combinatorial analysis for histone marks localization

To quantify all possible combinations of epigenetic modifications (Figures 5B, 5E, and S5F), we transformed our epigenetic data to a binary code in each genomic region (promoter/enhancer). Each epigenetic mark in promoter or enhancer was considered high (value = 1) if its z-score was above 1.5. For each sample, the percentage of each combination is presented. Combinations which are less than 3% of the total combinations in every sample are presented as “other” (gray color). Upregulated ESPGs were selected if their fold-change was as follows: $\text{mean}(\text{Day8,iPS})/\text{mean}(\text{MEF,Day1}) > 4$. Downregulated ESPGs were selected if their fold change was as follows: $\text{mean}(\text{MEF,Day1})/\text{mean}(\text{Day8,iPS}) > 4$.

Motif analysis

Enriched binding motifs were searched in various genomic intervals using findMotifsGenome function from homer software package version 4.7, using the software default parameters.

Motif analysis in open versus closed binding targets

In order to find binding motifs in open versus closed binding targets (Figure 2G) we followed the analysis outline presented by Soufi et al. (2015): We considered binding peaks of O/S/K/M in Day1, identified by MACS as explained. We calculated nucleosome occupancy in a 200bp window in the summit of the peak, and in two 100bp flanking regions on the two sides of the central window. Nucleosome occupancy was estimated from ATAC-seq data, measured in Day1, using nucleoatac occ software. Top 2000 binding sites with highest center/flanking ratio were selected as closed sites (as long as ratio > 1), and bottom 2000 sites were selected as open sites (as long as ratio < 1). Next, motif search and annotation was done as in Soufi et al. (2015), using DREME, Centrimo and TOMTOM software, of MEME suite.

Boxplot analysis

Box-plots show 25-th and 75-th percentile of the represented distribution values, with median marked by the mid-line. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box, and outliers are not presented.

Translation Analysis

Coding sequences

The coding sequences of *M. musculus* were downloaded from the Consensus CDS (CCDS) project (<ftp://ftp.ncbi.nlm.nih.gov/pub/CCDS/>).

tRNA gene copy numbers

The tRNA gene copy numbers of *M. musculus* were downloaded from the Genomic tRNA Database (<http://lowelab.ucsc.edu/GtRNADB/>).

Estimating translational efficiency by chromatin modification signature in the vicinity of tRNAs

We estimated translation efficiency of genes using the “tRNA activation index” (tACI) which was introduced previously by Gingold et al. (2014). This measure is calculated similarly to the tRNA Adaptation Index (tAI) measure of translation efficiency (Dos Reis et al., 2004), with one change—tRNA availabilities are determined based on chromatin modification in the vicinity of the tRNA genes rather than by gene copy numbers. Specifically, we set the activation score of each individual tRNA gene to be the maximal read per mega-base (RPM) value of the activation-associated modification H3K4me3 across a region spanning the 500 nucleotides upstream to the first nucleotide of the mature tRNA. Individual tRNA genes, for which no signal enrichment was found, were classified as “not activated.” Next, we defined the activation score of each tRNA type (anticodon) by the sum of the activation scores of its gene copies. Then, we determined the translation efficiency of each of the 61 codon types by the extent of activation of the tRNAs that serve in translating it, incorporating both the fully matched tRNA as well as tRNAs that contribute to translation through wobble rules.

Formally, the translation efficiency score for the i -th codon is

$$w_i = \sum_{j=1}^{n_i} (1 - s_{ij}) tCME_{ij}$$

where n is the number of types of tRNA isoacceptors that recognize the i -th codon, $tCME_{ij}$ denotes the sum of the chromatin modification scores of the activated copies of the j -th tRNA that recognizes the i -th codon, and s_{ij} corresponds to the wobble interaction, or selective constraint on the efficiency of the pairing between codon i and anticodon j , as was determined and implemented for the original tAI measure. As done in the original tAI formalism by the scores of the 61 codons are further divided by the maximal score (yielding w_i as the normalized scores for each codon type), and finally, the tACI value of a gene with L codons is then calculated as the geometric mean of the w_i 's of its codons

$$tACI(g) = \sqrt[L]{\prod_{c=1}^L w_c}$$

DATA AND SOFTWARE AVAILABILITY

All RNA-seq, ATAC-seq, ChIP-seq and methylation data are available to download from NCBI GEO, under super-series GEO: GSE102518.